

Morphology and Linguistic Typology

On-line Proceedings of the
Fourth Mediterranean Morphology Meeting (MMM4)
Catania, 21-23 September 2003

*Edited by Geert Booij, Emiliano Guevara,
Angela Ralli, Salvatore SgROI & Sergio Scalise*

URL: <http://morbo.lingue.unibo.it/mmm/>

2005 Università degli Studi di Bologna
On-line Proceedings of the Mediterranean Morphology Meetings (MMM)
ISSN: 1826-7491

Contents

<i>Foreword</i>	5
<i>Morphological Typology and First Language Acquisition: Some Mutual Challenges</i> , Wolfgang U. Dressler	7
<i>Typology, Diachrony, and Universals of Semantic Change in Word Formation: A Romanist's Look at the Polysemy of Agent Nouns</i> , Franz Rainer	21
<i>Resources for Suppletion: A Typological Database and a Bibliography</i> , Greville G. Corbett, Dunstan Brown, Marina Chumakina & Andrew Hippisley	35
<i>Morphological Universals and Diachrony</i> , Stephen R. Anderson	45
<i>Universals and Grammatical Categories: A Distributed Morphology Analysis of Spanish Colour Terms</i> , Antonio Fábregas	59
<i>On Diminutive Plurals and Plural Diminutives</i> , Ivan A Derzhanski	73
<i>Roots, Deverbal Nouns and Denominal Verbs</i> , Jan Don	91
<i>Hausa Final Vowel Shortening: Phrasal Allomorphy or Inflectional Category?</i> , Berthold Crysmann	105
<i>Explaining some Structural and Semantic Asymmetries in Morphological Typology</i> , Marian Klamer	127
<i>Prefix-Suffix Neutrality In Evaluative Morphology</i> , Nicola Grandi & Fabio Montermini	143
<i>Word Formation and Typology: Which Language Universals?</i> , Livio Gaeta	157
<i>Rule Counting vs. Rule Ordering: Universal Principles of Rule Interaction in Gender Assignment</i> , Tore Nessel	171
<i>The Morphological Typology of Change of State Event Encoding</i> , Andrew Koontz Garboden & Beth Levin	185
<i>Morphemes and Lexemes versus "Morphemes or Lexemes ?"</i> , François Nemo	195
<i>Typology and Boundaries: The Acquisition of a New Morphological Boundary by Modern Hebrew</i> , Irit Meir	209

<i>Modifier-Head Person Concord</i> , Irina Nikolaeva	221
<i>Unexpressed Features of Verb Agreement in Signed Languages</i> , Christian Rathmann & Gaurav Mathur	235
<i>The Morphosemantics of Transnumeral Nouns</i> , Paolo Acquaviva	251
<i>Loan Words and Declension Classes in Czech</i> , Darya Kavitskaya	267
<i>The Challenge of Typologically Unusual Structures</i> , Alice C. Harris	277

Foreword

The Mediterranean Morphology Meetings (MMM) are organized by a committee of three morphologists: prof. Geert Booij (Vrije Universiteit Amsterdam), prof. Angela Ralli (Univ. of Patras, Greece) and prof. Sergio Scalise (University of Bologna). They do this work together with a local organizer. MMM 4, held in Catania was a very successful meeting, not in the least thanks to the efforts of our colleague and local organizer, prof. Salvatore Sgroi.

The aim of these meetings is to bring together linguists who work on the morphology of (mainly, but not exclusively European) languages in an informal setting which guarantees maximal interaction between researchers, and gives young linguists the chance to present their work at a conference of moderate size where fruitful contacts with senior linguists can be established. Thus, a European network of morphologists has developed.

The first four meetings, in 1997 in Mytilene (Lesvos, Greece), in 1999 in Lija (Malta), in 2001 in Barcelona, and the last one in 2003 in Catania (Sicily) - have proven the success of this formula: the interest in attending these meetings was high, many abstracts were submitted, and the abstracts were selected anonymously which gave young linguists the chance to present their work on the basis of quality, not primarily reputation. In addition, each meeting had a number of invited speakers, leading morphologists of the world.

Each MMM has a specific topic that forms one of the criteria for the selection of abstracts. The topic of the Catania meeting was 'Morphology and linguistic typology'. At first sight, this may look like a very obvious topic since morphological parameters have always played an important role in the classification of languages. We are all acquainted with labels such as 'isolating language' or 'polysynthetic language'. Indeed, morphological typology forms a long-standing and very fruitful research tradition. Yet, there were good reasons to have a fresh look at the relation between morphology and linguistic typology. For many years, debates on morphology focused on theoretical issues, such as its relation to phonology and syntax. There are many different views on the degree of autonomy of morphology, but it is clear by now that morphology is well-established subdiscipline of linguistics. Typological issues have also received new interest, and there is no longer a fruitless separation of typological and theoretical research. Therefore, the MMM committee wanted to put the relation between morphology and typology high on the agenda. Many of the papers in these proceedings show that comparative and typologically informed morphological research is essential for proper morphological analyses of individual languages, and for the development of an empirical adequate theory of morphology.

Morphological Typology and First Language Acquisition: Some Mutual Challenges

Wolfgang U. Dressler

*Kommission für Linguistik, Österreichische Akademie der Wissenschaften
& Institut für Sprachwissenschaft, Universität Wien
wolfgang.dressler@univie.ac.at*

If one believes that external evidence (e.g. from first language acquisition) is relevant for linguistic theory and that acquisition studies should be done in relation to linguistic theory and to theory-guided descriptions of adult input systems, then one finds many problems which are relevant both for morphological and typological theory and for acquisition but which have not been dealt with adequately in connection with both fields, if at all. This paper attempts to raise the level of awareness for such problems and to propose solutions for them.

My acquisition data come from the “Crosslinguistic Project on Pre- and Protomorphology in Language Acquisition” (cf. Dressler 1997, Dziubalska-Kołaczyk 1997, Gillis 1998, Voeikova & Dressler 2002, Bittner, Dressler & Kilani-Schoch 2003). This project studies in more than a dozen languages the acquisition of morphology up to the age of three years and collects, transcribes, codes (in CHILDES format, cf. MacWhinney 2000) and analyses longitudinal corpora in strictly parallel ways. I have to thank all researchers of this project, whose published and prepublished work I am using (and citing) here. The aims of this project are to arrive at universal, typological and language-specific generalisations, in parallel to the grammatical model espoused, which is Natural Morphology (Dressler et al. 1987, Kilani-Schoch 1988, Dressler 2000) with its three subtheories of universal preferences or universal markedness, of typological adequacy and of language-specific system adequacy.

The subtheory of typological adequacy has taken over many ideas from Skalička (1979, cf. 2002), notably the concepts of linguistic types as ideal constructs which natural languages approach to various degrees. Thus these ideal types, despite their largely identical names, are not classes, such as the morphological types of classical morphological typology (cf. Lehmann 1983, Ramat 1995). These ideal types are characterised by (mostly mutually) favouring properties. We have reanalysed many of these properties as mutually and symmetrically or asymmetrically favouring or disfavouring preferences and have based them on the premises of our first subtheory of universal morphological preferences (cf. Dressler 1985a: 227ff, 1985b). Moreover, Skalička’s concept that the inflectional and the word formation component of a language may behave different typologically, must be extended to the subcomponents or submodules of inflectional morphology. Thus noun inflection and verb inflection may have a different typological character within the same language and develop diachronically in typologically different directions (this answers the critiques of Plank 1998 and Wurzel 1996).

Due to the available child-language data of our project as well as of the literature, I will deal with the inflecting-fusional, the agglutinating and the isolating ideal language types of Skalička. The noun and verb inflection systems of the following languages can be ordered gradually in regard to inflectional morphology on the scales of

(a) isolating ↔ inflecting-fusional ideal type, (b) inflecting-fusional ↔ agglutinating ideal type:

- a. Noun inflection: French – Spanish – English – Dutch – Italian – German – Greek – Slavic languages – Lithuanian
- a'. Verb inflection: English – Dutch – German – Spanish – French – Italian – Slavic languages – Greek – Lithuanian
- b. Noun and verb inflection: Lithuanian – Slavic languages – Finnish – Hungarian – Turkish

The main typological criteria that I will use, are: morphological richness (amount of productive morphology), morphological complexity (morphological richness plus unproductive morphology), the universal preference parameters of morphosemantic and morphotactic transparency, constructional iconicity, preferred shape of morphological units, binarity and biuniqueness (as preferred over uniqueness and especially ambiguity).

The typological value of the concepts of agglutinating and inflecting-fusional morphology has been severely criticised by many specialists, such as Anderson (1985: 10), Bauer (1988: 170), Plank (1998), Haspelmath (2000), but cf. Bossong (2001), Plungian (2001). However these critics neither distinguished between morphological class and morphological type of languages nor did they consider the possibility that noun inflection, verb inflection, derivational morphology and compounding may be typologically different within the same language. Thus I claim that typology is more than cross-linguistic comparison and different from research in universals, both in general typology (here in accordance with Seiler's (2000) UNITYP model) and in child-language studies. This is linked to a claim, most vigorously defended by Coseriu (1970) within a structuralist framework, that typology is a basic and not an epiphenomenal level of accounting for linguistic generalisations.

Now, if we want to use evidence from child language as evidence for this claim, then we must construct a bridge-theoretical link (cf. Botha 1979) between the domain of external evidence (in our case: first language acquisition) and the domain of the other theory (in our case: theory of linguistic typology). The main bridge-theoretical hypothesis that makes external evidence from child language relevant for morphological typology is the following: if typology is more than cross-linguistic comparison and different from research in universals and if typology is a basic and not an epiphenomenal level of accounting for linguistic generalisations, then this means within a mentalist framework which takes morphological typology seriously, that typological generalisations of morphology are themselves basic. Since psycholinguistic claims by, e.g., Jakobson (1941) and accounts of empirical work on processing by Burani et al. (2001) postulate that what is acquired early by a child is more basically represented in the adult system of representation and processing than what is acquired later, one may expect that basic typological generalisations should emerge early in first language acquisition. Analogously, *ceteris paribus*, unmarked options should be acquired earlier than their respective marked correspondents, as already postulated by Jakobson (1941), cf. Mayerthaler (1981).

Our developmental approach is constructivist (cf. Maturana & Varela 1979, Karmiloff-Smith 1992, Karpf 1991, Iturrioz Leza 1998): we do not assume that grammatical (sub)modules are genetically inherited but that they are gradually constructed by the children themselves, i.e. they construct a primitive system of grammar. When this global system, by accumulation of acquired patterns, becomes too complex, then it dissociates into modules of syntax and morphology, and later on the latter into submodules of inflection and word formation. This developmental model is integrated with the linguistic model, insofar as children's pattern selection and self-organisation is considered to take the preferences of Natural Morphology into account (cf. Dressler & Karpf 1995).

Comparative acquisition studies in morphology are rare and nearly always simply crosslinguistic, i.e. juxtaposing acquisition studies of each single language without comparing them in the sense of comparative typological linguistics or, if at all, only according to one contrastive variable each time (e.g. word order), with the notable exceptions of Peters (1997) and parts of Slobin's (1985-1997a, 1997b) seminal work. Or they concern very few languages or just one or two special morphological or morphosyntactic areas (e.g. Tsimpli 1996 on optional infinitives and agreement markers in six different languages). Furthermore, the comparison of evolving systems must be conducted along several dimensions which comprise, in addition to chronological age, also lexical age (measured, for example, in terms of timing of lexical acquisition) and morphological age (measured in terms of timing of emergence of morphological patterns and form oppositions). Previous work within the "Crosslinguistic Project on Pre- and Protomorphology in Language Acquisition" has striven to advance from crosslinguistic studies (as in Dressler 1997, Dziubalska-Kořaczyk 1997, Gillis 1998) to truly comparative ones, such as in Kilani-Schoch et al. (1997), Stephany (2002, on number), Voeikova (2002, on case), Bittner et al. (2003, on verbs).

We divide early morphological development into three subsequent phases:

1. Premorphology, a rote-learning phase in which the child's speech production is limited to a restricted number of lexically stored inflectional forms. Extragrammatical morphological operations such as reduplicative onomatopoeics (cf. Dressler et al., in print) and truncations are flourishing. Thus this phase partially and superficially resembles a reduced version of an isolating language. Word classes are hardly differentiated, i.e. words may be polyvalent, as with the Austrian girl's Katharina's onomatopoeics (as analysed by Sabine Laaha):

- | | | | |
|-----|------|------|---|
| (1) | 1;6 | brm | for the noise made by toy cars (or nominal for car) |
| | 1;9 | kra | for crowing (or nominal for birds) |
| | 1;10 | wawa | for barking (or nominal for dog) |
| | 1;11 | ia | for neighing (or nominal for horse) |

A strict distinction between extragrammatical (or expressive) and grammatical (or plain) morphology (cf. Dressler 2000b, Zwicky & Pullum 1987) is relevant for our topic in two ways: First of all, only grammatical morphology appears to play a role in morphological typology. Thus very isolating languages, such as South East Asian languages, may have no inflection and little grammatical word formation, therefore may lack a real morphological module, but may abound in extragrammatical morphology, which is not curbed by a grammatical module. Second, something similar happens in

early language acquisition: many children have much extragrammatical morphology in their premorphological phase. But when they detect and start to construct grammatical morphology in protomorphology, extragrammatical morphological operations decrease dramatically (cf. Bittner et al. 2003, Dressler et al. in print). This is, for example, the case with the extragrammatical phenomenon of fillers (Kilani-Schoch & Dressler 2000: 96).

2. Protomorphology, a phase in which the child starts to generalise over rote-learned forms, thereby detecting the morphological principle of (de)composing form and meaning word-internally. (S)he begins to construct morphology but also to use morphology creatively in coining first analogical formations (cf., e.g., MacWhinney 1978, Dressler & Karpf 1995). This phase partially resembles, at least initially, the reduced version of a weakly agglutinating language. The universal preferences for morphotactic and morphosemantic transparency, for biuniqueness (with reservations below), for constructional iconicity and binary relations are largely followed. However the bases for the typological identity of the respective language are already laid.

3. Morphology proper or modularised morphology, where (according to Dressler & Karpf 1995) the child constructs (sc. non-innate) modules and submodules and acquires a qualitatively adult-like morphology which already possesses all of its basic typological properties.

The first typological differences emerge already in premorphology: in reaction to, and in accordance with, the maternal or other adult input, the child selects and stores morphological patterns of high token frequency and which occur in the basic syntactic patterns that the child has taken up from the input. These patterns largely consist of unmarked forms, such as nominative singulars of singular-dominant nouns, plurals of plural-dominant nouns (e.g. G. *Eier* ‘eggs’), infinitives, singular imperative, first or third singular present forms (particularly of atelic verbs). This has typological implications, for example whether these forms are zero-base forms (Peters 1997: 179f, cf. also Croft 2003: 162ff), such as 3.Sg.Pres. in Turkish, Hungarian, Finnish, Lithuanian, Polish, Croatian, Spanish, French and, partially, Italian, or not, as in German, Russian, Greek. If, however, the 3.Sg.Pres. form is affixed, whereas the first person is a zero-base form, as in English and Dutch, then the 3.Sg.Pres. emerges later (cf. Bittner et al. 2003). This is a crosslinguistic manifestation of the above-mentioned principle “unmarked before marked”.

There is a second typological option in the 3.Sg.Pres. as base form, namely whether it has a thematic vowel or other stem indicator (a property of the inflecting type), as in Lithuanian, Polish, Croatian, Spanish, Italian, or not, as in agglutinating Turkish, Hungarian, Finnish and in more isolating French (cf. Dressler, Kilani-Schoch, Spina & Thornton 2003, Dressler & Kilani-Schoch 2004). In the second case, the child can start and often does focus on an uninflected form, in the first case all first forms of verbs are in some way inflected (cf. Kilani-Schoch 2003: 288).

What is typologically most important, is the degree of morphological richness (not complexity!) of a language, i.e. of productive morphology. In morphologically rich languages morphology fulfils more functions, already visible in more form-meaning mappings (cf. Slobin 1973, 1985b, 2001) and hence is more “informative” (Wijnen et

al. 2001) than in morphologically poorer languages. This is most obvious in Turkish, where the role of morphology is much greater, and correspondingly the role of syntax smaller, than in inflecting-fusional languages and particularly in weakly inflecting-fusional languages which share properties of the isolating type. Children become aware of the respective role of morphology in the language they are acquiring, i.e. they are more “tuned” to morphology if they are acquiring a morphology-rich language. Thus we can expect (cf. Slobin 1985b) that such children should detect morphology earlier than children acquiring morphologically poorer languages.

But how can we identify detection of morphology by children (cf. Dressler, Kilani-Schoch & Klampfer 2003)? For this purpose Kilani-Schoch & Dressler (2002) have elaborated the concept of the emergence of miniparadigms, i.e. incomplete paradigms (from the adult perspective). For example, the first miniparadigm produced by the Viennese boy Jan at 1;10 (Klampfer 2003: 314) is:

- (2) Inf. machen, 3.Sg.Pres.Ind. macht, PPP gemacht ‘to make’

Whenever we find three lemmas of the same word class of which three morphotactically and morphosemantically clearly distinct paradigm members have emerged and recurred in spontaneous production in various contexts, then we can safely assume that such a child has enough pattern variety in its uptake in order to detect the morphological principle of (de)composing form and meaning word-internally. This principle then appears soon to be extended from bound morphology to the morphology of (generally) monomorphemic function words, i.e. bound morphology, especially productive (bound) morphology tends to develop faster than free morphemes (function words, cf. Dressler, Kilani-Schoch & Klampfer 2003, Peters 1997: 180). Thus we hypothesise that the time point of the emergence of form oppositions is determined by the degree of morphological richness of the respective target language.

In support of this hypothesis, for verb inflection the miniparadigm criterion has been observed to be fulfilled for Turkish at 1;7 (Aksu-Koç & Ketrez 2003; first verb oppositions even at 1;5), for English after 2;5 (Gülzow 2003, cf. de Villiers & de Villiers 1985), cf. for the early emergence of Turkish morphology in general Aksu-Koç & Slobin (1985), Stephany (2002), Voeikova (2002). At first sight this result may seem paradoxical, because it should be much easier to acquire the very poor inflectional systems of English than the very rich inflectional systems of Turkish. What appears to be much more important for the child than superficial simplicity, is the much greater usefulness of acquiring inflectional morphology in Turkish than in English, plus the great difference in orderly variation available in the respective inputs.

If we now compare agglutinating with strong and weak inflecting-fusional languages, we must keep in mind that for each language the miniparadigm criterion has been investigated only for very few children (Bittner et al. 2003) and that no such gross differences have been found as between Turkish and English (cf. also Stephany 2002 for the emergence of nominal number). Still it is compatible with our hypothesis that the miniparadigm criterion has been fulfilled at the same age as with Turkish only with one Finnish child (at 1;8 with the other), i.e. solely for the other strongly agglutinating language of the language sample in Bittner et al. (2003). These two agglutinating languages are rather closely followed (at 1;10) by Lithuanian nouns and verbs, Croatian and Spanish verbs (rather strongly inflecting systems), and for verbs by one Italian boy (Dressler, Tonelli et al. 2003), by one French-learning girl (the other at 2;1), followed

by Greek (1;11). Yucatec Maya, Russian, Italian (except the afore-mentioned boy), German and Dutch come later.

These results presuppose, that we distinguish, as proposed above, within each language, different morphological systems. For example, in French, the noun system is of a very isolating type, the verb system much less. Thus it is French verbs where children first must detect morphology, whereas in German it is noun inflection and noun-compounding as well (cf. Dressler, Kilani-Schoch & Klampfer 2003), since much more different patterns in noun morphology are productive and show inflecting-fusional characteristics than in verb morphology. For example, the Austrian boy Jan produces at the onset of protomorphology (1;8) first oppositions between compounds and their members:

- (3) *Feuer(wehr)auto* ‘fire(brigade)-car’ and simplicia *Auto* ‘car’, *Feuer* ‘fire’, compound *Doppeldeckerbus* ‘double-decker-bus’ and its member *Doppeldecker* ‘double-decker’, compound *Segelschiff* ‘sailing-boat’ and simplex *Schiff* ‘boat’

At 1;9, a first example of analogy appears:

- (4) **Laster+wagen* ← *Laster* = *Last+wagen* ‘truck’,

evidence for the child’s creative use of compound formation. The recurrence of nouns within compounds and as autonomous words must have induced him to identify the basics of compounding. Among all languages of our project, only German compounding appears rich enough for stimulating children to use them productively at an early age. This represents further evidence for Skalička’s (1979) view that different subcomponents of morphology may approach different ideal types.

Such a morphological difference, as between German and French, appears to have even repercussions for the relative timing or preponderance of the emergence of word classes. The acquisition of many languages shows a “noun bias”, i.e. nouns emerge earlier and first expand faster than verbs. This is explained as being due to the greater cognitive ease of identifying reference to (especially concrete) objects than to actions or states (Gentner 1982, Gentner & Boroditsky 2001). And if in Mayan languages or Korean, etc. verbs emerge earlier, then this is explained by semantic and pragmatic (incl. cultural) properties underlying greater “verb friendliness”, plus by prosodic factors (De León 1998: 154ff, cf. the summary in Pfeiler 2002). Also syntactically nouns are easier to grasp, as they are, on the average, much less relational than verbs. But words and word classes are not only defined pragmatically, semantically and syntactically, but also morphologically (cf. Dixon & Aikhenvald 2002). And since, in French, nouns have a largely isolating morphology, but verbs a fair amount of properties of the inflecting-fusional type, this can be assumed to curb the noun bias in the acquisition of French (cf. Kilani-Schoch 2003: 288). Similarly in incorporating languages where verb morphology is central, verb morphology develops earlier than noun morphology (cf. Slobin 1992: 9, Fortescue & Olsen 1992). In accordance with this claim nouns have a priority in both lexical and, later, inflectional development in German (Dressler, Kilani-Schoch & Klampfer 2003), Lithuanian (Savickiene 2003, Wójcik 2003), Italian (Noccetti 2002), Yucatec Maya (Pfeiler 2002) among the languages of our project.

The inflectional-fusional type differs from the agglutinating type in having a complex hierarchical branching system of inflection classes (cf. Dressler 2003), whereas the ideal agglutinating type has none. Thus (cf. Pöchtrager et al. 1998) nearly all Turkish nouns and verbs inflect each according to a single type, Hungarian has few and hierarchically rather shallow class differences, Finnish already more, whereas Estonian is also in this respect rather an inflecting-fusional language, similar to Italian verb inflection. As a consequence, in Turkish, diminutives inflect in the same way as any other common noun, whereas in all the other diminutive-rich languages (derivationally) productive diminutives belong to the productive inflectional classes. For example, It. *tribù* ‘tribe’ is indeclinable, *poeta* ‘poet’, *amico* ‘friend’, and *pelle* ‘skin’ belong to unproductive classes, whereas their diminutives *tribu-ina*, *poet-ino*, *amich-etto*, *pell-icina* belong to productive classes. Since, *ceteris paribus*, productive patterns have a higher chance to be taken up by children than unproductive ones (cf. Dressler et al. 1996, cf. Smoczyńska 1985: 624ff, Peters 1997: 180f), diminutives of simplicia belonging to unproductive classes emerge earlier than their simplex bases and thus diminutives appear to help children to acquire inflection (cf. Gillis 1998, Savickiene 2003). This effect does not exist in Turkish.

In inflecting-fusional languages diminutives tend to belong to classes which are morphotactically very transparent, cf. It. *ami[k]o*, Pl. *ami[č]-i*, *uomo*, Pl. *uom-ini* ‘man’, but the respective diminutives have more transparent plurals: *amichetto*, *amichetti*; *omicino*, *omicini*. Again morphotactic transparency, similar to productivity, is known to facilitate early acquisition (cf. Slobin 1985b: 1216, Peters 1997: 181, Savickiene 2003, Aksu-Koç & Slobin 1985: 847).

A result of these intralinguistic and crosslinguistic differences in morphotactic transparency is that children acquiring inflecting-fusional and introflecting languages, after having detected morphology, tend to overgeneralise productive, and later even unproductive but more transparent patterns, such as Fr. *prendre* ‘take’, PP *pris* → *prend-u*, after *rendre*, *rendu* (cf. Kilani-Schoch 2003), a procedure which is scarcely possible in an agglutinating language.

Another instance where a specific morphological category helps to develop morphology is prefixation. Lithuanian is very rich in productive prefixation in the verbal system, and prefixation interacts with placement of clitic reflexives, as in:

- (5) kel-ti-s vs. at- si- kel-ti
 V-Inf.-Refl. Prefix-refl.-V-Inf.
 ‘to get up’ (imperfective vs. perfective)

Thus, on the one hand, reflexivity, which is a pervasive category for expressing passive voice in Lithuanian, can only be successfully handled in reference to prefixes. On the other hand, the insertion of the reflexive between prefix and verb root, renders prefixes easier to identify and to segment. And, indeed, Lithuanian prefixes emerge relatively early in protomorphology (Wójcik 2000, 2003), in contrast to the other languages of our project. The only, and superficial, exceptions are German and Dutch whose separable prefixes (or verb particles) are even easier to identify and to segment, e.g. German:

- (6) aúf-dreh-en vs. aúf-ge- dreh-t vs. ich dréh-e aúf
 P- V -Inf P- PP prefix-V-PP suffix I V-1.Sg. P
 ‘to turn on’ ‘turned on’ ‘I’m turning on’

Here main stress on the prefix/particle and change of position in finite forms render it even more salient. As a consequence such verb particles are with many German- and Dutch-learning children the first “verb forms” to emerge, as in German:

- (7) ab! ‘off!’ = adult: mach ab! ‘make off, separate it!’, cf. Inf. ab-mach-en (Dressler, Kilani-Schoch & Klampfer 2003, cf. Bennis et al. 1995)

A typological variable is also relevant for the acquisitional phenomenon of inflectional imperialism, identified by Slobin (1985b: 1216, cf. MacWhinney 1985) as the total or nearly total substitution of competing morphological patterns by a single one of them. This phenomenon, however, occurs less often in acquisition than postulated by Slobin. And this gives further evidence against the overestimation of the importance of the default concept by, e.g., Pinker (1984) and relegates it to a typological variable: it is irrelevant in a very agglutinating language such as Turkish, but is important in weakly inflecting-fusional and introflecting languages as well as in languages which approach both the ideal agglutinating and the inflecting-fusional type only to some extent. In strongly inflecting-fusional languages there is often no default or only a weak default among competing morphological patterns (cf. Dressler 1999), e.g. among verb macroclasses in Slavic languages or Lithuanian. Thus in the acquisition of Lithuanian, no instance of inflectional imperialism has been found for noun inflection (Savickiene 2003), whereas one instance has been found in verb inflection at the age of 1;8 (Wójcik 2003: 414; 2000: 111f):

- | | | | | |
|-----|---------------------------|-----------------------------|---|-------|
| (8) | Inf. sed-e-ti
‘to sit’ | 3.Sg.Pres. sed-i
‘sit-s’ | → | sed-a |
| | nor-e-ti
‘to want’ | nor-i
‘want-s’ | → | noj-a |

The macroclass with the thematic vowel /a/ is the richest and most frequent of the three Lithuanian macroclasses. Moreover small children, probably due to phonological reasons, seem to prefer the thematic vowel [a] to other thematic vowels, as we have found also in Slovene and Polish children.

In general, the higher amount of productivity, constructional iconicity and transparency favours acquisition of agglutinating morphology. Note the lack of infixes (which make the lexical root discontinuous and thus less transparent) in the agglutinating type. Note, for example, the lack of infixes in the agglutinating type: infixes make the lexical root discontinuous and thus less transparent. As to acquisition, infixes appear to be acquired late. They first lack in Tzeltal child speech (Brown 1998: 141), for Lithuanian cf. Wójcik (2000: 115). Note also the early reduction of the equally discontinuous and thus opacifying German past participle *ge-...-t/n* via loss of the prefix part (cf. Klampfer 2003: 306), as in:

- (9) runter-(ge)-fall-en ‘fallen down’, um-(ge)-d(r)eh-t ‘turned over’

Another favouring property of the agglutinating type is the greater preference for biuniqueness over uniqueness and ambiguity. Compare the following fragment of declension of the word ‘room’ in Turkish and Russian:

	Nominative		Genitive		Locative	
	Turk.	Russ.	Turk.	Russ.	Turk.	Russ.
Sg.	oda	komnat-a	oda-nIn	komnat-y	oda-da	v komnat-e
Pl.	oda-lar	komnat-y	oda-lar-In	komnat	oda-lar-da	v komnat-ax

Here, Turkish only uses affixation (highest degree of constructional iconicity), whereas Russian has no affix in the genitive plural of this feminine noun, with the effect that, in a non-iconic way, the morphosemantically marked plural form is shorter than the unmarked singular form. This results in children substituting the non-iconic zero form with the genitive plural suffix *-ov* of masculine nouns (cf. Smoczyńska 1985: 627f), thus obtaining an iconic, affixed form which is, in a non-iconic way, even longer than the respective singular form. Moreover agglutinating languages have no gender, a category which contributes to morphosemantic and morphotactic opacity in inflecting-fusional languages. Finally Turkish expresses Plural, Genitive and Locative in a biunique way, whereas Russian, in a non-unique way, expresses number and case cumulatively and has allomorphs for each number-case form.

Biuniqueness corresponds to the acquisitional principles of contrast in lexical acquisition (cf. Clark 1993) and Slobin's (1985b: 1227f) unifunctionality operating principle. In the acquisition of morphology, this principle can be easily followed in agglutinating morphology and when a default can be overgeneralised via inflectional imperialism (see above). But when acquiring a weakly inflecting language with many homophonies and syncretisms, as they exist, e.g., in the verbal systems of French and German, trying to introduce biuniqueness would require enormous efforts. Thus children at first settle for the second-best solution and rather exploit homophony and syncretism (cf. Kilani-Schoch & Dressler 2000: 102, 107), i.e. they focus on such widely usable homophonous forms. In other words, again, children adapt very early to typological characteristics of the language they acquire.

The early exploitation of homophony and syncretism may explain "why infinitives emerge earlier when they are homophonous with other verb forms" (Bittner, Dressler & Kilani-Schoch 2003: xviii), as in English, German (infinitive *-en* = 1. & 3.Pl. = suffix of strong past participles) and French (only productive inf. /e/ = only productive past participle = 2.Pl.). And this, in turn, (better than syntactic accounts), may explain the presence of bare infinitives in the child productions of these languages as opposed to their absence or near-absence in other languages (cf. Phillips 1995, Wijnen et al. 2001, Kilani-Schoch 2003: 289). As to typology, this supports the importance of syncretism and certain types of homophony as an instance of paradigm economy (cf. Carstairs 1987) in weakly inflecting-fusional morphologies.

Now in Italian verb inflection biuniqueness plays a bigger role than in strongly inflecting-fusional languages, insofar as in each tense at least the 1st and 2nd persons singular and plural are expressed by a superstable marker in the indicative. Thus it makes sense for children to pursue biuniqueness and to overextend it also to the 3rd person indicative, as in the early forms *ved-a* for *ved-e* 'sees'. Much more common are 2nd person singular imperatives of the first macroclass in *-a*, which are substituted by the form of the second macroclass in *-i*, which is homophonous with the superstable marker of the 2nd person singular indicative, e.g. in the protomorphological phase of several children (Berretta 1993: 163, Noccetti 2003: 368):

(10)

Child:	tir-i!	legh-i!	lav-i!	mang-i!	telefon-i!	guard-i!
Adult:	tir-a!	leg-a!	lav-a!	mangi-a!	telefon-a!	guard-a!
	‘draw!’	‘bind!’	‘wash!’	‘eat!’	‘phone!’	‘look!’

The agglutinating type presents two difficulties for the acquisition of morphology: (i) the occurrence of long sequences of suffixes within the same polysynthetic word form (cf. Pöchtrager et al. 1998: 60), (ii) the variable order of certain morphemes (with only pragmatic meaning differences), as described for Turkish, Mari and Quechua (cf. Sebüktekin 1974) in long polysynthetic word forms. In contrast, the inflecting-fusional and the introflecting type are devoid of these acquisitional difficulties because of the preference for having just one affix (with cumulative meaning, see above) per inflectional form, an instantiation of the universal preference for binary relations. Now, when starting to acquire morphology, even Turkish and Hungarian children greatly prefer having just one suffix per word form (cf. Peters 1997: 180f, and Fortescue & Lennert Olsen 1992: 141ff & Slobin 1992 for Greenlandic, Iturrioz Leza 1998: 108ff and Gómez López 1998: 187 for Huichol) and thus follow this universal preference for binary relations. In this preference they are supported by adults in early child-directed speech, which also appears to prefer short word forms, especially in motherese (cf. Iturrioz Leza 1997, 1998: 10ff, 51, 110ff, De León 1998: 159). Thus the two difficulties of the agglutinating type may become relevant much later in acquisition, but this has not yet been investigated.

The agglutinating type presents another property which facilitates acquisition: in an iconic and transparent way, morphology is word-based, i.e. typically declension is built up on a zero form of the morphosemantically unmarked elementary form of the Nom.Sg., conjugation analogously on the 3.Sg.Pres. form. In contrast, the inflecting-fusional and the introflecting type prefer in a morphology-pervasive way stem-based and root-based morphology. Although this diminishes iconicity and morphotactic transparency, children cannot avoid acquiring stem- and root-based paradigms. This starts, for inflecting-fusional languages, already in the protomorphological phase. Thus these basic and all-pervasive typological properties emerge early in language acquisition.

A noteworthy result of our project has been the finding (Voeikova 2002, cf. Stephany 2002) that case distinctions appear to emerge in agglutinating languages before number distinctions, whereas in inflecting languages case distinctions emerge after number distinctions (exception: Lithuanian, s. Savickiene 2003). Number is a more basic category than case. One can propose even an implication: if a language has case distinctions, it also has number distinctions, but not vice versa. Thus we can expect number to emerge earlier in acquisition than case. Now why is there the reverse order in the acquisition of, at least, Turkish, Finnish and Hungarian? Note that plural and case are marked separately in these languages and case after plural, e.g. Turkish:

(11) N.sg. ev ‘house’, N.Pl. ev-ler, Loc.Sg. ev-de, Loc.Pl. ev-ler-de

Thus, in case of a plural oblique case form it is easier for the child to strip off the case suffix than the plural suffix. Compare. also the well-known recency effect, which makes ends of words easier to segment than beginnings (cf. Slobin 1973: 191f, Peters 1997:

181f) and is the main reason for the suffixing preference (cf. Hall 1992). Also in our project languages suffixes emerge earlier than prefixes, except in the above-mentioned cases of separable prefixes.

There is a close connection between morphological and syntactic properties in linguistic typology, and consequently Skalička's ideal types are characterised both by morphological and syntactic properties. Such close connections, provided that they are viewed as pertaining to a basic typological level of language, present a problem to a nativist modular approach whereby morphology and syntax are identified as different innate modules, because interaction between different modules is limited to their superficial outputs but banned from their basic design. However, since we assume, and have found evidence for, that basic typological properties are acquired in the protomorphological phase, i.e. before the modules of morphology and syntax are dissociated, this problem does not exist for our constructivist approach. In protomorphology the emerging but not yet modularised components of morphology and syntax can freely interact. In this way data from first language acquisition can support the approach to typology as representing a basic level of language.

References

- Aksu-Koç, A. & F.N. Ketrez (2003), *Early verbal morphology in Turkish: Emergence of inflection*, in Bittner, D. et al. 27–52.
- Aksu-Koç, A. & D.I. Slobin (1985), *The acquisition of Turkish*, in D.I. Slobin (1985-1997), (ed.) vol I, 839–878.
- Anderson, S.R. (1985), *Typological distinctions in word formation*, in T. Shopen (ed.) *Language Typology and Syntactic Description*, III, Cambridge, Cambridge University Press, 3–56.
- Bauer, L. (1988), *Introducing Linguistic Morphology*, Edinburgh, Edinburgh University Press.
- Bennis, H., M. den Dikken, P. Jordens, S. Powers & J. Weissenborn (1995), *Picking up particles*, in *Proceedings of the 19th Annual Boston Conference on Language Development* 1, 70–81.
- Berretta, M. (1993), *L'imperativo italiano: problemi morfologici in L1 and L2*, "Quaderni del Dipartimento Linguistico di Bergamo", 9, 157–185.
- Bittner, D., W.U. Dressler & M. Kilani-Schoch (2003) (eds.), *Development of Verb Inflection in First Language Acquisition: A Cross-Linguistic Perspective*, Berlin, Mouton de Gruyter.
- Bossong, G. (2001), *Ausdrucksmöglichkeiten für grammatische Relationen*, in Haspelmath et al. (2001) 657–668.
- Botha, R. (1979), *External evidence in the validation of mentalistic theories*, "Stellenbosch Papers in Linguistics", 2, 1–38.
- Brown, P. (1998), *La identificación de las raíces verbales en tzeltal (maya)*, "Función", 18, 121–146.
- Burani, C., L. Barca & L.S. Arduino (2001), *Una base di dati sui valori di età di acquisizione, frequenza, familiarità, immaginabilità, concretezza, e altre variabili lessicali e sublessicali per 626 nomi dell'italiano*, "Giornale Italiano di Psicologia", 28, 839–854.
- Carstairs, A. (1987), *Allomorphy in Inflexion*, London, Croom Helm.
- Clark, E. (1993), *The Lexicon in Acquisition*, Cambridge, Cambridge University Press.
- Coseriu, E. (1970), *Synchronie, Diachronie und Typologie*, in E. Coseriu (1970), *Sprache, Struktur und Funktionen*, Tübingen, Narr, 235–266.
- Croft, W. (2003), *Typology and Universals*, Cambridge, Cambridge University Press.

- De León, L. (1998), *Raíces verbales tempranas en maya tzotzil: factores del input vs. restricciones cognoscitivas*, "Función", 18, 147–174.
- De Villiers, J.G. & P. de Villiers (1985), *The acquisition of English*, in D.I. Slobin (1985-1997), (ed.) vol I, 27–139.
- Dixon, R.M.W. & A.Y. Aikhenvald (2002) (eds.), *Word: a cross-linguistic typology*, Cambridge, Cambridge University Press.
- Dressler, W.U. (1985a), *Morphonology*, Ann Arbor, Karoma Press.
- Dressler, W.U. (1985b), *Typological aspects of Natural Morphology*, "Acta linguistica Hungarica", 35, 51–70.
- Dressler, W.U. (1997)(ed.), *Studies in Pre- and Protomorphology*, Wien, Verlag der Österreichischen Akademie der Wissenschaften.
- Dressler, W.U. (1999), *Why collapse morphological concepts?*, "Behavioural and Brain Sciences", 22, 1021.
- Dressler, W.U. (2000a), *Naturalness*, in G. Booij et al. (eds.), *Morphologie/Morphology I*, Berlin, de Gruyter. 288–296.
- Dressler, W.U. (2000b), *Extragrammatical vs. marginal morphology*, in U. Doleschal & A.M. Thornton (eds.), *Extragrammatical and Marginal Morphology*, München, Lincom. 1–10.
- Dressler, W.U. (2003), *Latin inflection classes*, in C. Kroon et al. (eds.), *Theory and Description in Latin Linguistics*, Amsterdam, Gieben. 93–110.
- Dressler, W.U., R. & D. Drążyk, K. Dziubalska-Kołaczyk & E.Jagła (1996), *On the earliest stages of acquisition of Polish declension*, in Ch. Koster & F. Wijnen (eds.) *Proceedings of the Groningen Assembly on Language Acquisition*, Groningen, Centre for Language and Cognition. 185–195.
- Dressler, W.U., K. Dziubalska-Kołaczyk, N. Gagarina & M. Kilani-Schoch (in print), *Reduplication in child language*, in B. Hurch (ed.) *Reduplication*, Berlin, de Gruyter.
- Dressler, W. & A. Karpf (1995), *The theoretical relevance of pre- and protomorphology in language acquisition*, "Yearbook of Morphology" 1994, 99–122.
- Dressler, W.U. & M. Kilani-Schoch (2004), *Iconicité dans la flexion du verbe français*, in G. Fenk (ed.) *Gedenkschrift W. Mayerthaler*.
- Dressler, W.U., M. Kilani-Schoch & S. Klampfer (2003), *How does a child detect bound morphology?*, in R.H. Baayen & R. Schreuder (eds.) *Morphological Structure in Language Processing*, Berlin, Mouton de Gruyter. 391–425.
- Dressler, W.U. & M. Kilani-Schoch, R. Spina & A.M. Thornton (2003), *Le classi di coniugazione in italiano e francese*, in M. Giacomo-Marcellesi & A. Rocchetti (eds.) *Il verbo Italiano, Atti 35° congresso SLI*, Roma, Bulzoni. 397–416.
- Dressler, W.U., W. Mayerthaler, O. Panagl & W.U. Wurzel (1987), *Leitmotifs in Natural Morphology*, Amsterdam, Benjamins.
- Dressler, W.U., L. Tonelli, S. Klampfer, M. Fabris & M. Magris (2003), *L'acquisizione della morfologia verbale del tedesco e dell'italiano: un progetto contrastivo*, in R. Bombi & F. Fusco (eds.) *Sguardi reciproci: Parallela 10*, 247–268. Udine, Forum.
- Dziubalska-Kołaczyk, K. (1997)(ed.) *Pre- and Protomorphology in Language Acquisition*, "Papers and Studies in Contrastive Linguistics", 33.
- Fortescue, M. & L.L. Olsen (1992), *The acquisition of West Greenlandic*, in D.I. Slobin (1985-1997)(ed.) vol III. 115–220.
- Gentner, D. (1982), *Why nouns are learned before verbs: linguistic relativity versus natural partitioning*, in S.A. Kuczaj (ed.) *Language Development*, 2, Hillsdale, Erlbaum. 301–334.
- Gentner, D. & L. Boroditsky (2001), *Individuation, relativity, and early word learning*, in M. Bowerman & S.C. Levinson (eds.) *Language Acquisition and Conceptual Development*, Cambridge, Cambridge University Press. 215–256.
- Gillis, S. (1998)(ed.), *Studies in the Acquisition of Number and Diminutive Marking*, "Antwerp Studies in Linguistics", 95.

Morphological Typology and First Language Acquisition: Some Mutual Challenges

- Gülzow, I. (2003), *Early development of verbal morphology in an English-speaking child*, in D. Bittner et al. 205–238.
- Hall, C.J. (1992), *Morphology and Mind*, London, Routledge.
- Haspelmath, M. (2000), *The agglutination hypothesis: a belated empirical investigation*, paper given at the Vienna International Morphology Meeting.
- Haspelmath, M., E. König, W. Österreicher & W. Raible (2001)(eds.), *Language Typology and Language Universals*, Berlin, de Gruyter.
- Iturrioz Leza, J.L. (1998), *Acoplamiento estructural y adquisición del huichol*, “Función”, 17.
- Jakobson, R. (1941), *Kindersprache, Aphasie und Allgemeine Lautgesetze*, Stockholm, Almqvist & Wiksell.
- Karmiloff-Smith, A. (1992), *Beyond Modularity: a developmental perspective on cognitive science*, Cambridge (Mass.), MIT Press.
- Karpf, A. (1991), *Universal Grammar needs organization*, “Folia Linguistica”, 25, 339–360.
- Kilani-Schoch, M. (1988), *Introduction à la morphologie naturelle*, Bern, Lang.
- Kilani-Schoch, M. (2003), *Early verb inflection in French*, in D. Bittner et al. 269–295.
- Kilani-Schoch, M. et al. (1997), *On the demarcation of phases in early morphology acquisition in four languages*, in Dziubalska-Kołaczyk (ed.), 15–32.
- Kilani-Schoch, M. & W.U. Dressler (2000), *Are fillers as precursors of morphemes relevant for morphological theory?* in W. Dressler et al. (eds.), *Morphological Analysis in Comparison*, Amsterdam, Benjamins. 89–111.
- Kilani-Schoch, M. & W.U. Dressler (2002), *The emergence of inflectional paradigms in two French corpora: an illustration of general problems of pre- and protomorphology*, in M. Voeikova & W.U. Dressler (eds.), 45–59.
- Klampfer, S. (2003), *Emergence of verb paradigms in one Austrian child*, in D. Bittner et al. (eds.), 297–321.
- Lehmann, C. (1983), *The present state of linguistic typology*, *Proceedings of the International Congress of Linguists* 13, 950–956.
- MacWhinney, B. (1978), *The acquisition of morphophonology*, Chicago, University of Chicago Press.
- MacWhinney, B. (1985), *Hungarian language acquisition as an exemplification of a general model of grammatical development*, in D.I. Slobin (1985-1997)(ed.) vol II, 1069–1155.
- MacWhinney, B. (2000), *The CHILDES Project*, Mahwah, Erlbaum.
- Maturana, H.R. & F.R. Varela (1979), *Autopoiesis and Cognition*, Boston, Reidel.
- Mayerthaler, W. (1981), *Morphologische Natürlichkeit*, Wiesbaden, Athenäum.
- Nocchetti, S. (2002), *Pre- and Protomorphology in Language Acquisition: an Italian case study*, Pisa, Edizioni Plus.
- Nocchetti, S. (2003), *Acquisition of verb morphology in Italian: a case study*, in D. Bittner et al. (eds.) 350–378.
- Peters, A.M. (1997), *Language typology, prosody, and the acquisition of grammatical morphemes*, in D.I. Slobin (1985-1997)(ed.) vol V. 135–197.
- Pfeiler, B. (2002), *Noun and verb acquisition in Yucatec Maya*, in M. Voeikova & W.U. Dressler (eds.), 75–81.
- Phillips, C. (1995), *Syntax at age 2: Crosslinguistic differences*, “MIT Working Papers in Linguistics”, 26, 325–382.
- Pinker, S. (1984), *Language Learnability and Language Development*, Cambridge, Harvard University Press.
- Plank, F. (1998), *The co-variation of phonology with morphology and syntax*, “Linguistic Typology”, 2, 195–230.
- Plungian, V.A. (2001), *Agglutination and flection*, in Haspelmath et al. 669–678.
- Pöchtrager, M.A., Cs. Bodó, W.U. Dressler & T. Schweiger (1998), *On some inflectional properties of the agglutinating type illustrated from Finnish, Hungarian and Turkish inflection*, “Wiener linguistische Gazette”, 62/63, 57–92.

- Ramat, P. (1995), *Typological comparison: towards a historical perspective*, in M. Shibatani & T. Bynon (eds.) *Approaches to Linguistic Typology*, Oxford, Clarendon Press. 27–48.
- Savickiene, I. (2003), *The Acquisition of Lithuanian Noun Morphology*, Wien, Verlag der Österreichischen Akademie der Wissenschaften.
- Sebüktekin, H. (1974), *Morphotactics of Turkish verb suffixation*, “Boğaziçi University Journal”, 2, 87–116.
- Seiler, H. (2000), *Language Universals Research: a synthesis*, Tübingen, Narr.
- Skalička, V. (1979), *Typologische Studien*, Braunschweig, Vieweg.
- Skalička, V. (2002), *Die Typologie des Ungarischen*, “Travaux du cercle linguistique de Prague” 4, 101–108.
- Slobin, D.I. (1973), *Cognitive prerequisites for the development of grammar*, in C. Ferguson & D. Slobin (eds.), *Studies of Child Language Development*, New York, Holt, Rinehart and Winston. 175–208.
- Slobin, D.I. (1985-1997)(ed.) *The crosslinguistic study of language acquisition*, Hillsdale, Erlbaum. 5 volumes.
- Slobin, D.I. (1985b), *Crosslinguistic evidence for the language-making capacity*, in D.I. Slobin (1985-1997)(ed.) vol II, 1157–1256.
- Slobin, D.I. (1992), *The crosslinguistic endeavor*, in D.I. Slobin (1985-1997)(ed.) vol III, 1–13.
- Slobin, D.I. (1997b), *The universal, the typological, and the particular in acquisition*, in D.I. Slobin (1985-1997)(ed.) vol V, 1–39.
- Smoczyńska, M. (1985), *The acquisition of Polish*, in Slobin I. 595–686.
- Stephany, U. (1997), *The acquisition of Greek*, in Slobin IV. 183–333.
- Stephany, U. (2002), *Early development of grammatical number – a typological perspective*, in M. Voeikova & W.U. Dressler (eds.) 7–23.
- Tsimpli, I.-M. (1996), *The prefunctional stage of first language acquisition: a crosslinguistic study*, New York, Garland.
- Voeikova, M. (2002), *The acquisition of case in typologically different languages*, in M. Voeikova & W.U. Dressler (eds.), 25–44.
- Voeikova, M.D. & W.U. Dressler (2002)(eds.), *Pre- and Protomorphology: early phases of morphological development in nouns and verbs*. München: Lincom.
- Wijnen, F., M. Kempen & S. Gillis (2001), *Root infinitives in Dutch early child language: an effect of input*, “Journal of Child Language”, 28, 629–660.
- Wójcik, P. (2000), *The Acquisition of Lithuanian Verb Morphology: a case study*, Warsaw, Universitas.
- Wójcik, P. (2003), *Early verb inflection in Lithuanian*, in D. Bittner et al. (eds.), 401–420.
- Wurzel, W.U. (1996), *Morphologischer Strukturwandel: Typologische Entwicklungen im Deutsch*, in E. Lang & G. Zifonun (eds.), *Deutsch – typologisch*, Berlin, de Gruyter. 492–524.
- Zwicky, A.M. & G.K. Pullum (1987), *Plain morphology and expressive morphology*, “Berkeley Linguistic Society Papers”, 13, 339–342.

Typology, Diachrony, and Universals of Semantic Change in Word-Formation: A Romanist's Look at the Polysemy of Agent Nouns

Franz Rainer

Vienna University of Economics and Business Administration

franz.rainer@wu.wien.ac.at

1. Introduction

It may be considered as part of the common body of knowledge of students of word-formation that agentive suffixes tend to have at the same time instrumental and, to a lesser extent, locative and other meanings. By the end of the nineteenth century, for example, Meyer-Lübke (1890) explained this polysemy as a consequence of the metaphorical use of agent nouns as designations of instruments (§ 498) and pointed out the conceptual ambiguity of containers between instrumental and locative nouns (§ 497). Similar observations on the polysemy of agent nouns can be found over and over again in the literature, but we have to wait until the 1970s in order to see appear the first studies dedicated specifically to this putative language-universal. In those times, Harald Haarmann and Oswald Panagl independently published several articles on the topic, presented as preludes to in-depth typological studies that they had the intention to undertake, intentions, unfortunately, never realised. But due to Panagl's pioneering study – Haarmann's articles, as far as I can see, have gone totally unnoticed – the subject had been effectively placed on the agenda of students of word-formation, sparking off a considerable amount of contributions up to the present day.

It is my intention here to review this by now conspicuous literature, to single out the main hypotheses and to assess their validity, especially on the background of the Romance languages. We will thus be concerned, on the one hand, with empirical issues, but on the other our discussion will always be guided, in accordance with the general theme of the Catania meeting, by the question of what typological research may contribute to our understanding of word-formation, and what methodology it should (not) adopt. The order of presentation will be, by and large, chronological, which allows us to draw, at the same time, a genre picture of research styles and habits in this area of linguistics.

2. Delimiting the Object of Study: Haarmann (1975)

Haarmann's study is presented as part of a larger project aiming at describing the "polyfunctionality" of certain suffixes which may refer at the same time to living beings (*Lebewesen*) and to material objects (*Sachobjekte*). His category of living beings, apart from prototypical human agents also includes animals and plants, while his category of material objects includes instruments and places. These two categories are artificial constructs defined *a priori* for the sake of typological comparability, but have no direct correspondence in the system of derivational categories of the language described, *viz.* Spanish. It is unclear to me what insights a typological analysis could yield that in a first step arbitrarily distorts the facts of the single languages that are going to be compared. I

would like to argue that typological studies of this kind should be based on accurate descriptions of the semantics and productivity of all relevant word-formation patterns. This does not exclude, of course, that in the second phase, where the different languages are compared, some conscious idealisation of the data may be in order, as long as this way of proceeding is carried out under controlled conditions and warranted by the purpose of the study.

Its misguided semantic analysis and neglect of productivity are not the only weak points in Haarmann's analysis. Other problematic aspects include the purely synchronic nature of the description (*cf.* p. 111), which proves insufficient as soon as one begins to ask the crucial question of the origin of this kind of polyfunctionality, or the lumping together of deverbal and denominal formations. Since both aspects will be taken up later on, we may dispense ourselves from dwelling on them here.

3. Metaphoric or Metonymic Extension: Panagl (1975-78)

According to Panagl (1977: 6–7), there are fundamentally two alternative ways of conceiving of the relation between the agentive and the instrumental reading of suffixes, a lexicalist and a transformationalist one.

From a traditional perspective, the instrumental use is viewed as the result of a meaning extension of the corresponding agentive formation, either through metaphor or through metonymy. Though the latter idea seems quite natural – the lighter, for example, in the frame of lighting a cigarette, is in an obvious relationship of contiguity to the person carrying out the action, while a metaphorical relationship is less straightforward –, Panagl seems to have been the first scholar to take into consideration this possibility. The reason why Panagl nevertheless rejects both of these possibilities is his observation that in many cases German instrumental formations in *-er* are not accompanied by homonymous agentive formations. Now, Panagl argues (*cf.* 1977: 13), if the instrumental use is considered as the result of a semantic extension, one should expect that every instrumental formation or at least an overwhelming majority be accompanied by agentive formations, since these form the bases of the semantic extensions. E. *lighter*, for example, would be a problematic case in point, as there is no established agentive formation *lighter* referring to a person who lights. This correct observation indeed excludes the possibility of explaining all instrumental formations as semantic extensions, metaphoric or metonymic, of corresponding established agentive formations. It does not exclude, however, another interpretation, where the mechanism of semantic extension is used only to explain the *rise* of the instrumental use, while later on instrumental neologisms may be coined in direct analogy to the existing instrumental formations. Under such an interpretation, instrumental uses without corresponding agentive formations would no longer be problematic, since they are attributed to an independent instrumental pattern, only diachronically linked to the agentive one. To be precise, the rise of the instrumental pattern is the result of one or several cases of meaning extension followed by a reinterpretation of the agentive suffix as instrumental: the meaning 'instrument used by the agent designated by V + suffix' (metonymic variant) or 'instrument similar to the agent designated by V + suffix' (metaphoric variant), which are the result of meaning extensions applied to single agentive formations, are reinterpreted as 'instrument used for V-ing'.

But there seems yet to be a third possible interpretation of how metaphor or metonymy may transform agentive into instrumental formations. In order to understand how it works, we first have to introduce the concepts of reinterpretation and approximation as they are defined in Rainer (2005). In this study, I claim that semantic change in word-formation, apart from conscious manipulation of the meaning of a pattern, may be due to two fundamentally different mechanisms, *viz.* reinterpretation and what I propose to call *approximation*. Reinterpretation is the mechanism we have described above as an alternative to Panagl's conception, and according to Jaberg (1905) this would be the only mechanism bringing about semantic change in word-formation. Contrary to this position, where all cases of semantic change in word-formation are seen as the result of lexical semantic change in individual complex words followed by reinterpretation, I argue that semantic change in word-formation may also occur at the very moment of the creation of a neologism, without the mediation of lexical semantic change. In such cases, the coiner of a neologism uses a word-formation pattern in an approximate way, hence the term *approximation* I have chosen to refer to this mechanism. The deviance between pattern and neologism is generally bridged by metaphor or metonymy, which in this instance apply to patterns of word-formation and not to single complex words.¹

The following simple example may serve to illustrate how approximation works. Marchand (1969: 150) notes that the English locative prefix *cis-* has also been used, occasionally, in a temporal sense: "The words *cis-Elisabethan* 1870 and *cis-reformation (time)* 1662 transfer the notion of place into that of time. The meaning here is 'belonging to the time after –, subsequent to –'." This semantic change of the prefix *cis-* from its proper spatial meaning to a temporal one cannot be accounted for in terms of lexical semantic change followed by reinterpretation. It was not the case that some individual adjective of the locative type *cisalpine* underwent a semantic change from the realm of space to that of time – no such case is documented nor is it easy to imagine how such a change could come about –, with subsequent irradiation of the new temporal meaning to the prefix *cis-*; the temporal meaning must have arisen at the very moment of the creation of the adjectives *cis-reformation* and *cis-Elisabethan*. The speakers or writers simply used the pattern itself in a metaphoric manner, relying on the pervasive conceptual metaphor TIME-RELATIONS AS SPACE-RELATIONS.

If one is willing to accept the existence of these two fundamental mechanisms of semantic change in word-formation, the question arises with respect to the agent-instrument polysemy whether the extension occurred according to one or the other. The question cannot be answered from a purely synchronic perspective, but as far as diachrony is concerned, the two mechanisms, reinterpretation and approximation, make somewhat different predictions. Reinterpretation predicts the existence of three phases in the passage from agentive to instrumental usage: at stage 1, there are only agentive formations, at stage 2, one or several of these agentive formations acquire a secondary instrumental use through lexical semantic change, and at stage 3 these secondary formations are reinterpreted as directly formed according to an instrumental pattern, which may now be used for the creation of neologisms independently of the existence of corresponding agentive formations. Approximation, on the other hand, does not require the existence of stage 2, i.e., there need not have been at any moment formations with both an agentive and an instrumental reading. Our theory thus leads us

¹ Similar ideas are also put forward in Panther / Thornburg (2002).

to pay particular attention to the earliest instrumental formations and to look whether this early set is a subset of the agentive formations or whether the two sets are complementary right from the beginning.

Noun in <i>-dor</i>	Instr. use	Agent. use
<i>pisador</i> ‘pestle’ (< <i>pisar</i> ‘to tread’)	1268	1200
<i>foradador</i> ‘drill’ (< <i>foradar</i> ‘to drill’)	1277	—
<i>asador</i> ‘spit’ (< <i>asar</i> ‘to roast’)	1295	1450
<i>tajador</i> ‘carving board, plate’ (< <i>tajar</i> ‘to cut’)	1295	—
<i>rascador</i> ‘scraper’ (< <i>rascar</i> ‘to scrape’)	1330–43	—
<i>follador</i> ‘tub for treading grapes’ (< <i>follar</i> ‘to tread’)	1380–85	1400 (1280?)
<i>alimpiador</i> ‘cleansing agent’ (med.) (< <i>alimpiar</i> ‘to clean’)	1381–1418	—
<i>menador</i> ‘cooking spoon’ (< <i>menar</i> ‘to stir’)	1385	—
<i>picador</i> ‘carving board’ (< <i>picar</i> ‘to chop’)	1423	1400
<i>pasador</i> ² ‘arrow’ (< <i>pasar</i> ‘to pass’)	1427–28	1280
<i>partidor</i> ‘some instr. of women’s toilet’ (< <i>partir</i> ‘to divide’)	1438	1180
<i>pelador</i> ‘depilatory’ (< <i>pelar</i> ‘to depilate’)	1438	1400
<i>bastidor</i> ‘frame’ (< <i>bastir</i> ‘to construct, to prepare’)	1440	—
<i>colador</i> ‘strainer’ (< <i>colar</i> ‘to strain’)	1450	—
<i>lamedor</i> ‘medicine to be licked’ (< <i>lamer</i> ‘to lick’)	1450	—
<i>majador</i> ‘pestle’ (< <i>majar</i> ‘to crush’)	1450	—
<i>aparador</i> ‘sideboard’ (< <i>aparar</i> ‘to set (table)’)	1477–96	—
<i>tapador</i> ‘stopper’ (< <i>tapar</i> ‘to close’)	1486–99	—
<i>cerrador</i> ‘lock’ (< <i>cerrar</i> ‘to lock’)	1492	—
<i>purgador</i> ‘screen’ (< <i>purgar</i> ‘to purify’)	1493	1494 (adj.)
<i>mosqueador</i> ‘fan’ (< <i>mosquear</i> ‘to chase away flies’)	1495	—
<i>raedor</i> ‘scraper’ (< <i>raer</i> ‘to scrape’)	1495	1256

Table 1: The oldest instrumental usages of Spanish *-dor*

The rise of the temporal use of *cis-* described above is a neat instantiation of approximation. Traditional descriptions of the rise of the instrumental use of agentive suffixes, however, are insufficiently detailed and reliable to allow to decide the question of what mechanism was responsible in our case (provided that metaphor and metonymy

² A loan word from Catalan, Provençal or Italian.

have played a role at all; *cf.* below). An investigation of the oldest instrumental usages of Spanish *-dor* carried out with the help of the historical corpus of the Real Academia Española (*CORDE*, see <http://www.rae.es>) yields the results displayed in table 1. This table contains, in chronological order, all the Medieval examples of the *CORDE* corpus. The last column indicates whether there was, at the moment of the first documented use of the instrumental formation, a corresponding agentive formation. Note, however, that the existence of a corresponding agentive formation is no proof that the instrumental formation was actually formed by a meaning extension on the basis of the corresponding agentive formation, since not all agentive formations qualify as plausible vehicles for a metaphorical or metonymic transfer. *Pisador*, for example, is attested from 1200 onwards in the agentive meaning ‘person treading grapes’, before the instrumental meaning ‘pestle’ appears in 1268. Now, may the pestle be viewed as a figurative treader of grapes? It does not strike me as particularly plausible, and this is the most plausible case in our data. The subjective element in assessing the existence of proper agentive vehicles at the moment of the creation of the corresponding instrumental formations makes the decision whether the Spanish data of table 1 better corresponds to reinterpretation or approximation a difficult one. My impression is that it better corresponds to approximation, though the complementary distribution is not perfect.

Independently of whether one thinks that the mechanism at work was reinterpretation or approximation, we still have to decide between metaphor and metonymy. As we have already seen, there can be no doubt that agent and instrument show a relationship of contiguity in the action frame. Nevertheless, I would like here to put forward one general argument against a metonymic interpretation of the relationship between agent and instrument nouns. We start from the observation that not all relations of contiguity that one can establish in the real world serve as the base for metonymies with the same frequency in the languages of the world. Languages seem to privilege certain relationships of contiguity, a subject which, unfortunately, has not, until now, attracted the attention that it deserves. In the absence of a full-blown theory of what constitutes a possible or preferred metonymic relationship in natural language in general or in specific languages, the following argument must be considered of a rather tentative nature, but nevertheless could possibly constitute a clue for deciding between the metaphoric and the metonymic account. The argument is simple and relies on the observation that, apart from morphologically complex agent and instrument nouns, the metonymic relationship between agents and instruments seems to have a clear directionality, the vehicle always being the instrument and the target the agent. With non-derived nouns or nouns not derived by agentive suffixes, in fact, it is quite common to find cases where an agent is designated by the name of the instrument he typically uses, but not *vice versa*. It is common in many languages, for example, to refer to the trumpeter as the trumpet, but not to the trumpet as the trumpeter.³ Another piece of evidence comes from onomasiological studies of designations for tools, where I have found no trace of agents as a possible diachronic source-domain. According to Gade (1898), for example, of the 40 Latin names of tools contained in Georges’ dictionary, no single one is an extension of the name of the worker that used it (*cf.* pp. 9–11), and the same is true of French (*cf.* pp. 75–76). If this generalisation turned out to be valid for

³ On the other hand, Panagl (1977: 13), followed on this point by Dressler (1980: 113), notes that he does not know of any case where an agentive pattern of word-formation developed out of an instrumental one. For a possible counter-example from Serbo-Croatian, *cf.* Beard (1990: 119).

languages in general, it would constitute an effective argument against the metonymical account of the origin of the agent-instrument polysemy with suffixes, since semantic extensions are a conceptual phenomenon and so should not distinguish between simple and complex bases.⁴

Summing up what we have said up to now about the issues metaphor *vs.* metonymy and reinterpretation *vs.* approximation, we have to admit that no definitive conclusion has been reached as to which of the four logically possible combinations is or are correct. We have put forward a possible argument against metonymy, and the Spanish data of table 1 appears to favour approximation over reinterpretation. This would point to metaphoric approximation as the most probable candidate. A metaphoric explanation would have the advantage of explaining the directionality of the agent instrument polysemy as a natural consequence of the anthropomorphism so typical of metaphor in general. But the evidence in favour of metaphor was also rather shaky in the Spanish case. This could even mean that in the end we are not dealing with a problem of semantic change at all. As we will see below, there are indeed some arguments that point in this direction. But even though we have not been able to reach conclusive evidence, it seems important to me that we begin to put the right questions about this unexpectedly complex problem, questions that may guide further research.

4. Synchrony and Diachrony in Typology: Dressler (1980)

While Haarmann (1975) is an exclusively synchronic study, Panagl, a student of Indoeuropean, also dedicated some reflections to diachronic aspects. In Panagl (1977: 4), for example, he notes that the pervasiveness of the agent-instrument polysemy in Indoeuropean could be due either to a Proto-Indoeuropean origin or, alternatively, a “drift” in the Sapirean sense of the term (an interpretation favoured, according to Panagl, by its absence from Hittite).⁵ In his endeavour to arrive at a cognitive foundation of the polysemy of agent nouns, Dressler (1980) also transcends the purely synchronic typological approach and includes some remarks on acquisition, aphasia and diachrony. As far as diachrony is concerned, he notes (*cf.* p. 113) that semantic extension in our domain seems to have been strictly directional: Agent patterns, according to him, may turn into Instrumental or Locative ones, and Instrumental patterns into Locative ones, but not *vice versa*. It is not made clear, however, how exactly such diachronic generalisations – provided that they turn out to be correct – or the observations about acquisition and aphasia might contribute to our understanding of the nature of the phenomenon under consideration. The “cognitive embedding” (*kognitive Verankerung*, p. 114) of the process is left for future research.⁶

⁴ For a recent defence of the metonymic nature of the agent-instrument relationship, *cf.* Panther / Thornburg (2002: 292, 298). In fact, they defend the Salomonic position that a metaphoric and a metonymic account are not mutually exclusive.

⁵ In Tichy’s (1995) study of Vedic agent nouns in *-tar-*, no mention is made of an instrumental extension either.

⁶ In Dressler (1986: 527), a relation is established between the unidirectionality of our polysemy and the animacy hierarchy: “This agent hierarchy seems to correspond to the animacy hierarchy [...]. Most central events of human life prototypically have a human agent; next come animal agents [...]; then plants which produce fruit [...]; then impersonal agents [...]; then instruments; and finally local conditions of events or states [...]. In other words, the conceptual basis of the agent hierarchy seems to lie in the prototypical human interpretation of events.”

1.	Deverbal agentive and instrumental formations:	
1.1.	Agent – Predicate	E. <i>cutt-er</i>
1.2.	Instrument – Predicate:	E. <i>cutt-er</i>
1.3.	Ag. / Instr. – Predicate – Object	E. <i>dress-mak-er, salt-shak-er</i>
2.	Denominal formations:	
2.1.	Ag. / Instr. (makes/typically deals with) Object	E. <i>garden-er</i>
2.2.	Agent (comes / is from) Object	E. <i>island-er, London-er</i>
2.3.	Agent (is typical for) Predicative Adjective ⁷	E. <i>foreign-er</i>
3.	Predicate – Place:	
3.1.	Recipients (instrumental reading possible)	L. <i>mulc-trum</i> ‘pail’ (deverbal) F. <i>salad-ier</i> ‘salad bowl’ (denominal)
3.2.	Locative meaning developed from instrumental one	G. <i>Ordn-er</i> ‘file’ (deverbal) Fr. <i>encl-ier</i> ‘inkpot’ (denominal)
3.3.	Truly locative	Fr. <i>dort-oir</i> ‘bedroom’ (deverbal) Fr. <i>guêp-ier</i> ‘wasp’s nest’ (denominal)

Table 2: The polysemy of agent suffixes according to Dressler (1980)

In the typological part of his study, Dressler, like Haarmann, includes both deverbal and denominal formations,⁸ which he classifies as subsets of the semantic formula Predicate – Agent – Instrument – Locative – Object as illustrated in table 2. I would like to argue now that such a classification, similar to the one presented by Haarmann, is not very suitable to gain deeper insights into the nature of the polysemy of agent nouns. A strictly diachronic approach, it seems to me, will yield better results, since it shows that what looks similar from a purely synchronic perspective often corresponds to entirely different phenomena when viewed from a diachronic one. Once again, I will use Romance data to illustrate this point.

The origin and history of the instrumental extension of Romance deverbal agent nouns is still not definitively settled. As we saw in the introduction, the most popular view attributes it to metaphor. While this venerable view may be open for discussion,

⁷ Conceived of as a kind of Object of the copulative verb.

⁸ In Dressler (1986: 527) the following explanation is given for the choice of the category of the base of agent nouns: “Since events are prototypically symbolized by verbs, it must come as no surprise that verbs are the preferred bases of agent nouns. Nouns are preferred when the conceptually ‘underlying’ verbs are semantically underspecified, or not distinct enough”. A recent case for a unified treatment of English deverbal and denominal formations in *-er* is Panther / Thornburg (2002: 284–285). This is too complex an issue to be addressed here. One problem for a unified treatment, however, is already pointed out by Panther and Thornburg themselves (*cf.* pp. 312–313): deverbal and denominal suffixes do not seem to behave alike with respect to semantic extension (for example, denominal agentive *-ist* does not show any semantic extensions).

there can be no doubt that some cases of polysemy of deverbal formations at least are attributable to other reasons.

Already in Darmesteter (1877), the first comprehensive treatment of French word-formation, nineteenth-century instrumental nouns in *-eur* are considered as “tirés d’adjectifs”, i.e., derived from adjectives (cf. pp. 47–49). And Darmesteter was certainly right in considering many names of tools and machines coined in the 19th century as the result of the ellipsis of the head noun in noun phrases of the form *appareil* + adjective in *-eur* or *machine* + adjective in *-euse*. Since at that time, French had already an established nominal instrumental pattern in *-eur*, it is often difficult in single instances to decide whether we have to do with the result of ellipsis or with a direct nominal formation, overall however there can be no doubt that both means were productively used (cf. also Spence 1990: 32–33). Apart from Darmesteter’s intuition we can also rely on the testimony of nineteenth-century texts, where often the noun phrase is documented before the short form, or side by side. This elliptical mode of forming instrument nouns in French seems to have arisen or at least gained momentum in the 19th century. Pharies (2002: 170) has recently proposed to extend this elliptical explanation to the rise, in the Middle Ages, of the instrumental and locative uses of the corresponding Spanish suffix *-dor*. However, as I have shown in Rainer (2004a), such a move is unwarranted, since we do not find any parallel syntagmas in Spanish up to the 19th century, when this mode of formation was probably imported from France along with a large number of names of tools and machines. Ellipsis is obviously *a priori* restricted to languages where, like in Romance, agent nouns in *-eur*, *-dor*, etc. have parallel adjectival formations, i.e., where ‘cutter’ and ‘cutting’ (adjective) are formed by one and the same suffix.⁹

Another source of instrumental nouns in Romance which has nothing to do with meaning extension is homonymisation. In Provençal (cf. Adams 1913: 54) and in Catalan¹⁰ (cf. Moll 1952: § 429), as well as in some other areas, among them Romania (cf. Graur 1929) and some Italian dialects (cf. Rainer 2004b), as a consequence of phonetic change the result of the Latin instrumental suffix *-torium* ended up identical with the one of the Latin agentive suffix *-torem*. L. *operatorium* ‘workshop’ (< *operari* ‘to work’), for example, became *obrador* in Provençal and in Catalan, with a suffix *-dor* identical to the one we find in agent nouns. The rise of the agent-instrument and agent-place “polysemy” is therefore due to pure accident in those languages. If we had no historical records of the Romance languages, the temptation would no doubt be great to give a semantic or “cognitive” interpretation of the formal identity of the agentive, instrumental and locative suffix.

The agent-place “polysemy”, however, someone might object, is also found in Spanish, where L. *-torium* and *-torem* did not become homophonous, but remained distinct as, respectively, *-dero* and *-dor*. But, as Malkiel (1988: 238) has shown convincingly, the first Spanish locative formations of the type *comedor* ‘dining room’ were borrowings from Provençal (or Catalan), where the locative use, as we have just seen, was due to phonetic change. The same hypothesis, by the way, had already been taken into consideration by Meyer-Lübke (1921: § 66) with respect to some surprising Old French instrumental and locative formations in *-eour*, the regular outcome of L. *-torem*, like *tailleour* ‘carving board, plate’ (← *tailler* ‘to cut’) or *ovreour* ‘workshop’

⁹ Adjectival usage of *-tor* was already common in Late Latin (cf. Fruyt 1990).

¹⁰ In the light of this fact, the purely semantic interpretation of the origin and development of Catalan nouns in *-dor* in Grossmann (1998: 390) is surprising.

(< *ovrer* ‘to work’), surprising because the French instrumental or locative suffix to be expected would have been *-oir*, the outcome of L. *-torium*. In these cases, too, no semantic or “cognitive” explanations are needed. Borrowing is a sufficient explanation.

We have thus identified, for Romance, three uncontroversial origins of instrumental or locative uses of deverbal agentive suffixes which have nothing to do with semantics or “cognition”, namely ellipsis, homonymisation and borrowing. I am firmly convinced that these examples are a strong caveat against overly rash semantic or “cognitive” speculations on the basis of purely synchronic data. The fact that meaning M^1 and M^2 of a suffix may plausibly be viewed as polysemous by an observer on purely synchronic grounds does not entail that we are really dealing with meaning extension from a diachronic perspective. The lesson to be drawn from this fact, it would seem to me, is that it is much more fruitful to study paths of semantic change with sound diachronic method than to extrapolate them from purely synchronic data.¹¹ This is not meant to deny any usefulness to typological studies in this domain, but one has to be extremely careful about their interpretation. The best thing to do would be to use as the basis of typological studies well-established diachronic paths of change rather than synchronic polysemies.

Many readers may accept this conclusion in principle, but will object that one should not overestimate the Romance evidence adduced in the face of the many cases of polysemous agent nouns documented for other languages. This is an argument that may be right, but for the moment being we simply cannot say how many of the cases adduced in the literature – which, after all, are not so great in number as the universalist rhetorics might make one believe – are genuine cases of semantic extension and how many are due to ellipsis, homonymisation and borrowing. We still don’t have even an approximate idea about how frequent our polysemy really is in the languages of the world, since all the typological studies published up to now have a very preliminary character and work with relatively few illustrative examples, mostly taken from Indo-European and supplemented with scattered exotisms that serve to suggest universality.

Things get even worse when we turn to denominal formations. The pronounced polysemy of denominal nouns like those in *-ier*, as is well-known (cf. Roché in press), is due to the fact that the etymon, Lat. *-ariu*, was a suffix forming relational adjectives that ended up as a nominal suffix after the ellipsis of the head nouns. Here again, it would be misleading to use just the synchronic data for speculations about the semantic or “cognitive” foundation of this agent-inhabitant / place / tree / set “polysemy”.

5. Extension Schemes: Booij (1986)

Booij (1986) proposed to account for the polysemy of agent nouns with what he called an extension scheme, which in our case takes the form Personal Agent > Impersonal Agent > Instrument. All three of these meanings, for example, are present in Dutch *zender* ‘sender’, which may refer to a person who sends (Personal Agent), a radio/tv station (Impersonal Agent), or a transmitter (Instrument). “The category Impersonal Agent”, according to Booij, “is not the same as Instrument, but an intermediate and mediating category” (p. 509). It roughly corresponds to automatic devices (cf. p. 510).

¹¹ Jurafsky (1996), one of the most detailed studies of universals of semantic change, is not exempt of this extrapolatory tendency.

The description of Impersonal Agent as an intermediate or mediating category somewhat closer to Agent than to Instrument seems intuitively appropriate from a synchronic perspective, since an impersonal agent shares the feature ‘autonomous movement’ with a human agent, and the feature ‘inanimate’ with an instrument. If Booij’s scheme, however, were meant to describe how the instrumental use of a pattern may arise from an agentive one in diachrony – which is nowhere explicitly claimed in Booij’s article, but seems to be an invited inference –, this prediction would be clearly wrong. In my diachronic study of the passage from agent to instrument in Spanish (*cf.* Rainer 2004a), for example, I have found that up to the 19th century, that is during the first 500 years of the suffix’s productivity, one only finds Instruments in Booij’s restricted sense, with the possible exception of *despertador* ‘alarm clock’, already attested in the 16th century, while Impersonal Agents are only attested after the Industrial Revolution, which of course is only to be expected, since automatic devices are typical products of this period. At least for the pre-industrial age, thus, one would have to postulate an extension scheme Agent > Instrument, without intermediate category.

Another prediction of the extension scheme, according to Booij, is that the agentive interpretation of Dutch nouns in *-er* “is always possible, although it may not be an established use of a certain noun” (p. 510). It is not clear whether, in the light of the admission of possible but not attested agent nouns, this prediction has any empirical content. A fair interpretation, probably, would be that there should not be too many missing agent nouns beside attested instrument nouns in *-er* or equivalent suffixes in other languages. Now, at least in present-day Spanish, most of the instrumental formations in *-dor* are not accompanied by a corresponding agentive formation. Of the 48 nouns in *-dor* contained under the letter *D* in the Spanish dictionary I have at hand, 24 are exclusively agentive, 21 exclusively instrumental, while only three have both meanings. As one will recall, the same point has also been made with respect to German by Panagl, who based his rejection of a semantic extension account of German *-er* precisely on this tendency towards a complementary distribution of agent and instrument nouns.

Another prediction is formulated as follows by Booij: “if [the extension scheme] is correct, the polysemy that we find for *-er* nouns should also be found for other types of derived words with an Agent interpretation. Moreover, since the structure of conceptual categories is presumably language-independent, we expect the same polysemy to exist for agent nouns in other languages” (p. 511). Both predictions, according to Booij, “are confirmed by the facts” (p. 511). The confirming evidence presented consists essentially in a short reference to the typological data presented by Panagl (1978) and Dressler (1980). Booij is aware of the fact that there are languages such as Finnish or Latin which have agentive patterns without instrumental extensions, but this is said to be a consequence of the blocking effect of rival instrumental patterns. Support for this argument could come from Spence’s (1990: 35) hypothesis that the instrumental extension of *-eur* in French was the consequence of the loss of productivity of the instrumental suffixes *-oir* and *-oire*, but more research is needed in order to gain certainty about the history of French instrumental suffixation. On the other hand Beard (1990: 118) notes that in Serbo-Croatian the existence of a productive instrumental suffix does not block the instrumental use of the agent suffix. The most crucial counter-evidence would seem to consist of languages without an instrumental pattern, but a productive and exclusively agentive pattern. As we have seen above, the

descriptive coverage of the existing typological literature is rather restricted, for the moment being, so that I will not venture a definitive assessment of Booij's prediction here. Beard (1990) has presented what he considers to be falsifying instances with respect to Booij's hypothesis, but more evidence is needed to arrive at a definitive settlement of this question.

6. Prototype Reanalysis: Ryder (1991)

One problem that has been left undecided by Booij, the exact nature of the passage from the agentive to the instrumental meaning, has been tackled some years later by Ryder (1991), a study couched in the framework of Californian-style cognitive linguistics. Her approach is based on the three basic notions *semantic case*, *event structure* and *prototype reanalysis*. Semantic cases like Agent, etc. are said to have a prototypical organisation (*cf.* p. 300). Complex events may be broken down into smaller units, the exact organisation depending very much on the point of view of the speaker. One may view, for example, the breaking of a glass with a hammer as one holistic event or divide it into smaller sub-events such as the act of seizing the hammer, the act of throwing the hammer and the splintering of the glass. A series of such minimal events is called *event chain*. With the help of this conceptual framework, the nature of the semantic extensions of agent nouns is interpreted as “the result of shifts in the construal of the defining episode” (p. 303).¹²

As the agent and instrument become more separated from each other in time, and the instrument's action becomes increasingly independent of the agent, the agent's action may be construed as outside the episode, leaving the instrument as the most agent-like participant remaining.¹³ [...] It is the shift of the agent to outside the boundaries of the episode that motivates the extension of agentive *-er* forms to include *instrument Er's*. (pp. 303–304)

Ryder's account resembles Booij's extension scheme – in my diachronic interpretation – in predicting that the instrumental use occurred when instruments became more and more autonomous, automatic, Impersonal Agents in Booij's terminology. And it fails for the same reasons that were advanced against Booij's hypothesis. With the possible exception of *clipper*, all nouns from the 15th to the 17th century mentioned by Ryder in support of her account (*viz.* *lighter*, *poker*, *scraper*, *snuffer*, *borer*, *knocker*, *grinder*, and *toaster* ‘toasting fork’), refer to instruments that may be characterised as simple tools and do not show any autonomy or automaticity. If Ryder's list of early instrument nouns proves anything, this is the extent to which perception may be distorted by theoretical expectations.

Ryder does not tell us how she arrived at her list. What is clear is that it is not an exhaustive enumeration of the earliest English instrument nouns. According to Marchand (1969: 275), “the oldest coinage appears to be *slipper* 1478”. Old English deverbial nouns in *-er* “are all agent nouns” (p. 275). In his detailed 1971 study of the Old English suffix *-er(e)*, Kastovsky, a pupil of Marchand's, arrives at the conclusion

¹² Note that in Ryder's approach the passage from agent to instrument does not involve metaphor, where a source domain is consciously projected on to a target domain.

¹³ Essentially the same explanation had already been proposed by Panagl (1975: 239).

that his teacher's statement is slightly too apodictic (p. 295). Kastovsky's Old English data (*cf.* pp. 294–295), in fact, contains one uncontroversially instrumental example, namely *pūnere* 'pestle' (< *pūnian* 'to pound'), which occurs in a gloss (unfortunately we are not told of what Latin word). To this example one might add *scēawere* 'mirror' (← *scēawian* 'to look at'), which translates Latin *speculum*.¹⁴ The third example, *word-samnere* 'catalogue, collection of words' (← *samnian* 'to collect') has a meaning somewhere between instrumental and locative. A more neatly locative meaning is present in the fourth of Kastovsky's examples, namely *scēawere* 'watch-tower'. Dalton-Puffer, herself a pupil of Kastovsky's, has returned to the problem in her 1996 study of the French influence on Middle English Morphology, where she comes to the conclusion that "there is only *one* word in the data¹⁵ that really answers the description of Modern English *cooker, opener*, namely *calculer* (ME3)¹⁶ 'computing, calculating device'" (p. 139). Interestingly, in the *OED* I have accidentally come across a semantically similar Middle English instrument noun documented somewhat earlier, in 1310, namely *counter*, defined as 'a round piece of metal, ivory, or other material, formerly used in performing arithmetical operations'.

Now, do these six non-agentive formations attested prior to Marchands *slipper* and Ryder's *lighter*, which, as expected, do not designate autonomous, automatic devices either, but traditional tools or places, allow us to infer how the passage from agent to instrument might have occurred in English? Personally, I can't see any obvious hint in this data, which I can only urge Anglicists to complete. What catches my attention, however, is that some words have interesting Romance or Latin parallels. *Scēawere* 'mirror' corresponds exactly to Old French *mirreur* 'mirror' (← *mirer* 'to look at'), first attested in 1180 according to *FEW* VI 149a (the Modern form *miroir* is first attested in 1260), which had already ousted the type SPECULUM in preliterary French (*FEW* VI 155b). The meaning 'watch-tower' does not seem to have existed in Old and Middle French, but is attested for Spanish *mirador* (← *mirar* 'to look at'), which must be a loan translation from Catalan or Provençal, as early as 1250 in *CORDE* (*cf.* Rainer 2004a). *Counter* is paralleled by French *comptoer* 'jeton pour compter', first attested in 1359 (*FEW* II 992b). Though the French word is slightly posterior to the English one, it seems quite obvious that French must have been the donor language. *Calculater* has no Middle French parallel, but could simply be an analogical formation on the model of *counter*. These parallelisms, I believe, might warrant a closer examination of the possible influence of French in the development of the instrumental and locative use of English *-er*. Foreign influence, finally, also seems possible in the rise of *word-samnere*, whose ending may have been influenced by the denominal collective *-er* loan-translated from Latin *-arium*, as in Old English *antefnere* 'antiphoner' (Kastovsky 1971: 295, fn. 23), a clear loan-translation of Medieval Latin *antiphonarium*. Though *word-samnere* is a deverbal formation, it fits perfectly into this semantic field. It could thus be worthwhile for Anglicists to pursue the hypothesis that the rise of non-agentive uses of *-er* was due – at least partially – to Latin and Romance influence.

The possible influence of loan-translations in the rise of non-agentive meanings of agent nouns should also be analysed with respect to other European languages. This

¹⁴ This is also the only instrumental formation Zbierska-Sawala (1993: 43) has found in her Early Middle English corpus.

¹⁵ *Sc.* the Helsinki corpus.

¹⁶ *ME3* refers to the Middle English period going from 1350 to 1420.

might help to explain at least part of a startling conspiracy in Medieval Europe: while Latin and, as it seems, Proto-Germanic agent nouns seem to have lacked non-agentive uses, in the Middle Ages *all* European languages seem to acquire such readings within several centuries. This could, of course, be an extreme case of polygenesis, since semantic extension is a universally available pattern, but the spatio-temporal coincidence makes it too strange for me to swallow this explanation without first checking the alternative hypothesis of inter-European loan-translation. Both explanations, of course, are not mutually exclusive, but may have reinforced each other. If this were the case, historical linguists would nevertheless have the task of establishing the specific mixture of both factors for any individual language.

My insistence on non-semantic or non-“cognitive” interpretations of the fragmentation process of agent nouns should not be misinterpreted as a general, *a priori* rejection of their importance. It is quite obvious that they do play an important role, in the derivational categories dealt with here (Agent, Instrument, Place), but also the other categories that are sometimes found with agent nouns, such as Action, Object, etc.¹⁷ The problem is rather that their obvious importance has obscured most researchers’ view of the other factors – ellipsis, homonymisation, borrowing, loan-translation – which seem to play an equally important role, at least in Romance. A fully satisfactory account of the polysemy of agent nouns cannot escape coping with this complexity, and only such detailed accounts will form a reliable basis for typological and semantic studies.

References

- Adams, E.L. (1913), *Word-Formation in Provençal*, New York, Macmillan.
- Beard, R. (1990), *The nature and origins of derivational polysemy*, “Lingua” 81, 101–140.
- Booij, G.E. (1986), *Form and meaning in morphology: The case of Dutch ‘agent nouns’*, “Linguistics” 24, 503–517.
- Dalton-Puffer, C. (1996), *The French Influence on Middle English Morphology*, Berlin / New York, de Gruyter.
- Darmesteter, A. (1877), *De la création actuelle de mots nouveaux dans la langue française et des lois qui la régissent*, Paris, Vieweg.
- Dressler, W.U. (1980), *Universalien von Agens-Wortbildungen*, in G. Bretschneider & C. Lehmann (eds.), *Wege zur Universalienforschung: sprachwissenschaftliche Beiträge zum 60. Geburtstag von Hansjakob Seiler*, Tübingen, Narr, 110–114.
- Dressler, W.U. (1986), *Explanation in natural morphology: illustrated with comparative and agent-noun formation*, “Linguistics” 24, 519–548.
- FEW* = Wartburg, W. von (1922 ff.), *Französisches Etymologisches Wörterbuch*, Bonn / Leipzig / Paris / Basel: changing editors.
- Fruyt, M. (1990), *La plurivalence des noms d’agent latins en -tor: lexique et sémantique*, “Latomus” 49, 59–70.
- Gade, H. (1898), *Ursprung und Bedeutung der üblicheren Handwerkzeugnamen im Französischen*, Diss. Kiel, Vollbehr & Riepen.
- Georges, K.E. (1995), *Ausführliches lateinisch-deutsches Handwörterbuch*, 2 vol., Hannover, Hahn.
- Graur, A. (1929), *Nom d’agent et adjectif en roumain*, Paris, Champion.
- Grossmann, M. (1998), *Formazione dei nomi di agente, strumento e luogo in catalano*, in G. Ruffino (ed.), *Atti del XXI Congresso Internazionale di Linguistica e Filologia*

¹⁷ Interestingly, these further categories are far more common in Germanic than in Romance.

- Romanza (Palermo 18–24 settembre 1995). Sezione 2: Morfologia e sintassi delle lingue romanze*, Tübingen, Niemeyer, 383–392.
- Haarmann, H. (1972), *Zur Polyfunktionalität finnischer Suffixe*, “Ural-Altäische Jahrbücher” 44, 144–152.
- Haarmann, H. (1973), *Polyfunktionale Suffixe im Neurussischen*, “Zeitschrift für den Russischunterricht” 8, 70–82.
- Haarmann, H. (1975), *Polyfunktionale Suffixe im Spanischen*, in *Filología y didáctica hispánica. Homenaje al profesor Hans-Karl Schneider*. Hamburg: Buske, 87–114.
- Jaberg, K. (1905), *Review of ‘Roediger, M. (1904) Die Bedeutung des Suffixes –ment’*, Diss. Berlin, “Archiv für das Studium der neueren Sprachen und Literaturen” 114, 458–462.
- Jurafsky, D. (1996), *Universal tendencies in the semantics of the diminutive*, “Language” 72, 533–578.
- Kastovsky, D. (1971), *The Old English suffix -er(e)*, “Anglia” 89, 285–325.
- Malkiel, Y. (1988), *Las peripecias españolas del sufijo latino -ōriu, -ōria*, “Revista de Filología Española” 68, 217–255.
- Marchand, H. (1969), *The Categories and Types of Present-Day English Word-Formation*, 2nd ed., Munich, Beck.
- Meyer-Lübke, W. (1890), *Italienische Grammatik*. Leipzig.
- Meyer-Lübke, W. (1921, ²1966), *Historische Grammatik der französischen Sprache. Zweiter Teil: Wortbildungslehre*, edited by J. Piel, Heidelberg, Winter.
- Moll, F. de B. (1952), *Gramática histórica catalana*, Madrid, Gredos.
- OED = Oxford English Dictionary*, Oxford, Clarendon Press (1991).
- Panagl, O. (1975), *Kasustheorie und Nomina agentis*, in H. Rix (ed.), *Flexion und Wortbildung*, Wiesbaden, Reichert, 232–246.
- Panagl, O. (1977), *Zum Verhältnis von Agens und Instrument in Wortbildung, Syntax und Pragmatik*, “Wiener linguistische gazette” 16, 3–17.
- Panagl, O. (1978), *Agens und Instrument in der Wortbildung*, in W.U. Dressler & W. Meid (eds.), *Proceedings of the Twelfth International Congress of Linguistics*, Innsbruck, 453–456.
- Panther, K.-U. & L.L. Thornburg (2002), *The roles of metaphor and metonymy in English -er nominals*, in R. Dirven & R. Pörings (eds.), *Metaphor and Metonymy in Comparison and Contrast*, Berlin / New York, Mouton de Gruyter, 279–319.
- Pharies, D. (2002), *Diccionario etimológico de los sufijos españoles*, Madrid, Gredos.
- Rainer, F. (2004a), *Del nombre de agente al nombre de instrumento en español: ¿cuándo y cómo?*, to appear in “Iberoromania” 59.
- Rainer, F. (2004b), *L’origine dei nomi di strumento italiani in -tore*, to appear in T. Krisch, T. Lindner & U. Müller (eds.), *Analecta homini universali dicata. Festschrift für Oswald Panagl zum 65. Geburtstag*, vol. 2. Stuttgart: Akademischer Verlag Hans-Dieter Heinz.
- Rainer, F. (2005), *Semantic change in word-formation*, to appear in “Linguistics” 43/2.
- Roché, M. (in press), *La dérivation en -ier(e) en ancien français*, to appear in C. Buridant (ed.) *La morphologie dérivationnelle dans l’ancienne langue française et occitane*, Villeneuve d’Ascq, Presses Universitaires du Septentrion (Lexique 17).
- Ryder, M.E. (1990), *Mixers, mufflers and mousers: The extending of the -er suffix as a case of prototype reanalysis*, “Berkeley Linguistic Society” 17, 299–311.
- Spence, N.C.W. (1990), *Instrumental -eur/-ateur/-euse/-atrice*, “The Modern Language Review” 85, 29–35.
- Tichy, E. (1995), *Die Nomina agentis auf -tar- im Vedischen*, Heidelberg, Winter.
- Zbierska-Sawala, A. (1993), *Early Middle English Word Formation*, Frankfurt am Main, Lang.

Resources for Suppletion: A Typological Database and a Bibliography¹

Greville G. Corbett, Dunstan Brown, Marina Chumakina & Andrew Hippisley
Surrey Morphology Group - University of Surrey
<http://www.surrey.ac.uk/LIS/SMG/>

1. The Special Interest of Suppletion

The phenomenon of suppletion, as found in English *go~went* where different inflectional forms of the same lexical item are not related phonologically, has a special place in morphology. Part of its importance is that it sets one of the outer bounds for the notion ‘possible word’ in a human language. It provokes questions about how such forms are to be treated in our theories, and how they are stored (Carstairs-McCarthy 1994). There has been considerable work on suppletion, particularly from Osthoff (1899) onwards. Current interest in the topic is shown by the recent appearance of two dissertations (Veselinova 2003 and Veselinovič 2003). While the body of research is extensive, the range of languages investigated is rather restricted in many publications. In order to stimulate further progress, we have constructed and made available a database (Brown, Chumakina, Corbett and Hippisley 2004). We hope this will help to put future research on a broader empirical base. An annotated bibliography is now available (Chumakina 2004); it contains over seventy entries on works written in five different languages (English, French, German, Italian and Russian) and this will give the reader a view of the literature.

2. The Canonical Approach in Typology

At the Barcelona meeting (MMM3), the first author outlined a ‘canonical’ approach to typology. In a canonical approach, we take definitions to their logical end point and build theoretical spaces of possibilities. Only then do we ask how this space is populated. The canonical instances, which are the best examples, those most closely matching the canon, may well not be the most frequent. Rather they may be rare, or even non-existent. They serve to fix a point from which occurring phenomena can be calibrated, and it is then significant and interesting to investigate frequency distributions. This approach was worked out with regard to agreement (Corbett 2003). It is an interesting issue how such an approach can be viable for a phenomenon like suppletion, which may be thought of as an ‘extreme’ phenomenon. It is clear that the

¹ The support of the AHRB under grant B/RG/AN4375/APN10619 and of the ESRC under grant R00027135 is gratefully acknowledged. The first author wishes to thank the organizers of the Fourth Mediterranean Morphology Meeting (MMM4), Catania, Italy, 21-23 September 2003, for the invitation to present this research, and to the participants for useful discussion. Constraints of space mean that this printed version covers only one part of the Catania presentation, namely the most collaborative part, hence it is co-authored by those involved in the Surrey Suppletion Database. At the editors’ invitation the bibliography is also included here (without the abstracts). A version of this paper was presented shortly after MMM4 at the Workshop on "Database-driven linguistic typology" at the Language Typology Research Centre Annual Meeting, Estoril.

object of such a typology will be lexemes, rather than constructions or languages. A helpful start for such an approach is offered by a part of Mel'čuk's definition of suppletion:

For the signs X and Y to be suppletive their semantic correlation should be maximally regular, while their formal correlation is maximally irregular.
Mel'čuk (1994: 358)

Beginning from this suggestion, we can establish dimensions along which the phenomenon may vary. We can establish the canonical instances, namely those which are maximally transparent in semantic terms and maximally opaque in formal terms (cf. Mel'čuk 1994: 342). As part of this we can recognise, for example, that some restrict suppletion to inflectional morphology, while others including Mel'čuk allow for suppletion in derivational morphology. Semantic correlations are typically more regular (more transparent) in inflectional than in derivational morphology, hence the clearer (and for some linguists the only) instances of suppletion will be found in inflectional morphology. Only instances of inflectional morphology are included in the database. To date fifteen criteria for canonical suppletion have been proposed in what is ongoing research.

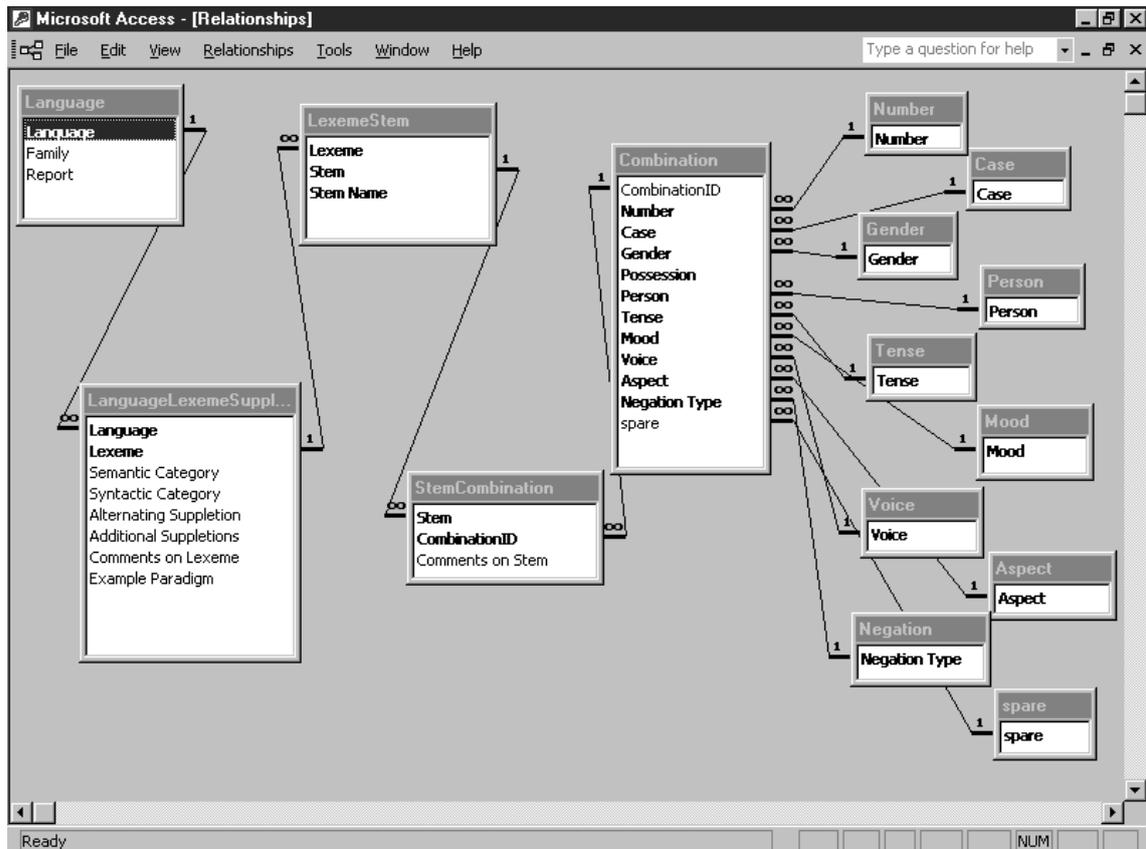
3. The Surrey Database of Suppletion

The database was designed and implemented both to inform our research and to make available to other linguists the data which we had collected and analysed. It allows a range of queries, and can be searched on-line over the web (at www.smg.surrey.ac.uk).

3.1 Structure of the Database

The design, due primarily to the second author, is indicated in Figure 1. Each table is motivated by a possible query.

Figure 1: Design of the Database



The design of the database in this figure allows for detailed description of the environments in which suppletion occurs and for the non-redundant storage of the information. On the right side of the figure there are the tables for feature sets (such as Case, Number, Person etc). Any feature in a feature set table occurs once in that table, but many times in the Combination table to the left of the feature sets. Feature combinations are then paired with stems (in the StemCombination table). The stem in a lexeme-stem pairing (in LexemeStem table) may occur with more than one feature combination. The relationship between the stem field in the LexemeStem table and the stem field in the StemCombination table is therefore one-to-many. The LanguageLexemeSuppletion table brings all the information together, combining the languages from the Language table with the lexemes from the LexemeStem table. The database has been implemented using Microsoft Access.

3.2 *The Data*

Languages were selected to ensure genetic and areal diversity. In addition, languages had to have the potential for inflectional suppletion (hence those with no inflectional morphology were not included). The data were derived from published grammars and dictionaries, and in many cases were checked with specialists. We are very grateful to Willem Adelaar, Nicholas Evans, George Hewitt, Paulette Levy, Marianne Mithun and Larry Trask for their help. The data on two languages, Komi and Xakass, were obtained

on field trips. The database records all instances of suppletion that were found in the following languages:

Language	Family
!Xóõ	Khoisan
Arapesh	Toricelli
Archi	Nakh-Daghestanian
Basque	Basque
Chichewa	Niger-Congo
Georgian	Kartvelian
Guaraní	Tupí
Hebrew	Semitic
Hua	Trans-New Guinea
Hungarian	Ugric
Itelmen	Chukotko-Kamchatkan
Jacalteco	Mayan
Japanese	Japanese
Kannada	Dravidian
Kayardild	Tangkic
Ket	Yenisei -Ostyak
Koasati	Muskogean
Komi	Finno-Permic
Limbu	Sino-Tibetan
Mayali (Bininj Gun-wok)	Gunwinyguan
Tetelcingo Nahuatl	Uto-Aztecan
Navajo	Athabaskan
Nishnaabemwin	Algonquian
Palauan	Austronesian
Qafar	Cushitic
Russian	Indo-European
Tariana	Arawak
Tarma Quechua	Quechuan
Totonac	Totonacan
Turkana	Nilo-Saharan
Xakass	Turkic
Yimas	Sepik-Ramu
Yukaghir	Yukaghir
Yup'ik	Eskimo-Aleut

For each example we present the phonologically distinct stems that belong to the same paradigm, and define the categories according to which the suppletion can be delineated. The database contains pointers to examples, illustrating each instance of suppletion in a particular language. In addition, there is a link to a report for each language, giving sources and enabling the user to see how decisions were made. We describe briefly for each language the morphological processes relevant for defining suppletion, and the inflectional system (major word classes and the categories they inflect for). We list the instances of suppletion and give examples of regular inflected items for contrast. In the cases where our analysis of the language material differs from

that of the source, we present both views and give our reasons for deciding whether or not to include this particular example in the database.²

Users can query the database online. Besides obvious searches, such as by language, it is also possible to do cross-linguistic searches in terms of semantic and morpho-syntactic categories. The web interface provides the user with pulldown menus for each of the relevant categories. There are three readme files with the database to aid initial searching.

3.3 *Some Initial Results*

A first observation is that suppletion is relatively common cross-linguistically. Out of the 34 language surveyed, in only four could we find no instances of suppletion (recall that in order to be included a language had to have inflection and hence the theoretical possibility of inflectional suppletion). The four languages are: Navajo, Tarma Quechua, Yukaghir and Yup'ik.

The database contains 178 lexical items and 417 stems. Among the morphological features involved in suppletion it is interesting to note person (in verbs) in Totonac; possession in Jacalteco and Nishnaabemwin; politeness in Japanese and Tetelcingo Nahuatl; and negation in Russian, Limbu and Hua.

It is true that the lexical items involved are usually frequent items like 'go' and 'child'. But that is not invariably the case. The Nakh-Daghestanian language Archi has the following remarkable suppletive item: *bič 'ni* (SG) / *boždo* (PL) 'corner of a sack' (Kibrik 1977: 46). Some results are presented in Hippiisley, Chumakina, Corbett & Brown (2004); we intend to continue exploiting the database, in parallel with other researchers.

4. **A Bibliography of Suppletion**

A second resource relating to suppletion is the annotated bibliography whose entries we give here. The related abstracts are accessible at <http://www.surrey.ac.uk/LIS/SMG/>

- Aski, J.M. (1995), *Verbal suppletion: an analysis of Italian, French and Spanish 'to go'*, "Linguistics" 33. 403–432.
- Beckmann, W. (2002), *Suppletion in Niederdeutschen*, *Niederdeutsche Studien*, 47, Köln, Böhlau.
- Bhat, Shankara D.N. (1967), *Lexical suppletion in baby talk*, "Anthropological Linguistics" 9, No 5. 33–36.
- Bittner, A. (1988). *Reguläre Irregularitäten: Zur Suppletion im Konzept der Natürlichen Morphologie*, "Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung" 41, 416–425.
- Börjars, K., C. Chapman, & N. Vincent (1997), *Paradigms, periphrases and pronominal inflection: a feature-based account*, "Yearbook of Morphology" 1996, ed. by G. Booij and J. van Marle, 155–180, Dordrecht, Kluwer Academic Publishers.

² The issue of reproducibility is discussed in Corbett (2004).

- Bortnik, N. (1979), *Nekotorye osobennosti osnov, obrazujuščix dvojnye formy množestvennogo čisla* (Some peculiarities of stems that have two alternating forms for the plural), "Russian Language Journal", Vol. 32, part 112, 43–58.
- Boyd, J. (2003), *Cognitive limits on natural language suppletion*, (Not published).
- Carstairs, A. (1988), *Some implications of phonologically conditioned suppletion*, "Yearbook of Morphology" 1988, ed. by G. Booij and J. van Marle, 67–94, Dordrecht, Kluwer Academic Publishers.
- Carstairs, A. (1990), *Phonologically conditioned suppletion*, in *Contemporary morphology*, ed. by W.U. Dressler, H.C. Luschützky, O.E. Pfeiffer & J.R. Rennison, Berlin - New York, Mouton de Gruyter, 17–23.
- Carstairs-McCarthy, A. (1994), *Suppletion*, in *The Encyclopedia of Language and Linguistics*, Vol. VIII, ed. by R. E. Asher, Oxford, Pergamon, 4410–4411.
- Deshpande, M.M. (1992), *Justification for verb-root suppletion in Sanskrit*, "Historische Sprachforschung" 105, Part 1, 18–49.
- Dressler, W.U. (1985), *On the predictiveness of Natural Morphology*, "Journal of Linguistics" 21. 321–337.
- Dressler, W.U. (1985), *Sur le statut de la suppléance dans la morphologie naturelle*, "Langages" 78. 41–56.
- Dressler, W.U. (1985), *Suppletion in word formation*, in *Historical semantics – historical word-formation*, ed. by J. Fisiak, Berlin: Mouton de Gruyter, 97–112.
- Dressler, W.U. (1986), *Inflectional suppletion in natural morphology in Languages in Global Perspective*, ed. by B.F. Elson, Dallas, Summer Institute of Linguistics, 97–111.
- Fertig, D. (1998), *Suppletion, natural morphology and diagrammaticity*, "Linguistics" 36. 1065–1091.
- Gorbačevskij, A.A. (1967), *K voprosu o putjax vzniknovenija suppletivnyx form v slavjanskix jazykax*, Dušanbe.
- Grasserie, R. de la (1895), *Du syncrétisme pronominal*, Paris, Jean Maisonneuve. [Reprinted from: "Revue de Linguistique et de Philologie Comparée" 28, 47–72, 113–157.]
- Greenberg, J.H. (1997), *The Indo-European first and second person pronouns in the perspective of Eurasiatic, especially Chukotkan*, "Anthropological Linguistics" 39. 187–195.
- Hale, K., J. LaVerne & P. Pranka (1990), *On suppletion, selection and agreement*, in *Interdisciplinary approaches to languages*, ed. by C. Georgopoulos & R. Ishinara, Dordrecht, Kluwer, 255–270.
- Hendrick, R. (1996), *Some syntactic effects of suppletion in the Celtic copulas*, in *The Syntax of the Celtic Languages: A comparative perspective*, ed. by R. Borsley & I. Roberts, Cambridge, Cambridge University Press, 75–96.
- Hippisley, A., M. Chumakina, G.G. Corbett & D. Brown. (2004), *Suppletion: frequency, categories and distribution of stems*, "Studies in Language" 28, 387–418.
- Jackendoff, R. (1975), *Morphological and semantic regularities in the lexicon*, "Language" 51, 639–671.
- Johnson, D. B. (1972), *Semi-suppletive aspect pairs in Slavic*, in *The Slavic Word*, ed. by D.S. Worth, The Hague, Mouton, 169–188.

- Juge, M.L. (1999), *On the rise of suppletion in verbal paradigm*, in *Proceedings of the twenty-fifth annual meeting of the Berkeley Linguistics Society*, ed. by S.S. Chang, L. Liaw & J. Ruppenhofer, Berkeley, BLS, 183–194.
- Koneckaja, V.P. (1973), *Suppletivizm v germanskix jazykax*, Moskva, Nauka.
- Lamarche, J. (1996), *Gender Agreement and Suppletion in French*, in *Grammatical theory and Romance languages: selected papers from the 25th Linguistic Symposium on Romance Languages*, ed. by K. Zagana, Amsterdam, Benjamins, 145–157.
- Letoublon, F. (1985), *Il allait, pareil à la nuit. – Les verbs de mouvement en grec: suppletivisme et aspect verbal*, Paris, Klincksieck.
- Levin, S. (1972), *Non-paradigmatic forms: suppletion or preemption?* “Foundations of Language” 8. 346–359.
- Loi Corvetto, I. (1989), *Anomalia e paradigmi. Il suppletivismo nelle lingue romanze*, “Annali della Facoltà di Magistero” 34, Cagliari, Università di Cagliari.
- Maiden, M. (1992), *Irregularity as a determinant of morphological change*, “Journal of Linguistics” 28, 285–312.
- Maiden, M. (1995), *A proposito dell'alternanza esce, usciva, in italiano*, “Lingua Nostra” 56, 37–41.
- Maiden, M. (2001), *Morphological bootstrapping? Paradigmatic autonomy and the rise of suppletion*, Paper read at ICHL, Melbourne.
- Mańczak, W. (1966), *La nature du supplétivisme*, “Linguistics” 28, 82–89.
- Markey, T.L. (1969), *The verbs varda and bliva in Scandinavian: With special emphasis on Swedish*, “Studia philologiae Scandinavicae Upsaliensia” 7, Acta Universitatis Upsaliensis, Uppsala Universitet.
- Markey, T.L. (1985), *On suppletion*, “Diachronica” II:1. 51–66.
- Mel’čuk, I.A. (1976), *On suppletion*, “Linguistics” 170. 45–90.
- Mel’čuk, I.A. (1983), *Suppletivologičeskie ètjudy*, “Russian Linguistics” 7. 207–218.
- Mel’čuk, I.A. (1992), *Suppletion*, in *International Encyclopedia of Linguistics*, ed. by W. Bright, New York, Oxford University Press, 97–99.
- Mel’čuk, I.A. (1994), *Suppletion: Toward a logical analysis of the concept*, “Studies in Language” 18, 339–410.
- Mel’čuk, I.A. (2000), *Suppletion*, in *Morphologie / Morphology: Ein internationales Handbuch zur Flexion und Wortbildung / An International Handbook on Inflection and Word-Formation 1*, ed. by G. Booij, C. Lehmann & J. Mugdan, Berlin - New York, Mouton de Gruyter, 510–524.
- Myhill, J. (2001), *Suppletion, lexical meaning, semantic primitives, and translation data*, “Linguistics” 39, 761–802.
- Nicholson, J.G. (1988), *Correlative dependencies of a suppletive type in the East Slavic verbal paradigm*, “Canadian Slavonic Papers” 30, part 1, 34–40.
- Nicholson, J.G. (1989), *The incidence of suppletion within inflection*, “Lacus Forum” 16, 250–259.
- Nübling, Damaris. (2000), *Prinzipien der Irregularisierung. Eine kontrastive Analyse von zehn Verben in zehn germanischen Sprachen*. Tübingen: Max Niemeyer Verlag.
- Osthoff, H. (1899), *Vom Suppletivwesen der indogermanischen Sprachen*, Heidelberg, Kommissionsverlag von Alfred Wolff.
- Plank, F. (1994), *Homonymy vs. suppletion: A riddle (and how it happens to be solved in ...)*, “EUROTYP Working Papers” VII/23. 81–86.

- Ronneberger-Sibold, E. (1987), *Verschiedene Wege zur Entstehung von suppletiven Flexionsparadigmen: deutsch gern – lieber – am liebsten*, in *Beiträge zum 3. Essener Kolloquium über Sprachwandel und seine bestimmenden Faktoren*, ed. by N. Boretzky, W. Enninger & T. Stolz, Bochum, Brockmeyer, 243–264.
- Ronneberger-Sibold, E. (1988). *Entstehung von Suppletion und Natürliche Morphologie*, “Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationforschung” 41, 4, 453–462.
- Ronneberger-Sibold, E. (1990), *The genesis of suppletion through morphological Change*, in *Proceedings of the Fourteenth International Congress of Linguistics. Berlin/GDR, August 10-August 15, 1987*, Vol. I, ed. by W. Bahner, J. Schildt, & D. Viehweger, Berlin: Akademie-Verlag, 628–631.
- Rudes, B.A. (1980), *On the nature of verbal suppletion*, “Linguistics” 18, 655–676.
- Schwarze, C. (1970), *Suppletion und Alternanz in Französischen*, “Linguistische Berichte” 6, 21–34.
- Sokolova M.A. (1952), *Iz istorii slov osnovnogo slovarnogo fonda russkogo jazyka*, “Doklady i soobščeniya Instituta Jazykoznanija AN SSSR” 2. 8–19.
- Swiggers, P. (1985), *How to order eggs in French*, “Folia Linguistica” 19. 63–66.
- Tanaka, T. (2002), *The Origin and Development of the *es vs. *wes suppletion in the Germanic copula: from a Non-Brugmannian Standpoint*, “NOWELE” 40 (April), 3–27.
- Testen, D. (1997), *The suppletive imperative of Arabic ‘come’*, in *Perspectives on Arabic Linguistics X: Papers from the Tenth Annual Symposium on Arabic Linguistics*, ed by M. Eid & R.R. Ratcliffe, Amsterdam, Benjamins, 173–187.
- Theakston, A.L., E.V.M. Lieven, J.M. Pine & C. F. Rowland. (2002), *Going, going, gone: the acquisition of the verb ‘go’*, “Journal of Child Language” 29. 783–811.
- Tranel, B. (1993), *On Suppletion and French Liaison*, in *Romance languages and modern linguistic theory: 20th Linguistic symposium on Romance languages*, ed. by P. Hirschbühler and K. Koerner, Amsterdam, Benjamins, 269–308.
- Tuggy, D. (1997), *Rule-governed allomorphy can be suppletive also*, in *Work papers of the Summer Institute of Linguistics, University of North Dakota Session 41*. 65–69.
- Unbegaun, B.O. (1969), *Les anciens Russes vus par eux-memes*, in *Selected Papers on Russian and Slavonic Philology*, Oxford, Clarendon Press, 272–286.
- Veselinova, L. (2000), *Suppletion in the paradigms of copula verbs*, in *Cognition and function in language*, ed. by B.A. Fox, D. Jurafsky & L.A. Michaelis, Stanford, Center for the Study of Language and Information, 65–83.
- Veselinova, L. (2003), *Suppletion in verb paradigms: Bits and pieces of a puzzle*, Ph.D. dissertation, Department of Linguistics, Stockholm University.
- Veselinovič, E. (2003), *Suppletion im irischen Verb*, (Philologia – Sprachwissenschaftliche Forschungsergebnisse, 55), Hamburg, Verlag Dr. Kovač.
- Vincent, N. & K. Börjars. (1996). *Suppletion and syntactic theory*, in *Papers from the 1st LFG Colloquium. Grenoble*, ed. by M. Butt and T. Holloway King, 448–462.
- Welna, J. (2001), *Suppletion for suppletion, or the replacement of ēode by went in English*, “Studia Anglica Posnaniensia” 36, 95–110.
- Werner, O. (1977), *Suppletivwesen durch Lautwandel*, “Salzburger Beiträge zur Linguistik” 3. 269–283.

- Werner, O. (1987), *Natürlichkeit und Nutzen morphologischer Irregularität*, in *Beiträge zum 3. Essener Kolloquium über Sprachwandel und seine bestimmenden Faktoren*, ed. by N. Boretzky, W. Enninger & T. Stolz, Bochum, Brockmeyer, 289–316.
- Werner, O. (1987), *The aim of morphological change is a good mixture – not a uniform language type*, in *Papers from the 7th International Conference on Historical Linguistics. Pavia, Italy*, ed. by A. Giacalone Ramat, O. Carruba and G. Bernini, Amsterdam, Benjamins, 591–606.
- Werner, O. (1991), *The incorporation of Old Norse pronouns into Middle English. Suppletion by loan*, in *Language Contact in the British Isles: Proceedings of the Eighth International Symposium on Language Contact in Europe, Douglas, Isle of Man*, ed. by S. Ureland and G. Broderick, Tübingen: Max Niemeyer Verlag, 369–401.
- Wurzel, W.U. (1985), *Die Suppletion bei den Dimensionsadjektiven*, “Linguistische Studien” (Zentralinstitut für Sprachwissenschaft, Akademie der Wissenschaften der DDR) A126, 114–143.
- Wurzel, W.U. (1987), *Zur Morphologie der Dimensionsadjektive*, in *Grammatische und konzeptuelle Aspekte von Dimensionsadjektiven*, ed. by M. Bierwisch and E. Lang, Berlin, Akademie-Verlag, 459–516.
- Wurzel, W.U. (1990), *Gedanken zu Suppletion und Natürlichkeit*, “Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung” 43, 86–91.

5. Conclusion

Suppletion is indeed a challenge for morphologists and typologists. There are some remarkable instances, which push back the boundary of what is a ‘possible word’. By constructing the database, compiling the bibliography, and making both generally available, we hope to contribute to a better understanding of this extreme phenomenon.

References

- Brown, D., M. Chumakina, G.G. Corbett and Andrew Hippisley. (2004), *The Surrey Suppletion Database*, available at: <http://www.smg.surrey.ac.uk/>
- Carstairs-McCarthy, A. (1994), *Suppletion*, in R. E. Asher (ed.) *Encyclopedia of Language and Linguistics*: 8, 4410–4411, Oxford, Pergamon.
- Chumakina, M. (2004), *An annotated bibliography of suppletion*, available at: http://www.surrey.ac.uk/LIS/SMG/Suppletion_BIB/WebBibliography.htm
- Corbett, G.G. (2003), *Agreement: Canonical instances and the extent of the phenomenon*, in G. Booij, J. DeCesaris, A. Ralli and S. Scalise (eds.) *Topics in Morphology: Selected papers from the Third Mediterranean Morphology Meeting (Barcelona, September 20-22, 2001)*, 109–128, Barcelona Universitat Pompeu Fabra.
- Corbett, G.G. (2004), *Suppletion in personal pronouns: theory versus practice, and the place of reproducibility in typology*, “Linguistic Typology” 8.3.
- Hippisley, A., M. Chumakina, G.G. Corbett and D. Brown. (2004), *Suppletion: frequency, categories and distribution of stems*, “Studies in Language” 28, 387–418.
- Kibrik, A.E. (1977), *Opyt strukturnogo opisaniya arčinskogo jazyka: II: Taksonomičeskaja grammatika*. (Publikacii otdelenija strukturnoj i prikladnoj lingvistiki 12), Moscow, Izdatel'stvo Moskovskogo universiteta.

- Mel'čuk, I.A. (1994), *Suppletion: toward a logical analysis of the concept*, "Studies in Language" 18, 339–410.
- Osthoff, H. (1899), *Vom Suppletivwesen der indogermanischen Sprachen*. Heidelberg, Kommissionsverlag von Alfred Wolff.
- Veselinova, L. (2003), *Suppletion in verb paradigms: bits and pieces of a puzzle*. PhD dissertation. Department of Linguistics. Stockholm University.
- Veselinovič, Elvira M. (2003), *Suppletion im irischen Verb* (Philologia 55), Hamburg: Verlag Dr. Kovač.

Morphological Universals and Diachrony

Stephen R. Anderson*
Yale University
stephen.anderson@yale.edu

Although linguistics is plausibly taken to be “the science of language”, the actual object of inquiry in the field has changed considerably over time. Prior to the influence of de Saussure in the first part of the twentieth century, linguists concerned themselves primarily with the ways in which languages have developed historically. For the next several decades, they devoted their attention to the external facts of sounds, words and sets of utterances. With the advent of the cognitive (or “Chomskyan”) revolution around 1960, however, they came increasingly to see themselves as studying the human language faculty: speakers’ knowledge of language and the cognitive capacity that makes this possible (Anderson & Lightfoot 2002), *Universal Grammar*. This is what our theories attempt to represent nowadays.

Unlike the documented facts of language history or the measurable properties of sounds and utterances, such a cognitive faculty is not directly observable, so the question naturally arises of how we might study it empirically. Two important modes of argument have emerged that are generally taken to aid in this enterprise. First, if we can show that speakers know something about their language for which relevant evidence is not plausibly present in the input on the basis of which they learned the language, we assume that this knowledge must be a consequence of the structure of the ‘language organ’. This is the argument from “the poverty of the stimulus”, and (despite the skepticism of some: e.g. Pullum & Scholz 2002) it has proven to have wide applicability, especially with respect to speakers’ knowledge of syntax.

A second line of argument is to assume that when we find that something is true of all (or at least nearly all) of the languages we can observe, it must be true of Language more generally, and thus a property of the human language faculty. The assumption that valid generalizations about language typology must be reflected in constraints within linguistic theory is widely agreed to, but is it really valid? Why should we believe that observed regularities across languages are a good guide to the structure of the language organ?

We can note that knowledge of language arises in the individual through the application of some learning strategy – a strategy that may be partly specific to the domain of language, and partly more general – to the data available during a sensitive period in early life. As a result, regularities which we find in the grammars attained by human speakers might have a variety of sources:

The Input Data: Only systems that correspond to the evidence available can be acquired.

* I am grateful to the participants in the Catania meeting, especially Paul Kiparsky and Alice Harris, for comments, questions, and suggestions relevant to this paper. The influence of the recent work of Juliette Blevins will be apparent.

The Learning Process: Only languages that are accessible through the procedure employed can be attained, so some cognitively possible grammars might not be learnable.

The Language Faculty: Only cognitively possible languages can be acquired, whatever abstract regularities may exist in the data.

The argument that cross-linguistic regularities provide us with evidence for the structure of Universal Grammar rests on the assumption that only the last of these is relevant. It assumes that a complete range of input data is (at least in principle) available, and that the learning system can (again, in principle) consider any possible account of those data, so that the only filter on the class of grammars acquired is the nature of the cognitive system, or Universal Grammar. But surely this is extremely implausible.

To provide a serious theory of the regularities we find across the languages of the world, we need not only a theory of the language faculty but also theories of the learning system and of various sources for regularities in the input data. In connection with the latter, an important source of regularities in the input is the nature and working of historical change. A variety of linguists from Baudouin de Courtenay to the present have suggested that many of the regularities we find in the grammars of the world's languages actually result from the fact that historical change tends to produce certain configurations and not others, rather than from cognitive limitations that would exclude the unobserved systems.

This paper examines the force of this argument as it applies to morphology. We look first at what seems to be a general correlation between case marking and verbal aspect, one which has been suggested to reflect a property of Universal Grammar, and show that the connection here is an adventitious effect of several converging patterns of diachronic change rather than a systematic property of human language. We look next at the claim that morphological theory should exclude a particular formal device, metathesis, as the marker of morphological information, and show that the observed rarity of this device has plausible roots in the pathways of historical change rather than in a limitation of the language faculty. Finally, we consider the claim that morphological information should be biuniquely related to the markers that express it, as is implicit in morpheme-based models of word structure, and find that the general tendency to such isomorphism of form and content is again a reflection of plausible historical patterns, rather than being inherent in the structure of the language organ. We then briefly draw some broader conclusions.

Case 1: Split Ergativity and Aspect

Many of the world's languages display a pattern of nominative vs. accusative marking for the subject and (direct) object of a clause only under some circumstances, while other conditions result in ergative vs. absolutive marking. Such split ergative patterns are not distributed randomly, however. Typologists have observed that in a number of such cases, nominative/accusative marking is associated with a main verb bearing imperfective aspect (or some form derived from that source), while ergative/absolutive marking is associated with perfective aspect or its descendents. It

has been widely assumed (Delancey 1981, Dixon 1994, Tsunoda 1985) that Universal Grammar should account for this correlation by positing some sort of privileged link between ergativity and perfectivity, accusativity and imperfectivity.

An alternative possibility, however, is that this apparent connection actually results from a quite different source, the pathways of historical change that produce innovations or shifts in case marking patterns. This was the conclusion of an earlier paper (Anderson 1977), in which I investigated several established sources for ergative case marking in natural language, as well as one source that leads to innovative accusative marking.

It has long been known that perfective verbal forms in many languages are historical innovations. Benveniste (1952) studied this process in a number of branches of Indo-European, and documented one common source of such perfects in the re-analysis of originally passive forms. The semantics of a sentence such as *The fish was cooked (by Julia Child)* typically includes the interpretation that the cooking in question is a *fait accompli*, and thus it is entirely plausible that the use of passives should be generalized as a way to focus on perfectivity. If the morphology of the passive is then re-interpreted as a signal of the perfect, the result is a construction in which the original, notional subject is marked with a special form (instrumental, or with a preposition such as English *by*) while the original, notional direct object appears in the same form as an intransitive subject:

- | | | | | | | | |
|-----|--------------|------------------------|---|----------------------|---|--------------------------|----|
| (1) | (Original) | NP _{Obj} -NOM | – | Verb _{Pass} | – | NP _{Sbj} -INSTR | => |
| | (Innovative) | NP _{Sbj} -OBL | – | Verb _{Pass} | – | NP _{Obj} -NOM | |

This development is widely considered to be the source of the ergative constructions found in the modern Indic languages, such as Nepali:

- | | | | | | |
|-----|---------------------------------------|------------|----------|------------|--|
| (2) | Sita-le | aluma | nun | haleko chə | |
| | Sita-ERG | potato-LOC | salt-NOM | put AUX | |
| | ‘Sita (has) put salt in the potatoes’ | | | | |

While there is still much to be said about the precise sequence of developments by which passives can give rise to later perfects, the possibility of such a development is not seriously in question for a number of languages. The perfects thus derived may themselves be re-analyzed subsequently as simple past tenses.

Assuming the original state of affairs within which this innovation takes place had a nominative/accusative system of case marking, the result is one in which (the new) perfect or past tense forms are associated with an ergative construction, while the (unchanged) non-perfect forms are associated with an accusative construction. This is a standard sort of split-ergative system, but we should note that the parameters of the split are determined by the case marking properties of the (passive) ancestor of the new perfect, not by some constraint of Universal Grammar.

In other languages, though, Benveniste (1960) documents a different source for innovative perfects. He notes that in language after language, whatever verbal expression serves to express possession is also pressed into service as a marker of the perfect – as is the case, indeed, in English, where *have* serves both functions. The expression of possession is often a transitive verb (like English *have*, Spanish *tener*,

Latin *habēō* – not cognate with *have*, etc.). In some languages, however, a distinct prepositional construction is used:

- (3) Russian: U menya Ø kniga
 at me (is) book
 ‘I have a book’
- Breton: Eur velo c’hlas am eus
 A bicycle blue at-me is
 ‘I have a blue bicycle’

In case a construction of this type comes to be employed as a marker of the perfect, note the consequences. The subject of a transitive perfect verb will be marked with some oblique (originally locative) case, while the object will be marked in the same way as the subject in copular constructions: as a nominative. But as in the case of perfects descended from passives, the result is a situation in which the new perfects are associated with what is formally an ergative constructions, while non-perfects are associated with the original (presumably accusative) construction. Benveniste argues that this can be seen in the origin of the Armenian perfect. Here the subject appears in the genitive, betraying the possessive origin of the construction, while the object appears in the accusative, presumably by a later extension of this case to all objects.

- (4) zayn nšan arareal ēr nora
 that miracle-ACC performed AUX he-GEN
 ‘He performed that miracle’

Benveniste proposes that the Old Persian form *ima tya manā krtam* ‘that is what I have done’ represents this same evolution of a perfect from a possessive in a “pure” form (i.e., without extension of the accusative to the object).

Again, we have a split ergative system in which the perfect is associated with ergative marking, the imperfect with accusative marking. The two developments (from passives and from possessive constructions) have nothing to do with one another, and in neither instance is the case marking of the original construction mandated by Universal Grammar. The two developments happen to converge however, on systems with the same inherited, synchronically accidental) correlation of case marking and verbal aspect.

A third, completely independent, development can also lead to the same result. Suppose that instead of innovating a perfect, a language were to reanalyze some construction as an imperfect verbal form. What original structure might be appropriate for this purpose? A plausible candidate would be a structure in which the object of a transitive verb, instead of being marked with a direct case such as the accusative, appears as a prepositional adjunct. English has a number of contrasting pairs of this sort:

- (5) a. i. Jones read *War and Peace* to his wife.
 ii. Jones read to his wife from *War and Peace*.
- b. i. Fred shot my cat.
 ii. Fred shot at my cat.

In each of these pairs, the (ii) example is interpreted as an action not necessarily completely carried out, the object not completely affected, etc. Similar pairs form the basis of comparable contrasts in a wide range of languages, as discussed in Anderson (1988). The constructions in question clearly overlap semantically with the verbal notion of an ‘imperfective’. It would therefore be plausible for a language wishing to develop such a category to take as the starting point a structure in which a transitive verb is constructed intransitively, with its notional object appearing in an oblique or prepositional form.

This is exactly what has happened in the history of Georgian, according to a suggestion originating with Braithwaite (1973), developed in Anderson (1977), and made much more precise in Harris (1985). On this account Georgian was originally a consistently ergative language. In the course of its history, a new series of imperfective forms developed from an ‘object demotion’ construction similar to (5). These forms underlie what are now called the ‘series I’ tenses, in which case marking is nominative/accusative. A different set of forms, the ‘series II’ tenses, continues the original situation.

Roughly, the division between series I and series II tenses can be seen as (originating in) a difference between imperfective and perfective forms. Again, as with the two paths of development for new perfects summarized above, the result is a split between ergative perfects and accusative imperfects. Again, however, this split should not be seen as mandated by Universal Grammar, but rather as the accidental consequence of the formal properties of the earlier construction on which the innovated forms – here the imperfectives, as opposed to the perfectives in the earlier cases – are based.

These completely independent developments all happen to converge on the same kinds of data. Each results in a state of affairs in which perfective forms (or their descendants) are associated with an ergative pattern, while imperfectives (or their later reflexes) are associated with nominative/accusative patterns. This is not, however, due to some regularity stipulated by Universal Grammar which relates case marking and verbal aspect: rather, it is an epiphenomenal regularity that emerges from a number of unrelated lines of development. This should suggest to us that not every pattern we can find in the data of language typology reflects the structure of the language faculty directly.

Case 2: Morphological Metathesis

Another set of issues revolves around the question of whether morphological theory should countenance the possibility of rules of metathesis: rules which simply re-arrange the sequence of segmental material in a form to mark a grammatical category, with no concomitant addition of an affix or other marker. Some morphologists have argued that morphological metathesis rules ought to be excluded in principle from the theory, because such rules are (by definition) unformulable as concatenative affixes. Accommodating them would seem to entail a theory involving the full power of “the extremely rich transformational notation” (McCarthy 1981:373), an undesirable result if we hope to provide a restrictive account of the notion “possible morphological system”.

The possibility of metathesis (by itself) as a grammatical mechanism was first raised as a theoretical issue in Thompson & Thompson (1969), who cited a small

number of potential cases. Although some of these have resisted all attempts to reduce them to affixal morphology, the number of cases is undeniably quite small, and this has led researchers to hope that the remaining ones would eventually yield to re-analysis as well, allowing for the preservation of the notion that all morphology is affixation.

Arguing that although rare, morphological metathesis must nonetheless be accommodated by a general theory of morphology, Janda (1984) proposes that the explanation for the very small number of plausible cases is rooted in facts about historical change. He argues that morphological metathesis is rare because historical changes that might lead to such a situation are rare. Non-affixal morphology arises when an originally phonological alternation is reanalyzed as morphologically conditioned. But Janda argues that *phonological* metathesis processes are quite rare, and thus the opportunity for a language to morphologize such a rule is hardly ever presented.

This argument has an affinity with the program of Evolutionary Phonology proposed recently by Juliette Blevins (to appear). She argues that much of what we find (or fail to find) in synchronic phonologies is not a product of the basic structure of the human language faculty (as represented by linguistic theories of various domains). Instead, many (perhaps most) typological generalizations result from the pathways of historical change and their results. If historical change operates in such a way as to favor or disfavor certain situations, its results are what we will find, and such generalizations are thus at best a poor guide to the structure of the language faculty itself.

Going back to Baudouin de Courtenay (1895 [1972]), still one of the most comprehensive reviews of the processes governing the “life cycle” of alternations, we see that the main path by which morphological processes emerge is when an originally phonological regularity becomes increasingly opaque as a result of other changes. When the phonological conditioning factors for an alternation become lost (or at least difficult to recover from surface forms), it may be reinterpreted as aligned with morphological factors. To the extent phonological bases for such a change are lacking, we would expect the corresponding morphological rules to be rare or absent, regardless of the character of morphological theory *per se*.

Unfortunately for the viability of this explanation, phonological rules of metathesis are actually not rare. In a series of papers devoted to this subject, Blevins and Garrett (1998, to appear) have shown that there are several systematic types of sound change that can result in phonological metathesis rules, and that a substantial number of such processes do in fact exist in a wide variety of languages. If morphological metathesis is rare, then, it cannot be because there are no phonological processes to serve as its precursors.

Given that synchronic phonological metathesis is a real (and not especially rare or exotic) phenomenon, a historical explanation for the rarity of corresponding morphology must take some form other than the one proposed by Janda. Let us ask how morphological metathesis might be expected to arise in a grammar. As noted above, this is most likely where antecedent phonological processes have become opaque as a result of later changes. Eventually, language learners come to align the alternation with some grammatical category, rather than with a phonological trigger whose presence in the environment is highly abstract or perhaps no longer visible at all. On that basis, we can ask how plausible it is for phonological metathesis to be reanalyzed as morphological in this way.

Blevins and Garrett, in the works cited above, identify four categories of phonological metathesis processes:

Perceptual metathesis, in which a phonetic property realized over a multisegmental span of the utterance becomes misallocated and is attributed to a segment other than the one from which it originates in the sequence.

Compensatory metathesis, in which a foot-peripheral syllable node is lost and the phonetic content originally assigned to it is re-assigned in a way that does not respect the original phonetic sequence.

Coarticulatory metathesis, in which overlap of gestures in adjacent segments leads to ambiguity with respect to their original order.

Auditory metathesis, in which fricative noise becomes decoupled from the sequential speech stream and re-assigned to a location other than its original one.

Of these possibilities, compensatory metathesis does not really count, because the primary operation involved is not a re-ordering but rather the loss of prosodic structure, with “metathesis” emerging as a concomitant. One of the instances cited both by Thompson & Thompson (1969) and Janda (1984), the formation of the incomplete phase in Rotuman, has been shown conclusively by McCarthy (2000) to have this character, and Blevins & Garrett (1998) exclude it from the class of true phonological metathesis processes on that account.

The remaining three types of metathesis are each limited to specific combinations of segment types: laryngeal, rhotic, etc. and vowel for the perceptual type; $p+k$ (becoming $k+p$) for the coarticulatory type; and sibilant plus stop for the auditory type. Crucially, in all three varieties, the conditioning factors are entirely internal to segments undergoing the positional interchange. That is, there is no external conditioning factor for any of these processes, such that that aspect of the structural description could become opaque or be lost altogether. Since the elements that undergo the change are themselves its trigger, the normal historical processes of morphologization can gain no foothold.

Compare this situation with processes such as Umlaut, for example, in which some element (e.g., a high front vowel or glide in a succeeding syllable) conditions the change but is not part of it. When this element itself undergoes change (e.g., reduction to schwa in unstressed syllables), the alternation can persist in morphologized form. No such development is possible for the well established types of phonological metathesis, however.

If there is no natural path by which phonological rules of metathesis can be morphologized, does this mean that metathesis is confined to the phonological domain? No, for while the re-analysis of a corresponding phonological rule may be the most straightforward source for a morphological rule, it is not the only one. In fact, the case which was first cited (by Thompson & Thompson 1969) in this regard, the relation between the “non-actual” and the “actual” forms of the verb in Northern Straits Salish languages like Klallam and Saanich, turns out to be a valid instance of “metathesis as a grammatical device”.

In Klallam pairs like those in (6), for example, a sequence of consonant plus vowel in the “non-actual” form is inverted to produce the “actual” (a form with a semantic interpretation that includes that of the English present progressive), with no accompanying affix or other factor that could be said to condition the change.

(6) Klallam: CCV → CVC

<u>Non-Actual</u>	<u>Actual</u>	<u>gloss</u>
qq'í-	qíq'-	'tie up, restrain'
pk ^w ǎ-	pǎk ^w -	smoke
čk ^w ú-	čúk ^w -	shoot

Where does such a relation originate, if not in an originally phonological rule of metathesis? Demers 1974 argues that in the related language Lummi, the original process involved a rule copying vowels (converting CCV into CVCV), followed by a shift of stress in the resulting forms (converting CVCV to CVCV̄), and finally loss of the unstressed vowel to yield CVC. This sequence is plausible as a historical account of the origins of the form of the “actual”, and may even be valid as a synchronic analysis of the facts of Lummi. Unfortunately, however, the crucial rules are not operative in Klallam, or in another relevant language, Saanich (Montler 1986, 1989):

(7) Saanich: CCəC → CəCC

<u>Root</u>	<u>Non-Actual</u>	<u>Actual</u>	<u>gloss</u>
θk ^w -	θk ^w ət	θək ^w t	straighten (something)
t's-	t'sət	t'əst	break (something)
t ^θ łək ^w ,	t ^θ łək ^w ,	t ^θ əłk ^w ,	pinch (something)
ʒ'pəx	ʒ'pəx	ʒ'əpx	scatter (something)
x ^w q'p'ət	x ^w q'p'ət	x ^w q'əp't	patch (something)

The Saanich facts are discussed by Stonham (1994), who offers an analysis on which the alternations in (7) do not instantiate grammatically conditioned metathesis, but are rather the result of the addition of a mora in the actual forms with concomitant re-organization of segmental material. Stonham's account involves unusual assumptions about the nature of the association between segmental and prosodic structure, but in any event it does not extend to a full range of the relevant cases. As he notes (Stonham 1994:175f.), metathesis of a CCV root to CVC would close the syllable, thus plausibly satisfying a constraint that the “actual” should have one mora more than the “non-actual” (assuming it could be shown that Saanich and Klallam are languages in which coda consonants are moraic, which is not obvious from the rest of their phonology). But the forms in (7) do not conform to this description. Montler (1989) shows that roots like the first two are actually vowel-less in their basic form, and become eligible for conversion to an “actual” form through the addition of a stressable suffix such as -'ət 'control transitive' which already has a closed syllable. Metathesis would thus not have the desired effect of adding a mora to such stems. The same is true of any root whose basic form already contains a coda consonant, such as the last three in (7), where the transposition of a prevocalic consonant into the coda cannot be said to

satisfy such a prosodic requirement for an additional mora. We could only reconcile these examples with Stonham's analysis by assuming that multiple coda consonants can contribute multiple moras to the prosodic weight of a form, something that has not been claimed for any language and which would be extremely hard to justify. See also Kurisu (2001) for discussion of this case, which we must conclude is a genuine (if isolated) instance of "metathesis as a grammatical device."

Cases of this sort do not counter-exemplify the claim above that natural processes of historical change do not produce morphological metathesis rules from originally phonological metatheses. The reason is that the origin of the non-actual metathesis in Salish is apparently something like the path identified by Demers. As such, it is a matter of restructuring rather than simply morphologization. Processes of rule inversion, telescoping, and the like were identified at least as early as Bach & Harms (1972) as the source of "crazy rules," rules cut off from their original phonetic motivation through the ongoing reanalysis of alternations by successive generations of speakers. This is a known source of grammatically conditioned metathesis: Garrett & Blevins (2004) discuss other instances in which metathesis rules have arisen within the Lexical Phonology of a language through restructuring without having a source in a phonetically natural metathesis process.

However inconvenient this may be for theories that assume all morphology to be based on affixation, then, it is necessary for morphological theory to recognize purely non-affixal markers for grammatical categories. If such markers are rare, the explanation for that fact is to be sought not in the nature of the human cognitive capacity for language, but rather in the paucity of historical scenarios that could yield such a process in practice.

This should not be particularly surprising, if we look at a broad range of evidence for the nature of the capacity with whose structure we are concerned. Language games, secret languages, and similar systems show widespread use of re-ordering, as is evident from a systematic survey such as that of Bagemihl (1988). These often instantiate processes which are extraordinarily unlikely ever to be found in any naturally occurring language. One might claim, of course, that such systems are outside the scope of normal language, but the facility with which they are acquired and used in a wide range of the world's cultures makes that unlikely. Indeed, Bagemihl shows that the processes that set them apart from "normal" systems can be precisely placed with respect to the rest of the grammar, and that it is really only their unusual content that differentiates them from other rules of phonology and morphology.

We should probably conclude that the rules of such systems display a freedom not available to naturally occurring languages precisely because they are not constrained to arise through the usual processes of historical change. Their rules need not originate in perceptual or articulatory effects of the sort argued by Blevins (to appear) to underlie changes of the more familiar sort, but are constrained only by the imaginations of speakers. Further, since there is no "intelligibility constraint" on the relation between the base language and a secret or language-game variant (indeed, precisely unintelligibility is sometimes the essence of this relation), these can differ much more dramatically than in the case of systems developed through transfer of a language across generations. These examples provide us with a kind of laboratory, then, in which we can observe some of the differences between what is "natural" (in terms of our phonetically based expectations) and what occurs in nature. The existence of grammatically conditioned metathesis rules is not at all unexpected in this context.

Case 3: Multiple Exponence

A number of views of morphology assert, as a matter of theoretical necessity, that a single category of content which is reflected in a given word must be indicated by exactly one formal marker (Halle & Marantz 1993, Noyer 1992, Steele 1995). That is, they deny the possibility of what some (e.g. Matthews 1972) refer to as “extended” or “multiple exponence”, in which the same category is reflected formally in two or more distinct components of the word’s morphology. The more seriously one is attached to a model based on the classical notion of the “morpheme” (an irreducible one-to-one association of a piece of form with a piece of content, the minimal Saussurean sign), the more important this matter becomes.

A historical perspective might suggest that the requirement of simple or unique exponence of morphological categories is a plausible one. Morphological markers typically represent pieces of form that have gradually shifted in status over time from fully independent words through phonological reduced forms (“simple” clitics) to clitics more intimately associated with their host, eventually becoming affixes. If this path of development is indeed the origin of all morphological markers, it makes sense that the components of content within a given word should be bi-uniquely related to the components of its form.

Apparent counter-examples to the requirement of uniqueness of exponence are typically dismissed by designating one of the markers as the “real” one, and assigning other formal reflections of the same category the status either of special stem forms associated (non-distinctively) with certain categories, or of morphophonemic changes triggered by the primary marker. For instance, in German *Kraft/Kräfte* ‘strength(s)’ the category of plural appears to be marked twice, once by the ending *-e* and again by Umlaut of the stem vowel. One might say that Umlaut is a “morphologically conditioned phonological rule,” or that Umlaut is a property of a special variant of the noun’s stem; and that only the ending is a genuine plural marker. At minimum this analysis is not obvious, given the existence of other words such as *Tag/Tage* ‘day(s)’, *Jahr/Jahre* ‘year(s)’ in which the ending *-e* alone marks the plural, without Umlaut, and *Apfel/Äpfel* ‘apple(s)’, *Graben/Gräben* ‘ditch(es)’ in which Umlaut alone serves this function.

I have argued (Anderson 2001) that it is impossible to maintain the constraint of “one category, one marker” as a requirement on morphological theory in this way without completely trivializing it (as Distributed Morphology does, for instance, with its array of post-syntactic morphological manipulations including fission, fusion, impoverishment, arbitrary and stipulated morpheme-to-morpheme concord, etc.). Despite the fact that morphological categories and markers line up in a one-to-one fashion in the vast majority of cases, this cannot be a requirement on morphological structures, because in at least some cases, it is violated without any evidence that the result is ill-formed or unstable.

A particularly robust system displaying such multiple exponence is that of verbal agreement in the Kiranti languages of Nepal and neighboring areas (van Driem 1990, 1997). In a form such as Dumī *dza-ŋ-pə-t-ə* ‘I’m going to eat’ both the *-ŋ* and the final *-ə* are markers of the first person subject. Such multiple marking of the categories of a verb’s arguments is very widespread in all of these languages – indeed, it is the exception, rather than the rule, that a given argument is marked only once in a language like Dumī.

Again, we can look to historical change for the bases of (at least some) instance of multiple exponence. In *Dumi* or, somewhat more perspicuously, *Limbu* (van Driem 1987), the verbal agreement markers (apart from a limited set of prefixes) group themselves into two suffix clusters, each of which may contain markers for the same or similar properties of the same argument(s). What is responsible for this state of affairs is clear, on van Driem's reconstruction of the family.

A reasonably common historical source of agreement markers in a language is an original inflected auxiliary. Such an auxiliary may be associated with some or all (lexical) main verb forms; like other words, it may undergo reduction to a simple (and later a special) clitic, thus coming to be attached to an associated uninflected form of the lexical verb. This reduced form of the auxiliary may then come to be reinterpreted as morphology on the verbal base, rather than a separate element. The Muskogean languages, for instance, have undergone such a development, as argued originally by Haas (1969) and subsequently confirmed in the study of several of the individual languages.

What has happened in the Kiranti languages is that this developmental pattern has occurred not just once, but (at least) twice in the history of languages like *Limbu* and *Dumi*, each time leaving a new set of inflectional markers on the verb. When one examines the patterns of marking within each subset of the suffixes, it becomes clear that the pattern of marking was not the same in the two historical inflected auxiliaries that are now reflected on the verb, but the arguments with which they show agreement are the same, and many of the same category distinctions are made in both cases. The result is a pattern that displays (at least) two distinct markers on the verb corresponding to the same agreement information relevant to a given argument.

While this repeated process of auxiliary reduction is obviously unusual, it does not seem theoretically problematic, and thus the clear instance of multiple exponence to which it gives rise should not be rejected either. Though inconvenient for morpheme-based models of word structure, many-to-many relations between a word's formal markers and the categories they reflect are simply a fact of linguistic structure. Just as the predominance of one-to-one marking has its explanation in the paths of historical change (along which markers typically originate in the progressive reduction of full words), so also the exceptions to this principle have a clear motivation in the historical morphology of individual languages.

Conclusion

We conclude that what we find in language is only partially explained by what is "natural". Some things that we find in the morphology of a language are there not because the language faculty requires them but because change tends to create them for independent reasons; while some things that are rare or perhaps even non-existent are not to be found because there are few if any pathways that could produce them from an available source. These observations have surprisingly important consequences: they mean that our account of the human cognitive capacity for language cannot be based simply on generalizations about what we find in the languages of the world, or on what can be grounded in some other domain, such as phonetics. The cognitive capacity we hope to capture may well be much

more flexible than we might think at first glance, and as a result, it may be considerably harder to determine its properties than has been assumed.

References

- Anderson, S.R. (1977), *On mechanisms by which languages become ergative*, in C. Li (ed.), *Mechanisms of syntactic change*, Austin, Texas, University of Texas Press, 317–363.
- Anderson, S.R. (1988), *Objects (direct and not so direct) in English and other languages*, in C. Duncan-Rose, T. Vennemann & J. Fisiak (eds.), *On language: A Festschrift for Robert Stockwell*, Beckenham, Kent, Croom-Helm Publishers, 279–306.
- Anderson, S.R. (2001), *On some issues of morphological exponence*, “Yearbook of Morphology” 2001, 1–18.
- Anderson, S.R. & D.W. Lightfoot (2002), *The language organ: Linguistics as cognitive physiology*, Cambridge, Cambridge University Press.
- Bach, E. & R. Harms (1972), *How do languages get crazy rules?* in R.P. Stockwell & R. Macaulay (eds.), *Linguistic change and generative theory*, Bloomington, Indiana University Press, 1–21.
- Bagemihl, B. (1988), *Alternate Phonologies and Morphologies*, Doctoral dissertation, University of British Columbia.
- Baudouin de Courtenay, J. (1895 [1972]), *An attempt at a theory of phonetic alternations*, in E. Stankiewicz (ed.), *A Baudouin de Courtenay anthology*, Bloomington, Indiana University Press, 144–212.
- Benveniste, E. (1952), *La construction passive du parfait transitif*, “Bulletin de la Societe de Linguistique de Paris” 48, 52–62.
- Benveniste, E. (1960), “*Etre*” et “*avoir*” dans leurs fonctions linguistiques, “Bulletin de la Societe de Linguistique de Paris”.
- Blevins, J. (to appear), *Evolutionary phonology: The emergence of sound patterns*, Cambridge, Cambridge University Press.
- Blevins, J. & A. Garrett (1998), *The origins of consonant-vowel metathesis*, “Language” 74, 508–556.
- Blevins, J. & A. Garrett (to appear), *The evolution of metathesis*, in B. Hayes, R. Kirchner & D. Steriade (eds.), *Phonetically driven phonology*, Cambridge, Cambridge University Press.
- Braithwaite, K. (1973), *Case Shift and Verb Concord in Georgian*, PhD thesis, University of Texas at Austin.
- Delancey, S. (1981), *An interpretation of split ergativity and related patterns*, “Language” 57, 626–657.
- Demers, R. (1974), *Alternating roots in Lummi*, “International Journal of American Linguistics” 40, 15–21.
- Dixon, R.M.W. (1994), *Ergativity*, Cambridge, Cambridge University Press.
- Garrett, A. & J. Blevins (2004), *Morphophonological analogy*, in K. Hanson & S. Inkelas (eds.), *The nature of the word: Essays in honor of Paul Kiparsky*, Cambridge (Mass.), MIT Press.
- Haas, M.R. (1969), *The prehistory of languages*, The Hague, Mouton.
- Halle, M. & A. Marantz (1993), *Distributed morphology and the pieces of inflection*, in K. Hale & S.J. Keyser (eds.), *The view from building 20*, Cambridge, MIT Press, 111–176.
- Harris, A. (1985), *Diachronic syntax: The Kartvelian case*, Syntax & Semantics, vol. 18, Orlando, Academic Press.
- Janda, R.D. (1984), *Why morphological metathesis rules are rare: On the possibility of historical explanation in linguistics*, “Proceedings of the Annual Meeting of the Berkeley Linguistics Society” 10, 87–103.

- Kurisu, K. (2001), *The Phonology of Morpheme Realization*, PhD thesis, University of California, Santa Cruz.
- Matthews, P.H. (1972), *Inflectional morphology*, Cambridge, Cambridge University Press.
- McCarthy, J.J. (1981), *A prosodic theory of non-concatenative morphology*, "Linguistic Inquiry" 12, 373–418.
- McCarthy, J.J. (2000), *The prosody of phase in Rotuman*, "Natural Language and Linguistic Theory" 18, 147–197.
- Montler, T. (1986), *An outline of the morphology and phonology of Saanich, north straits Salish*, "Occasional Papers in Linguistics" 4. University of Montana.
- Montler, T. (1989), *Infixation, reduplication and metathesis in the Saanich actual aspect*, "Southwest Journal of Linguistics" 9, 92–107.
- Noyer, R.R. (1992), *Features, Positions and Affixes in Autonomous Morphological Structure*, PhD thesis, MIT.
- Pullum, G.K. & B.C. Scholz (2002), *Empirical assessment of stimulus poverty arguments*, "The Linguistic Review" 19, 9–50.
- Steele, S. (1995), *Towards a theory of morphological information*, "Language" 71, 260–309.
- Stonham, J.T. (1994), *Combinatorial morphology*, Amsterdam, John Benjamins.
- Thompson, L.C. & M.T. Thompson (1969), *Metathesis as a grammatical device*, "International Journal of American Linguistics" 35, 213–219.
- Tsunoda, T. (1985), *Split case-marking patterns in verb types and tense/aspect/mood*, "Linguistics" 21, 385–396.
- van Driem, G. (1987), *A grammar of Limbu*, New York, Mouton de Gruyter.
- van Driem, G. (1990), *An exploration of proto-Kiranti verbal morphology*, "Acta Linguistica Hafniensia" 22, 27–48.
- van Driem, G. (1997), *A new analysis of the Limbu verb*, in D. Bradley (ed.), *Tibeto-burman languages of the Himalayas*, Vol. 14 of "Papers in Southeast Asian Linguistics", Canberra, Pacific Linguistics, 157–173.

Universals and Grammatical Categories: A Distributed Morphology Analysis of Spanish Colour Terms*

Antonio Fábregas

Instituto Universitario Ortega y Gasset / Universidad Autónoma de Madrid
webfabregas@hotmail.com

1. Two Ways to Define Grammatical Categories

A universal property of natural languages that has become well-established as a result of the typological studies of the last century is that every language has grammatical categories. This explains that every grammatical theory has been concerned with the existence of different word classes – each one with distinguishing properties – that establish among them formal and conceptual relations. Therefore, one of their aims is to provide an adequate description of grammatical categories that gives an account of what the possible relationships are. In addition to this, some theories also try to propose an explanation of how a word is assigned to a particular grammatical category.

There are two approaches to explain categorisation. One answer to this question, which is rooted in philosophical tradition and can be traced back to as far as Aristotle's *Poetics*, argues that the grammatical category of a word is dependent on the meaning expressed by it. The basic tenet of this semantic approach is that there is a restricted universal set of non-definable concepts that are stored and combined in the human conceptual apparatus; from this level they are somehow projected as grammatical objects and they take a morpho-syntactic disguise. Consequently, syntax / morphology is a level that interprets semantic information, which has neither generative nor explanatory power. Word classes are the result of the grammaticalisation of notional or cognitive constructs. Semantics – and, perhaps, pragmatics – is the only autonomous level, and morpho-syntax is just a formal device to embody meaning.

This idea has been recently renewed in the morphological and syntactic literature (cf. Dixon 1982, Wierzbicka 1980, 1987, 1996, Langacker 1999, Anderson 1997 and references therein). To have one explicit statement of the contemporary tenets of this view, let us consider the following quotes:

- (1) a. I reject, however, the assumption that semantic representations, to be plausible, must be postulated jointly with rules for translating those representations into surface syntax. Recent modes favouring “autonomous syntax” notwithstanding, I would suggest that **it is semantics, not syntax, which has the right to autonomy**. The task of uncovering semantic structures is **locally prior to the task of postulating syntactic rules**.

[Apud Wierzbicka 1980: 31; emphasis mine]

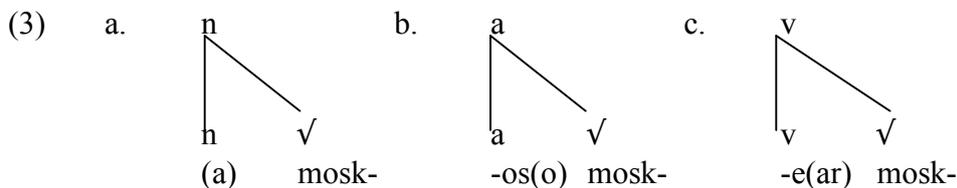
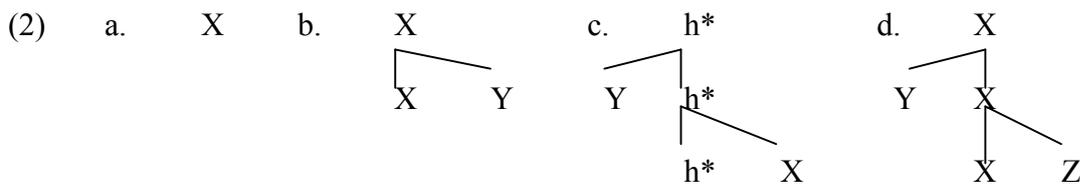
* I would like to thank Soledad Varela, Violeta Demonte, Carlos Piera and Luis Eguren for judgements and useful discussion; I am indebted to Leanna Smith for linguistic advice in the English translation of the paper. The research reported here is partially supported by a grant FPU AP2001-0299 of the Spanish MEC and the project BFF 2000 – 1307 – C03 – 02, *La variación gramatical. Variación micro y macro paramétrica en morfología y en sintaxis. Teoría, descripción, aplicaciones*. This paper would have never been made were it not for Carlota.

b. We work from the assumption that **the syntactic properties of a lexical item can largely be predicted from its semantic description**. Semantics is thus held to be prior to syntax. The ways in which syntactic properties can be predicted on the basis of semantic representations are complex, and are not yet fully understood.

[Apud Dixon 1982: 8; emphasis mine]

The other view held in contemporary linguistics roots in the development of formal syntax. The syntactic approach tries to get a definition of grammatical categories without reference to their conceptual import. In this view, grammatical categories are defined through formal means. Syntax and morphology have only very restrained access to semantic information, if they have some access at all. Due to the modularity hypothesis, grammar is blind to concepts; therefore, they cannot be invoked to explain formal properties of grammar. Consequently, the only level able to explain why a word is included in a particular grammatical category is the morpho-syntactic level. Moreover, as one of its strongest statements, this theory predicts that the independently motivated morpho-syntactic operations must be able to explain the categorisation of a word.

In the last ten years, two independently developed theories, both of them rooted in the generative framework, have argued for a formal distinction of grammatical categories. These are the Distributed Morphology framework (Halle & Marantz 1993, Marantz 1997) and the works on argument structure by Hale & Keyser (1993, 1998). The two theories agree in the following fact: syntax alone determines the category of an element, so no element belongs to a grammatical category prior to its syntactic projection. H&K admit the existence of a lexical-syntactic level where argumental structure is defined; DM argues that there is only one syntax, which is able to generate both sentences and words. H&K propose that the argumental structure of a head determines its grammatical category: there are only four lexical categories, which correspond to the four logically possible combinations of heads with a specifier and a complement (2). As for the categorisation in DM, it is claimed that category-less roots acquire their category through merge with a functional head (3).



(2a) corresponds to a non relational category, a noun; (2b) defines a head with complement and without specifier, a verb; (2c) defines a head in need of a specifier that has to merge with a head able to provide it with that specifier, an adjective; finally, (2d)

defines a relational category with both specifier and complement, a preposition¹. As for (3), there is a root without category that is defined as a noun in (3a), through merging with the lexical head *n*, as an adjective by *a* in (3b) and as a verb by *v* in (3c). Note that lexical heads materialise as affixes.

What the semantic view and the syntactic view have in common is that they are attempts to avoid the stipulation of category labels for every single morpheme of a language. In contrast, Lexicalist Morphology approaches need to state the category of every individual element in the lexicon (cf. Chomsky 1970, Siegel 1976, Lieber 1980, Selkirk 1982; note that Jackendoff 1990 also has to employ stipulative category labels).

Trying to choose between these two views on conceptual grounds may be a scholastic exercise. However, they make different predictions concerning the data. The syntactic view predicts that an element that expresses a certain concept may project in different categories, without change of conceptual meaning, depending on the formal requirements of the syntactic configuration. In other words: as what counts is syntax, it predicts that we will find the very same concept projected in different morpho-syntactic categories provided that the syntactic configurations are different. This type of mismatch will be problematic for the semantic view, for it predicts that syntax is not independent of concepts and, unless implemented with additional machinery, it will be expected that a concept will determine the syntactic configuration. Therefore, every change in syntax must be rooted on a change in conceptualisation (cf. Langacker 1999).

In this paper we will argue that there are empirical cases that confirm the predictions of the syntactic view and cast doubt on the accuracy of the semantic view. The relevant data are taken from Fábregas (2002) and regard the formal behaviour of Spanish Colour Terms (SCT).

2. The Puzzling Behaviour of Spanish Colour Nouns

Morphological properties of Spanish adjectives are quite clear. In the first place, adjectives show agreement in gender and number with a noun that must be interpreted as its semantic subject. In (4a), where the A shows feminine singular agreement, the only available reading of the sentence is that the event of outrunning the boys took place when Juana was exhausted: in (4b), where A shows masculine plural agreement, the event takes place when the boys are exhausted.

- (4) a. Juan-a adelant-ó a l-os muchach-os agotad-a
 Juan-f.sg out.run-PT.3SG (ac) the.M.PL boy-M.PL exhausted-F.SG
- b. Juan-a adelant-ó a l-os muchach-os agotad-os
 Juan-f.sg out.run-PT.3SG (ac) the.M.PL boy-M.PL exhausted-M.PL

Adjectives may combine with syntactic and morphological devices to express grade. Therefore, they may be modified by *muy*, *bastante* and *demasiado* as well as by the comparative adverbs *más* and *menos*, which license a comparative phrase (5). Adjectives in Spanish may also exhibit grade morphology, as the suffix *-isim-* (6).

¹ In H&K framework, it is possible for the two languages to parameterise the argument structure configurations in different categories. It is plausible, though, that English and Spanish have selected the same equivalences (cf. Mateu 2002), which we will assume.

- (5) a. Pedro es {muy / bastante / demasiado} alto
 Pedro is {very / quite / enough} tall
 ‘Pedro is very tall, quite tall, tall enough’
- b. Pedro es {mas / menos} alto (que Teresa)
 Pedro is {more / less} tall (than Teresa)
 ‘Pedro is taller / less tall than Teresa’

On the other hand, nouns do not show either of these properties. Noun inflection in gender and number implies a semantic difference, and therefore N’s do not agree (6).

- (6) un gato ≠ una gata, un gato ≠ unos gatos

As for grade syntax and morphology, ungrammaticality usually arises when an N is combined with adverbs such as *muy* and *más* (7a)², and with *menos*, *bastante* and *suficiente* when they do not stand for noun-modifying quantifiers (7b). Grammaticality judgements are even clearer with the morphological superlative *-isim-* (7c).

- (7) a. *muy mesa, *más choza³
- b. #bastante despertador, #suficiente arroz, #menos lobo
- c. *reloj-isim-o, *carter-isim-a...

With these facts in mind, let us consider the following set of Spanish Colour Terms (SCT) data:

- (8) a. un-as cas-as roj-as
 some-f.pl house-f.pl red-f.pl
 ‘Some red houses’
- b. un-as cas-as roj-isim-as
 some-f.pl house-f.pl red-SPL-f.pl
 ‘Some very red houses’
- c. un-as cas-as más roj-as que la sangre
 some-f.pl house-f.pl more red-f.pl than the blood
 ‘Some houses redder than blood’
- (9) a. un-as cas-as {roj-o / *roj-as} sangre
 some-f.pl house-f.pl {red-m.sg / red-f.pl} blood
 ‘Some blood-red houses’

² A very reduced group of these combinations is possible, but note that in those cases the N has to be interpreted as a property, like in *muy hombre*, which *grosso modo* corresponds to *muy masculino*, very masculine.

³ Unless reinterpreted as properties, which is semantically implausible.

- b. *un-as cas-as roj-sim-o sangre
 some-f.pl house-f.pl red-SPL-m.pl blood
 ‘*Some very blood red houses’
- c. *un-as cas-as más rojo sangre
 some-f.pl house-f.pl more red blood
 ‘*Some more blood red houses’

In (8) the colour term behaves as expected from an A: (8a) shows that it agrees in gender and number with the N whose property it denotes. It can also combine with the superlative morpheme, as (8b) witnesses, as well as with a grade adverb that licenses a comparative clause, as (8c) shows. However, the very same element, in (9) does not have adjectival qualities. (9a) shows that agreement with the head noun is prohibited and causes ungrammaticality; note that the indefinite determiner still has to show agreement with the same head noun. As can be seen in (9b), the colour term is no more combinable with a superlative morpheme and in (9c) it can be seen that the comparative adverb is no longer available.

Actually, the (negative) properties that the colour term displays in (9) are those that one would expect from an N. In (10) it is demonstrated that Spanish CoTs also have the positive properties of N’s. Namely, the colour term is combinable with determiners and quantifiers (10a, 10b), and can be the complement of a P⁰ (10c).

- (10) a. Este rojo oscuro no me gusta nada
 ‘I don’t like this (tone of) dark red’
- b. Hay dos azules distintos en este cuadro
 ‘In this painting, there are two different blues’
- c. Lo pintó de verde
 ‘She painted it [P, of] green’

As Ns, SCTs show the regular behaviour of Mass Nouns, denoting a shapeless non-delimited substance. When inflected in plural, they express taxonomic differences between tones of that particular colour: *varios azules* may mean various types of blue.

We find the same pattern in other languages. For an illustration, consider the following data from Italian (11) and English (12).

- (11) a. una giacca grigia
- b. una giacca {grigio / *grigia}scuro
- c. una giacca {grigio / *grigia} perla
- (12) a. a red(der) carpet
- b. a dark red(*der) carpet
- c. a yellow(er) carpet

- d. a sulphury yellow(*er) carpet

The traditional analysis of sentences such as (10a) and (10b) was given by Bello (1847). This grammarian argued that in the constructions of (9) and (10) the colour term is actually an A that agrees with an elided N, *colour*. This analysis cannot be maintained for a number of reasons. First of all, note that this situation wouldn't preclude the colour term to take a superlative morpheme or to be combined with a comparative adverb, for it would still be an A. Secondly, if this analysis is correct, we would expect the colour term to surface in feminine in those languages – such as French – where the N *colour* is feminine. This prediction is not confirmed, though (13).

- (13) un jaune clair / * une jaune claire

Finally, it is a fact of the structure of Spanish NP's that the indefinite determiner *un* must surface as *uno* when followed by an empty noun (14a) (Bernstein 1993). If we had an elided noun we wouldn't expect sentences such as (14b) to be grammatical, but they are.

- (14) a. Un libro de matemáticas y un-*(o) de literatura
'One book of maths and another one of literature'
- b. Un rojo brillante
'A bright red'

Sentences where the indefinite must appear as *uno* in front of the N do exist, but, crucially, they have a different meaning. In (15a) the speaker refers to a certain tone of blue; in (15b), he or she refers to a certain individual, whose type must be inferred, with the distinguishing property that it is blue.

- (15) a. un azul 'lit. a blue'
- b. uno azul 'one blue'

Therefore, we must admit that SCTs surface as Ns and As.

The context where the CT will appear as an N can be determined on syntactic grounds. Colour terms manifest themselves as Ns if they are modified by adjectives that denote the hue or the intensity of the colour (16).

- (16) a. unas alfombras {rojo brillante / *rojas brillantes}
[lit. *some carpets* {*red.MASC.SG bright.MASC.SG* /
**red.FEM.PL bright.FEM.PL* }]
- b. unas alfombras azul verdoso oscuro
'dark greenish blue'
- c. unas alfombras amarillo grisáceo pálido
'pale greyish yellow'

- d. unas alfombras verde amarillento brillante
‘bright yellowish green’
- e. unas alfombras azul eléctrico
‘electric blue’

Among the adjectives that can modify CT we find two groups. In the first group we find non-basic colour terms, usually morphologically derived from basic colour terms, such as *amarillento*, *verdoso*, *rojizo*, *blanquecino*, *negruzco* and *grisáceo*. These precise the hue of the colour expressed by the colour. In the second group there are those adjectives that denote the intensity or the brightness of the hue expressed by the colour noun and the optional hue adjective, such as *brillante*, *pálido*, *oscuro*, *claro*, *apagado*, *eléctrico* and *intenso*. The unmarked order between these two types of adjectives is that in which the hue adjective precedes the intensity adjective.

The second syntactic context where they show as N's is when accompanied by another noun specifying the hue of the colour (17).

- (17) a. unas paredes {blanco hueso / *blancas hueso}
[lit. *some walls* {*white.M.SG bone* / **white.FEM.SG bone*}]
- b. unas paredes azul cielo
‘sky blue’
- c. unas paredes verde manzana
‘apple green’
- d. unas paredes rojo fuego
‘fire red’
- e. unas paredes gris perla
‘pearl grey’

Only nouns that express substances or entities which are straightforwardly and recognisably characterised by a particular colour can participate in this construction (Fernández Ramírez 1951).

Finally, this same situation takes place when colour terms are selected by a preposition (18).

- (18) a. teñir el jersey de {rojo / *rojísimo}
‘lit. *to dye the jersey of* {*red* / **red.SUPERLATIVE*}’
- b. pintar la pared de negro
‘to paint the wall [of] black’
- c. hacer verde con azul y amarillo
‘to make green with blue and yellow’

Colour Terms project as A elsewhere⁴.

If we want to avoid the mere stipulation that there is a process of conversion here that applies to colour terms and transforms adjectives into nouns, we have to attempt another analysis. To our knowledge, just positing a rule that takes colour adjectives and turn them into nouns does not explain what is happening here, but only highlights the fact that in a given context adjectives cannot appear and, in their place, nouns are placed. Although this is a logically possible analysis, we think that it actually means to give up trying to find an explanation. In the next section I provide an attempt of finding an explanation within the Distributed Morphology framework.

3. Minimalist Colour Terms

In the Minimalist Framework (Chomsky 1995, 1999, 2000, 2001) syntactic operations are feature-driven. There are two different types of features: interpretable features and uninterpretable features. While the former are necessary in LF, the latter cannot be read in this level and therefore must be eliminated before the syntactic derivation is transferred. If an uninterpretable feature fails to be erased, the derivation crashes, which means ungrammaticality.

Feature erasure is accomplished through agreement. Agreement is, actually, a two-fold operation. In the first place, it requires identification of an element that contains interpretable features of the same kind than those in need to be erased, and accord of the uninterpretable feature, which is unvalued, with the interpretable one. Secondly, the uninterpretable feature is checked and erased (19).

- (19)
1. [uR] ... [iR]
 2. Accord ([uR], [iR])
 3. Check [uR] with [iR] and Erase [uR].

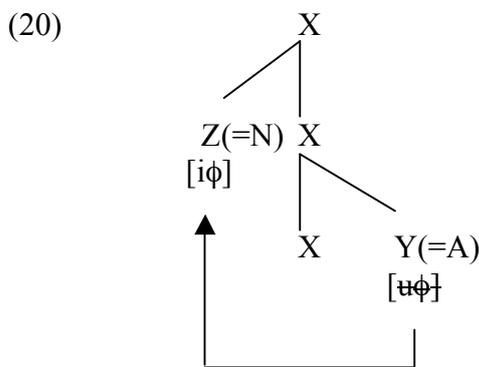
Spanish As contain, at least, a set of uninterpretable features related to the nominal properties gender and number. We will represent this technically as an uninterpretable set of phi-features or [ϕ]. This forces the A to check those features with

⁴ A possible analysis of these data that could explain this behaviour cannot be maintained. To our mind, these constructions are clearly not compounds. Their behaviour, at least in the dialect of Spanish that the people that I have tested – and my own dialect –, has nothing to do with what we expect from compounds. These structures can be coordinated (i), noun elipsis is possible with them (ii), alpha-movement is possible with a part of the construction (iii) and it is also possible to modify only part of the structure that is formed (iv). Given the Lexical Integrity Hypothesis, this proves that they are not compounds.

- (i) a. un amarillo oscuro y verdoso.
'Lit. a yellow dark and greenish'
- (ii) b. un amarillo oscuro y un-o claro.
'Lit. a yellow dark and another light'
- (iii) c. lo verdoso, que es este amarillo t_i
'Lit. how greenish is this yellow'
- (iv) d. un amarillo [terriblemente [verdoso]]
'lit. a yellow [terribly [greenish]]'

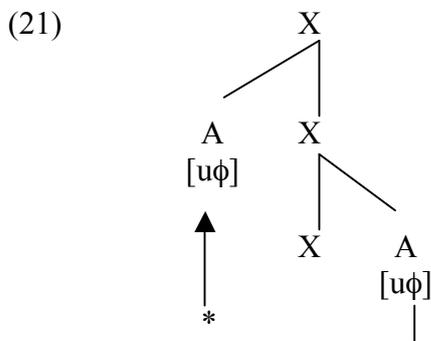
a legitimate element, that is, an element which contains an interpretable set of phi-features, $[i\phi]$. In Spanish, only N's contain $[i\phi]$. This means that A agrees with N.

This structure is represented in (20). Note that the element to be interpreted as A needs a specifier to satisfy its semantic conditions; this is provided by X, a relational element, closely following H&K's proposal (cf. Mateu 2002). This spec position is occupied by N. A's $[u\phi]$ enters, then, in an Accord relation with N's $[i\phi]$, their value is assigned to A's features and they are checked. The derivation will converge in FL. Following the spirit of Chomsky's (2001) proposal about the necessity of u-features, namely, that their checking gives rise to semantic relations, we propose that, as a result of this checking operation, the At(tributive) categorial role of A is saturated through theta identification with the R(eferential) categorial role of the N (Spencer 1999), which means that it will be interpreted as its subject.



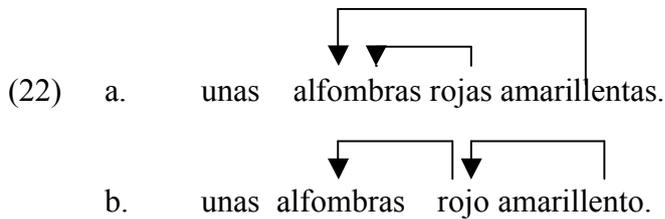
This structure explains the close connection between agreement and adjectival predication – remember the data in (4) –.

Now let us consider the first context where CT must obligatorily project as N. Remember that in those situations they are modified by an A expressing hue or intensity. Crucially, the logical subject of that property is the CT. The hue is a property of the colour denoted, (and so it is the intensity) not of the head noun to which the CT refers. The CT, then, must occupy [Spec X] position in the tree. However, if CT is an A, checking of the hue / intensity A's u-features won't be possible, for Accord must be established as a prerequisite to checking, and Accord takes place only between i-features and u-features belonging to the same class. Therefore, (21) will crash at the Interfaces, for there are u-features unvalued and unchecked.



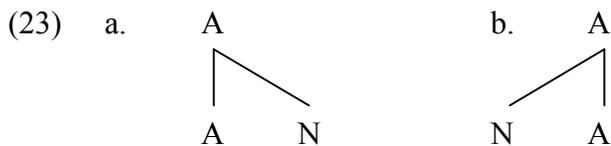
The subject of the A must contain $[i\phi]$ for the derivation to be convergent; therefore, the category of the subject must be N. Note that we will expect the CT to

surface as A if every other A in the NP referred to other elements in the construction. This prediction is confirmed. Consider the minimal pair in (22).



There is a slight difference between (22a) and (22b). In (22a) the hue A *amarillentas* takes the N *alfombras* as its subject, and so it does the CT. Therefore, the meaning of (22a) is a carpet that is both red and yellowish. In contrast, in (22b) the hue A is predicated of the CT, which must surface as N, and therefore the expression denotes a carpet which is red, and the hue of that red is yellowish.

As for the second context, that in which an N modified the CT, it can be explained provided that we take seriously the role of features in syntactic operations. Through languages, adjectives are modifiers of nouns, and not the opposite. We will show that merge operations correctly predict this. Consider (23).



Feature driven operations are automatic, compulsory and cannot be directed by semantic requisites. Then, as a result of its merge with N, A unavoidably checks each one of its phi u-features.

Following Chomsky (2000, 2001), when a head has erased every one of its uninterpretable features, it becomes inert for further operations, which means that it becomes inactive. What this means is that when A is merged with N, A becomes inactive because it has automatically checked all its uninterpretable features.

The head of a construction is the element that projects its label in the construction. As the label is the only information available to merge, the label must be syntactically active. If it were inert, merge won't be able to apply to it, because inert elements are inactive to syntactic operations.

This somehow oblique reasoning is actually deriving a very intuitive statement (24):

(24) Heads must be syntactically active in their own projections.

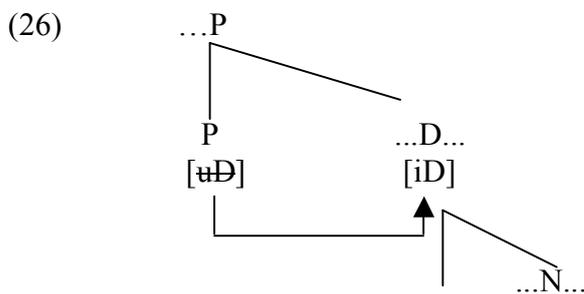
This is why A must be a modifier of N and not the opposite, which explains why the structures in (24) are ungrammatical.

Crucially, (23) has to merge with some element. Why? Because it contains one constituent, N, that has not checked its u-feature Case, and A is not a legitimate probe for that operation. Therefore, what we have in (23) is a structure that will crash, and, consequently, is agrammatical.

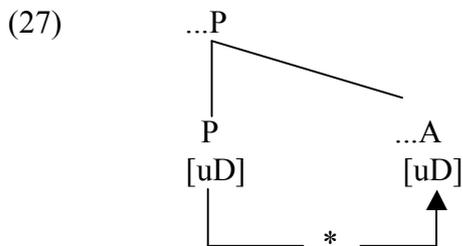
Note that the structure in (20) actually predicts that N, and not A, will be the syntactically active constituent in further operations. N is active, so it can transmit its

features to the head X through standard spec-head agreement. Therefore, when merged with another element, N will be capable of entering in a checking relation with that head, and not A. It is predicted, then, that every extended projection of (20) will count as an extended projection of N.

The third context can be explained in a similar vein if we assume that Spanish prepositions contain an uninterpretable D feature. This feature is motivated by the fact that P's can denote referential entities, but, as opposed to deictic adverbs, only when they take a nominal complement. What this means is that for the PP constituent to be convergent, P must combine with an element that contains among its i-features [iD]. Obviously, D itself must contain such a feature. However, in Spanish, D can only merge with an NP, not with an AP. The only category that contains such a feature in Spanish is N. Therefore, a convergent derivation for a projection headed by P will be as in (26).



Let us consider now what would happen if the CT projected as A in this configuration. As APs do not contain [iD], for they are never referential nor combinable with D, [uD] would never get checked and, as in the other case, the derivation will crash when transferred to the interfaces (27).



We have intended to show that a syntactic explanation can give account of the puzzling behaviour of Spanish CT in a principled manner. Semantically driven theories of categorisation cannot explain these data in an accurate manner. Note that the conceptual meaning of the CT does not change when projected as an N and when projected as an A. Therefore, if conceptual semantics is prior to syntax, the different categorisation of CT is unpredicted and remains unexplained. As for structural differences in meaning, they are actually predicted by syntax, for each syntactic configuration has a specific semantic import when interpreted in LF (H&K 1993, 1997, Mateu 2002).

However, conceptual semantics does play a role in the construction, but its intervention takes place once the syntactic structure has been built and its constituents have been categorised. The relevant question at this point, obviously, is why colour terms can behave in such a way, while other elements – such as those denoting shapes

or psychological states – cannot. The answer to this is in the Encyclopaedia. In DM, vocabulary items are inserted post-syntactically and then the conceptual non-structurally predictable information associated to those items is accessed. This information is listed in the Encyclopaedia, where the entries would contain every kind of cultural information. The encyclopaedic entry of a CT would give information concerning the special conceptual status of colours in the human mind. As Quine (1970) pointed out, every substance is characterised by a certain colour. This invites us to regard colour not exactly as an accidental property of substances, but as a component of substances. Due to this ambiguous conceptual status, colour can be regarded as a potentially referential entity as well as a quality of referential entities. Almost every other nominal concept would be regarded as either a quality of entities, without independent existence out of those entities, or as an entity, and, if syntax categorised it in a different class, a pragmatically marked reading would arise.

These facts are related to other aspects of the behaviour of Spanish CT to which we will not have time to make justice here, such as the use of CT to define political, ethnic and professional groups of people, in a manner that reminds us of relational adjectives.

References

- Anderson, J.M. (1997), *A Notional Theory of Syntactic Categories*, Cambridge, Cambridge University Press.
- Bello, A. (1847 [1970]), *Gramática de la lengua castellana destinada al uso de los americanos*, Sopena, Buenos Aires.
- Bernstein, J.B. (1993), *Topics in the Syntax of Nominal Structure across Romance*, Doctoral Dissertation, The City University of New York.
- Chomsky, N. (1995), *The Minimalist Program*, Cambridge, MIT Press.
- Chomsky, N. (1998), *Minimalist Inquiries: The Framework*, MIT Working Papers in Linguistics 15.
- Chomsky, N. (2000), *Derivation by phase*, in M. Kenstowicz, (ed.), *Kenneth Hale: A Life in Language*, Cambridge (Mass.), MIT Press.
- Chomsky, N. (2001), *Beyond Explanatory Adequacy*, MIT Working Papers in Linguistics.
- Dixon, R.M.W. (1982), *Where have all the Adjectives Gone?*, The Hague, Mouton.
- Fábregas, A. (2002), *Gramática de los Nombres de Color en Español Actual*, Trabajo de Investigación Avanzado, Instituto Universitario Ortega y Gasset / Universidad Complutense de Madrid / Universidad Autónoma de Madrid.
- Fábregas, A. (2003), *Los verbos de realización gradual: estructura léxica*, “RSEL - Revista española de lingüística” 32/2, 475–506.
- Fernández Ramírez, S. (1951), *Gramática española. El Nombre*, Madrid, Arco-Libros.
- Hale, K. & S.J. Keyser (1993), *On argument structure and the lexical representation of syntactic relations*, in K. Hale & J. Keyser (ed.), *The View from Building 20*, Cambridge (Mass.), MIT Press, 53–110.
- Hale, K. & S. J. Keyser (1998), *The Basic Elements of argument structure*, in H. Harley (ed.), *Papers from the Upenn/ MIT roundtable on argument structure and aspect*, Cambridge (Mass.), MIT Working Papers in Linguistics, 73–118.
- Halle, M. & A. Marantz (1993), *Distributed Morphology and the Pieces of Inflection*, in K. Hale & S. J. Keyser (eds.), *The View from Building 20*, Cambridge (Mass.), MIT Press, 111–176.
- Higginbotham, J. (1985), *On Semantics*, “Linguistic Inquiry” 16, 395–420.

Universals and Grammatical Categories:

- Langacker, R.W. (1999), *Grammar and Conceptualization*, Berlin/New York, Walter de Gruyter.
- Marantz, A. (1997), *No Escape from Syntax: Don't Try Morphological Analysis in the Privacy of Your Own Lexicon*, "UPenn Working Papers in Linguistics" 4:2, 201–225.
- Mateu, J. (2002), *Argument Structure: Relational Construal at the Syntax-Semantics Interface*, Doctoral dissertation, UAB.
- Quine, W.O. (1970), *Las raíces de la referencia*, Madrid, Revista de Occidente.
- Spencer, A. (1999), *Transposition and Argument Structure*, in G. Booij & J.van Marle (eds.) "Yearbook of Morphology" 1998, Dordrecht, Kluwer Academic Publishers.
- Wierzbicka, A. (1980), *Lingua Mentalis. The Semantics of Natural Language*, Sydney / New York, Academic Press.

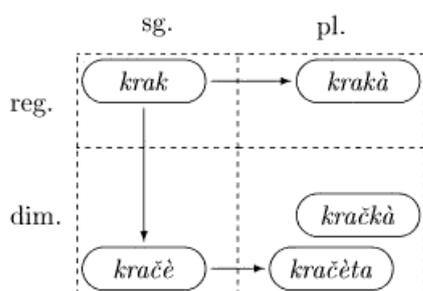
On Diminutive Plurals and Plural Diminutives*

Ivan A Derzhanski
Bulgarian Academy of Sciences
iad@math.bas.bg

The present book may not be snapped up by a public mistakenly eager for the latest contribution to number theory. But if a few stray mathematicians read it, I hope they will find that the linguistic number systems analysed here show the elegance and complexity they are accustomed to in their area of enquiry. – Greville G. Corbett, *Number*

‘What is the singular of *kračkà*?’

A mathematician of my acquaintance asked this question of another in the course of a long train journey that I chanced to be sharing with them. I was too tired to join the conversation at the time, but the matter rested in my mind.



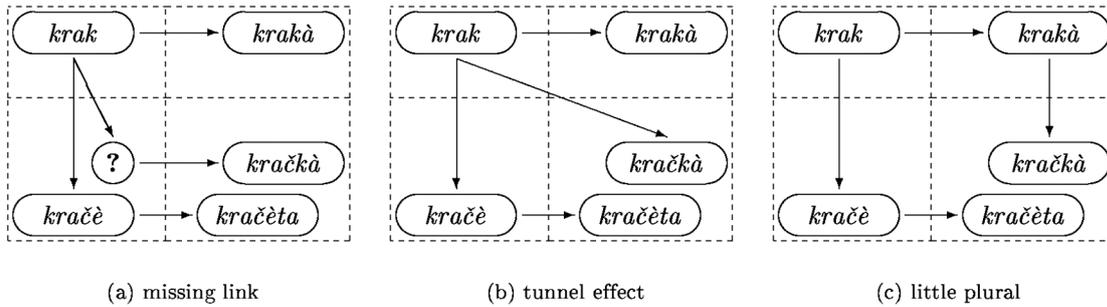
The word in question means ‘little legs/feet’, and it has, in fact, no apparent singular correlate. In this it differs from *kračèta*, the plural form of *kračè*, which is a diminutive derived from *krak* ‘leg/foot’. In most contexts the two are freely interchangeable. The form *kračèta* is more common except in the context of cooking, where *kračkà* is used as the technically correct term for trotters of pork or lamb. On the other hand, *kračkà* does not cooccur easily with cardinal numerals, so if one is present, *kračèta* is preferred even in that sense: *tja nosi 4 [...] kračeta ot svinče* (HS) ‘she is carrying four pig’s legs’. In other words, *kračkà* acts as a collective plural and *kračè* as the corresponding singulative.

The figure doesn’t try to show the full array of diminutives and plural forms, and it is conceivable that *kračkà* is the plural, or more likely the erstwhile dual, of another diminutive of *krak*, whose singular is perhaps unattested (the circle with the question mark in the diagram labelled ‘missing link’ on the next page).¹ If so, we are dealing

* My main sources of data are Arnott (1995) (Fula), Elanskaja (1980) (Coptic), Green www (Dakelh), Hemon (1995) (Breton), Koval’ (1997) (Fula), Leont’ev (1974) (Asmat), Maslova (2002) (Kolyma Yukaghir), Sova (1989) (Bantu), Stump (2001) (Southern Barasano, Yiddish), Sylestine et al. (1993) (Alabama), Volodin (1976) (Itelmen), Wolgemuth (2002) (Isthmus Nahuatl), Wright (1981) (Classical Arabic). The authors of Bulgarian texts identified by their initials are Kiril Hristov, Hristo Smirnenki and Peyo Yavorov.

¹ This possibility was suggested to me by Vladimir Plungian (p.c.).

with a highly abnormal development. The Proto-Slavic diminutive suffix **-ik-* yields Old Bulgarian *-ič-* (owing to the Third Palatalisation), Modern Bulgarian *-ec*, in all forms of masculine nouns. A case in point is *kračec*, a rare hypocoristic derivative of *krak*, which is *singulare tantum*, like most diminutives in *-ec*. If this existed in Old Bulgarian, it must have had the form **kračiči* in the singular and **kračica* in the dual, the latter being close to both *kračkà* and *kračicà* (another *plurale tantum* diminutive of *krak*, an obsolete one), but still significantly different from both.



Dictionary entries for *kračkà* label it as ‘dim. pl. of *krak*’ or ‘pl., dim. of *krakà*’. Taken literally, the former implies that the two operators, derivation of a diminutive and inflexion for plural number, are applied cumulatively, in a single morphological process (‘tunnel effect’), whereas the latter suggests that *kračkà* is not the outcome of the pluralisation of a diminutive noun, but is itself a diminutive derived from a plural noun form (‘little plural’). Either way imaginary (and aberrant) forms are eschewed, but an unusual mechanism is assumed.

This makes three hypotheses. The uncountability of the term can’t help us to choose among them, because they all correlate with it. The plurals of non-human masculine nouns don’t normally cooccur with cardinal numerals, as those nouns have corresponding count forms, whose purpose is to do exactly that (cf. *dva_M kràk-a_{CT}* ‘two legs, two feet’). On the other hand, a noun that has no singular form is *plurale tantum*, and by virtue of that fact uncountable.

At this point it is expedient to ask two questions:

What other lexical items in Bulgarian behave in similar ways (that is, what other *pluralia tantum* diminutives are there, and if they have synonyms that do have singular correlates, are there any more or less consistent differences in usage as between *kračkà* and *kračèta*)?

What will a search for comparable phenomena elsewhere yield?

1. The Bulgarian Data

Bulgarian is a highly fusional language, in which a word form’s morpheme structure can be controversial. For most categories of stems from which diminutives can be formed it has a variety of diminutive suffixes, some with a marked preference for a certain denotative (undersize entity, young of a species) or connotative (hypocoristic, pejorative) interpretation. Diminutivisation may preserve gender, or it may involve conversion from masculine or feminine to neuter gender. Some suffixes permit the further formation of secondary and even tertiary diminutives: *momà* f. ‘lass, maiden’ >

mom-ič-a f. dto. (a rare hypocoristic diminutive) > *mom-ič-e* n. ‘girl’ > *mom-ič-e-nc-e* n. ‘little girl’.

The words from which *pluralia tantum* diminutives are derived fall into the following groups, which shall be considered in order:

- masculine and neuter nouns with irregularly formed plurals;²
- other masculine nouns with regularly formed plurals, almost all of which fall into two semantically motivated subgroups (*viz.*, edible stuffs and kinds of footwear);
- *pluralia tantum* nouns, also including some semantically motivated subgroups (e.g., kinds of legwear);
- numerals.

1.1 Masculine Nouns

As I said in the Introduction, the plurals of non-human masculine nouns don’t cooccur with cardinal numerals or with *kòlko* ‘how many?’. However, the diminutives formed from them, which correspond to no singular or count forms, are not countable either.

There are four masculine nouns in the language with plurals (erstwhile duals or collectives) in *-à*; three of them have corresponding diminutive plurals (1–3). (The fourth one is *gospodìn* ‘gentleman, mister’, pl. *gospodà*, from which no diminutives are derived, evidently for semantic reasons.)

The noun *čovèk* ‘person, human being’ (4) is exceptional in having three plural forms. The regular plural *čovèci* is used seldom, and only in the sense ‘human beings *par excellence*’ (as in the adage *xora mnogo, no čoveci malko* ‘[the] people [are] many, but [the] human beings [of any virtue are] few’) or occasionally ‘humans as opposed to other sentient beings’ in fictional settings (as Rudyard Kipling uses the English plural *men* in *The Jungle Books*, where there are numerous non-human species of *people*³). One of the suppletive plural forms, *ljùde*, is antiquated (and stylistically marked). The commonly used plural is *xòra*, from which the diminutive *xòrica* ‘poor, harmless people’ is derived. Since the hypocoristic diminutive *čovèčec* ‘poor, harmless person’ has no regular plural, it effectively forms a suppletive paradigm with *xòrica*.

The noun *bodil* (5) means ‘thorn’ in the sense of either ‘thistle’ or ‘prickle’, but the two meanings are differentiated in the plural, and from *bodli* ‘prickles’ a diminutive can be formed. Depending on how one looks at it, *bodil* : *bodli* can be considered as one of the two instances of fleeting *i* in Bulgarian (the other one is in the numeral *edìn* : *edn-* ‘one’) or a case of partial suppletion. (Diachronically the latter is correct: originally ‘prickle’ was *bodèl*, but as that word went out of use, *bodil* took over both its meaning and its regularly formed plural).

As I said above, hypocoristic diminutives in *-ec* don’t usually have plural forms. But in some speakers’ usage some of those that are formed from monosyllabic nouns do (6). The plural diminutive form *grošòvce* is more readily used metaphorically for ‘little money, small change’ than literally for ‘(dear) little piastres’, though the latter may also have been likely whilst the piastre was in circulation. There is a theory that the

² Indeed, the more unlike a plural form something is, the more likely it is to manifest behaviour not normally associated with plural forms, such as feeding derivation.

³ Tsvetan Stoyanov aptly renders *men* as *čoveci* in his partial Bulgarian translation of *The Jungle Books* (1967).

morpheme *-ovce* is composed nonlinearly from the diminutive suffix *-ec* and the plural ending *-ove*.⁴

Diminutive plurals (nearly always in *-ki*) are also derived from masculine nouns with regular plurals (in *-i*). Some of these are names of edibles⁵: *domàt* ‘tomato’ (7), *kartòf* ‘potato’, *mòrkov* ‘carrot’, *badèm* ‘almond’, *lèšnik* ‘hazelnut’, *òrex* ‘walnut’, *fòstók* ‘peanut’; also *makaròn* ‘strand of macaroni’, where the singular form is a back-formation from the collective *makaròni* (originally a *plurale tantum*). Others are kinds of footwear: *botùš* ‘boot’ (8), *nalóm* ‘patten’, *čoràp* ‘sock, stocking’. The plural of *čèxøl* ‘slipper (without back)’ (9), namely *čèxli*, forms the diminutive *čèxlički*. In all cases there is a plural diminutive as well, e. g., *domàtčeta* ‘little tomatoes’, which tends to describe the size of the individual vegetables, as opposed to *domàtki*, which conveys the speaker’s attitude to a salad of them perhaps; such differences in the likely interpretation obtain throughout.

Two names of body parts, one paired (10), the other one plural (11), also belong here; the latter also has the diminutive plural form *zòbìci*, but that one hardly ever occurs except in poetry: *da bjaxa margar mønista tvoite beli zòbìci* (PY) ‘would that thy (dear) white teeth were pearl beads’.

	sg.	dim.	pl. dim.	pl.	dim. pl.	
1	<i>krak</i>	<i>kračè</i>	<i>kračèta</i>	<i>krakà</i>	<i>kračkà</i> , <i>kračicà</i>	leg, foot
2	<i>rog</i>	<i>rògče</i>	<i>rògčeta</i>	<i>rògòve</i>	—	horn
				<i>rogà</i>	<i>rogcà</i>	
3	<i>nòmer</i>	<i>nòmerče</i>	<i>nòmerčeta</i>	<i>nomerà</i>	<i>nomercà</i>	(ordinal) number
4	<i>čovèk</i>	<i>čovèče</i>	<i>čovèčeta</i>	<i>čovèci</i>	—	person, human being
		<i>čovèčec</i>	—	<i>xòra</i>	<i>xòrica</i>	
5	<i>bodil</i>	<i>bodilče</i>	<i>bodilčeta</i>	<i>bodili</i>	—	thorn, thistle
				<i>bodli</i>	<i>bodlički</i>	thorn, prickle
6	<i>groš</i>	<i>gròšec</i>		<i>grošòve</i>	<i>grošòvce</i>	piastre, obsolete Lv
		<i>gròšče</i>	<i>gròščeta</i>			0.20 coin
7	<i>domàt</i>	<i>domàtče</i>	<i>domàtčeta</i>	<i>domàti</i>	<i>domàtki</i>	tomato
8	<i>botùš</i>	<i>botùšče</i>	<i>botùščeta</i>	<i>botùši</i>	<i>botùški</i>	boot
9	<i>čèxøl</i>	<i>čèxølče</i>	<i>čèxølčeta</i>	<i>čèxli</i>	<i>čèxlički</i>	slipper
10	<i>mustàk</i>	<i>mustàče</i>	<i>mustàčeta</i>	<i>mustàci</i>	<i>mustàčki</i>	moustache
11	<i>zòb</i>	<i>zòbče</i>	<i>zòbčeta</i>	<i>zòbi</i>	<i>zòbki</i>	tooth
				<i>zòbì</i>	<i>zòbìci</i>	
12	<i>okò</i>	<i>očè</i>	<i>očèta</i>	<i>očì</i>	<i>očìci</i>	eye
13	<i>uxò</i>	<i>ušè</i>	<i>ušèta</i>	<i>ušì</i>	<i>ušìci</i>	ear
14	<i>detè</i>	<i>detènec</i>	? <i>detènca</i>	<i>decà</i>	<i>dečìca</i>	child
15	<i>nèšto</i>	<i>nèštičko</i>	—	<i>neštà</i>	<i>neštìcà</i>	(some)thing

⁴ ‘It can be said that the diminutive marker is inserted into the plural marker in these rare forms’ (Maslov 1981: 137). Historically the *ov* in both *-ove* and *-ovce* is a vestige of the fact that in Proto-Slavic *u*-stems ended in *-aŭ* before vowel-initial suffixes and endings.

⁵ Note that *kračkà* ‘trotters of pork or lamb’ is one also.

1.2 Neuter Nouns

The diminutives formed from the plurals of neuter nouns are countable (that is, they can cooccur with cardinal numerals), but it is difficult to draw any conclusions from this, due to the scanty number of nouns involved.

There are two neuter nouns with plurals (erstwhile duals) in *-i* (12–13). The hypocoristic forms *očici* and *ušici* are rare, though they do occur, esp. in poetry: *da bjaxa ogøn elmazi tvoite černi očici* (PY) ‘would that thy (dear) black eyes were fiery diamonds’. However, the secondary diminutive *očički* is common enough.

The noun *detè* ‘child’ (14) was originally a singulative (*dæte* from the collective *dæti* ‘children’). Its partially suppletive plural *decà* is a contraction of Old Bulgarian *dætica*, attested in the thirteenth century (Mirčev 1963: 57). The regular plural diminutive *detènca* is very rare, so for most practical purposes *detènca* and *dečica* form a (partially) suppletive paradigm. Of some interest is the expression *màmino detènca* ‘Mummy’s little child; mother’s darling, milksop, mollycoddle’, whose plural is *mamini dečica* in the literal sense and *mamini detenca* in the idiomatic one; the derivation through *deca* ‘children’, which mollycoddles are not almost by definition, would be inappropriate.

The indefinite pronoun *nèšto* ‘something’ (< *næ-* ‘some-’ + *što* ‘what’) has been degrammaticalised to mean ‘thing’ (15) and inflects as a noun when so used. As such it forms the plural *neštà* ‘things, stuff’, whence the diminutive *neštica*. The singular *neštice*, as in *tam ni edničko neštice ne sveti* (KH) ‘there [*sc.* in the skies] not a single (little) thing is shining’, is quite rare, and is as likely to be a back-formation of *neštica* as a diminutive of *nèšto*. The *singulare tantum* form *nèštičko* ‘little something’ is an adjectival diminutive, and more readily used as a pronoun than as a noun.

1.3 Pluralia Tantum

Semantically speaking, the relatively restricted class of *pluralia tantum* nouns in Bulgarian presents no surprises, compared to other languages. It includes the names of numerous kinds of legwear (16–18; also *potùri* ‘breeches’, *šalvàri* ‘shalwars’, *šòrti* ‘shorts’ etc.) as well as the word *obùšta* ‘footwear, shoes’ (19), twosome tools (20–22) and mass terms (23). There are also names of mountains, diseases, festivals and financial terms, but those are outside our present scope, as they form no diminutives.

The language finds such nouns an inconvenience and strives to eliminate them, either by back-forming singulars from them, with the same meaning or a different one, or, when the phonological shape permits it, by reinterpreting them as singulars (the modest size of the nominal paradigm, given the loss of case marking, makes this a good deal easier than it is in other Slavic languages). Examples of the former scenario are *nòžica* ‘scissors’ from *nòžici* *dto.*, *pantalòn* ‘trousers’ from *pantalòni* *dto.* and *očilò* ‘spectacle lens’ from *očilà* ‘spectacles’. The latter accounts for *vratà* ‘gate; door’ (24), *kolà* ‘waggon, ox-cart; car’ (25) and *ustà* ‘lips, mouth’ (26), originally *pluralia tantum* after the manner of plural neuters, but currently feminine nouns with plurals in *-i*. (In the glosses of the three words the semicolons separate the older meanings from the newer ones.) However, their old diminutives have not been so reinterpreted; rather, they have been superseded by new ones, with the suffix *-ic(a)*.

The cardinal numerals from two onwards, general and masculine personal, constitute a special class of *pluralia tantum* words. A few of them have diminutive forms (27–31).

	pl.	dim. pl.			pl.	dim. pl.	
16	<i>gàšti</i>	<i>gaštèta, gàštički</i>	pant(ie)s	24	<i>vratà</i>	<i>vratcà</i>	gate
17	<i>pantalòni</i>	<i>pantalònki</i>	trousers	25	<i>kolà</i>	<i>kolcà</i>	ox-cart
18	<i>plùvki</i>	<i>plùvčici</i>	swimming trunks	26	<i>ustà</i>	<i>ust(i)cà, ustènca</i>	mouth
19	<i>obùšta</i>	<i>obuštèta</i>	shoes, footwear	27	<i>dve</i>	<i>dvèčki, dvènki</i>	2 (gen.f./n.)
20	<i>klèšti</i>	<i>klèštički</i>	pincers	28	<i>tri</i>	<i>trički, trìnki</i>	3 (general)
21	<i>nòžici</i>	<i>nòžički</i>	scissors	29	<i>četiri</i>	<i>četirki</i>	4 (general)
22	<i>očilà</i>	<i>očilcà</i>	spectacles	30	<i>dvàma</i>	<i>dvàmca, dvàmka</i>	2 (m. pers.)
23	<i>trìci</i>	<i>trički</i>	bran	31	<i>dvamìna</i>	<i>dvamìnka</i>	2 (m. pers.)

1.4 Patterns

Three of the most opaque plural nouns and the masculine personal numerals form their diminutives as singular feminine nouns do, except that they have no secondary diminutives (there are such words as *kóštička*, *rekička*, *živinčica*, but no **xòrička* etc.), and the nouns that *krakà* patterns with are all formed from adjectives by the suffix *-in(à)*.

	reg.	dim.			reg.	dim.	
m. pl.	<i>xòr-a</i>	<i>xòr-ic-a</i>	people	f.	<i>kóšt-a</i>	<i>kóšt-ic-a</i>	house
n. pl.	<i>dec-à</i>	<i>deč-ìc-a</i>	children	f.	<i>ovc-à</i>	<i>ovč-ìc-a</i>	sheep, ewe
					<i>rek-à</i>	<i>reč-ìc-a</i>	river
m. pl.	<i>krak-à</i>	<i>krač-k-à</i>	legs, feet	f.	<i>živin-à</i>	<i>živin-k-à</i>	live being, animal
num.	<i>dvàm-a</i> <i>dvam-in-a</i>	<i>dvàm-k-a</i> <i>dvam-in-k-a</i>	two (people)	f.	<i>žil-a</i>	<i>žil-k-a</i>	tendon, vein

Now *xòra* is a loan from Greek, where *χώρα* is the citation (singular) form of a feminine noun meaning ‘country, nation’, *dèca* ‘children’ can behave as a singular feminine noun in Serbo-Croat, and *-in(a)* in *dvamìna* etc. is a derivational (usually augmentative) suffix. This puts the erstwhile dual *krakà* with the associated diminutive *kračkà* in unusual company.⁶

Most other diminutive plurals have the form of plural diminutives, except that they have no corresponding singular forms. They can be divided into four groups.

⁶ The final *-ma* in *dvàma* etc. is also in origin an Old Bulgarian dual ending, but of the dative and instrumental cases. With the disintegration of the case system it ceased being associated with any particular syntactic functions, then was copied from ‘two’ to several higher numerals.

	reg.	dim.	redim.	
f. sg.	<i>žil-a</i>	<i>žil-k-a</i>	<i>žil-č-ic-a</i>	tendon, vein
f. pl.	<i>žil-i</i>	<i>žil-k-i</i>	<i>žil-č-ic-i</i>	
m. pl.	<i>zób-i</i>	<i>zób-k-i</i>		teeth
pl. t.	<i>pantalòn-i</i>	<i>pantalòn-k-i</i>		trousers
pl. t.		<i>plùv-k-i</i>	<i>plùv-č-ic-i</i>	swimming trunks
num.	<i>čètir-i</i>	<i>čètir-k-i</i>		four
	reg.	dim.	redim.	
f. sg.	<i>dàrb-a</i>	<i>dàrb-ic-a</i>	<i>dàrb-ič-k-a</i>	talent
f. pl.	<i>dàrb-i</i>	<i>dàrb-ic-i</i>	<i>dàrb-ič-k-i</i>	
m. pl.	<i>čexl-i</i>		<i>čexl-ič-k-i</i>	slippers
pl. t.		<i>nòž-ic-i</i>	<i>nòž-ič-k-i</i>	scissors
pl. t.	<i>gàšt-i</i>		<i>gàšt-ič-k-i</i>	pant(ie)s
pl. t.	<i>klèšt-i</i>		<i>klèšt-ič-k-i</i>	pincers
	reg.	dim.	redim.	
f. sg.	<i>glav-à</i>	<i>glav-ìc-a</i>	<i>glav-ič-k-a</i>	head
f. pl.	<i>glav-ì</i>	<i>glav-ìc-i</i>	<i>glav-ič-k-i</i>	
m. pl.	<i>zòb-ì</i>	<i>zòb-ìc-i</i>		teeth
m. pl.	<i>bodl-ì</i>		<i>bodl-ič-k-i</i>	prickles
n. pl.	<i>oč-ì</i>	<i>oč-ìc-i</i>	<i>oč-ič-k-i</i>	eyes
pl. t.		<i>tr-ìc-i</i>	<i>tr-ič-k-i</i>	bran
num.	<i>tr-i</i>		<i>tr-ič-k-i</i>	three
pl. t.	<i>klešt-ì</i>		<i>klešt-ič-k-i</i>	pincers

The first and largest group is composed of those that look like plurals of feminine diminutives formed from feminine nouns. The various types are illustrated in the table; they employ the suffixes *-k(a)*, unstressed and stressed *-ic(a)* and their combinations *-[k>č]-ic(a)* and *-i[c>č]-k(a)*. The inclusion of the numeral *tri* ‘three’ is provisional; I shall return to this point later.

In fact some of the corresponding singular forms do exist. Compare *bonbòn* ‘sweet, candy’, whose extant (though dated) alternative form *bonbòna* (with the same plural form *bonbòni*) and its diminutive *bonbònka* might explain the plural diminutive *bonbònki* even in the speech of those who don’t use the two feminine singulars, to *pantòf* ‘slipper (with back)’, which lacks the first of the two ‘intermediate’ forms, and to *botùš* ‘boot’, which lacks both.

m.	f.	dim. f.	dim. pl.	
<i>bonbòn</i>	<i>bonbon-a</i>	<i>bonbòn-k-a</i>	<i>bonbòn-k-i</i>	sweet, candy
<i>pantòf</i>	—	<i>pantòf-k-a</i>	<i>pantòf-k-i</i>	slipper
<i>botùš</i>	—	—	<i>botùš-k-i</i>	boot

The second group is made up of the diminutive derivative of the *plurale tantum* noun *gàšti* ‘pant(ie)s’, which has the form of the plural of a neuter diminutive derived from a feminine noun, and of *obùšta* ‘footwear, shoes’, which is exceptional in that the diminutive is related to the base as the plural of the neuter diminutive is to the singular of the feminine noun from which it is derived.

	reg.		dim.	redim.	
f. sg.	<i>kóšt-a</i>	n. sg.	<i>køšt-è</i>	<i>køšt-è-nc-e</i>	house
f. pl.	<i>kóšt-i</i>	n. pl.	<i>køšt-è-ta</i>	<i>køšt-è-nc-a</i>	
pl. t.	<i>obùšt-a</i>		<i>obušt-è-ta</i>	<i>obušt-è-nc-a</i>	footwear, shoes
pl. t.	<i>gàšt-i</i>		<i>gašt-è-ta</i>	<i>gašt-è-nc-a</i>	pant(ie)s

The diminutive plurals in the third group are shaped as plurals of neuter diminutives formed from neuter nouns. The unusual case is that of the masculine personal numerals: the words they pattern with have more than two syllables, whereas *dvàma* ‘two (people)’ and *trìma* ‘three (people)’ contain precisely two each.

	reg.	dim.			reg.	dim.	
n. sg.	<i>mjàst-o</i>	<i>mest-ènc-e</i>	place	n. sg.	<i>kopit-o</i>	<i>kopit-c-e</i>	hoof
n. pl.	<i>mest-à</i>	<i>mest-ènc-a</i>		n. pl.	<i>kopit-a</i>	<i>kopit-c-a</i>	
pl. t.	<i>ust-à</i>	<i>ust-ènc-a</i>	mouth	num.	<i>dvàm-a</i>	<i>dvàm-c-a</i>	two (people)
	reg.	dim.			reg.	dim.	
n. sg.	<i>per-ò</i>	<i>per-c-è</i>	feather	n. sg.	<i>lic-è</i>	<i>lič-ic-è</i>	face
n. pl.	<i>per-à</i>	<i>per-c-à</i>		n. pl.	<i>lic-à</i>	<i>lič-ic-à</i>	
m. pl.	<i>rog-à</i>	<i>rog-c-à</i>	horn	m. pl.	<i>krak-à</i>	<i>krač-ic-à</i>	legs, feet
m. pl.	<i>nomer-à</i>	<i>nomer-c-à</i>	numbers	n. pl.	<i>nešt-à</i>	<i>nešt-ic-à</i>	things
pl. t.	<i>očil-à</i>	<i>očil-c-à</i>	spectacles	pl. t.	<i>ust-à</i>	<i>ust-ic-à</i>	mouth
pl. t.	<i>vrat-à</i>	<i>vrat-c-à</i>	gate				
pl. t.	<i>kol-à</i>	<i>kol-c-à</i>	cart				

The diminutive plurals or plural diminutives in *-ovce* constitute a class of their own.

The last case to consider is that of the cardinal numeral *dve* ‘two’ (feminine or neuter) with its diminutives *dvèčki* and *dvènki*, where the initial vowel of the diminutive suffix *-ičk-* or *-ink-* (an uncommon suffix generally restricted to adjectives) is missing, as though it has been reanalysed as something other than part of the suffix—and in this case the only other thing it could be a part of is an inflected stem *preceding* the suffix. The same analysis can arguably be applied to the diminutives of *tri* ‘three’, as an alternative to the classification proposed above.

2. The Crosslinguistic Situation

This section reports the results of my search of the world’s languages for diminutive plural forms that are not obtained by pluralisation of diminutives.

2.1 Missing Links

I don’t have many examples of missing link derivations. My best example is from Polish.⁷ In that language diminutives in *-e*, pl. *-et-a*, and singulatives/rediminutives *-qt-k-o*, pl. *-qt-k-a*, are formed from names of animal species and a few ethnic and racial groups (and then denote young animals and children, respectively) as well as some

⁷ There are exact parallels in Ukrainian and Belorussian (but not Russian).

other words for live beings (e.g., *wnuk* ‘grandson’, *wnucz-ę* ‘grandchild’; *zwierz* ‘beast’, *zwierz-ę* ‘animal’). However, the plural form in *-ęt-a* (with no corresponding rediminutive) is used as a *plurale tantum* diminutive of the names of some body parts (*oko* ‘eye’, *ręka* ‘arm, hand’, *noga* ‘leg, foot’, colloquially a few other body part and paired clothing items as well), especially when referring to a child’s or a woman’s eyes or limbs, and only in the literal (anatomical) sense, never for any metaphorical meanings that the base noun or other diminutives may have.

		reg.	dim.	dim.	redim.
Polish	sg.	<i>kot</i>	<i>kot-ek</i>	<i>koci-ę</i>	<i>koci-ąt-k-o</i>
	pl.	<i>kot-y</i> cat	<i>kot-k-i</i> little cat	<i>koci-ęt-a</i> kitten	<i>koci-ąt-k-a</i> little kitten
	sg.	<i>ok-o</i>	<i>ocz-k-o</i>	—	—
	pl.	1. <i>ocz-y</i> , 2. <i>ok-a</i> 1. eye; 2. cell (of net)	1. <i>ocz-k-i</i> , 2. <i>ocz-k-a</i>	<i>ocz-ęt-a</i>	—
	sg.	<i>ręk-a</i>	<i>rącz-k-a</i>	—	—
pl.	<i>ręc-e</i> arm, hand	<i>rącz-k-i</i> 1. little arm, hand; 2. handle	<i>rącz-ęt-a</i>	—	
	pl.	<i>but-y</i> shoes	<i>but-k-i</i>	<i>buci-ęt-a</i>	—
	pl.	—	<i>port-k-i</i> pants	<i>porci-ęt-a</i>	—
Isthmus Nahuatl	sg.	—	<i>tao-tz̄in</i>	<i>tao-l̄in</i>	
	pl.	—	<i>tao-tzi-tz̄in</i> little girl	<i>tao-li-l̄in</i>	
	sg.	<i>chacalin</i>	<i>chacal-tz̄in</i>	—	
	pl.	<i>chacalimej</i> prawn	<i>chacal-tzi-tz̄in</i> little prawn	<i>chacal-li-l̄in</i>	

The addition of the data from Isthmus Nahuatl (Uto-Aztecan) is provisional: there is the form *chacal̄in* ‘prawn(s)’, which can be considered a variant of *chacalin* or a diminutive; in the latter case *chacal-li-l̄in* would not appear to be a missing link derivation. (The diminutive suffix *-l̄in* occurs only in a few nouns; beside *tao-l̄in* ‘little girl’ there are *choo-l̄in* and *huen-l̄in* ‘little boy’, all *diminutiva tantum*.)

2.2 Tunnel Effects

It is rare for a language to express diminution and plurality cumulatively, but it does happen. In Fula (Atlantic-Congo), as well as Swahili and many other Bantu languages, number marking can’t be separated from the formation of evaluatives, which is done by conversion, so that the forms in the four positions in the paradigm are equally distant from one another. Anderson’s (1985: 177) statement made in regard to Fula: ‘This process is (in principle given – semantic limitations) completely productive, and its full integration into the noun-class system [...] makes its inflectional status clear’ is applicable to the Bantu languages as well.

In Asmat (Trans-New Guinea) regular nouns do not distinguish number (*pok* ‘thing, things’), as is generally the case in the Papuan languages, but the diminutive markers express singularity (*mu* ‘water’, *mu-nakap* ‘a little water’) or plurality.

Diminutives can be formed from phrases as well as words, which Leont'ev 1974: 65 brings up as evidence of their non-derivational status (*amas* 'sago', *amas nec* 'raw sago', *amas nét-nakap* 'some raw sago').

	reg.	pl.	dim.	dim. pl.	
Fula	<i>wur-o</i>	<i>gur-e</i>	<i>gur-el</i>	<i>ngur-on</i>	compound
Swahili	<i>m-nyama</i>	<i>wa-nyama</i>	<i>ki-nyama</i>	<i>vi-nyama</i>	animal
Asmat	<i>pok</i>		<i>pok-nakap</i>	<i>pok-nakas</i>	thing

2.3 Little Plurals

The idea that *kračkà* and some of the other *pluralia tantum* diminutives in Bulgarian are derived from plural forms is in line with the peculiarities of their semantics and usage. It is, however, at variance with Greenberg's Universal 28: 'If both the derivation and inflection follow the root, or they both precede the root, the derivation is always between the root and the inflection' (Greenberg 1966: 93). By extension, all derivation ought to take place before the word can be inflected.

Croft (1990: 176) comments:

Derivational morphology alters the lexical meaning of the root, sometimes drastically, whereas inflectional morphology only adds semantic properties or embeds the concept denoted by the root into the larger linguistic context.

The formulation allows for exceptions if a token derivational process does not alter the lexical meaning. This is arguably the case with the formation of connotational (as opposed to denotational) evaluatives: the size of an entity is a more substantial property than its quantity, but the latter is, in turn, more stable than the speaker's attitude. Thus it is to be expected that evaluatives will time and again give occasion for digressions from the universal, as indeed they do.

In the course of his discussion of the Nootka (Wakashan) stem *inikw-ihl-'minih¹-'is²*- 'little² fire-s¹ in the house, burn plurally¹ and slightly² in the house' Sapir (1921: 104–105) comments:

the plural element precedes the diminutive in Nootka [...], which at once reveals the important fact that the plural concept is not as abstractly, as relationally, felt as in English [...]; and may not the Nootka diminutive have a slenderer, a more elusive content than our *-let* or *-ling* or the German *-chen* or *-lein*?⁸

The question is asked on behalf of the reader, but the author agrees, in a footnote:

The Nootka diminutive is doubtless more of a feeling-element, an element of nuance. This is shown by the fact that it may be used with verbs as well as with nouns. In

⁸ It is remarkable that Nootka is here contrasted to German, whose diminutive markers share at least one prominent feature with the Nootka one, that of being able to stand closer to the periphery of the word form than the plural marker (cf. Subsection 2.4). Besides, the German diminutives surely 'have a slenderer, more elusive content' (that is, are more readily used to impart the speaker's attitude) than the English ones have.

speaking to a child, one is likely to add the diminutive to any word in the sentence, regardless of whether there is an inherent diminutive meaning in the word or not.⁹

In other words, in Nootka it is not the case that diminutive formation and pluralisation are ordered as instances of derivation and inflexion, respectively. Sapir also makes the point that in Nootka ‘neither the plural nor the diminutive affix corresponds to anything else in the sentence’, which might have argued for their derivational character.

The same morpheme order is also obligatory in Dakelh, also known as Carrier (Athabaskan), and in Southern Barasano (Tukanoan): evaluative (diminutive and augmentative) markers are located closer to the periphery than number markers. This is what Stump (2001: 98f) calls head marking, not an uncommon phenomenon on a global scale, though most often observed in compounding or derivation by means of word-like affixes (that is, such as retain their adverbial, pronominal etc. character to a greater or lesser extent), and, as he acknowledges (p. 283, n. 6), seldom where an inflexional marker ends up linearly between the root and a derivational formative, as in this case.

In Kolyma Yukaghir (Paleo-Siberian) the diminutive marker *-die/-tie* follows the plural marker *-p(ul)/-pe*. Maslova (2000: 91) calls this relative order of the two markers a ‘noteworthy distributional feature’. She also notes that in many cases the diminutive is used to express affection, so that, if the intended meaning is ‘little’, forms of the verb *juko:-* ‘be little’ are used in conjunction with diminutive marking. There is also a diminutive form of the negative pronoun *n’e-leme* ‘nothing’ which has ‘emphatic impact’: *n’e-leme-die* ‘nothing at all’ (p. 92; cf. Bulgarian *ništičko*, diminutive of *ništo* ‘nothing’ < *ni-* ‘no-’ + *što* ‘what’). A further use of the diminutive marker is to merely make recent Russian loans ‘more Yukaghir-like’, as in Russian *šuka* ‘pike’ > Yukaghir *šu:ka:-die* ‘pike’, and in this case the plural marker follows the diminutive one (p. XXIV). Thus the relative position of the two markers is influenced by the function of the diminutive.

Classical Arabic¹⁰ is another language in which the use of the diminutive is by no means restricted to size.¹¹ Its nominal morphology is notorious for its large variety of plural formations, with many nouns exhibiting alternative plurals. Diminutive plurals are derived from the four ‘broken’ (transfixal) plurals which, when they are not the only plural form of a noun, have a paucal interpretation (being used with numerals in the range 3–10, etc.).¹² None of the other plurals are diminutivised; however, singular diminutives can form ‘sound’ (suffixal) plurals. Remarkably, Brockelmann (1985: 100) states that both plural diminutives (*sunayyāt* “Ge. ein Paar Jährchen”, ‘a few years’, from *sunayya*, diminutive of *sana* ‘year’) and diminutive plurals (*nusayya* “Ge. ein Paar Weiber”, ‘a few women’, from *niswa*, suppletive paucal plural of *imra’a* ‘woman’) can

⁹ And also, as he attests elsewhere (Sapir 1915), in speaking about children or speaking to or about people with various bodily deformities or disabilities. Another similar suffix, namely *-aq*, is used when addressing or discussing excessively tall or overweight people. Clearly any denotational interpretation is out of the question.

¹⁰ I thank Ali Idrissi for drawing my attention to this language and Tat’jana Frolova for providing excerpts from Wright (1981).

¹¹ Witness its formation from the demonstrative pronoun *ḏā* ‘this’, dim. *ḏayyā*, and Wright’s (1981: 167) testimony that diminutives ‘cannot be formed from nouns which have already the measure of a diminutive, as *ḡumayl* “a kind of a small bird”, *kumayt* “a bay horse”, implying that from all others they can.

¹² Since the exponent of the diminutive is also a transfix, the vowels of the paucal plural are lost; however, the prefix *’a-* in those forms that have it contributes an additional radical consonant, and the ending *-a* is retained.

express the same meaning as paucal plurals. This is an uncommon case of a reference grammar calling attention to what is beyond doubt a common phenomenon (cf. Bulgarian *godinki* ‘little years’, obviously used, like German *Jährchen*, only for pragmatic impact), but one that is seldom brought up,¹³ conceivably because the paucal plural is not a self-sustained category in most languages.

This subsection started with a generalisation based on an intuition formulated in Croft (1990). To my knowledge, the closest thing to a counterexample to that is found in Itelmen (Chukotko-Kamchatkan), in whose noun the number marker (a suffix of order 13 in Volodin 1976’s model) is located farther from the root than any of the several unproductive pejorative or hypocoristic diminutive suffixes (order 3), but closer to the root than the productive denotational diminutive suffix *-c[(a)χ]* (order 14) and the pejorative augmentative suffix *-aj* (order 15). (The two derivational processes can take place together: *qow-sk’ele*_{PEJ-CχDIM} ‘little good-for-nothing deer skin jacket’, pl. *qow-sk’ele*_{PEJ-7h_{PL}-C_{DIM}}.)

Although the central meaning of the diminutive in *-c[(a)χ]* is stated to be smallness, words such as *laccax* ‘little sun’ (cf. *lac* ‘sun’), *juñjucχ* ‘whale’ (lit. ‘little whale’, but the non-diminutive noun **juñjuñ* is never used), *qisχcaχ* ‘sky’ (lit. ‘little sky’) show that there is more to it than meets the eye. (Volodin 1976: 133 attributes the high productivity of the diminutive to the speakers’ desire to lessen at least the perceived size of large objects in their environment.)

In Alabama (Muskogean) the diminutive suffix *-(o)s(i)* (which can be repeated to form secondary diminutives: *poskòosi* ‘child, baby’, *poskòososi* ‘infant’) and the plural marker for human nouns *-ha* can occur in either order (a kind of variation seldom seen in the morphology in any language). Admittedly pluralisation and diminutivisation are not quite on a par, since only the former can correlate with something else in the sentence (to wit, the plural distributive form of the verb, marked by *ho-*, if the term is its subject). However, neither the noun suffix *-ha* nor the verb prefix *ho-* are obligatory, and their co-occurrence hardly constitutes agreement.

¹³ In Jurafsky (1996) it is only cursorily alluded to, and illustrated by Zulu pl. *amazwi* ‘words’, pl. dim. *amazwana* ‘a few words’, cf. the corresponding sg. *i(li)zwi* ‘voice; order, command; word’, dim. *i(li)zwana* ‘word’.

	reg.	dim.	pl. dim.	pl.	dim. pl.	
Nootka	<i>inikw-ihl-</i>	<i>inikw-ihl-'is-</i>	—	<i>inikw-ihl- minih-</i>	<i>inikw-ihl-'m inih-'is-</i>	fire in the house
Dakelh	<i>lhi</i>	<i>lhi-yaz</i>	—	<i>lhi-ke</i>	<i>lhi-ke-yaz</i>	dog
South. Bar.	<i>wi</i>	<i>wi-aka</i>	—	<i>wi-ri</i>	<i>wi-ri-aka</i>	house
Kol. Yuk.	<i>terike</i>	<i>terike-die</i>	—	<i>terike-pul</i>	<i>terike-p-tie</i>	wife, old woman
	(Russian) <i>šuka</i>	<i>šu:ka:-die</i>	<i>šuke-die-pe</i>	—	—	pike
Class. Ar.	<i>bayt</i>	<i>buyayt</i>	<i>buyayt-āt</i>	<i>buyūt</i>	—	house
	<i>bayt</i>	<i>buyayt</i>	—	<i>'abyāt</i>	<i>'ubayyāt</i>	verse
	<i>fatā</i>	<i>futayy</i>	<i>futayy-ūn</i>	<i>fīty-ān</i> (usual) <i>fīty-a</i> (paucal)	<i>futayy-a</i>	young man
Itelmen	<i>quwa</i>	<i>quwa-cχ quwa-sk'el</i>	<i>quwa-sk'el-7</i>	<i>quwa-7n</i>	<i>quwa-7ñ-c</i>	trousers
Alabama	(<i>posko-</i>)	<i>poskò-osi</i>	<i>poskò-osi-ha</i>	<i>poskoo-ha</i>	<i>poskoo-ha-s i</i>	child

2.4 Double Plurals

In some languages evaluatives are pluralised twice, both before and after the derivation. In Breton diminutive plurals are formed by adding the diminutive suffix *-ig* followed by *-où*, a productive plural ending characteristic of inanimate nouns¹⁴, to the plural form of the noun, whether the formation of the latter is productive, unproductive or suppletive. A similar situation obtains in Yiddish, where plurals are by and large formed as in German, although nouns of Hebrew origin retain the plural forms they have in the source language, which are suppletive from the point of view of Yiddish. The diminutive suffix is *-l(e)* (cf. German *-lein*); diminutive plurals also acquire the ending *-ex* of unknown origin, perhaps another diminutive suffix (cf. German *-chen*).

Another parallel, if only a superficial one, is found in many Bantu languages (the examples in the table are from Lamba and Mabiha), where there are different diminutive markers for the two numbers, but the original class and number marker is retained (in a reduced form or in its entirety), effectively becoming part of the stem of the diminutive noun, so that the latter has different stems for the two numbers.¹⁵

In Isthmus Nahuatl this affects one noun, – *piltzīn* ‘son, daughter’ (never used without a possessive prefix). This word is also unusual in that it has a diminutive suffix in the singular even without diminutive semantics, though this is not so in the plural.

¹⁴ Note that inanimacy is correlated with diminutivity in Breton as the feminine and especially the neuter gender are in Bulgarian.

¹⁵ This is potentially an unstable situation. In some other languages of the same family the singular prefix is retained within the forms of the diminutive noun for both numbers, so the double number marking is eliminated, and the plural diminutive correlates only with the corresponding singular, cf. Nsenga *mu-ntu* ‘person’, pl. *wá-ntu*, but dim. *ka-mu-ntu*, pl. dim. *tu-mu-ntu*. A similar development takes place occasionally in Fula as well, cf. *kor-dó* ‘slave girl’, pl. *hor-ḥe*, but dim. *kor-d-el*, pl. dim. *kor-d-on*.

The German form *Kinderchen* ‘little children’ is a classic example of a diminutive plural derivation, though there is a case for considering it a double plural (*Kind-er_{PL}-chen-0_{PL.NOM}*, gen. *Kind-er_{PL}-chen-0_{PL.GEN}*, cf. sg. *Kind-chen-0_{SG.NOM}*, gen. *Kind-chen-0_{SSG.NOM}*). Although the contrary is stated sometimes in the literature (e. g., Bauer 1983: 26), in contemporary German such diminutive plurals in *-er-chen* and *-er-lein* can be formed (without necessarily being very common) from many nouns that pluralise by *-er*, neuter as well as masculine.¹⁶ Some of these nouns have another plural form as well. One such word is *Wort* ‘word’, pl. *Worte* (mostly collective) or *Wörter* (mostly distributive), dim. *Wörterchen*.

The availability of the plurals in *-er* for subsequent morphological processes has parallels elsewhere in the languages that constitute German’s close kin, where they acquire further plural marking (cf. Middle English *child-er*, Modern English *child-r-en*, African American Vernacular English *child-r-en-s* > *chilluns*). In Dutch the old plurals of such words, reinterpreted as uninflected stems, give rise not only to new plural forms, but also to alternative diminutive plurals, used side by side with the ones obtained by pluralisation of the diminutives. In a sense what has happened here is just the opposite to what we saw in the Bulgarian diminutive plurals in *-ovce* as per fn. 4: there a part of one form of the stem has been reinterpreted as a part of a compound ending, whereas in Dutch an ending has been reinterpreted as part of an allomorph of the stem.

Many speakers perceive no semantic difference between *kindjes* and *kindertjes*; there is, however, a tendency for the former to be preferred as an individualising plural, esp. when talking of someone’s offspring, and for the latter to be interpreted as a collective form, a fact arguably related to its derivation from a plural.¹⁷ An unusually complex case is that of the noun *kleet* ‘cloth, (rarely) garment’. This word has three plural forms: *kleden* ‘cloths’, *klederen* ‘garments’ (an archaic or elevated form) and *kleren* ‘clothes’ (etymologically a syncopated version of the former, but now effectively a *plurale tantum* lexeme). The diminutive plural *kleertjes* corresponds to *kleren*; the plural diminutive *kleedjes*, to *kleden*.

A remarkable situation arises in Portuguese, where evaluatives formed by /z/-initial suffixes (diminutive *-zinh-* or *-zit-*, augmentative *-zão*) from nouns and adjectives whose stem undergoes one of several kinds of morphophonological change before plural *-s* (also /z/) have alternative plural forms in which the same changes take place before the evaluative suffix. In light of the existence of corresponding /z/-less evaluative suffixes in the language (diminutive *-inh-* and *-it-*, augmentative *-ão*) it is tempting to think that the standard orthography is misleading, and that the /z/ in *florezinhas* is neither the /z/ (written *z*) of *-zinh-* nor ‘a formative which does not realise a morpheme’ (as according to Bauer 1983: 26), but the /z/ (written *s*) of *flores*.

¹⁶ It is noteworthy that the masculine nouns involved tend to be animate (*Geist* ‘ghost’, *Gott* ‘god’, *Mann* ‘man’, *Wurm* ‘worm’). This suggests that the language sees in these forms a remedy for the conflict between animacy and the number syncretism that is characteristic of diminutives in all cases except the genitive. Another kind of remedy is explored with overt double plurals such as *Kinderchens* and *Kinderleins* (much less often formed from other nouns); a further one with *Fräulein* ‘young lady, miss’ (formally a diminutive from *Frau* ‘lady, woman’, pl. *Frauen*), which forms in the colloquial language the plural *Fräuleins*, being thus the only noun with a diminutive suffix to get the plural ending *-s* in the absence of another plural marker.

¹⁷ ‘Since the word is derived from a diminutive and has no singular, it refers to a group (e. g., a class in kindergarten)’ (Alexander Lubotsky, p.c.).

	reg.	dim.	pl. dim.	pl.	dim. pl.	
Breton	<i>bag</i>	<i>bag-ig</i>	—	<i>bag-où</i>	<i>bag-où-ig-où</i>	boat
	<i>merc'h</i>	<i>merc'h-ig</i>	—	<i>merc'h-ed</i>	<i>merc'h-ed-ig-où</i>	daughter girl
	<i>den</i>	<i>den-ig</i>	—	<i>tud</i>	<i>tud-ig-où</i>	person
Yiddish	<i>xet</i>	<i>xet-l</i>	—	<i>xatoim</i>	<i>xatoim-l-ex</i>	sin
	<i>kind</i>	<i>kind-l</i>	—	<i>kind-er</i>	<i>kind-er-l-ex</i>	child
Lamba	<i>umu-si</i>	<i>ka-mu-si</i>	—	<i>imi-si</i>	<i>tu-mi-si</i>	village
Mabiha	<i>mu-uto</i>	<i>ka-mu-uto</i>	—	<i>mi-uto</i>	<i>tu-mi-uto</i>	river
Isth. Nahuatl	<i>-pil-</i>	<i>-pil-tzīn</i>	—	<i>-pil-ohuān</i>	<i>-pil-ohuān-tzi-tzīn</i>	child
German	<i>Kind</i>	<i>Kind-chen</i>	<i>Kind-chen</i>	<i>Kind-er</i>	<i>Kind-er-chen-(s)</i>	child
Dutch	<i>kind</i>	<i>kind-je</i>	<i>kind-je-s</i>	<i>kind-er-en</i>	<i>kind-er-tje-s</i>	child
	<i>kleed</i>	<i>kleed-je</i>	<i>kleed-je-s</i>	<i>kled-en</i> <i>kled-er-en</i> <i>kler-en</i>	<i>kleer-tje-s</i>	cloth garment clothes
Portuguese	<i>flor</i>	<i>flor-zinha</i>	<i>flor-zinha-s</i>	<i>flor-es</i>	<i>flor-ez-inha-s</i>	flower
Italian	<i>bracci-o</i>	<i>bracc-in-o</i>	<i>bracc-in-i</i>	<i>bracc-i</i> <i>bracci-a</i>	<i>bracc-in-e</i>	arm
Egyptian > Coptic	<i>šm(ə)</i>	<i>šm-εə</i> <i>šimmo</i>	<i>šm-εə.w</i> <i>šimmōou</i>	<i>šm(ə).w</i>	<i>šm(ə).w-εə.w</i> <i>šimmoi</i>	stranger

In Italian¹⁸ there is a group of nouns which are masculine (and have the ending *-o*) in the singular, but can be pluralised into either gender, typically with a differentiation in the meaning: the masculine plural (ending *-i*) may have an abstract, figurative or idiomatic sense and the feminine (ending *-a* or, more rarely, *-e*) a concrete (frequently anatomical) one, or the former may be distributive and the latter collective. An example is *braccio* 1. (pl. *braccia* or occasionally *bracce*) ‘arm (of human body)’, 2. (pl. *bracci*) ‘arm (of chair), protruding part of a building etc.’. The plural form of the diminutive derivative *braccino*, namely *braccini* (m.), can have both meanings, as Merlini-Barbaresi (2004) attests. There is also a diminutive formed, in her analysis, from the feminine plural: it is *braccine*, which can be considered a double plural (once pluralised by the conversion to feminine gender and once by the regular ending *-e*).¹⁹

In Coptic some descendants of Egyptian noun-adjective compounds with *εə* ‘great’ in second position (in effect, augmentatives, though not all of them have recognisable augmentative semantics) have two different plural forms. An example is *šmmo* ‘stranger’ (from Egyptian *šm-εə* = *šēmēōə > *šēmōə), plural *šmmōou* [-o:w] or *šmmoi* [-oj]. Elanskaja (1980: 100f) argues that the Egyptian prototype of *šmmōou* is a plural form treated as a unit, whereas in the prototype of *šmmoi* both the noun and the adjective are pluralised: the former is descended from *šm-εə.w* = *šēmēōəēw > *šēmōə(ēw) and the latter from *šm.w-εə.w* = *šēmōəēw > *šēmōəjj(ēw), with loss of the Egyptian plural ending *-ēw* in both cases (as always in Coptic). To this she attributes the lower frequency of most forms in *-oi* as compared to their correlates in *-ōou*: ‘the doubly marked forms are, in a manner of speaking, twice as inflecting and by

¹⁸ I am indebted to Franz Rainer for bringing the facts of this language to my attention and for providing the relevant passage from Merlini-Barbaresi (2004).

¹⁹ The plural form *braccina* (also f.), though judged incorrect, also occurs in contemporary usage.

virtue of that are more archaic'. Already in Ancient Egyptian, that is, the lexicalisation of a compound such as *šm-εə* would have made the plural form *šm-εə.w* more common and *šm.w-εə.w* less so. This example is particularly interesting in that it lets us trace the making of an evaluative along with the variation in its plural form, which is why I am taking the liberty of including it here, although it is not about diminutives.

3. Conclusions

The languages in which parallels can be found to the several unusual diminutive plural formations in Bulgarian are not very many, but neither are they trivially few. There may be only one or two such forms (as in Isthmus Nahuatl), or this may be the general rule (as in Nootka); however, in the languages that are between these extremes the lexical items involved tend to form morphologically or semantically delineated classes (Portuguese is an example of the former, Polish of the latter, and Bulgarian of both).

The opposition between the distributive interpretation of plural diminutives and the collective interpretation of diminutive plurals (cf. especially the comments to examples (1, 6, 7, 14, 17), as well as the Dutch, Polish and Yukaghir evidence), though rarely sharp, is also revealing.²⁰ It supports the idea that these enigmatic forms are indeed connotational diminutives formed from plurals, which contrast with plurals formed from primarily denotational diminutives. This ambivalent interpretation of the diminutive, a derivational category, arguably leads to the apparent conflict with Greenberg's Universal 28.

References

- Anderson, S.A (1985), *Inflectional Morphology*, in T. Shopen (ed.), *Language Typology and Syntactic Description*, Vol. III, 150–201.
- Arnott, D.W (1995), *Fula*, in W. Bright (ed. in chief), *International Encyclopedia of Linguistics*, New York, Oxford, Oxford University Press, 2, 31–37.
- Bauer, L. (1983), *English Word-Formation*, Cambridge, Cambridge University Press.
- Brockelmann, C. (1985), *Arabische Grammatik*, Leipzig, VEB Verlag Enzyklopädie.
- Croft, W. (1990), *Typology and Universals*, Cambridge, Cambridge University Press.
- Elanskaja, A.I. (1980), *Složnye sušestvitel'nye s komponentom εə na koptskoj stadii drevneegipetskogo jazyka*, in I.S. Kacnel'son (ed.), *Drevnij Vostok: Sbornik 2*, Moscow, Nauka, 98–104.
- Green, J. (www), *Plurals of Nouns: Nak'albun/Dzinghubun Dialect*, The Yinka Déné Language Institute, Electronic publication: <http://www.ydli.org/dakinfo/plnonak.htm>.
- Greenberg, J.H. (1966), *Language Universals, with Special Reference to Feature Hierarchies*, *Janua Linguarum, Series Minor*, 59, The Hague, Mouton.
- Hemon, R. (1995), *Breton Grammar*, Dublin, Everson Gunn Teoranta.
- Jurafsky, D. (1996), *Universal Tendencies in the Semantics of the Diminutive*, "Language" 72.3, 533–578.
- Koval', A.I. (1997), *Imennye kategorii v pular-ful'ful'de*, in V.A. Vinogradov (ed.), *Osnovy afrikanskogo jazykoznanija. Imennye kategorii*, Moscow, Aspekt press, 92–220.
- Leont'ev, A.A. (1974), *Papuasskie jazyki*, Moscow, Nauka.
- Maslov, Y.S. (1981), *Grammatika bolgarskogo jazyka*, Moscow, Vysshaja shkola.
- Maslova, E.Y. (2000), *A Grammar of Kolyma Yukaghir*, Mouton de Gruyter.

²⁰ Remarkably, all types of diminutive plurals (missing link, tunnel effect, little plural and double plural derivations) behave alike in this respect.

On Diminutive Plurals and Plural Diminutives

- Merlini-Barbaresi, L. (2004), *Alterazione*, Chapter 5.1.1.7 in M. Grossmann & F. Rainer (eds.), *La formazione delle parole in italiano*, Tübingen, Niemeyer.
- Mirčev, K.S. (1963), *Istoričeska gramatika na bŕlgarskija ezik*, Sofia, Nauka i izkustvo.
- Sapir, E. (1915), *Abnormal Types of Speech in Nootka*, Canada Department of Mines Geological Survey, Memoir 62, Anthropological Series No. 5. Ottawa: Government Printing Bureau.
- Sapir, E. (1921), *Language: An Introduction to the Study of Speech*, New York, Harcourt, Brace and Company.
- Sova, L.Z. (1989), *Sinxronija i diaxronija jazykov bantu*, in A.V. Desnickaja (ed.), *Aktual'nye voprosy sravnitel'nogo jazykoznanija*, Leningrad, Nauka, 203–238.
- Stump, G.T. (2001), *Inflectional Morphology: A Theory of Paradigm Structure*, Cambridge, Cambridge University Press.
- Sylestine, C., H.K. Hardy & T. Montler (1993), *Dictionary of the Alabama Language*, Austin, University of Texas Press. (<http://www.ling.ut.edu/~montler/Alabama/Dictionary>).
- Volodin, A.P. (1976), *Itel'menskij jazyk*, Leningrad, Nauka.
- Wolgemuth, C. (2002), *Gramática náhuatl (mela'tájtŕl) de los municipios de Mecayapan y Tatahuicapan de Juárez*, Veracruz, Instituto lingüístico de verano. An electronic edition is available: <http://www.sil.org/mexico/nahuatl/istmo/G020a-DiccNahIst-NAU.htm>.
- Wright, W. (1981), *A Grammar of the Arabic Language*, Beirut, Librairie du Liban.

Roots, Deverbal Nouns and Denominal Verbs

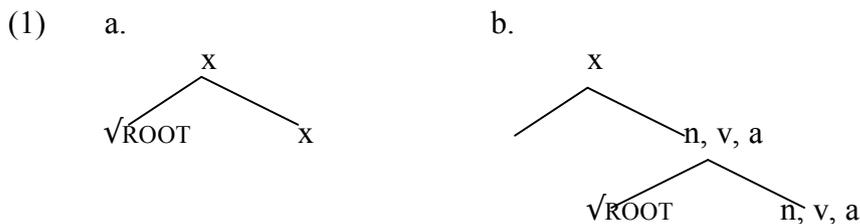
Jan Don
University of Amsterdam
j.don@uva.nl

0. Introduction

In this paper I want to show that we must distinguish between nouns derived from verbs and verbs derived from nouns. In a theory proposed by Marantz (1997) the noun *destruction* and the verb *destroy* do not stand in a derivational relationship. Neither the verb is derived from the noun, nor the other way around; but, both are derived from an underlying root $\sqrt{\text{DESTROY}}$. Categories like Verb, Noun and Adjective do not come from the lexicon under this view, but originate in syntax. Marantz' proposal is interesting because it starts from the (minimal) assumption that there is only a single device in the grammar which actually constructs larger units from smaller ones. This assumption has the immediate consequence that words cannot be built in a different place, or by a different set of rules, than sentences. Put differently, word-formation cannot take place in the lexicon but must take place in syntax.

This single-engine model is somewhat counterintuitive to the morphologist who has been happy all these years in knowing that there are two places where words can be constructed. In the lexicon, where complex words often receive an idiosyncratic interpretation and where lexical phonology may change the form of words and syntactic word-formation which is far more regular in nature both with respect to semantic interpretation and with respect to the phonological form of words.

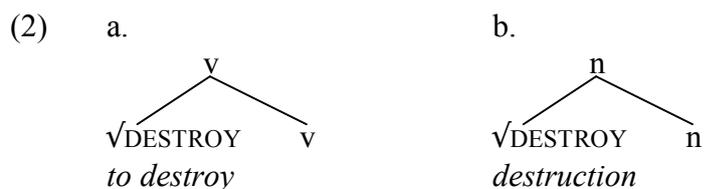
Marantz reconstructs this "two-places" idea as follows. Rather than assuming that there are two separate locations where words are formed, Marantz assumes that words can be built by combining a category-less root with a syntactic head, thus turning the root in an n, v or a, but also by combining a thus constructed word with a new syntactic head. The representations in (1) illustrate this idea:



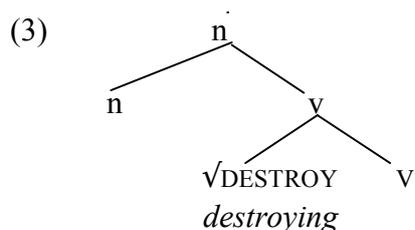
Under this reconstruction (1a) corresponds to what we would call lexical word formation, and (1b) corresponds to syntactic word-formation. But, crucially both (1a) and (1b) are syntactic constructs, but with an important difference. Marantz assumes that the structure in (1a) forms a phase (in the sense of Chomsky 1999) and consequently that the root with its syntactic head is immediately interpreted semantically and phonologically. This interpretation may be different depending on the root. That is, the information contained in the root, influences this interpretation. However, contrary to this, the word-formation depicted in (1b) is not sensitive to information contained in the root. The outer head cannot access information contained

within the root. Moreover, any interpretation given to the root in combination with its first phase head is necessarily carried over to the second. Therefore, we expect that words formed through (1b) receive an interpretation that entails the interpretation given to the root-cum-first head.

Turning now to *destroy* and *destruction*: the idea is that both result from the word-formation process depicted in (1a). That is both are root-derivations. The gerund *destroying* however is the result of a word-formation process like (1b). First, the verb *destroy* is built (through (1a)) and after that this form is combined with a nominal head *-ing*.



We can now see what we mean by saying that categories are not specified in the lexicon but originate in the syntax: *destruction* and *destroy* derive from the same category-less root.



Given such a model it is a small step to assume that conversion, or zero-derivation, is an instantiation of root derivation. Moreover, such a step answers potentially tricky questions about zero-morphemes, since we do not need such morphemes if we derive both the noun *hate* and the verb *hate* from an underlying root $\sqrt{\text{HATE}}$. However, we will demonstrate that the linguistic data point towards a more complex situation.

First, we will demonstrate that in Dutch there are good reasons to believe that the relation between some nominal forms and their verbal counterpart is *directional*. That is, one form is derived from the other. Therefore, these noun-verb pairs cannot be treated as root-derivations, although the nominal members of these pairs are not gerund-like.

Second, we distinguish between root-derivations and word-derivations by looking at the phonological and semantic properties of the derivations involved. The non-root derivational status of these derivations is confirmed by looking at their semantic and phonological properties.

Third, this predicts that apart from the word-derivations we should also be able to find true root-derivations. We will argue that some data can be better understood by assuming that they are indeed root-derivations. Thus, we conclude that Marantz' model makes the correct predictions with respect to the situation in Dutch.

1. Zero derivation in Dutch

Before turning to a detailed discussion of the Dutch data, let us briefly go into the different ways in which a relation between a root and a word can be conceived in Marantz' model.

As noted above, we can assume that a verb and a noun are derived from a common root (as in the case of *destroy* and *destruction* above). However, we may also assume that a noun is derived from a verb (which in its turn is derived from a root). As an example, Marantz gives the gerund form *destroying* but we may think of other examples in which the relation between the derived noun and the verb is directional. The same holds for verbs that might be derived from nouns (or adjectives) rather than from roots. So, in fact, under Marantz' view we may expect three different types of noun-verb pairs: nouns and verbs directly derived from roots (represented in (4a) on the hand-out), verbs derived from nouns (which are themselves derived from roots) ((4b) on the hand-out) and nouns derived from verbs (which are derived from roots) ((4c) on the hand-out). Arad (2003) shows that for a Semitic language like Hebrew such distinctions make sense and even explain some of the phonological and semantic properties related to root derivations and word-derivations. In this paper I will show that the same distinctions between root derivations and word-derivations can be made in a Germanic language like Dutch, although some of the typological differences between Hebrew and Dutch make the system unfold in a slightly different fashion.

- (4) a. [root] → [x]_n
 [root] → [x]_v
- b. [x]_n → [x]_v
 [x]_v → [x]_n

In Dutch many stems may be used either as verbs or nouns. Examples are in (5):

- (5) a. Jan *val-t* uit de boom Jan's *val*
 'John fall-s from the tree' 'John's fall'
- b. Jan *koop-t* een huis de *koop* werd gesloten
 'John buy-s a house' 'the buy was closed'
- c. Jan *feest-t* de hele nacht Jan's *feest*
 'John party-s all night' 'John's party'
- d. Jan *water-t* in de gracht het *water* in de gracht
 'John water-s in the canal' 'the water in the canal'

We will first argue on the basis of several empirical observations that the relation between the verbs and nouns in (5) is *directional*. That is, while not in every given phonologically identical Dutch noun-verb pair, it can be decided whether the verb is derived from the noun or vice versa, in many such cases it is either the verb or the noun which has to be considered as derived from the other. If we fail to recognize this directional property of the conversion-pairs in question certain generalizations about the

grammar of Dutch will be missed. The following arguments are partly taken from Don (1993) and Don (to appear). We have split the arguments in morphological arguments (section 2.1) and in phonological arguments (section 2.2).

2. Directionality of conversion

2.1 Morphology: gender and inflection type

Dutch has a gender distinction between neuter and non-neuter. The latter often called “common” gender. The gender of a noun can be seen from the choice of definite article, which is either *het* for neuter nouns, or *de* for non-neuters:¹

- | | | | | | |
|-----|----|-------------------|-------|-------------|----------|
| (6) | a. | het | huis | ‘the house’ | *de huis |
| | | art.def.neut. | house | | |
| | b. | de | weg | ‘the road’ | *het weg |
| | | art.def.non-neut. | road | | |

Dutch verbs also fall into two main classes: regular verbs, using the same stem in all tenses; and the so-called “strong” or irregular verbs which have different stems in different tenses and in some cases deviant inflectional endings:²

(7) a. Regular pattern:

	Pres. Ind.	Past	Past Participle	
sing.	noem (1 person) noem-t (2/3 person)	noem-de	ge-noem-d	‘to name’
plur.	noem-en	noem-den		

b. Irregular verbs:

Pres.	Past	Past Part.	
spijjt(-t)(-en)	speet	ge-speet-en	‘to regret’
val(-t)(-en)	viel(-en)	ge-val-en	‘to fall’
bind(-t)(-en)	bond(-en)	ge-bond-en	‘to bind’
sla(-t)(-en)	sloeg(-en)	ge-slag-en	‘to beat’

Given these two classes of nouns and two classes of verbs, without further assumptions we expect four types of conversion pairs to occur:

¹ Only in a very limited number of cases the noun seems to have a double gender status since it can be combined with both the neuter and the non-neuter definite article, e.g. *de / het prospectus* ‘the leaflet’; these cases should not be confused with nouns like *de / het slag* ‘hit’/‘kind’ or *de / het hof* ‘garden’ / ‘court’ which have different meanings in their neuter and non-neuter forms.

² The different endings are: no past tense affix (*-de* or *-te*) and in some cases the *-en* suffix appears in the past participle rather than the regular *-t* or *-d*: For example: *loop* [1st person, pres. ind.]; *liep* [1st person, past] **liep-te*; *ge-loop-en* [past participle] **ge-loop-t*.

- (8) a. regular verb – non-neuter noun
 b. regular verb – neuter noun
 c. irregular verb – non-neuter noun
 d. irregular verb – neuter noun

Interestingly, examples of the first three types of conversion pairs can be easily found. See the examples in (9a), (9b) and (9c) respectively. However, no convincing examples of the fourth type can be given.^{3, 4}

- (9) a. fiets – de fiets ‘bike’ b. werk – het werk ‘work’
 ren – de ren ‘run’ deel – het deel ‘part’
 tel – de tel ‘count’ feest – het feest ‘party’
 twijfel – de twijfel ‘doubt’ slijm – het slijm ‘slime’
- c. val – de val ‘fall’
 wijk – de wijk ‘flee’
 loop – de loop ‘walk’
 kijk – de kijk ‘look’

This lack of data of type (8d) can be easily explained under a directional view of conversion. Let us assume that the noun-producing morphological process (call it V→N-conversion) renders “common” nouns. We will motivate this specific assumption below. For now, note that in general the idea that morphological processes determine the gender of the output-class is a phenomenon we encounter in many languages. (cf. Beard 1993 for several examples from different languages). Furthermore, we assume that the verb-producing morphological process (call it N→V-conversion) renders regular verbs. This again seems a natural assumption, since irregular verbs consist of a closed class of stems. From these two independent assumptions, the systematic gap, i.e. the lack of examples of type (8d), automatically follows. The verbs in (9b) can only result from N→V conversion, since conversion in the opposite direction would render the nouns non-neuter. Similarly, the nouns in (9c) can only result from V→N-conversion, since conversion in the opposite direction would render the verbs regular. Therefore, if these processes are the only way to make new words from phonological identical forms then pairs of a irregular verbs (which cannot be the product of a conversion process) and neuter-nouns (which cannot be the output of a conversion process either) are expected to be non-existent.

³ There are two marginal examples: *het blijk – blijken* and *het spuug – spugen*. With respect to the first, we must say that the noun *blijk* only occurs in idiomatic expressions without the definite article and w.r.t. the second, we must note that for most native speakers *spugen* is a regular verb.

⁴ There is a small class of verbs that do have nominalizations with neutral gender: *sluit – het slot*, *zuig – het zog*, *bied – het bod*, *duik – het dok*, *spuug – het spog*, etc. However, for most speakers of Dutch these forms are not recognized as being morphologically related. Moreover, so far we only have looked at cases in which the verbal stem and the noun are phonologically identical. In these cases the stem vowel is changed (from [i] or [oey] to [o]). We assume that these forms are historically related by a different type of derivation.

Under a directional analysis of conversion, the systematic gap follows from independently motivated assumptions about the grammar of Dutch. The fact that there is a deverbal morphological process creating [-neuter]-nouns is independently motivated by the observation that there is a class of nouns with the same semantics as the deverbal conversions, but marked by the affix *-ing*, which also take the [-neuter] gender. Some examples are listed in (11):

- | | | | | |
|------|----------|-----------|----------------|---------------|
| (11) | verwoest | ‘destroy’ | de verwoesting | ‘destruction’ |
| | weiger | ‘refuse’ | de weigering | ‘refusal’ |

These data lend support to the assumption that noun-forming conversion in Dutch renders [-neuter]-nouns. The *-ing*-nominalizations are in complementary distribution with converted forms supporting the idea that both conversions and *-ing*-nominals are derived through the same morphological process.

Further support for the idea that V→N-conversion produces non-neuter nouns comes from nouns such as in (12):

- | | | | | | |
|------|-------------|-----------------|-------|-------------|--------|
| (12) | de aan-vang | ‘beginning’ | vang | ‘to catch’ | *vang |
| | de aan-voer | ‘supply’ | voer | ‘to supply’ | *voer |
| | de aan-hef | ‘beginning’ | hef | ‘to lift’ | *hef |
| | de in-breng | ‘participation’ | breng | ‘to bring’ | *breng |

The argument is straightforward and quite simple. These nouns, consisting of a particle and a verbal stem, are converted from the phonologically identical verbs, which consist of a left-hand particle (often a prepositional type element) and a verb as a right-hand member. As can be seen from the right-hand column in (12) the isolated nouns do not exist; so the fact that these nouns are all non-neuter is further evidence for the correctness of our hypothesis.

If we do not assume the directionality of conversion, these data become coincidental. Not only would (8d) present us with a gap that we cannot account for, also the gender of the nouns in (12) would be unaccounted for.⁵

⁵ The proposed directional analysis of conversion in Dutch is at first sight problematic in view of the following data, which all pair a (prefixed) neuter noun with an irregular (prefixed) verb:

- | | | | | |
|-----|----------------------------|----------------|----------------------------|---------------|
| (i) | be-houd _{NEUTER} | ‘preservation’ | be-houd _{STRONG} | ‘to preserve’ |
| | ver-val _{NEUTER} | ‘decay’ | ver-val _{STRONG} | ‘to decay’ |
| | ont-werp _{NEUTER} | ‘design’ | ont-werp _{STRONG} | ‘to design’ |

These data seem to fill the systematic gap of (6d) the existence of which forms one of the main arguments for the assumption that verb- and noun-forming conversion in Dutch is directional. However, in Don (1990) I have shown that the nouns in the left-hand column of (i) can be analyzed as resulting from an underlying structure as in (ii):

- (ii)
- | | | |
|--|--------------|-----|
| | N | |
| | / \ | |
| | / \ | V |
| | / \ | / \ |
| | pref. pref. | V |
| | | |
| | ge- be- houd | |

The prefix *ge-* derives, contrary to the general Right-hand Headedness of the language (cf. Trommelen & Zonneveld (1986)), neuter nouns from verbs. Furthermore, the prefix *ge-* is also used in the formation of

2.2 Phonology: syllable-structure

The idea that the systematic gap in conversion pairs can be explained by assuming two directional processes of conversion is further supported by several observations that relate the syllable-structure of underived words to their morphological category. Trommelen (1989) demonstrates that Dutch nouns may have far more complex syllable structures than verbs. According to Trommelen, the relation between syllable-structure type and category of the word is such that we might even want to say that the lexical category of an underived word can be derived from its syllable structure.⁶ This is slightly overstated but for at least a subset of underived words, it is certainly true that their syllable structure can be used as a litmus-test for their categorial status.

Let us make a distinction between words having so-called complex syllable structures, and words having simple syllable structures. Complex structures are the ones in (13), having a syllable rhyme, consisting in a long vowel, followed by a consonant, followed by two (coronal) consonants.

- | | | | |
|------|--------|---------|-----------|
| (13) | gierst | [girst] | ‘millet’ |
| | koorts | [korts] | ‘fever’ |
| | oogst | [oxst] | ‘harvest’ |

Another set of words having complex syllable structures is formed by those having final rhymes consisting of either short vowels followed by 3 consonants (of which the last one is always coronal), or long vowels followed by two consonants (again with the final one being restricted to coronals). Examples are in (14):

- | | | | | | | |
|------|---------|----------|------------|-------|--------|----------|
| (14) | worst | [wɔrst] | ‘sausage’ | koord | [kort] | ‘rope’ |
| | schurft | [sxœrft] | ‘scabies’ | reeks | [reks] | ‘series’ |
| | hengst | [hɛŋst] | ‘stallion’ | hoofd | [hoft] | ‘head’ |
| | inkt | [ɪŋkt] | ‘ink’ | | | |

With this division in mind, Trommelen now observes that there are no verbs displaying a complex syllable structure that also lack a nominal counterpart.

Following Trommelen, these examples, and many more could be given, suggest that the situation in Dutch can be characterized as follows: verbs have a very limited

past participles. Interestingly, it is absent from these participles, if the verbal stem contains a (stressless) prefix:

- | | | | | |
|-------|----------|-------------|------------|----------------|
| (iii) | maak | maak-te | ge-maak-t | |
| | haal | haal-de | ge-haal-d | |
| | ver-maak | ver-maak-te | ver-maak-t | *ge-ver-maak-t |
| | be-haal | be-haal-de | be-haal-d | *ge-be-haal-d |

This property of *ge-* was already observed and analysed by Schultink (1973), following Kiparsky’s (1966) analysis of a similar phenomenon in German. By assuming that *ge-* is deleted under exactly the same conditions (before a stressless prefix) as in the participles, the nouns in (i) can be given the structure in (ii). In doing so, they are no longer filling the systematic gap in (6d) since they are not cases of conversion, but derivations with the prefix *ge-*.

⁶ Trommelen (1989,65): “[...] [this paper] wants to give arguments for the position that in Dutch for a large part by the sound form [Du: klankvorm] of an underived word its category can be deduced, and more specifically: the degree of complexity of the syllable structure can be indicative for the morphological category of the word.” (my translation, JD)

phonological make-up: they are restricted to monosyllables, with a heavily constrained rhyme structure, or to bi-syllabic forms with the same restrictions on the rhyme of the first, and of which the last syllable contains a schwa.⁷ Nouns have far greater possibilities with respect to syllable structure and the number of syllables per stem. All verbs with a complex syllable structure and truly multi-syllabic verbs (i.e. containing at least two full vowels) have a nominal counterpart while only verbs with simple syllable structure have no nominal counterparts. So, as in the case of gender and irregular inflection, here again we are confronted with a systematic gap in the set of lexical items, as illustrated in the diagram in (15):

(15)

	Simple Syllable Structure	Complex Syllable Structure
with identical noun	numerous examples: <i>bal, lepel, kat, etc.</i>	some examples: <i>oogst, feest, hoofd</i>
no identical noun	numerous examples: <i>win, kom, vang, etc.</i>	No examples

This situation seems to call for an analysis along the following lines: Dutch, as many other languages, has phonological restrictions on the type of syllables allowed. However, in Dutch these restrictions seem to be specific for categories: the syllabic restrictions on a potential (underived) verb are far more restrictive than similar constraints on nouns.⁸ The gap in the diagram in (15) can then be easily explained if we assume that there is a lexical restriction that forbids verbs with complex syllable structures of the above-mentioned type.

These phonological generalizations concerning the syllable structure in relation to category distinctions cannot be accounted for without the assumption of categories in the lexicon. Or, to put it differently, if we suppose that in the pairs *oogst_V – oogst_N* and *hengst_N – hengst_V* the noun nor the verb are to be considered as “basic”, but that both the noun and the verb are instantiations of the same root, we cannot uphold the generalization that verbs only have rhymes consisting of lax vowels followed by two consonants, or tense vowels followed by a single consonant, since both *oogst* and *hengst* (and many more) would be counterexamples. If, as proposed, we assume that the noun is basic in these pair and the verb is derived the generalization does not face any counterexamples. So, it is impossible to account for these generalizations in a theory that does not make a distinction between nouns and verbs in the lexicon.

In order to rule out a potential diachronic explanation for these generalizations, and to establish that these generalizations belong to the knowledge of native speakers, we ran a small test. This test contained nonsense words of two types: words with complex syllable structures and multi-syllabic forms on the one hand, and words with simple syllable structures and monosyllabic forms, or bi-syllabic forms with schwa on the other. The nonsense words were read to the subjects, and the subjects were asked to choose whether these nonsense words were (stems of) verbs or nouns. All subjects

⁷ Kager & Zonneveld (1985) argue that Dutch bisyllabic words ending in a schwa-syllable should be considered as phonologically monosyllabic. That may allow for a more generalizing formulation of the constraint under scrutiny.

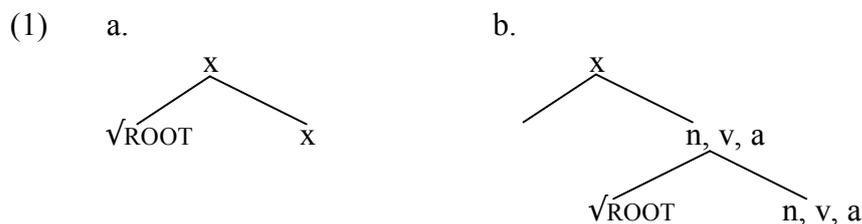
⁸ Adjectives seem to occupy a position between verbs and nouns with respect to their potential syllable structure. However, we focus the discussion here on the distinction between verbs and nouns.

without hesitation classified the words with complex syllable structures as nouns. Similarly all multi-syllabic forms were without exception classified as nouns. While mostly they hesitated for the stems with simplex syllable structures and often categorized these as verbs. For example, the test contained the nonsense word *dónkam*. This word was categorized a noun by all subjects without hesitation. A word like *dreup* on the other hand was categorized a verb by some subjects, while others reported that they could not choose. This small test with nonsense words indicates that native speakers have clear knowledge of the relation between the form of words and their category and that speakers are able to use this knowledge once they are asked to, thus ruling out any potential diachronic explanation.

Combining the above generalization with respect to syllable-structure with the discussion of morphological arguments for directionality of conversion, we also predict that verbs with complex syllable-structures and multi-syllabic verbs are regularly inflected. This is indeed the case: there are no irregular verbs that have syllable structures with these types of syllable structure.

3. Root-derivation versus word-derivation

At first sight we might be inclined to think that the above arguments for directionality go against a view in which categories arise only in the syntax. The relevant derived nouns and verbs cannot be the result of root derivation. Let us briefly turn back to the representations in (1): (repeated here for convenience)



Now, Marantz seems to claim that only so-called gerunds are formed through (1b). Only those nominalizations, contrary to derived nominals (to borrow terminology originally due to Chomsky 1972), display the syntactic behavior which we may expect from nominalizations that are created post-lexically.

If derived nominals are formed through word formation of the type represented in (1a), and if Dutch noun-forming conversion are derived nominals then there is no way to make a distinction between nouns derived from verbs and verbs derived from nouns, and thus we would have to reject this theory. To this, we would have to show that Dutch deverbal conversions are derived nominals rather than gerunds.

However, assuming for the moment that gerund-like behavior surely indicates a word-derivation, but that *vice versa* not every word-derivation necessarily displays gerund-like behaviour, we may analyse the zero-derived forms in Dutch as word-derivations in Marantz' framework. Similarly, in a recent paper Arad (2003), building on observations by Kiparsky (1982), Myers (1984) and others, shows that there is a clear distinction between the denominal verbs such as *to tape* and root derived verbs such as *to hammer*. The first class of verbs necessarily implies the use of tape (and hence the ungrammaticality of (16a)), while *to hammer* does not necessarily imply the

use of a hammer (and hence the grammaticality of (16b)): (Examples from Kiparsky 1997)

- (16) a. *She taped the picture to the wall with pushpins.
 b. She hammered the nail with a rock.

So, according to Arad we can make a clear distinction between root-derived verbs (with the structure (1a)) and noun derived verbs (which have the structure (1b)).

Applying the same argument to derived nouns, Arad claims that nouns like *kiss*, *roast*, *walk* and *slap* to give just several examples, are verb-derived nouns, since their semantics necessarily implies a kissing, roasting, walking and slapping event respectively, while nouns such as *tape* and *hammer* do not necessarily involve a taping or hammering event. Note that *kiss*, *roast* etc. are not gerunds or gerund like constructions.

Interestingly, Arad shows that also phonological properties of the nouns support the analysis. The generalization is that when the noun and the verb have strictly identical phonological properties (like e.g. stress), this goes hand-in-hand with a semantics that suggests a derivational relationship with the word rather than with the root. So, e.g. the noun *defeat* necessarily implies an act of defeating, and the stress is the same in both the noun and the verb. However, in the pair *pérmit* – *permít* the noun and the verb have a more distant semantics, suggesting a root derivation, corresponding to different stress properties.

Turning to Dutch again, this would predict that the noun-verb pairs that we have argued to stand in a directional relationship should also have the same phonology (which is the case), and the semantic directionality. Moreover, we should be able to find examples of root derivations, i.e. of related pairs, not necessarily having strictly identical phonology that stand in a looser semantic relationship to each other.

With respect to the first prediction, we should note that the phonology is identical in all data discussed so far since that was a criterion for selecting them as potential instantiations of zero-derivation. Considering the semantic relationship, let us look at the examples in (17):

- (17) [regular verb; denominal interpretation of V; neuter gender]

	N		V
a.	krijt	‘chalk’	krijten ‘to use a piece of chalk’
b.	kwijl	‘drewl’	kwijlen ‘to drewl’
c.	prijz	‘price’	prijzen ‘to put a price on’
d.	ring	‘ring’	ringen ‘to put a ring on (of a bird)’

The examples in (16) are regular verbs with a phonological make-up that we often find among the irregulars. Therefore, these verbs are claimed by us to be denominal for phonological reasons (if they were root derivations, the verb would have been an irregular verb). Interestingly, this correlates exactly with the denominal interpretation of the verbs involved. All these verbs entail the use or presence of the noun.

(18) [irregular verb; deverbal interpretation of N; non-neuter gender]

- a. kijk ‘watch’
- b. wijk ‘flee’
- c. strijk ‘clothes for ironing’

Conversely, in those cases where the verb is irregularly inflected (examples in (18)) we find a deverbal interpretation on the noun.

Also, those verbs that show a complex syllable structure (and thus are denominal according to the above given arguments for directionality) also seem to have a denominal semantics. Although it is not easy to find the relevant examples, some are in (19):

- (19) a. olie ‘oil’ olieën ‘to smear with oil’
 *Hij olie-de de pan met boter
 ‘He oil-ed the pan with butter’
- b. blinddoek ‘blindfold’ blinddoeken ‘to blindfold’
 blinddoeken implies the use of a *blinddoek*

Let us briefly summarize the argument so far. We have shown that Dutch has at least two types of noun-verb pairs: nouns derived from verbs and verbs derived from nouns. At first sight, it seems as if Marantz’ theory cannot account for these two types since categorial distinctions are not made within the lexicon and by some criterion for root-level derivation the data involved seem to be requiring a lexical analysis, i.e. they are derived nominals rather than gerunds. However, putting aside this criterion and accepting that not only gerunds but also at least some classes of derived nominals may be derived from verbs (rather than from roots), a different picture emerges. Under such a view, we expect three types of noun-verb pairs: nouns derived from verbs, verbs derived from nouns and derivations of nouns and verbs form a single root. So far, we have given examples from Dutch for the first two types but not for the latter, i.e. the root derivations. What properties are they supposed to have? Arad shows that in Hebrew roots can receive quite different interpretations depending on whether they are verbal or nominal. English seems to differ in this respect that roots are semantically related whether they turn up in verbal or nominal contexts. Dutch not surprisingly mirrors the situation in English in the sense that no widely different interpretations are given to roots in nominal and verbal environments. Apart from the semantic difference between root derivations and word-derivations, we may also expect a difference in phonology. Where the word-derivations are characterized by the fact that they so to speak “inherit” the semantic and phonology of the first phase, the root derivations are characterized by the fact that information contained in the root is available in the first phase. Therefore, different root derivations may alter the exact contents of the root. So, more or less deviant semantic interpretations for root derivations should go hand-in-hand with deviant root-phonology.

We believe that Dutch, like English and Hebrew provides some interesting examples of root derivations. Consider for example the pair *slot* – *sluit*. They are evidently related, although exhibiting a different phonology; their semantics is also clearly related but far less predictable than in the derivational cases. For example, *sluiten* not necessarily involves a *slot* (see (20)).

(20) Jan sluit het raam **is not** Jan doet het raam op slot

Also, the use of a slot not always can be described by the verb sluiten:

(21) Jan zet zijn fiets op slot **is not** *Jan sluit zijn fiets

A further piece of evidence for the different status of *slot* comes from the fact that also the nominal *sluiting* exists. *Sluiting* can be argued to be a truly deverbal noun since in all uses of the verb *sluit*, we can make the nominalization *sluiting*:

(22) a. Jan sluit het raam ‘John closes the window’
 => There is something as ‘een sluiting op het raam’

 b. Jan sluit zijn broek ‘John closes his trousers’
 => There is something as ‘een sluiting aan zijn broek’

Similar considerations hold for the pairs in (23).

(23) a. stof ‘dust’ stuiv ‘to fly (of dust)’

 b. dok ‘dock’ duik ‘to dive’

 c. zog ‘mother-milk’ zuig ‘to suck’

Conclusion

In this paper I have argued that there is a distinction between verbs derived from nouns and nouns derived from verbs. Several morphological and phonological generalizations in Dutch cannot be understood in case we fail to acknowledge directionality of derivation. At first sight this seems to be problematic for Marantz’ single engine hypothesis, since this theory does not allow for categorial distinctions in the lexicon, which seem required if we want to uphold a directional analysis. However, we may also interpret the derived nominals in Dutch in a similar way as Arad (2003) analyzes derived nominals in Hebrew. That is, the derived nominals are formed on the basis of verbal constructions that are made by merging a category-less root with a category-bearing syntactic head.

This view of things predicts that there are three types of derivations to be distinguished: verbs and nouns derived from roots, which do not show evidence for directionality, and verbs derived from nouns and nouns derived from verbs, which do show evidence for directionality.

References

- Anderson, S.R. (1982), *Where’s Morphology?*, “Linguistic Inquiry” 13, 571–612.
Arad, M. (2003), *Locality Constraints on the Interpretation of Roots: the Case of Hebrew Denominal Verbs*, “Natural Language and Linguistic Theory” 21, 737–778.
Beard, R. (1993), *Lexeme-Morpheme Base Morphology*, Albany, SUNY Albany Press.

- Booij, G. (1995), *The Phonology of Dutch*, Clarendon Press, Oxford.
- Chomsky, N. (1972), *Remarks on Nominalization*, “Studies on Semantics in Generative Grammar”, Mouton, The Hague, 11–61.
- Chomsky, N. (1995), *Bare Phrase Structure*, in G. Weibelhuth (ed.), *Government and Binding Theory and the Minimalist Program*, Oxford, Blackwell, 385–439.
- Don, J. (1990), *Prefixed nouns in Dutch*, in R. Bok-Bennema & P. Coopmans (eds.), “Linguistics in the Netherlands” 1990, 33–42.
- Don, J. (1993), *Morphological Conversion*, PhD dissertation, Utrecht University.
- Don, J. (to appear), *Categories in the Lexicon*, “Linguistics”.
- Haaften, T. van, et al. (1985), *Nominalisaties in het Nederlands*, “Glot” 8, 67–104.
- Hale, K. & S.J. Keyser (1993), *On the Argument Structure and the Lexical Expression of Syntactic Relations*, in K. Hale & S.J. Keyser (eds.), *The View from Building 20, Essays in Linguistics in honor of Sylvain Bromberger*, MIT Press, Cambridge, Mass., 53–109.
- Halle, M. & K.P. Mohanan (1985), *Segmental Phonology of Modern English*, “Linguistic Inquiry” 16, 57–116.
- Kager, R. & W. Zonneveld (1985), *Schwa, Syllables, and Extrametricality in Dutch*, “The Linguistic Review” 5, 197–221.
- Kiparsky, P. (1966), *Über den Deutschen Akzent*, “Studia Grammatica” 7, 69–97.
- Kiparsky, P. (1982), *From Cyclic Phonology to Lexical Phonology*, in H. van der Hulst & N. Smith (eds.), *The Structure of Phonological Representations*, Foris, Dordrecht.
- Kiparsky, P. (1997), *Remarks on Denominal Verbs*, in A. Alsina, J. Bresnan & P. Sells (eds.), *Argument Structure*, Stanford, CSLI, 1997, 473–499.
- Lieber, R. (1981), *On the Organization of the Lexicon*, Bloomington, IULC.
- Marantz, A. (1997), *No Escape from Syntax: Don't try morphological analysis in the privacy of your own lexicon*, in A. Dimitriadis et al. (eds.), *Proceedings of the 21st Annual Penn Linguistics Colloquium*, “UPenn Working Papers in Linguistics” 4, 201–225.
- Marantz, A. (2001), *Words*, handout of a paper presented at West Coast Conference on Formal Linguistics, 23–25 February, University of Southern California. [<http://web.mit.edu/~marantz/Public/WCCFL.doc>]
- Schultink, H. (1973), *Het prefix ge- in Nederlandse (en Duits) verleden deelwoorden*, “De Nieuwe Taalgids” 66, 101–111.
- Trommelen, M. (1983), *The syllable in Dutch with special reference to diminutive formation*, Foris, Dordrecht.
- Trommelen, M. (1989), *Lettergreepstructuur en woordcategorie*, “De Nieuwe Taalgids” 82, 64–77.
- Trommelen, M. & W. Zonneveld (1986), *Dutch Morphology: Evidence for the Right-hand head Rule*, “Linguistic Inquiry” 17, 147–169.
- Williams, E. (1981), *On the notions ‘lexically related’ and ‘head of a word’*, “Linguistic Inquiry” 12, 245–274.
- Wunderlich, D. (1997), *Cause and the Structure of Verbs*, “Linguistic Inquiry” 28, 27–68.
- Verkuyl, H. (1972), *On the compositional nature of aspects*, Kluwer, Dordrecht.
- Zubizarreta, M.L. & T. van Haaften (1988), *English -ing and Dutch -en Nominal Constructions: a Case of Simultaneous Nominal and Verbal Projections*, in M. Everaert et al. (eds.) *Morphology and Modularity, in honour of Henk Schultink*, Foris, Dordrecht, 361–393.

Jan Don

Hausa Final Vowel Shortening: Phrasal Allomorphy or Inflectional Category?

Berthold Crysmann*

German Research Center for Artificial Intelligence (DFKI GmbH)
crysmann@dfki.de

1. Introduction

In this paper, I will address the phenomenon of final vowel shortening (FVS) in Hausa¹. Based on detailed morphological evidence, I shall argue that FVS is but one exponent of a systematic morphosyntactic distinction in the language. Given the systematicity of the distinction together with the diversity of exponence, I shall conclude that a treatment in terms of inflectional morphology is to be preferred over Hayes (1990)'s analysis as Precompiled Phrasal Phonology (PPP). The morphological view will furthermore enable us to connect the Hausa data to a typologically well-established inflectional category, namely marking of the mode of argument realisation, a perspective that will deepen our understanding of Hausa syntax and morphology.

The paper is organised as follows: after a brief introduction to the basic pattern and a discussion of Hayes' account in terms of phrasal allomorphy, I shall present additional data to the extent that FVS cannot be singled out as an isolated allomorphic process. Rather, we shall see that vowel length alternation is subject to close interaction with Hausa stem morphology. Moreover, under a broader empirical perspective, the twofold length distinction will turn out to be only one of many patterns in which an underlyingly tripartite distinction is morphologically neutralised.

Next, I shall submit Hayes's surface-oriented adjacency requirement – a necessary criterion for precompiled phonologies – to some further scrutiny and show that Hausa provides a body of evidence against such a surface-oriented view, supporting instead an analysis in terms of argument structure and lexicalised traceless extraction. In section 4, I shall connect Hausa to strikingly similar phenomena in Chamorro and French, all displaying morphological sensitivity to extraction contexts (Bouma *et al.*, 2001). Furthermore, we shall see that Hausa already provides independent evidence for its membership in the typological class of extraction-marking languages.

* I am greatly indebted to my former Hausa teacher Joseph McIntyre for helping me with various empirical issues in the initial stages of this paper. I would also like to thank the audience of the 4th Mediterranean Morphology Meeting (Catania, Sep 2003), and, in particular Bernard Fradin, Joan Mascaró, and Andrew Spencer for helpful suggestions on different aspects of the proposal.

¹ Hausa is a Chadic language spoken by some 30 million speakers in Northern Nigeria and bordering areas of Niger. Hausa is a tone language, featuring 3 distinct surface tones: H, L, HL (=falling). Throughout this paper I will only mark L, using a grave accent, and falling tone, indicated by a circumflex. All syllables not marked with any diacritic are high. Vowel length, which is also distinctive, is marked by means of a colon.

The data in sections 2 and 3 of this paper are almost entirely taken from Newman's reference grammar of Hausa (Newman 2000), with glosses added by me. The Hausa data in section 4 are mainly reproduced from Davis (1986).

1.1 Hausa Final Vowel Shortening (FVS): The Basic Pattern

It is a well-known fact about Hausa that verb forms in certain lexical classes (traditionally called grades; see Parsons, 1960; Newman, 2000) undergo shortening of the final vowel, when followed by a full NP direct object: “A verb-final long vowel is shortened immediately before an object NP” (Hayes 1990: 87).

- (1) a. Na: ka:mà ki:fi:
1.S.CMPL.ABS catch fish
‘I caught fish’
- b. Na: ka:mà:
1.S.CMPL.ABS catch
‘I caught’
- c. Na: ka:mà: shi.
1.S.CMPL.ABS catch him
‘I caught it’
- d. Na: ka:mà: wà Mu:sa: ki:fi:
1.S.CMPL.ABS catch for Musa fish
‘I caught fish for Musa’
- e. ki:fɪn dà na ka:mà:
fish.DEF COMP 1.S.CMPL.ABS catch
‘The fish I caught’

The data in (1) illustrate the basic pattern with the regular grade 1 verb *ka:mà(:)* ‘to catch’: if the direct object NP is right-adjacent to the verb, as in (1a), the verb’s final vowel is short. If the direct object is unexpressed (1b) or realised as a pronominal clitic (or affix²) (1c), no shortening can be observed. The same holds, if an indirect object intervenes (1d), or if the direct object is extracted (1e).

In spite of the apparent sensitivity to phrase-structural context, Hayes (1990), however, argues that the rule of Final Vowel Shortening must apply in the lexicon, since it interacts with other lexical-phonological rules of the language, such as low-tone raising (Leben 1971).³ Low Tone Raising applies to heavy final syllables, realising an

² Although it is clearly beyond the scope of this article to engage into a full-fledged discussion of the clitic vs. affix status of Hausa direct object pronominals, there is, however, initial evidence in favour of an affixal analysis: first, they show a high degree of selection towards their host (Zwicky & Pullum 1983’s Criterion A), nothing can intervene between a direct object pronominal and its host, not even modal particles (Newman 2000:331), nor can they get fronted. Furthermore, these elements are segmentally and tonally weak, consisting of a single light (CV) syllable to which a polar tone is assigned. Choice of tone, however, does not depend on the preceding surface tone, but on the underlying tone, as detailed in the discussion of Low Tone Raising below. For the sake of this article, I conclude that an analysis of direct object pronominals as inflectional affixes is defensible on empirical grounds.

³ Besides word-boundedness, the main reason for regarding Low Tone Raising as a lexical rule is the existence of lexical exceptions. On the basis of these exceptions, Newman (2000:241f) even contests the status of Low Tone Raising as a productive synchronic rule of Hausa. See Newman & Jaggard (1989a,b); Schuh (1989) for detailed discussion.

underlying L as H, if preceded by another L. FVS can bleed Low Tone Raising, as witnessed by the following trisyllabic grade 1 verb:

- (2) a. Na: karànta:
1.S.CMPL.ABS read
'I read.'
- b. Na: karànta: littà:fii
1.S.CMPL.ABS read book
'I read the book.'
- c. Na: karànta: shi.
1.S.CMPL.ABS read it
'I read it.'

Besides interaction with other lexical-phonological rules, the shape of the pre-NP direct object form (or C-form) is not always fully predictable: some verbs, e.g., *gani*: 'see' or *bari*: 'leave', feature idiosyncratic C-forms, viz. *ga* or *bar*, respectively.

With a large number of stems, i.e. those in grade 2, shortening is accompanied by segmental change of the final vowel, which is *-i* in the C-form, *-e*: in the B-form, preceding pronominal direct objects, and *-a*: elsewhere (A-form).

- (3) a. Na: sàya:
1.S.CMPL.ABS buy
'I bought.'
- b. Na: sàye: shi
1.S.CMPL.ABS buy him
'I bought it.'
- c. Na: sàyi àbinci
1.S.CMPL.ABS buy food
'I bought food.'

Finally, in grade 2 one can find a few irregular A-forms (Newman 2000: 637), characterised by an exceptional tonal pattern (H-L instead of L-H) and/or segmental changes, e.g. *dî:bà*: (A), *dê:be*: (B), *dê:bi* (C) 'dip out, take'.

1.2 Precompiled Phrasal Phonology (PPP; Hayes 1990)

In order to reconcile the apparent sensitivity of the FVS phonological rule to phrase-structural contexts with basic tenets of both Prosodic Hierarchy Theory (Selkirk 1986; Nespor & Vogel 1982, 1986; Hayes 1989) and the *Principle of Phonology-free Syntax* (Pullum & Zwicky 1988), he suggests to preserve the restrictiveness of the indirect approach to phonology-syntax interaction offered by the theory of prosodic domains and complement it with what he calls Precompilation Theory (or Precompiled Phrasal Phonology; PPP), a kind of "phrasal allomorphy" (Hayes 1989: 92) reminiscent of Zwicky (1985)'s *Shape Conditions*.

He suggests that alternations such as Hausa FVS are allomorphic in nature, and should be derived in the lexicon. Sensitivity to syntactic context, however, is captured by means of “phonological instantiation frames”: in essence, the allomorphic variant is diacritically marked for a specific insertion context, and selection of a particular allomorph is handled by lexical insertion, subject to the Elsewhere Condition (Anderson, 1969; Kiparsky, 1973).

- (4) *Hausa shortening*:
 $V: \rightarrow V / [\dots _]_{[FrameI]}$
- (5) *Frame I*:
 $/ [_{VP} _ NP \dots]$
- (6) *Hausa raising*:
 $a \rightarrow i / [\dots _]_{[GradeII \& FrameI]}$

In the concrete case at hand, a (lexical) shortening rule (4) derives the C-form allomorph and diacritically annotates it with a reference to a particular phonological instantiation frame, as given in (5) above. Other morphophonological rules can make reference to this insertion frame as well, e.g., the grade 2 vowel raising rule in (6).

It should be clear from this very brief description that rules of allomorphy, under this approach, can make wild reference to heterogeneous types of information, namely morphological class, phonological shape and surface-syntactic and phrase-phonological environment. Furthermore, reference to surface context does not appear to be constrained by structural configurations, such as functor-argument relations, or even tree locality.

Although I have no reason to doubt, at least at this point, that Hayes’s proposal can successfully account for the empirical patterns encountered so far, there are nevertheless theoretical and methodological issues lurking here encouraging us to explore an alternative perspective on the data: first, the instantiation frames invoked by Hayes resemble very much the subcategorisation frames of Aspects-style lexical entries. However, as we have seen above, FVS only applies in the context of direct objects *in situ*. We are thus forced to assume that these instantiation frames are not meant to be reducible to ordinary subcategorisation. Under this perspective, we are confronted with a massive duplication problem: why should a language invoke two distinct, though strikingly similar, systems of subcategorisation? Moreover, if phonological instantiation frames are considered a mode of subcategorisation in its own right, PPP blurs the distinction between lexical and prosodic phonology, in that morphophonological idiosyncrasies, which were hitherto considered unambiguous evidence in favour of lexical status, do now receive an alternative interpretation as instances of PPP, a possibility that has already been exploited by Vigário (1999) to discuss away some of the evidence pointing towards a morphological analysis of European Portuguese clitics (see Crysmann, 2003 and Luís & Spencer, to appear for a detailed criticism). As a net effect, the scope of Zwicky & Pullum (1983)’s Criterion C will be severely limited.

There is, however, a theoretically less harmful interpretation of Hayes’s proposal, namely to assume that morphophonological alternations can (only) make reference to lexicalised syntactic context. Under this perspective, PPP will be reducible to standard notions of subcategorisation in lexicalist theories of syntax, e.g., HPSG or

LFG, essentially regarding phonological alternations as an exponent of morphosyntactic distinctions, or, in other words, as exponents of an inflectional category. It is of note that Selkirk has once proposed, in response to Hayes's proposal, to analyse all instances of precompiled phonologies as inflection (Hayes 1990: 106). I will argue, in the subsequent sections, that an interpretation along these lines will not only provide a theoretically cleaner solution to the paradox, but that it will also provide for a better understanding of Hausa morphosyntax, both language-internally and in a broader cross-linguistic, typological context.

2. Hausa FVS: Extending the Empirical Base

2.1 Neutral Paradigms

The perspective on Hausa FVS assumed by Hayes is essentially that of a syntactically conditioned allomorphy, described by means of a phonological rule, i.e. as a fossilised or lexicalised version of a phrase-phonological rule. This characterisation of precompiled phonology appears to me somewhat instrumental for setting apart this new device from standard notions of inflectional morphology, placing PPP halfway between true phrasal phonology and morphology. Yet, on closer inspection, this picture of a phonologically determined allomorphy seems to obscure the fact how tightly FVS is integrated with the morphological paradigms of the language.

A first piece of evidence pointing in this direction is the fact that entire classes of verbs are exempt from the application of the shortening rule. Among the 7 Hausa grades, "grade 6 is [...] very productive and commonly used" (Newman 2000: 663) indicating orientation towards the speaker. Also phonologically, verbs in this grade are highly regular, characterised by all H syllables and a final long theme vowel *-o:*.

Given Hayes's shortening rule, one would expect a short final vowel in the C-form. Yet, despite the fact that grade-6 verbs do match the structural description of the rule, the contrast is fully neutralised.

- (7)
- | | | | |
|----|--------------------|--------|--------|
| a. | ya: | sa:to: | |
| | 3.S.M.CMPL.ABS | steal | |
| | 'He stole (it)' | | |
| b. | ya: | sa:to: | shì |
| | 3.S.M.CMPL.ABS | steal | him |
| | 'He stole it' | | |
| c. | ya: | sa:to: | mo:tà: |
| | 3.S.M.CMPL.ABS | steal | car |
| | 'He stole the car' | | |

Newman (2000: 662) mentions that in Western Hausa dialects, some speakers tend to shorten the final vowel in the C-form. He adds, though, that this should be regarded as an innovation by analogy with grades 1, 2, and 4. Moreover, even for these speakers, shortening appears to be subject to an additional phonological restrictions,

namely the weight of the penultimate, a restriction that is not operative in any other grade.

- (8) a. ya: karanto là:ba:rì:
 3.S.M.CMPL.ABS read news
 ‘He read the news’
- b. sun harbo za:kì:
 3.P.CMPL.ABS shot lion
 ‘They shot a lion’
- c. mun baro: yâ:ra: à gida:
 1.P.CMPL.ABS leave children at house
 ‘We left the children at home’

If Newman’s interpretation is correct, we have good reason to question a phrase-phonological rule as the historical basis of current FVS.

Apart from grade 6, there is another set of verbs which fails to undergo FVS, all characterised by the subregular pattern *CiCa:*. Although verbs like *kiraa* ‘call’ and *jiraa* ‘wait’ are pretty similar to grade 1 and grade 2 verbs, as far as the segmental level is concerned, still no shortening applies.

- (9) ya: kira: mùtûm
 3.S.M.CMPL.ABS call man
 ‘He called the man’

Although I concur with Hayes in adopting the lexicon as the locus of rule application, I take the tight integration of this phenomenon with Hausa stem classes as an indicator of the morphological status of the alternation.

2.2 Tripartite Paradigms

We have already mentioned in passing that shortening is not the only device by which Hausa C-forms are marked: in grade 2 shortening is accompanied by vowel change. Moreover, unlike grade 1, not only the C-form is set apart, but rather three different situations are morphologically distinguished. Traditionally, Hausaists adopt (at least) a three-fold system to describe the verb forms in all Hausa grades. Under this perspective, the identity of A and B-forms in grade 1 can be regarded as another instance of neutralisation.

- (10) a. Na: sàya:
 1.S.CMPL.ABS buy
 ‘I bought’
- b. Na: sàyi àbinci
 1.S.CMPL.ABS buy food
 ‘I bought food’

- c. sàyi!
 buy.IMP
 ‘Buy!’
- d. sàyi àbinci!
 buy.IMP food
 ‘Buy food!’

Further evidence in favour of an essentially tripartite morphological system comes from grade 2 imperatives: here, the A-form of grade 2 verbs is identical to the C-form, displaying a short final *-i*. Selection of the C-form in the A-form context is probably best understood as a rule of referral, since identity does not only involve selection of the final vowel, but also selection of stem form.

- (11) a. ya: dí:bà:
 3.S.M.CMPL.ABS dip.out
 ‘He dipped (it) out’
- b. dè:bi!
 dip.out.IMP
 ‘Dip out!’

Taking together the evidence from grades 1, 2 and 6, we can conclude that what we find in Hausa is essentially a tri-partite system of morphological marking that displays different patterns of neutralisation (or syncretism): A-B-C (grade 6), A-B vs. C (grade 1), A-C vs. B (grade 2 imperative)⁴, and A vs. B vs. C (grade 2 “indicative”). The syncretism that can be observed between the A- and C-form cells in the grade 2 imperative yet again underlines the tight integration of vowel shortening with the overall morphological system: with bisyllabic grade 2 A-forms, the rule of referral constitutes the sole exponent of the morphological category imperative, as the typical L-initial tonal pattern of imperatives is effectively masked in this grade.

2.3 *Verbal Nouns (Gerunds)*

Verbal inflectional categories like tense and aspect are signalled by means of discrete markers, which are often fused with exponents of subject agreement. Typically these TAM markers select a verb in its base form. Exceptional in this respect are the continuative markers (absolute/relative/negative), where a gerundive form of the verb is chosen (see Tuller, 1986 and Davis, 1993 for detailed discussion of the syntactic properties of verbal nouns). These verbal nouns (VNs) come in essentially two forms: a regular, or weak VN, and a strong form, which morphologically behaves more or less like a noun.

In this section, I will show that the object-sensitive alternation found with verbs carries over to non-verbal categories as well, and that, in sum, these alternations, despite

⁴ As pointed out to me by Joe McIntyre (p.c.), irregular monosyllabic verbs of the Ci type also display neutralisation between A and C forms, e.g. *fí* ‘exceed’, *ci* ‘eat’, and *ji* ‘hear’.

clear difference in exponence, are far too pervasive to be regarded as a mere instance of allomorphy, at least not without missing a central property of Hausa morphology.

2.3.1 Weak Verbal Nouns

Verb in grades 1, 4, 5, 6 and 7 typically choose the regular weak VN as their gerundive form (see Newman 2000: ch. 77), although some verbs in these grades also possess (alternate) strong form VNs (e.g. *dinkà*: ‘sow’ – *dinkì*: ‘sowing (m)’).

Weak VNs in the A-form are derived by suffixation of -`*wa*:. In all other forms, the weak VN is identical to the corresponding form of the base verb.

(12)

grade	form	A	B	C	D/E
1	V	<i>karànta</i> :	karànta: shi	karàntà	karànta: wà/masà
	VN	<i>karàntà:wa</i> :	karànta: shi	karàntà rufè	karànta: wà/masà
4	V	<i>rufè</i> :	rufè: shi		rufè: wà/masà
	VN	<i>rufè:wa</i> :	rufè: shi	rufè	rufè: wà/masà
6	V	<i>ka:wo</i> :	ka:wo: shi	ka:wo:	ka:wo: wà/masà
	VN	<i>ka:wô:wa</i> :	ka:wo: shi	ka:wo:	ka:wo: wà/masà

Four things are worth noticing here: first, in the context of neutralisations within a basically tri-partite system, these data provide the missing type of neutralisation (A vs. B-C).

Second, and most importantly, overt marking of this deverbal form singles out the A-form. In contrast to the picture drawn by Hayes, where forms other than the C-form were regarded as default realisations, governed by the Elsewhere Condition, the above data appear to support the view that the A-form actually forms a natural class, comprising intransitives, suppressed direct objects, and non-locally realised direct objects.

- (13) a. *yanà*: *karàntà:wa*:
 3.S.M.CONT.ABS reading
 ‘he is reading’
- b. *litta:fin* *dà* *yakè*: *karàntà:wa*:
 book.DEF.M that 3.S.M.CONT.REL reading
 ‘the book he is reading’

Under Hayes’s account, which is confined to strict adjacency, the identical morphological marking in (13) must appear as purely accidental. Under a slightly different angle, we might as well take the non-locality of the relation as an indicator of this form’s inflectional status, following essentially the characterisation given in Hayes (1990: 106). There is, however, a way to save a Hayes-style precompilation account in the light of these data: if we assume that zero derivation or a rule of referral, rather than suffixation constitutes the more specific case, the marking patterns of weak VNs might be assimilated to that of grade 1 base verbs. Although technically surely viable, such a solution would stand in sharp contradiction to what is standardly assumed as a working principle of human language, namely that zero derivation is *the* default option in the

absence of any more specific marking, cf., e.g., Stump’s *Identity Function Default* (Stump 1993, 2001).⁵ Furthermore, such a solution would be highly uneconomical, owing to the fact that zero marking would involve three clearly distinct instantiation frames: unlike vowel shortening with base verbs, derivation of weak VNs treats the case of intervening indirect objects differently from other A-form environments, thereby strengthening the view of the A-form as a distinct class, not reducible to surface configurations.

Finally, the fact that marking of A-forms can even be attested for deverbal forms in grades that otherwise neutralise the distinction, should be taken as strong evidence both for the centrality of such an inflectional distinction and for the status of the A-form as a natural inflectional class.

2.3.2 *Strong Verbal Nouns*

Verbs in grade 2 and 3 typically use a subregular or irregular strong VN in the continuative. Newman (2000: ch. 77) subdivides strong VNs into two broader classes: regular stem-derived VNs, which are identical to the A-form in grade 2 and which are assigned mostly feminine gender, and base-derived VNs, which display a greater variation w.r.t. shape. Many grade-2 verbs, as well as verbs from other grades have an alternate base-derived VN, alongside the stem-derived or weak form. In a few cases, the irregular form has completely replaced the regular one. Although the forms of strong VNs, in particular base-derived ones, are morphologically quite heterogeneous, they all obligatorily take the “linker” *-n/-r* in the B and C-forms, thereby behaving essentially like nouns: within the NP, the head noun is suffixed with the linker preceding a pronominal or full NP complement. Choice of the linker depends on the inherent gender of the head noun or VN, i.e. *-n* for masculine and *-r* for feminine.

- (14) a. ta: kàrbi kuđi:
 3.F.S.CMPL.ABS receive money
 ‘She received money’
- b. ta: kàrbe: shi
 3.F.S.CMPL.ABS receive him
 ‘She received it’
- c. abîn dà ta kàrba:
 thing that 3.F.S.CMPL.ABS receive
 ‘The thing she received’
- (15) a. tanà: kàrban kuđi:
 3.F.S.CONT.ABS receive.M money
 ‘She is receiving money’

⁵ Even if we did not accept this argument – because the *Identity Function Default* might not be applicable to linguistic areas outside morphology –, a precompilation account will be equally hard pressed to explain that both B and C-forms invoke zero derivation, given that the syntactic environments in which these forms can surface are quite distinct: as argued in footnote 2, direct object pronominals display a good deal of properties that make them qualify as pronominal affixes. As a consequence, it will turn out to be difficult to provide a unified phonological instantiation frame for these forms.

- b. tanà: kàráansà
 3.F.S.CONT.ABS receive.M.POSS.M
 ‘She is receiving it.’
- c. abîn dà takè: kàrfa:
 thing.DEF.M that 3.F.S.CONT.REL receive
 ‘The thing she is receiving’
- (16) a. ta: karàntà litta:fin Audù
 3.S.F.CMPL.ABS read book.M Audu
 ‘She read Audu’s book’
- b. ta: karàntà litta:finsà
 3.S.F.CMPL.ABS read book.M.POSS.M
 ‘She read his book’
- c. Audù ne: ta karàntà litta:finsà
 Audu 3.S.F.CMPL.ABS read book.M.POSS.M
 ‘It’s Audu she read a book of’
- d. ta: karàntà litta:fi:
 3.S.F.CMPL.ABS read book
 ‘She read a book’

Several things are important here: first, despite the difference in major morphological class, the distribution of the A-form of strong VNs is identical, in all relevant aspects, to that of ordinary verbs. Second, we again find neutralisation, this time affecting frames B and C on the one side, and A, D, and E on the other. Thus, the contrast between A and C form that is so characteristic of FVS, is present here as well, although exponence is radically different. Third, under the broader perspective of a basically tripartite system for marking argument realisation, Hayes (1990)’s claim that X’-categories are treated differently cannot be maintained: while this may be true, if we regard FVS as an isolated phonological process, we have established in the preceding sections that this view has a very limited explanatory potential, already failing to account for the full range of variation and neutralisation within the verbal paradigms. As illustrated by the data in (14–16), marking of argument realisation not only generalises from verbs to verbal nouns (15), but also to ordinary common nouns like *litta:fi*: ‘book’ (16). Within proper NPs, not all environments for the A-form are attested, owing to the fact that extraction out of NPs is independently ruled out in Hausa. Instead, a resumptive (affixal) pronoun must be used. Still, in intransitive contexts, the partitioning is exactly parallel to that of VNs. With verbal nouns, where this island effect is not operative, A-frame environments are exactly those found with true verbs.

Summary

In this section, I have argued that Hausa FVS is but one exponent of a much more fundamental morphological distinction drawn in the language. To my mind, the alternation is far too pervasive to warrant an analysis in terms of (subregular)

allomorphy, at least not without missing an important property of the language. In particular, it affects the two major open class categories of Hausa, namely verbs and nouns in a similar way. Furthermore, we have seen that opposition w.r.t. vowel length, which is regarded as quite fundamental in Hayes's account, is but one way an at least threefold morphological distinction is neutralised, depending on a specific morphological class. Finally, we have established, mostly on the basis of the marking of weak VNs, that the A-form must be considered a natural morphological class in Hausa, ranging over intransitives as well transitives with unexpressed or non-locally realised direct objects. On the basis of the striking similarity of the distinctions involved, together with the degree of variation found in the set of exponents, I conclude that we are dealing here with an inflectional category.

3. Adjacency

In the preceding section, I have restricted myself to a discussion of the morphological aspects of Hausa FVS and related phenomena. The proposal to regard FVS as an instance of PPP, however, was mainly motivated by an apparent surface-syntactic constraint on the alternation. In order to maintain an essentially morphological analysis of the data, it is crucial, though, to determine what exactly the morphosyntactic property is that is morphologically expressed. Consequently, I will subject the syntactic environments of the alternation to some further scrutiny, showing that (a) the apparently surface-syntactic conditioning is but an artefact of canonical Hausa word order, and (b) that exceptions to a purely surface-oriented constraint can be found which point towards argument structure as the proper representation to formulate the contextual restrictions.

3.1 *Intervention*

3.1.1 *Indirect Objects*

One of the main pieces of evidence to motivate the surface-syntactic conditioning of FVS are the intervention data found in ditransitives (Hayes 1990: 93):

- (17) Na: ka:mà: wà Mu:sa: ki:fi:
1.S.CMPL.ABS catch for Musa fish
'I caught fish for Musa'

Here, shortening does not apply, even though *ka:ma:* does take a direct object complement (*ki:fi:*), realised in the local clause. At first blush, it appears that it is not transitivity *per se* that matters but surface adjacency of an NP complement.

However, a property of Hausa not taken into account by Hayes (1990) is the very strict word order in this language. As detailed by Newman (2000: ch. 39) (but cf. any learner's grammar of Hausa, e.g., Cowan & Schuh 1976) the canonical position of the indirect object, be it pronominal or not, is directly after the verb. Nothing save a few very light modal particles can intervene between the verb and the direct object marker *-wà*. Direct objects, in particular, canonically follow the indirect object. If, for reasons

of prosodic weight, an indirect object must be shifted to the right, it has to be expressed by means of a prepositional phrase *gà*⁶:

- (18) a. ya: fadā: wà mutànhèn làba:ri:
 3.S.M.CMPL.ABS tell men.DEF news
 ‘He told the men the news.’
- b. ya: fadí làba:ri: gà mutànhèn dà
 3.S.M.CMPL.ABS tell news to men.DEF that
 sukè: goyon ba:yansà
 3.P.CONT.REL supporting him
 ‘He told then the news to the men who were supporting him’

In this respect, basic Hausa ditransitives are quite similar to dative shift in English, where the indirect before direct object order is equally strict.

If we assume that word order in languages such as Hausa and English is determined by an obliqueness hierarchy on the argument structure of the verb (Pollard & Sag 1987), right dislocation of the indirect object will necessarily involve demotion to an oblique PP argument. Under this perspective, non-application of FVS with ditransitives can readily be accounted for at the level of argument structure, without any reference to surface adjacency.

In this context, it is of note that in the Kano dialect, the stranded IO marker *-wà* is lengthened whenever the IO itself is extracted. Newman (2000: 277) offers a potential explanation to the extent that speakers of this variety have reanalysed the almost inseparable IO marker as a verbal clitic (or rather affix [BC]).

- (19) Standard Hausa
- a. shi: nè: mùtumìn dà ya gayà: wà
 he COP man that 3.S.M.CMPL.REL tell IOM
 ‘He is the man I told it to’
- b. wà: ka ji: wà ciwo:
 who 2.S.M.CMPL.REL feel IOM injury
 ‘Whom did you injure?’
- c. ya ji: wà ya:rò: ciwo:
 3.S.M.CMPL.REL feel IOM boy injury
 ‘He injured the boy’

- (20) Kano dialect
- a. shi: nè: mùtumìn dà ya gayà:wà:
 he COP man that 3.S.M.CMPL.REL tell.IOM
 ‘He is the man I told it to’

⁶ Although historically, there is reason to believe that *wà* derives from *gà* (Newman 2000:276), synchronically, these two must be clearly distinguished, since *-wà*, unlike any other preposition is obligatorily stranded in extraction contexts, whereas stranding is ruled out for true prepositions.

b. wà: ka ji: wà: ciwo:
 who 2.S.M.CMPL.REL feel IOM injury
 ‘Whom did you injure?’

c. ya ji: wà ya:rò: ciwo:
 3.S.M.CMPL.REL feel IOM boy injury
 ‘He injured the boy’

With the IO marker being reanalysed as part of the verb, these speakers now choose short (=“C form”) *wà*, whenever the least oblique complement is locally realised, but lengthen it to “A-form” *-wà:*, if it is extracted. Note that presence or absence of a more oblique direct object does not have any impact on the lengthening. To summarise, these Kano dialect speakers have generalised FVS to be sensitive to the least oblique complement, regardless of function, whereas the Standard Hausa pattern can be reinterpreted in such a way that this sensitivity additionally takes into account the grammatical function of this complement.

3.1.2 Modal Particles

With the exception of the Kano dialect data, our discussion of word order and obliqueness in the preceding section has so far not been very conclusive, only offering an alternative interpretation of the data, i.e. in terms of argument structure rather than surface adjacency.

Clear evidence against the adjacency condition⁷ formulated by Hayes (1990) comes from modal particles (Schmalting 1991; Newman 2000). Although other modifiers cannot separate a verb from its direct object or indirect object complement (Joseph McIntyre, p.c.), modal particles can actually intervene.

(21) a. Ya: shuukà **kuma** audùga:
 he.CMPL.ABS planted also wheat
 ‘He also planted wheat’

b. *Ya: shuukà: **kuma** audùga:
 he.CMPL.ABS planted also wheat
 ‘He also planted wheat’

(22) a. ya: ga kuma irin ka:yàyya:kîn dà kè: ciki
 3.S.M.CMPL.ABS see also kind goods that CONT.REL inside
 ‘he saw also the kind of goods that were inside’

b. ta: tàmbàyi **kùwa** mà:târ
 3.S.F.CMLP.ABS ask moreover woman
 ‘She asked, moreover, the woman’

What is telling about these data is that surface intervention does not affect selection of the short vowel C-form, in any of the cases. Sure, one could try and refine the phonological instantiation frames to take these elements into account, but in doing

⁷ Hayes mentions these facts in a footnote, casually remarking that his Frame 1 needed to receive some refinement to take these elements into account.

so, the adjacency-oriented precompilation approach will lose much of its appeal: as Hayes claims himself (p. 106), strict adjacency is a defining property of precompiled phonologies and not so typical of inflection. If the adjacency requirements have to be relaxed, this can be taken as indirect evidence in favour of inflectional status.

3.1.3 Negation (Northern Dialects)

Similar evidence can be found in some Northern dialects of Hausa (Newman 2000). In Standard Hausa, sentential negation is expressed, in most tenses, by a discontinuous negative marker *bà...ba* where the first part immediately precedes the TAM marker (and sometimes fuses with it) and the second part is found VP-finally, either including (marked) or excluding complement sentences.

As noted by Newman (2000: 639), in some Northern varieties the second part of the discontinuous negation marker also appears directly after the verb, separating it from its direct object NP complement. With pronominal direct objects, such intervention is not possible, underlining the affixal status of the Hausa object pronouns (see footnote 2).

(23) Standard Hausa

a.	bàì	hàrbi	gi:wa:	ba
	3.S.M.CMPL.NEG	shoot	elephant	NEG
	'He didn't shoot an elephant'			

b.	bàì	hàrbe:	tà	ba
	3.S.M.CMPL.NEG	shoot	her	NEG
	'He didn't shoot it'			

(24) Northern dialects

a.	bàì	hàrbi	ba	gi:wa:
	3.S.M.CMPL.NEG	shoot	NEG	elephant
	'He didn't shoot an elephant'			

b.	*bàì	hàrbe:	ba	tà
	3.S.M.CMPL.NEG	shoot	NEG	her
	'He didn't shoot it'			

It should come as no surprise now that intervention does, again, not impede selection of the C-form (24). In contrast to modal particles, the marker of sentential negation cannot, under whatsoever circumstances, be reanalysed as part of the following NP. Thus, the Kano dialect data discussed above, together with the Northern dialect data presented here reveal, even more clearly than the standard variety, that surface adjacency is not the relevant concept to address the distribution of FVS in Hausa.

3.2 Double Accusatives

The finally conclusive piece of evidence on the issue comes from verbs taking two DO complements. Although, in these constructions, both complements are realised as direct objects (25), the first DO receives special status, being the “structural” object susceptible to promotion (in grade 7; see (26)):

- (25) a. sun biya: Mu:sa: kuđi:
 3.P.CMPL.ABS pay Musa money
 ‘They paid Musa money’
- b. kadà kà rò:ki Bàla: go:rò!
 2.S.M.NEG.SUBJ beg Bala cola nut
 ‘Don’t ask Bala for cola nuts!’
- (26) a. Abdù ba: yà: rò:kuwa: go:rò à ha:lin yànzù
 Abdu 3.S.M.CONT.NEG beg cola nut now
 ‘Abdu was asked for cola nuts.’
- b. *Go:rò ba: yà: rò:kuwa: Abdù à ha:lin yànzù
 cola nut 3.S.M.CONT.NEG beg Abdu now

However, if this first DO is extracted, as in (27), the verb (or VN) appears in its A-form, despite the presence of a right-adjacent direct object complement (Newman 2000).

- (27) a. su wà: kukè: biyà: kuđin?
 who.p 2.P.CONT.REL pay money.DEF.M
 ‘Who are you paying the money?’
- b. *su wà: kukè: biyàn kuđin?
 who.p 2.P.CONT.REL pay money.DEF.M

To conclude, these facts suggest, just like the intervention data, that surface adjacency fails to capture the full range of data and that reference to a privileged argument and its mode of realisation provide a more consistent picture of the Hausa data, a solution that I will explore in more detail in the following section. Moreover, this perspective will also align more neatly with the morphological facts established in the previous section, ultimately providing a definition of the inflectional category I consider FVS to be an exponent of.

4. Modes of Argument Realisation and Morphological Marking

In the preceding sections, I have argued that FVS in Hausa is but one exponent of a highly systematic distinction drawn in the language relating to the mode of realisation of some privileged argument, viz. the direct object. In particular, we have seen that the contexts in which A, B, and C-forms appear are highly consistent, even across major categories. As such, the underlying distinction is “based on a fairly restricted set of syntactic structural relations”, a property Hayes (1990: 106) takes as a defining property of inflectional morphology. Furthermore, the closer look at the full range of morphological alternation has revealed that, unlike Hayes’s characterisation of precompiled phonology, these data do not “involve rather haphazard environments that reflect [their] origin in true phrasal phonology” (Hayes 1990: 106). Furthermore, the

phenomena at hand are not “subject to a strict locality requirement” (Hayes 1990: 106) defined in terms of surface adjacency, as claimed by Hayes. Moreover, as evidenced by the morphology of weak VNs, reference to non-local realisation is a fundamental property of the system.

In this section I will review independent evidence both from Hausa and from language typology that underlines that the approach adopted here can not only do justice to the systematicity of the phenomenon, but that it will also further our understanding of Hausa morphosyntax in a broader cross-linguistic context.

4.1 Cross-linguistic Evidence

In their (2001) article, Bouma *et al.* propose a novel theory of extraction that operates crucially on argument structure: in this theory, which is developed within the framework of Head-driven Phrase Structure Grammar (Pollard & Sag 1987, 1994), both the introduction of a gap and the percolation of non-local information up the tree proceed via the argument structure of a lexical head. Thus, “information about the extracted element is locally encoded throughout the extraction path” (Bouma *et al.* 2001: 1).

What is important about this proposal in the present context, is that the authors motivate their approach on the basis of a wide range of extraction-sensitive morphological data. In particular, they discuss evidence from languages as diverse as Irish (Sells 1984; McCloskey 1989), Chamorro (Chung 1998), and French (Kayne & Pollock 1978; Kayne 1989; Miller & Sag 1997), all involving morphological marking of extraction contexts. The authors claim that similar evidence can be found in a number of other languages, including Palauan, Icelandic, Kikuyu, Ewe, Thompson Salish, Moore, Spanish, and Yiddish (see Bouma *et al.* 2001: 2 for references).

In Chamorro, as illustrated by the following data, verbs are morphologically marked depending on the mode of realisation of their subject, i.e. inflection signals whether or not a subject is extracted or contains a gap.

(28) Chamorro (Bouma *et al.* 2001: 27)

- a. Hayi *f-um-a'gasi* i kareta
 who WH.SU-wash the car
 ‘Who washed the car?’
- b. Hayi si Juan ha-sangan-i hao [*f-um-a'gasi* i kareta]
 who UNM Juan tell you WH.SU-wash the car
 ‘Who did Juan tell you washed the car?’
- c. Hafa *um-istotba* hao [ni malagao'-na i lahi-mu]
 what WH.SU-disturb you COMP WH.OBL-want-3SG the son-your
 ‘What does it disturb you that your son wants?’

These data show some striking similarity with what we found in Hausa: in both languages, verbal morphology is used to mark local vs. non-local realisation of some argument.

An even closer analogue to Hausa is French participle agreement (Kayne & Pollock 1978; Kayne 1989; Miller & Sag 1997): when used in conjunction with the

auxiliary *avoir*, past participles in this language may display agreement with the direct object. Presence vs. absence of agreement, however, depends on the way the direct object is realised: with locally realised direct object NPs, past participle is ruled out, and a default masculine singular form is selected. If, however, the direct object is extracted or realised as a pronominal affix on the auxiliary, the participle has to agree in number and gender with its direct object.

- (29) a. Marie a écrit / *écrite la lettre
 Marie has written the letter
 ‘Marie has written the letter’
- b. Marie l’a *écrit / écrite.
 Marie her-has written
 ‘Marie has written it (=the letter)’
- c. la lettre que Marie a *écrit / écrite.
 the letter that Marie has written
 ‘the letter that Marie wrote’
- (30) a. Marie s’est coupée/*coupé.
 Marie self.is cut
 ‘Marie has cut herself.’
- b. Marie s’est coupé/*coupée.
 Marie self.is cut
 ‘Marie has cut herself’
- c. la maison qu’il s’est construite/*construit.
 the house that he self.is built
 ‘the house he has built for himself’

If we compare now the French data with Hausa, we find that the former is actually a mirror image of the latter: while in French, presence of participle agreement morphologically expresses non-local realisation of a direct object complement, in Hausa, it is by-and-large local realisation of a direct object that receives morphological expression. Under this view, the role of the A-form, which is morphologically unmarked in the overwhelming majority of the cases, functions as a default form: in addition to non-local realisation, this form is used in all those cases where the distinction simply has no bearing.

4.2 *Further Evidence from Hausa: Marking of UDCs*

Although we cannot overestimate the role of the typological similarity between French and Hausa in our understanding of FVS and related phenomena, it would be even more satisfying, if we could find independent language-internal evidence, showing that Hausa is really an instance of this typologically well-attested type of languages, where morphological marking of extraction or unbounded dependency constructions (UDCs)

is a defining characteristic. As we will see shortly, exactly this type of evidence can in fact be found.

As we have already mentioned above, verbal inflectional categories such as marking for tense, aspect and mood are expressed, in Hausa, by a set of independent TAM markers, preceding the verb or VP. Often, these markers are fused with subject agreement and the marker of negation. Although neutralised in most tenses (including all negative “tenses”), continuative and completive aspect have two independent sets of forms, called absolutive (or general) vs. relative.

Although, in narratives, the relative completive has a secondary function for describing a series of events, in normal speech, choice between these sets is syntactically conditioned (Tuller 1986; Davis 1986; Newman 2000).

(31) Declaratives

- a. mutà:ne: sun zo: jiyà:
 people 3.P.CMPL.ABS come yesterday
 ‘The people came yesterday’
- b. mutà:ne: sunà: zuwà:
 people 3.P.CONT.ABS coming
 ‘The people are coming’

(32) Relative clauses

- a. mutà:nen dà sukà /*sun zo: jiyà:
 men.DEF.P that 3.P.CMPL.REL 3.P.CMPL.ABS come yesterday
 ‘the people who came yesterday’
- b. mutà:nen dà sukè: /*sunà: zuwà:
 men.DEF.P that 3.P.CMPL.REL 3.P.CMPL.ABS coming
 ‘the people who are coming’

(33) Wh questions

- a. mè: ya /*ya: gani:
 what 3.S.M.CMPL.REL 3.S.M.CMPL.ABS see
 ‘What did he see?’

(34) Topicalisation

- a. Kànde cè: ta /*ta: zo:
 Kande COP 3.S.F.CMPL.REL 3.S.F.CMPL.ABS come
 ‘It’s Kande who came?’
- b. cikin mo:tà: ne: mukà /*mun zo:
 in car COP 1.P.CMPL.REL 1.P.CMPL.ABS come
 ‘By car we came’

As illustrated by the data above, markers from the absolute set are chosen in ordinary sentences without any unbounded dependencies. Once a non-local dependency is present, forms from the relative set must be used instead.⁸

- (35) mè: sukè: fatan sun /*sukà gamà:
what 3.P.CONT.REL hoping 3.P.CMPL.ABS 3.P.CMPL.REL finish
'What did they hope they have finished?'

Although it is pretty evident that this alternation is sensitive to extraction contexts, the data in (35) reveal that selection of the relative set of TAM markers is only triggered at the point where the nonlocal dependency is bound off by a filler (Davis 1986; Newman 2000).

In sum, we can conclude that marking of nonlocal dependencies is a central property of Hausa morphosyntax. Marking of unbounded dependencies actually demarkates the two extreme points of a UDC, i.e. the filler and the gap: while the position of the former is morphologically signalled by the choice of TAM marker, position of the latter is marked, at least for direct objects, by selecting the A-form.⁹

Note further that in contemporary lexicalist frameworks such as Head-driven Phrase Structure Grammar (HPSG) or Lexical Functional Grammar (LFG), reference to local vs. non-local realisation of arguments can be straightforwardly expressed without any recourse to phrase-structural configurations, either by means of head-driven, traceless extraction (HPSG), or inside-out functional uncertainty (LFG).¹⁰ Under this perspective, the precompilation approach appears also to be an artifact of the descriptive devices offered by transformational syntax.

5. Conclusion

In this paper, I have argued that Hausa FVS is but one exponent of a systematic distinction drawn in Hausa morphosyntax, namely marking of argument realisation modes, ranging from direct local realisation, over pronominal affixation to extraction. This basic distinction, which has been shown to be highly characteristic of Hausa morphosyntax, receives a natural explanation, once we abandon the narrow perspective of an isolated rule of phrasal allomorphy in favour of a morphological perspective on the data, accounting for the tight integration of FVS with Hausa stem morphology, the diversity of exponence expressing the morphosyntactic distinction, as well as the class-specific and sporadic patterns of neutralisation, including rules of referral. This morphological perspective has also paved the way for a deeper understanding of Hausa morphosyntax, brought about by the connection we have established between the phenomenon at hand to the typologically well-attested pattern of morphologically marked extraction contexts, thereby characterising Hausa as the mirror image of French.

⁸ Embedded declaratives pattern with matrix declaratives, underlining that the sensitivity involves extraction paths, not merely a filled COMP position.

⁹ Within the context of long-distance extraction, marking of local vs. nonlocal realisation also receives a functional explanation: with transitives, choice of non-A forms (as witnessed by C-form *fa:tan* in (35) above) can provide a clue, during sentence processing, as to the location of the gap site.

¹⁰ Due to space limitations, the formal analysis had to be omitted. I therefore refer the reader to an extended version of this paper, currently under review, which is available from my homepage (<http://www.dfki.de/~crysmann/>).

Furthermore, we have investigated in some detail the syntactic environments defining the underlying inflectional categories and have found that simple surface-oriented adjacency requirements should be supplanted with reference to argument structure.

Finally, it is worth noting that a morphological analysis is not only to be preferred on empirical and typological grounds, but that it is also advantageous for methodological reasons: besides the usual Occamian arguments, which surely apply here as well, elimination of Precompiled Phrasal Phonology from the theory of grammar will ultimately provide for a more strengthened division between phrasal and lexical phonology. This goal seems actually quite attainable, given that a variety of seemingly precompiled phonologies has meanwhile been successfully reanalysed, e.g., the Mende and Kimatuumbi data (Cowper & Rice 1987), which, alongside Hausa, have formed the empirical base of Hayes's original proposal.

References

- Anderson, S. (1969), *West Scandinavian Vowel Systems and the Ordering of Phonological Rules*, Indiana University Linguistics Club, Bloomington.
- Bouma, G., R. Malouf & I.A. Sag (2001), *Satisfying constraints on extraction and adjunction*, "Natural Language and Linguistic Theory" 19 (1), 1–65.
- Chung, S. (1998), *The Design of Agreement: Evidence from Chamorro*, University of Chicago Press.
- Cowan, J.R. & R. Schuh (1976), *Spoken Hausa*, Spoken Language Services, Ithaca.
- Cowper, E.A. & K.D. Rice (1987), *Are phonosyntactic rules necessary?*, "Phonology Yearbook" 4, 185–194.
- Crysmann, B. (2003), *Constraint-based Coanalysis. Portuguese Cliticisation and Morphology-Syntax Interaction in HPSG*, no. 15 in Saarbrücken Dissertations in Computational Linguistics and Language Technology, Computational Linguistics, Saarland University and DFKI LT Lab, Saarbrücken.
- Davis, A. (1986), *Syntactic binding and relative aspect markers in Hausa*, in *Proceedings of the Fifteenth Annual Conference on African Linguistics, Los Angeles, CA, 1984*.
- Davis, A. (1993), *Empty heads and missing subjects: Underspecification in Hausa VPs*, in *Proceedings of the Chicago Linguistics Society* 28.
- Hayes, B. (1989), *The prosodic hierarchy in meter*, in P. Kiparsky & G. Youmans (eds.), *Rhythm and Meter*, Academic Press, Orlando.
- Hayes, B. (1990), *Precompiled phrasal phonology*, in S. Inkelas & D. Zec (eds.), *The Phonology-Syntax Connection*, University of Chicago Press, 85–108.
- Kayne, R. (1989), *Facets of Romance past participle agreement*, in P. Benincà (ed.), *Dialect Variation and the Theory of Grammar*, Foris, Dordrecht, 85–103.
- Kayne, R. & J.-Y. Pollock (1978), *Stylistic inversion, successive cyclicity and move NP in French*, "Linguistic Inquiry" 9, 93–133.
- Kiparsky, P. (1973), *'Elsewhere' in phonology*, in S. Anderson & P. Kiparsky (eds.), *A Festschrift for Morris Halle*, Holt, Rinehart and Winston, New York, 93–106.
- Leben, W. (1971), *The morphophonemics of tone in Hausa*, in C.-W. Kim & H. Stahlke (eds.), *Papers in African Linguistics*, Linguistic Research, Edmonton, 201–218.
- Lúis, A. & A. Spencer (to appear), *A paradigm function account of 'mesoclisis' in European Portuguese*, in G. Booij & J. van Marle (eds.), "Yearbook of Morphology" 2004, Springer.

Hausa Final Vowel Shortening: Phrasal Allomorphy or Inflectional Category?

- McCloskey, J. (1989), *Resumptive pronouns, \bar{A} -binding, and levels of representations in Irish*, in R. Hendrick (ed.), *Syntax and Semantics 23: The Syntax of Modern Celtic Languages*, Academic Press, New York, 199–248.
- Miller, P. & I. Sag (1997), *French clitic movement without clitics or movement*, “Natural Language and Linguistic Theory” 15 (3), 573–639.
- Nespor, M. & I. Vogel (1982), *Prosodic domains of external sandhi rules*, in H. van der Hulst & N. Smith (eds.), *The Structure of Phonological Representations*, vol. 1, Foris, Dordrecht, 225–255.
- Nespor, M. & I. Vogel (1986), *Prosodic Phonology*, Foris, Dordrecht.
- Newman, P. (2000), *The Hausa Language. An Encyclopedic Reference Grammar*, Yale University Press,
- New Haven, C.T., P. Newman & P. Jaggard (1989a), *Low tone raising in Hausa: A critical assessment*, “Studies in African Linguistics” 20, 227–251.
- Newman, P. & P. Jaggard (1989b), *LTR: A reply to Schuh*, “Studies in African Linguistics” 20, 263–264.
- Parsons, F.W. (1960), *The verbal system in Hausa*, “Afrika und Übersee” 44, 1–36.
- Pollard, C. & I. Sag (1987), *Information-Based Syntax and Semantics*, vol. 1, CSLI, Stanford.
- Pollard, C. & I. Sag (1994), *Head-Driven Phrase Structure Grammar*, CSLI and University of Chicago Press, Stanford.
- Pullum, G. & A. Zwicky (1988), *The syntax-phonology interface*, in F.J. Newmeyer (ed.), *Linguistics: The Cambridge Survey*, vol. 1, New York, Cambridge University Press, 255–280.
- Schmalting, C. (1991), *Modalpartikeln im Hausa: ‘Gishirin Hausa’*, Master’s thesis, Universität Hamburg.
- Schuh, R. (1989), *The reality of ‘Hausa low tone raising’: A response to Newman & Jaggard*, “Studies in African Linguistics” 20, 253–262.
- Selkirk, E. (1986), *On derived domains in sentence phonology*, “Phonology Yearbook” 3, 371–405.
- Sells, P. (1984), *Syntax and semantics of resumptive pronouns*, Ph.D. thesis, University of Massachusetts at Amherst.
- Stump, G.T. (1993), *Position classes and morphological theory*, in G. Booij & J. van Marle, (eds.), “Yearbook of Morphology” 1992, Kluwer, Dordrecht, 129–180.
- Stump, G.T. (2001), *Inflectional Morphology. A Theory of Paradigm Structure*, Cambridge University Press, Cambridge.
- Tuller, L.A. (1986), *Bijjective relations in Universal Grammar and the syntax of Hausa*, Ph.D. thesis, UCLA, Ann Arbor.
- Vigário, M. (1999), *Pronominal cliticization in European Portuguese*, “Catalan Working Papers in Linguistics” 7, 219–237.
- Zwicky, A. (1985), *Rules of allomorphy and phonology-syntax interactions*, “Journal of Linguistics” 21, 431–436.
- Zwicky, A. & G.K. Pullum (1983), *Cliticization vs. inflection: English n’t*, “Language” 59, 502–513.

Explaining some Structural and Semantic Asymmetries in Morphological Typology

Marian Klamer
Leiden University
marian.klamer@let.leidenuniv.nl

1. Introduction

This paper discusses two types of asymmetries in the typology of words. The first asymmetry concerns the morphological structure of words, the second type concerns their lexical-semantic properties. For both types of asymmetries I first present some empirical evidence, followed by a proposal on how the asymmetries can be explained.

My basic argument will be that the observed structural and semantic asymmetries are two sides of the same coin, and that they can be explained by referring to two quite general well-formedness constraints: Semantic Transparency and Structural Contrast, and one universal semantic principle on form-meaning relationships: Iconicity.

2. Evidence for the Structural Asymmetries

In this section I present some empirical evidence for the following three typological asymmetries in the morphological make-up of words: prefixing/suffixing is more common than circumfixing¹ (section 2.1); empty morphemes are always a minority in a language's morphology (section 2.2); and compounding is more common than conversion (section 2.3).

2.1 Prefixing/Suffixing Is more Common than Circumfixing

At least since Greenberg 1963, it has often been observed that pre/suffixes are more frequent than circumfixes, both within and across languages. Since they are typologically less marked than circumfixes, the following implicational hierarchy applies: a language with affixes will always have a pre/suffix, but not necessarily a circumfix. When a language has a circumfix, it will at least have one pre/suffix as well.

An example is Dutch, which has many productive and unproductive prefixes and suffixes (cf. Booij 2002), but only one (clear²) circumfix *ge__te*, which functions to derive collective nouns: *berg* 'mountain' > *ge-berg-te* 'mountain range'.

Kamera (an Austronesian language spoken on the island of Sumba in Eastern Indonesia; Klamer 1998) has one productive and many unproductive prefixes, as well as several suffixes, but only one circumfix *ka__k*. The circumfix derives verbs from ideophonic roots denoting sounds, motions and sights: *reu* 'sound of people talking' >

¹ Cf. Greenberg (1963: 92): If a language has discontinuous affixes, it always has either prefixing or suffixing or both.

² Dutch perfect participles may be formed by what looks like a circumfix: prefix *ge__t / g__d* (the voice of the final stop agrees with the voice of the final stem consonant), though various analyses of this affix are possible, see Booij (2002), section 2.4.3.

ka-reu-k ‘to talk’, *ndiku* ‘jerking motion’ > *ka-ndiku-k* ‘to jerk’, *bila* ‘light, brightness’ > *ka-bila-k* ‘to emit light, be bright’ (Klamer 1998: 245–247; 2001).

The exceptional status of circumfixes is also evident from the fact that many linguists would argue that circumfixes can (or should be) reduced to a combination of suffixing and prefixing (cf. Spencer 1991: 13), i.e. that they have a ‘derived’ status in the synchronic morphology of a language. In any case, it is remarkable that the two parts of a circumfix are often formally identical to affixes with other functions. For example, the prefixing part of the Dutch collective noun circumfix *ge__te* is formally identical to the productive nominalising prefix *ge-*, used as in *schrijf* ‘write’ > *ge-schrijf* ‘writing’, while its suffix *-te* is formally identical to the unproductive suffix *-te* that derives de-adjectival nouns (as in *leeg* ‘empty’ > *leeg-te* ‘emptiness’.) Observe also that both affixes are nominalising, just like the circumfix is. In other words, either part of *ge__te* is formally and functionally related to another affix, and their combination might be analysed as a derived structure in the synchronic morphology of Dutch.

Similar observations can be made about the Kambera circumfix, though here only the prefixing part is used elsewhere in the morphology as an unproductive prefix: *mboka* ‘be fat’ > *ka-mboka* ‘look healthy, prosperous’, *hilu* ‘language’ > *ka-hilu* ‘ear’, *beli* ‘go back’ > *ka-beli* ‘turn around; return’ (Klamer 1998: 254).

We conclude that circumfixes are less frequent and less common than pre/suffixes, and may often be analysed as derived, complex morphological units.

2.2 *Empty Morphemes are Always a Minority in a Language’s Morphology*

Morphemes such as the *cran* of *cranberry* or *ceive* in *conceive* and *perceive* are forms with no clear meaning of their own and they are not productive. Though we find such forms in probably every language, it is generally agreed upon that morphemes without meaning would never constitute the majority of a language’s morphology – they are always a minority class. Often they have special characteristics, for example because they refer to specific semantic domains (e.g. fruits), or because they are part of the non-native lexicon.

In other words, we do not expect to find a language whose morphology only, or mainly, consists of empty (cranberry) morphemes – if it has any of such morphemes, there will also be a class of productive, meaningful morphemes, and this class will be larger.

Similarly, in a language that employs reduplication, we often find empty or meaningless reduplicative elements. For example, the lexicalised relicts of reduplication processes that were productive in the past. Yet, we do not expect to find a language with only empty reduplicative elements. In other words, the existence of empty reduplicative elements implies the existence of productive, and meaningful, reduplicative elements.

2.3 *Compounding Is Typologically more Common than Conversion*

Compounding is a word-formation process that is distinct from other derivational processes, because it combines two lexemes into one new one while there is no bound morpheme involved in the process. Conversion (also referred to as zero-derivation) resembles compounding in that it is also a morphological process that does not involve any bound morphology (cf. Aronoff 1994: 15–16).

As a first step in the typological comparison of these processes, I would like to address the question which of the two is more commonly used in a language that has both of them, such as Dutch. In Dutch, the process of compounding goes in various directions: it is possible to productively derive nominal, adjectival, and numeral compounds on various types of bases (verbal compounds exist but are not productive). The base of a compound can be either a morphologically simple or a complex form (e.g. compounds can be derived from derived compounds, [*fiets-band*] [*ventiel- [dop-je]*] ‘small lid of a bicycle tyre valve’). In contrast, the direction of productive conversion is quite limited: N>V is the productive pattern (*zon* ‘sun’ > ‘to sunbathe’), while V > N (*kook* ‘cook’ > ‘boiling’), A > N (*gek* ‘mad’ > madman), A > V (*wit* ‘white’ > ‘to whiten’) are marginally productive, or have a restricted domain of application. There is no conversion of nouns or verbs into adjectives. The base for a conversion is preferably morphologically simple – it is not easy to find derivationally complex nouns that feed conversion. In other words, Dutch conversion is subject to a lot more structural restrictions than compounding is. In addition, the semantics of compounds in relation to their morphological structure is also more transparent for compounds than for converted forms.

In Kambera, compounding is a productive process, deriving both nouns and verbs (Klamer 1998: 40, 58, 115, 117) but conversion does not exist. Similarly, in standard Indonesian, compounding derives both nominal and verbal forms (Sneddon 1996: 23–25), but conversion is not mentioned as a derivational process in Indonesian reference grammars or textbooks. Note however, that in Kambera as well as in (substandard) Indonesian we often find words with no nominal or verbal affixes which are used as so-called ‘multifunctional’ items: lexemes without a clear lexical category/word class that function in both verbal and nominal contexts. For example, in Kambera *tanda* can be used as a noun ‘sign, symbol’, as well as a verb ‘to know, recognise’. Multifunctional items are distinct from words undergoing conversion, because the lexical category of their base form is unclear. In conversion, the lexical category of the base can usually be established, e.g. on semantic grounds.

In sum, while neither compounding nor conversion involves the addition of bound morphological material, we formulate the hypothesis that, if a language has both, compounding is more common than conversion. Why would this asymmetry exist?

3. Explanation of the Structural Asymmetries

In this section I present a proposal on how the three typological asymmetries discussed above might be explained. The basic idea behind my explanation is that structurally simple forms are cross-linguistically more common than complex ones. In this view, prefixes would then be structurally simpler than circumfixes, meaningful morphemes simpler than empty ones, and compounding simpler than conversion.

Why would this be the case? In what sense are circumfixes, empty morphemes and conversion structurally complex? If we envisage a linguistic system as a set of constraints on the wellformedness of utterances, we may say that those linguistic items that obey the constraints are structurally less complex than those which violate the constraints. Put in a different way, structurally complex items violate more wellformedness constraints than simple items do.

constraint.⁵ In general, then, compounding is structurally simpler than conversion, since it conforms more to the ideal of one form-one meaning matching.

We conclude that the cross-linguistically less common morphological patterns can be considered to be more complex forms because they violate the constraint on Semantic Transparency: they are structurally less ‘optimal’ than the forms that do comply to the constraint.

Now, if the typological asymmetries observed in section 2 are indeed correct, they indicate that cross-linguistically, morphologically simple structures are preferred over complex ones. Why would this be so? I suggest that the explanation might be sought outside language itself, in the language user. Most language users are both speaker and hearer. It is generally believed that economy considerations play a role in structuring linguistic communication. As speakers, we strive to be economical in speech production, so that we say as much as possible with as little effort as possible, i.e. we reduce formal contrasts (cf. (4a)). As hearers, on the other hand, we want to be economical in the processing of what we hear, so that utterances must be as distinct as possible. As hearers, then, we prefer reduced formal identity (cf. 4b). In other words, ‘economy’ concerns of hearer and speaker are the motivation of a second family of structural constraints, the constraints on structural contrasts between linguistic elements:

- (3) *Constraints on Structural Contrast between linguistic elements*
- a. “No formal contrast” (i.e., “Favour increased similarity”)
(Economy in production; speaker’s perspective)
 - b. “No formal identity” (i.e., “Favour increased dissimilarity”)
(Economy in processing; hearer’s perspective)

Constraints on structural contrasts between linguistic elements are well-known in phonology. Examples of constraints on formal contrast (4a) are constraints on certain complex segments or complex phonotactics, and examples of constraints on formal identity (4b) are the constraints on similar homorganic consonant pairs such as the OCP (Pierrehumbert 1993).

In morphology, an example of a constraint on formal contrast (> “Favour increased similarity”) would be one that penalises morphologically complex structures: an isolating language where every single linguistic unit represents a single meaning unit would then be the ideal.⁶ An example of a morphological constraint on formal identity would be a constraint on homophonous morphemes. If several distinct functions are expressed by one single form, processing becomes increasingly difficult; so our preference is to link different meanings to different morphemes. These constraints may be used to explain observed asymmetries, but they cannot be categorical: it is obvious that not all languages are isolating, and of course, homophonous morphemes do exist. I come back to this in section 7.

⁵ Here we refer to endocentric compounds and not to exocentric ones: endocentric compounds do not violate the constraint since their interpretation is a sum of their parts, while the interpretation of exocentric compounds is much less regular.

⁶ Note, however, that in such a language the structural contrast between individual *lexemes/words* is maximal; so minimal morphological complexity does not lead to minimal structural contrast in the overall make-up of a language.

4. Evidence for the Semantic Asymmetry

After discussing structural asymmetries in section 2 and 3, I now turn to an asymmetry that relates to the semantics of certain types of words. In the present section I present evidence that certain types of words show a remarkable semantic pattern. The types of words under consideration are not random: I only look at words that have a “complex” morphological make-up in the sense discussed in the previous section. That is, we look at the semantics of words which violate the structural constraints discussed above. The words under consideration are all from Austronesian languages: Kambera, Ilocano (Philippines, Rubino 2001), and Kéo (Flores, Baird 2002). We look at the semantics of words with a meaningless prefix in Kambera (4.1), words with a circumfix in Kambera (4.2), words with a meaningless reduplication in Ilocano (4.3) and lexicalised compounds in Kéo (4.4).

Since morphemes, like lexemes, are generally arbitrary signs (i.e. onomatopoeic morphemes hardly exist), we do not expect to find a direct correlation between the phonetic make-up of a morpheme and its meaning, or between its position (pre/suffix) and its meaning. For example, there is no a priori reason why a verbalizing affix should be a prefix rather than e.g. a circumfix, or why it would have the particular phonetic make-up it has (e.g. why a nominalising prefix in Dutch has the shape *ge-* rather than *li-*, *pa-* or any other string of sounds). In general, then, we say that the relation between the shape of a morpheme and its meaning is arbitrary.

I mention this very obvious generalisation here because in the cases discussed below we *do* find a direct correlation between the shape of the words and their semantics. We will see that these words have a “complex” morphological structure (in the sense of section 3) and tend to have the semantics of a particular, circumscribed, semantic domain: the domain of “expressives”. In other words: complex forms link to expressive semantics.

‘Expressive’ items belong to one of the semantic types of “Sense words”, “Names” or “Bad words”, as explained in Table 1 (for additional motivation, see the Appendix and Klamer 2002).

Table 1. The semantic types of expressive items

TYPE	EXPLANATION	EXAMPLES
Sense words	Words denoting sense impressions: sound, touch, taste, smell, feeling, emotion and sight, incl. movements of the body and/or body parts.	English: <u>tweet</u> , <u>blob</u> , <u>burp</u> , <u>bob</u>
Names	Personal or place names, nicknames, epithets, terms of endearment, names for plants and animals.	English: <u>Bob</u> , <u>baboon</u> , <u>moron</u>
Bad words	Taboo words, and lexical items with negative connotations or items that refer to undesirable states.	English: <u>boob(s)</u> , <u>tit(s)</u>

Expressive items are conceptually more complex, and more specific (less general) than common, prototypical referential lexemes. For example, *jabber* is

semantically more specialised, and conceptually more complex than *talk*: since *jabber* is a special kind of talking, *jabber* has at least one feature more than *talk*: an evaluative, subjective, and/or descriptive semantic feature. Expressive items are used less frequently than lexical items with more general meanings because they refer to very specific events or referents, (hence) they are not usually phonologically reduced, less easily accessible on-line, and never subject to grammaticalisation (cf. Hopper and Traugott 1993: 87, Slobin 2001: 432/3).

Having established what it means to say that an item has an ‘expressive’ semantics, let us now return to the semantic asymmetry that can be observed in the lexicon of a number of Austronesian languages.

4.1 The Semantics of Words with a Meaningless Prefix in Kambera

Kambera has a limited number of formally derived words with the prefix *la-*. They are listed in (5). With one exception, none of them has a root form that is still used independently. The prefix *la* has no independent meaning and does not occur elsewhere in Kambera morphology. The argument to analyse the words in (5) as morphologically complex forms is purely formal (cf. Klamer 1998 for further motivation).

(5) Kambera words with the ‘empty’ prefix *la-*

<i>la-lei</i>	‘be a husband’	<i>la-mbungur</i>	‘flower spec.’ <i>Datura factuosa</i>
<i>la-ngora</i>	‘wipe off’	<i>la-mboya</i>	‘name of medicinal plant’
<i>la-wihir</i>	‘turn one’s back, give way to X’	<i>la-wungu</i>	‘tree sp. with hard wood’
<i>la-mihi</i>	‘clean away X’	<i>la-wina</i>	‘bean sp.’ <i>Cajanus Cajan</i>
<i>la-manga</i>	‘be weak’	<i>la-nggapa</i>	1. ‘tree with thin bark’ 2. ‘very thin’
<i>la-mbiri</i>	‘look sleepy’	<i>la-ngira</i>	‘tree sp. used for canoes’
<i>la-muji</i>	‘suck’	<i>la-ngaha</i>	‘tree sp.’ <i>Barringtonia asiatica</i>
<i>la-nggori</i>	‘burp’	<i>la-yia</i>	1. ‘ginger plant’ 2. ‘brother in law’
<i>la-ngidip</i>	‘hickup, ‘gasp’	<i>la-hona</i>	‘red onion’
<i>la-ngudu</i>	‘be in a heap’	<i>la-bawa</i>	‘white onion’
<i>la-nggeha</i>	‘be thin’	<i>la-mbàku</i>	‘civet cat’
<i>la-wùjur</i>	‘with bended back’	<i>la-wora</i>	‘iguana’
<i>la-nggudu</i>	‘tie w. feet together’	<i>la-nggudu</i>	‘tuberous plant sp.’ <i>Toca palmata</i>
<i>la-mbonga</i>	‘deep large hole’	<i>la-ngàdi</i>	‘type of coral’
<i>la-mbaru</i>	‘centipede’	<i>la-ngiha</i>	‘gums’
<i>la-pàpu</i>	‘ulcer in armpit/groin’		

When we consider the semantics of these *la-*derivations, we observe that they are both verbs and nouns. The nouns are mostly plant or animal names (cf. the right column), whereas a sizable number of the verbal forms denote a position or state of the body, or movements/sounds that are related to the mouth. In other words, the nouns are Names, and quite a number of verbs are Sense words. The large majority of *la-*derivations can

thus be characterised as semantically “expressive” in the sense defined above. There is thus a remarkable semantic asymmetry to be observed in the class of words with the meaningless prefix *la-*.

4.2 *The Semantics of Words with a Circumfix in Kambera*

As mentioned in section 2.1, Kambera has one circumfix, *ka__k*, which derives verbs denoting sounds, motions and sights from ideophonic roots:

- (6) *mbùtu* ‘thud’ (sound) > *ka-mbùtu-k* ‘(fall) with a thud’
jila ‘flash’ (sight) > *ka-jila-k* ‘gleam; flash (as lightning)’

Morphologically, words derived with *ka__k* are special because they are the only Kambera forms that are derived by circumfixation. (In addition, they also have an exceptional phonotactic make-up, as well as special syntactic properties, not discussed here.) Both the root forms and the derived verbs denote sounds, motions and sights; and can thus all be classified as “Sense” words.

4.3 *The Semantics of Words with Meaningless Reduplication in Ilocano*

Ilocano, spoken in the Philippines and described by Rubino (1999, 2001) has a very elaborate morphology, including several morphemes and morphological processes that are especially related to sounds, i.e. derive onomatopoeic words.

Rubino (2001) presents an overview of the onomatopoeic morphology. His overview also contains a set of 45 lexical items of the shape $C_1V_1.C_2V.C_1V_1C_2$, for example *bu.ki.buk* ‘scatter, overturn’. Structurally, the roots in this set are made up of two identical CVC sequences separated by a vowel, resulting in a tri-syllabic lexical item, e.g. *bug-a-bug* ‘to be mixed (varieties of rice)’, *bas-i-bas* ‘hurl a long object’. Rubino analyses the derived items as “roots” (2001: 317), which I take to imply that there is no meaningful root unit *bug/bugi* or *bas/basi* etc. in Ilocano morphology. In other words, formally these items are reduplications, but the base of the reduplication is non-existent. Rubino further remarks that most of the words in this set are “no longer” onomatopoeic.

(7) *The semantics of words with a meaningless reduplication in Ilocano*

reduplication	non-existent base	meaning
<i>yaba-yap</i>	* <i>yab(a)</i>	‘flap (flags), flutter’
<i>ngasa-ngas</i>	* <i>ngas(a)</i>	‘wear out (shoes); suffer injury’
<i>pali-pal</i>	* <i>pal(i)</i>	‘black magic’
<i>wisa-wis</i>	* <i>wis(a)</i>	‘fishing tackle’
<i>guyu-guy</i>	* <i>guy(u)</i>	‘suggest; convince’
<i>bala-bal</i>	* <i>bal(a)</i>	‘scarf, muffler; wrap snugly’
<i>rangi-rang</i>	* <i>rang(i)</i>	‘dry, parched land’
<i>wida-wid</i>	* <i>wid(a)</i>	‘swing the arms when walking’
<i>nuru-nur</i>	* <i>nur(u)</i>	‘erode from water contact’

The translations suggest that most of them are not related to sounds, i.e. they are not onomatopoeic. In the list of (7), I would classify *yabayap* and *widawid* as “Sense” words, while *ngasangas* and *palipal* are “Bad” words, words with negative connotations. In a similar way, of the 45 examples given in the paper we can classify 16 as a “Sense” or a “Bad” word, i.e. about one third of the items are semantically expressive. This is a remarkable semantic asymmetry, considering the wide semantic range of the words (from ‘fishing tackle’ to ‘black magic’ to ‘suggest’...!). If the potential semantic range of the given forms is so wide, why would one third of them cluster in the particular, rather circumscribed, semantic domain of expressives?

4.4 *The Semantics of Compounds in Kéo*

A similar semantic asymmetry is found in a particular set of morphologically complex words in Kéo, a language spoken in Central Flores in Eastern Indonesia, and described by Baird (2002).

Kéo is an isolating language – it has no inflectional morphology and no productive morphological derivation. The only sub-lexical element in the language is the numeral clitic *ha-* ‘one’. Kéo has some lexicalised reduplicated forms and a limited number of compounds. Many of the Kéo compounds are semantically opaque, and Baird (2002: 182) suggests that they are lexicalised inheritances from ritual, parallel speech. The compounds attested by Baird are all listed in the grammar. There are 47 compounds listed, of which 20 items semantically belong to the class of Sense, Name or Bad words. For example (cf. Baird 2002: 171–182) (a question mark as gloss indicates lack of independent meaning of a word):

(8) *Kéo lexicalised compounds*

<i>da'é-dondo</i>	‘space’	<i>méké-suné</i>	‘flu’
place-place		cough-sniffle	
<i>mutu-tiwo</i>	‘gathering’	<i>pémba-jawa</i>	‘sit cross-legged’
gathering-meeting		hold on lap-corn	
<i>dera-kiri</i>	‘afternoon’	<i>mundé-mi</i>	‘place name Mundemi’
sun-slant		citrus.fruit-sweet	
<i>topo-dhupa</i>	‘machete sheath’	<i>méso-mélo</i>	‘sit restlessly’
machete-?		move-?	

Of the items given in (8), those in the left column are not expressive, while I would interpret those in the column on the right as expressive (from top to bottom: Bad, Sense, Name, Sense). Classifying the 47 items in Baird 2002 in a similar way, 20 turned out to be expressive. Considering the semantic range covered by the compounds (‘machete sheath’ to ‘afternoon’ to ‘flu’), it is again a remarkable asymmetry that 42,5 % of the items cluster in the semantic domain of expressives.

In sum, in this section I have presented four case studies from three Austronesian languages which illustrate that certain morphologically complex forms show a tendency to associate with expressive semantics. In the next section I suggest an explanation for this semantic asymmetry.

5. Explanation of the Semantic Asymmetry

The semantic asymmetry observed in the classes of words above can be explained when we consider the link between the meaning and the structural complexity of the items in question. Recall that expressive items (Sense, Name and Bad words) are assumed to be semantically or conceptually more complex than common, referential items: an expressive word has one or more evaluative, subjective, and/or descriptive semantic feature(s), and is more specific than a referential lexeme.

Recall also that the circumfix *ka__k* in Kambara is structurally complex because it violates the Semantic Transparency constraint, as does the empty prefix *la-* in this language. When derivations with *ka__k* and *la-* are semantically expressive, we observe a match between the structural complexity of these items and their semantic complexity, a matching that might be called an ‘iconic’ matching of form and function.

Turning now to the Ilocano words with an empty reduplicative element, we observed that these items are structurally complex for two reasons. Firstly, since they contain a reduplicative syllable they are longer than common roots: Ilocano roots usually have only two syllables – not three. Secondly, they contain a meaningless (reduplicative) morpheme, and therefore violate Semantic Transparency. We observed the asymmetry that one third of the items are semantically expressive. This asymmetry can be explained when we assume that these cases too show a preference for an iconic matching of complex form and complex semantics.

Finally, how to account for the skewed semantics of the Kéo compounds?

In the isolating language Kéo, there is obviously a very strong preference for morphologically simple forms, since – apart from the set of compounds discussed here and some lexicalised reduplications – morphologically complex forms do not occur in this language. In other words, the constraint on formal contrast in (3a), which disallows morphologically complex structures, is generally obeyed in Kéo. The compounds are exceptional because they are morphologically complex forms, and they show a tendency to match their structural complexity with the complex semantics of expressiveness.

In sum, the fact that a significant part of the complex forms discussed here are semantically expressive is not a coincidence, but something that can be explained: in many cases, the general principle of Iconicity seems to be applied, and a complex form is matched with a complex semantics.

Put differently, expressive words constitute a subclass in the lexicon which shows a non-arbitrary connection between form and meaning, and Iconicity is the principle steering the lexical semantic asymmetries observed. Note that this is a tendency observed for certain types of words; it does not apply categorically in all languages for all words: there are many morphologically complex items that are not expressive, and there are also many simple words with a complex semantics.

6. Conclusions and Discussion

In this paper I presented a number of structural and semantic asymmetries at the word level; some of them obvious, others perhaps more controversial.

I argued that pre/suffixes are crosslinguistically more common than infixes, that meaningless affixes are always a minority in a language, and that compounding is typologically less marked than conversion. I then explained the skewed distribution of

these morphological patterns as a cross-linguistic preference for simple morphological structures over complex ones. This preference can be expressed by formulating structural constraints on the wellformedness of linguistic forms.

I suggested that the relevant constraints are (a) constraints concerned with the one-to-one linking of form and meaning (maintaining Semantic Transparency), and (b) constraints on structural identity and structural contrast.

I then demonstrated a striking asymmetry in the semantics of certain classes of morphologically complex words in the Austronesian languages Kambera, Ilocano and Kéo. The four types of morphologically derived words all showed a strong preference for an expressive semantics. If it is correct to assume that expressive items are semantically more complex than common referential items, we can explain this semantic skewing as the outcome of the application of the universal principle of Iconicity: link a complex form to a complex meaning.

As a consequence, we understand why certain types of morphological processes are typologically less common than others, and why cross-linguistically, expressives appear to have a preference for complex structures.

The explanations proposed here are not new: economic and iconic motivations for certain linguistic forms or patterns have been proposed in various linguistic research traditions (both typological and generative), as well as for various sub-disciplines of linguistics, including morphology. For morphologists, the ideas presented here may sound similar to those presented as the theory of Natural Morphology (Mayerthaler 1981, Dressler 1985, 1987, and references cited there). Natural Morphology is a theory of what constitutes a natural, or unmarked morphological system, and how we can predict and explain deviations from that system. In this theory, the most 'natural' type of morphology is fully transparent: every morpheme has one form and one meaning, and every meaning corresponds to only one form (the 'bi-uniqueness' principle, e.g. Dressler 1987: 111 v.v.). 'Bi-uniqueness' is an explication of the intuition that has always been implicit in the classical notion of the morpheme as the minimal form-meaning unit. Natural morphology regards deviations from the most natural, transparent state as unnatural or marked, and the assumption is that cross-linguistic patterns, historical change, language acquisition, speech errors and language disorders show a statistical tendency to prefer the natural, unmarked state to the unnatural, marked one.

The present paper agrees with natural morphologists such as Dressler (1987) in that typological asymmetries in morphology can and should be explained with very simple, general constraints on the linking of form and function.

The constraints should, however, not be used to characterise possible (and impossible) morphological systems, but rather to calculate which systems are more probable than others (cf. Croft 2003: 283).

It is clear that the constraints discussed here cannot be categorical. For example, the Semantic Transparency constraint would exclude the existence of polysemous and synonymous affixes, as well as allomorphy, yet we find such morphemes in many languages. The Dutch diminutive is but one illustration of allomorphy, where one meaning is expressed through five forms: *koning-kje* 'king-DIM', *riem-pje* 'belt-DIM', *huis-je* 'house-DIM', *oven-tje* 'oven-DIM', *tong-etje* 'tongue-DIM'. An example of synonymy in English affixes (where various forms have one common meaning) is the fact that *-ship*, *-dom*, *-hood* all indicate a 'state or quality of being', compare *friendship*, *serfdom*, *motherhood*. And an illustration of polysemy in an English affix (where one

form has various meanings) is the suffix *-ist*. Canonically this suffix means ‘one who does X’, as in *rapist* ‘one who rapes’, but it also appears in words like *racist*, and *sexist* where it means something like ‘one who is prejudiced against a group’. The Semantic Transparency constraint does not exclude such phenomena, but is a way to express that such items are structurally more complex than morphemes with a one-to-one mapping of meaning and form.

Since there are various motivations for linguistic structure, both functional and formal, and since these motivations relate to distinct linguistic modules (phonetics, phonology, morphology, syntax, semantics), and/or to language external factors (sociolinguistic, psycholinguistic), etc., the motivations for certain linguistic structures compete with each other in many ways. It is impossible to predict the outcome of this competition for a language; indeed, it is usually assumed to be arbitrary (cf. Croft 1995: 504–509). In other words, the synchronic grammar of a particular language always involves a lot of arbitrariness, and not everything in language is explainable in terms of a completely deterministic set of formal or functional principles.⁷

Since morphological systems are the outcome of many different, competing synchronic and diachronic forces, historical developments may lead to a complex, ‘unnatural’ or ‘marked’ situation. For example, the occurrence of clitics within other morphemes is crosslinguistically very unusual, but such ‘endo-clitics’ are attested in Udi (Harris 2002) as the outcome of a unique combination of particular historical changes and certain morpho-syntactic features in this language. In other words, it cannot be maintained that diachronic change, language acquisition, speech errors etc. always strive towards more natural, less marked, simpler structures.

Yet, strikingly asymmetrical crosslinguistic patterns do exist, and call for explanation. In this paper I presented evidence for some of such asymmetrical patterns in morphological structure as well as in semantics. I argued that their skewed patterning might be explained with some very general constraints and principles. These constraints may also be used to calculate which morphological types are more probable than others.

References

- Aronoff, M. (1994), *Morphology by itself*, Cambridge (Mass.), MIT Press.
- Baird, L. (2002), *A grammar of Keo*, Ph.D. dissertation, Canberra, Australian National University.
- Booij, G. (2002), *The Morphology of Dutch*, Oxford University Press.
- Clynes, A. (1995), *Topics in the phonology and morphosyntax of Balinese*, Ph.D. dissertation, Canberra, Australian National University.
- Clynes, A. (1998), *The emergence of the marked: interactions of phonology and semantics in Balinese*, University of Brunei Darussalam, ms.
- Croft, W. (1995), *Autonomy and functionalist linguistics*, “Language” 71, 490–532.
- Croft, W. (2003), *Typology and universals*, [2nd edition], Cambridge, Cambridge University Press.
- Dressler, W.U. (1985), *Morphology*, Ann Arbor, Karoma Press.
- Dressler, W.U. (1987), *Word formation as part of natural morphology*, in W.U. Dressler, W. Mayerthaler, O. Panagl & W.U. Wurzel (eds.), *Leitmotifs in Natural Morphology*, Amsterdam, John Benjamins, 99–126.

⁷ If it were, all languages would be alike, internally invariant and no languages would change (Croft 2003: 282).

- Greenberg, J.H. (1963), *Some universals of grammar, with particular reference to the order of meaningful elements*, in J.H. Greenberg (ed.), *Universals of language*, Cambridge (Mass.), MIT Press, 73–113.
- Humboldt, W. von (1836), *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts*, Berlin, Dümmler.
- Harris, A.C. (2002), *Endoclititics and the origins of Udi morphosyntax*, Oxford, Oxford University Press.
- Hopper, P. & E. Traugott (1993), *Grammaticalization*, Cambridge, Cambridge University Press.
- Klamer, M. (1998), *A grammar of Kambara*, Berlin/New York, Mouton de Gruyter.
- Klamer, M. (2001), *Expressives and iconicity in the lexicon*, in F.K. Erhard Voeltz & C. Kilian-Hatz (eds.), *Ideophones*, Amsterdam, John Benjamins, 165–181.
- Klamer, M. (2002), *Semantically motivated lexical patterns: a study of Dutch and Kambara expressives*, “*Language*” 78.2, 258–286.
- Klinken, C. van (1999), *A grammar of the Fehan dialect of Tetun*, Canberra, Pacific Ling. 155.
- Mayerthaler, W. (1981), *Morphologische Natürlichkeit*, Wiesbaden, Athenäum.
- Pierrehumbert, J. (1993), *Dissimilarity in Arabic verbal roots*, in *Proceedings of NELS 23*, 367–381.
- Rubino, C. (1999), *Ilocano Grammar and Dictionary*, Honolulu, University of Hawaii Press.
- Rubino, C. (2001), *Iconic morphology and word formation in Ilocano*, in F.K. Erhard Voeltz & C. Kilian-Hatz (eds.), *Ideophones*, Amsterdam, Benjamins.
- Seuren, P., & H. Wekker (1986), *Semantic transparency as a factor in creole genesis*, in P. Muysken & N. Smith (eds.), *Substrata versus universals in creole genesis*, Amsterdam, John Benjamins, 57–70.
- Slobin, D.I. (2001), *Form-function relations: how do children find out what they are?*, in M. Bowerman & S.C. Levinson (eds.), *Language acquisition and conceptual development*, Cambridge, Cambridge University Press, 406–449.
- Sneddon, J.N. (1996), *Indonesian reference grammar*, St Leonards NSW, Allan & Unwin.
- Spencer, A. (1991), *Morphological theory*, Oxford, Blackwell.
- Williams, E. (1997), *Blocking and anaphora*, “*Linguistic Inquiry*” 28, 677–628.

Appendix 1. Motivation for Sense, Name and Bad words as Expressive or semantically complex

(See Klamer 2002 and the references cited there.)

Items from the Sense category in Table 1 are generally well-established expressives (see Hinton et.al. 1994). Items from the categories Name and Bad may be more controversial as ‘expressives’, but it should be noted that the distinction between sound symbolic forms on the one hand, and names and taboo words on the other, is not sharp. For example, names often derive from vocabulary used to refer to sounds, motions, and shapes, reflecting visible or audible characteristics of the named person, plant or animal (e.g. body shape, hair colour, bird’s call, animal movement). For example, in Mundang (Niger-Congo), animal and plant names are part of the same type of expressive vocabulary as ideophones (Elders 1999), in Estonian and Finnish, bird names are expressive forms to some extent (Antilla 1976), and in Greek, nicknames pattern with the other expressive forms (Joseph 1997). Bartens (2000: 166–169) explicitly discusses ‘de-ideophonic’ animal names in a number of Atlantic Creoles. This suggests that there is no categorical distinction between Sense items and Names in a language. With respect to the semantic type Bad (taboo words and words with negative connotations); there is cross linguistic evidence that words from the Bad type may pattern structurally and semantically with the Sense items (for Japanese: Kita 1997:98, Hamano 1998; for Balinese: Clynes 1995, 1998, and for Greek: Joseph 1994, 1997). In addition, there are cases where the distinction between the types Bad and Name is fluid (cf. English baboon as animal name and epithet in English), so if Name is a semantically complex type, then Bad is too.

Appendix 2. Additional evidence for the iconic matching of form and function in expressives

The evidence presented above concerned morphologically complex items that were semantically complex. Klamer (2002) contains quantitative evidence from other linguistic domains: there are certain classes of Kambara and Dutch words with a complex phonotactics or with complex segments that show a statistical tendency to match that formal complexity to expressive meanings. Below follows some additional phonological evidence from Austronesian languages that suggests a similar iconic patterning in the lexicon of these languages. This particular evidence shows that phonotactically/prosodically complex base forms in Tetun, Ilocano, and Balinese tend to be semantically expressive.

Fehan Tetun

Root forms in Fehan Tetun (Central Timor, Van Klinken 1999) are generally di- (55%) or tri-syllabic (43 %). Only 2% of the roots have 4 syllables.

Trisyllabic words are prosodically complex: they consist of one disyllabic foot and an extrametrical syllable. In general, we can say that a Tetun root is a disyllabic foot:

(a) Root = PrWd = F = $\sigma\sigma$

Four-syllable roots violate this constraint. 9 illustrations given by Van Klinken (1999:16):

(b)	<i>akitou</i>	‘dove’	<i>bibiliku</i>	‘drum’ (noun)
	<i>banokae</i>	‘kind of sea shell’	<i>labadain</i>	‘spider’
	<i>kaibók</i>	‘leaf vegetable’	<i>tualekik</i>	‘wake songs’
	<i>sibalebok</i>	‘parsley’	<i>liurai</i>	‘executive noble’
	<i>maufinu</i>	‘danger’		

Note that 7 of these forms are semantically expressive (Name, Bad). If this list is representative for the class of four-syllable lexemes in Tetun, it suggests that semantic expressiveness is matched with a complex form: a form that violates the constraint against prosodically complex roots.

Ilocano

Ilocano (Phillipines, Rubino 1999, 2001) roots are usually disyllabic CV(C).CV(C):

(c) Root = F = $\sigma\sigma$

There are less than 5 monosyllabic roots, e.g. *wak* ‘crow’ and *waw* ‘thirst’ and three- or four-syllable roots (all monomorphemic) are generally expressive: most of them represent sounds (repetitive or rustling), as in:

(d)	<i>sa.ʔid.dek</i>	‘hiccup’	<i>sa.ʔib.bek, sa.ʔin.nek</i>	‘sob’
	<i>ta.rat.tat</i>	‘sound of typing’	<i>ka.ra.sa.kas</i>	‘rustling sound of leaves’
	<i>dis.su.or</i>	‘waves breaking’	<i>ka.ra.si.kis</i>	‘rustling sound of bamboo’
	<i>sa.ra.i.si</i>	‘waterfall’	<i>ʔu.bu.ʔub</i>	‘fumigate’
	<i>dil.la.wit</i>	‘instant, brief period of time’		
	<i>sa.rung.kar</i>	‘visit’		

Balinese

In Balinese, semantic and formal markedness are also aligned, as argued by Clynes (1995, 1998). Balinese expressives violate at least one, but usually more of the six constraints listed below. Balinese nicknames are an especially clear instance of expressives in this language: they are meaningless but inelegant words that have ‘bad’ connotations. All of them violate at least one structural constraint that applies elsewhere in the language. Illustrations:

- (e) Constraints violated by Balinese ‘bad names’ (Clynes 1995: 51–52, 1998: 21–22)

Onset:

“Every syllable must have an onset”: violated by the bad names: *Cluit, Joet*.

* **Complex**^{ONS} :

“No complex onsets”: violated by the bad names *Klemug, Namprut, Gomblos, Cluit*.

* σ [h/ : “No /h/ as onset”:

violated by the bad name *Cibuhut*.

Root ^{$\sigma\sigma$} :

“Roots must be bisyllabic”: violated by the bad names *Cidaku, Cibuhut, Maseni*

Vowel harmony:

“Cooccurring [+ATR] vowels agree in height”: Violated by the bad names *Kedi, Keni, Maseni, Toti*.

Consonant disharmony:

“Two homorganic consonants do not cooccur in a root”: violated by the bad names *Cidaku, Namprut, Toti, Latep, Petet*.

Prefix-Suffix Neutrality In Evaluative Morphology

Nicola Grandi
University of Milan – Bicocca
nicola.grandi@unimib.it

Fabio Montermini
ERSS – CNRS & Université de Toulouse–Le Mirail
fabio.montermini@univ-tlse2.fr

1. Introduction: The “Suffixing Preference”¹

A well-known and often discussed issue in the studies on morphology, in particular on morphological typology, is the somehow puzzling asymmetry languages display in the use of the different morphological strategies they have at their disposal. We refer to the fact that suffixes are largely preferred to prefixes, and that these are preferred to other types of affixes (infixes, circumfixes, etc.). This has become a leitmotiv in the studies on morphological typology at least since Greenberg’s (1963) fundamental and seminal work on language universals. This asymmetry in the distribution of prefixes and suffixes among World’s languages, usually called ‘suffixing preference’, has been related to two typologically relevant parameters (use of prepositions vs. use of postpositions and VO vs. OV basic word orders).

The following table sums up the results of this correlation:

(1)

	Prefixes	Suffixes
VO / Pr	X	X
OV / Po	∅	X

(Hawkins/Gilligan 1988: 219)

As expected, among the World’s languages, all those which use prefixes are prepositional and VO (i.e., head-initial languages); and as expected as well, all postpositional and OV languages (that is, ‘head-final’ languages) use suffixes. Nevertheless, and unexpectedly, many prepositional and VO languages use suffixes too. This fact shows that many languages prefer using suffixes rather than prefixes *in any case* to express their grammatical relationships.

The first row of the table can be interpreted as follows:

- (a) intra-linguistic interpretation: a VO and prepositional language (or, in other words, a head-initial language) can have both inflexional prefixes and inflexional suffixes. For example, in Berber verbs, third person singular is

¹ We are grateful to Andrew McMichael for having reviewed the English text, to our informants for having filled up our questionnaire, and to the colleagues that participated in the ‘animated’ discussion which followed the presentation of this paper at the 4th Mediterranean Meeting on Morphology. Their questions, comments, criticisms and suggestions have been a great spur to correct and improve (we hope) some crucial points of our study.

marked by a prefix (*y/i-* for masculine and *tɔ-* for feminine), but third person plural is marked by a suffix (*-ɔn* for masculine and *-ɔn(t)* for feminine)

- (b) cross-linguistic interpretation: in VO and prepositional languages (that is, in head-initial languages), an inflectional category can be cross-linguistically expressed both by prefixes and suffixes. For example, number is marked by a suffix in Italian (*alber-o* ‘tree’ vs. *alber-i* ‘trees’) and by prefixes in Swahili (*m-tu* ‘man’ vs. *wa-tu* ‘people’).
- (c) a third possibility, though conceivable, is not attested within the World’s languages: in a single language, an inflectional category is always expressed either by prefixes or by suffixes, but never by both.

It should be specified that table (1) has been drawn up on the basis of a wide cross-linguistic comparison of inflectional categories. In fact, it is undeniable that inflectional categories are cross-linguistically more constant than derivational ones and so it is easier to compare inflection than derivation. Nonetheless, if derivation had been taken into account too, the table would probably not have an empty slot.

(2)

	Prefixes	Suffixes
VO / Pr	X	X
OV / Po	(X)	X

In fact, derivational prefixes are attested even in some OV languages, although they are rarer than in VO languages (as indicated by the round brackets).

As for inflection, also for derivation there are two possible interpretations of the two rows of the table:

- (a) intra-linguistic interpretation: a language (independently of the position of the head) can have both derivational prefixes and derivational suffixes (for example, in Romance languages relational adjectives are formed by adding a suffix to the base-word, but negative adjectives are formed by adding a prefix to the base-word);
- (b) cross-linguistic interpretation: a derivational category can be cross-linguistically expressed both by prefixes and suffixes. For example, agent nouns are formed by suffixes in English (i.e. *sing* > *sing-er*) and by prefixes in Malay (*nyanyj* ‘sing’ > *pe-nyanyj* ‘singer’).
- (c) even in this case, the other conceivable interpretation seems to be excluded or, at least, seems to show a very low degree of occurrence: in a single language a derivational category tends not to be expressed both by prefixes and suffixes.

Different explanations have been proposed to take into account this asymmetry, including psycholinguistic factors, such as the greater relevance of the beginning of a word for processing than of its end (cf. Cutler *et al.* 1995, Hawkins / Gilligan 1988); diachronic tendencies for grammaticalization of free elements (cf. Hall 1988), or

observations concerning the fact that prefixes are usually considered as non prototypical affixes (in comparison to suffixes), since they are learnt later in language acquisition, lost earlier in aphasia, etc. (cf. Mel'čuk 2000).

In this paper, we would like to investigate the real value of such an asymmetry, and the link it may have with the expression of different semantic and functional categories. It is uncontroversially assumed that prefixes and suffixes can equally serve to express both inflectional and derivational categories, and that the same semantic value can be expressed, in different languages, by either strategy (see, for example, Italian and Swahili plural forms or English and Malay agent nouns presented above). Nevertheless, some systematic relationships between semantic functions and the position occupied by the morphemes used to express them have been observed. For instance, case, mood, and valence are almost always expressed by suffixes², while there seems to exist a strong preference for negative morphemes to be prefixes (cf., for example, negative verbal prefixation in Indo-European languages). A major problem with all these generalizations is the fact that – as we stated above – they are almost all based on inflection, with no reference to derivation. Even regardless of the difference between inflection and derivation, as we have already pointed out, the crucial point is that in a single language one semantic value can be expressed either by prefixes or by suffixes never by both. It is usually assumed that this generalization is undoubtedly true for inflection, but it seems to hold also for derivation. There is, however, a remarkable exception that has rarely been taken into account by scholars: evaluative affixes, which, in many genetically and typologically unrelated languages, seem to disregard the suffixing preference, favouring – as we will see in the next paragraph – a sort of ‘prefix-suffix neutrality’.

2. Evaluative Derivation: A Case of Prefix-Suffix Neutrality?

If we go back to Tables 1 and 2, we can wonder which of them describes the situation for evaluative morphology³, since – as many scholars have pointed out – evaluative morphology usually lies on the borderline between inflexion and derivation.

A first cross-linguistic survey of the data suggests that evaluative affixes behave just like derivational affixes: both evaluative prefixes and suffixes are attested both in OV and VO languages. But if evaluative morphology is concerned, all three possible interpretations of the tables are widely attested. In fact,

- (a) a single language can have both evaluative prefixes and suffixes (e.g. Italian *gatto* ‘cat’ > *gattino* ‘kitten’ and *moto* ‘motor-cycle’ > *maximoto* ‘big motor-cycle’);
- (b) the same evaluative function can be cross-linguistically expressed by prefixes and suffixes (e.g. Italian diminutive suffix *-ino* in *ragazzino* ‘little boy’ and Shona diminutive prefix *ka-* in *kakomana* ‘little boy’);

² Hawkins & Gilligan (1988: 234) present some data on prefixal vs. suffixal marking of 11 different semantic and functional classes in some 220 different languages.

³ We recall that what we call ‘evaluative morphology’ is the morphological expression of semantic and functional relationships along the two axes SMALL ↔ BIG and GOOD ↔ BAD (see Grandi 2002 for details).

- (c) a single evaluative function can be formally expressed both by prefixes and suffixes even in the same language (e.g. Italian *appartamento* ‘flat’ > *appartamentino* / *miniappartamento* ‘small flat’).

So, if we pass from inflection (and derivation – although to a lower degree) to evaluative morphology, the so-called ‘suffixing preference’ seems to become a sort of ‘prefix-suffix neutrality’. What is really interesting from a typological point of view is that there seems to be some semantic functions that are insensitive to the well-know and cross-linguistically widespread prefix-suffix asymmetry. This typologically unusual situation is best exemplified by Indo-European languages:

- (3) i. Indo-European languages

Romance languages:

Diminutive prefixes:	It., Sp., Port., Fr. <i>mini-</i> , <i>micro-</i> etc.
Diminutive suffixes:	It. and Sp. <i>-ino</i> , Port. <i>-inho</i> , It. <i>-etto</i> , Fr. <i>-et(te)</i> etc.
Augmentative prefixes:	It., Sp., Port., Fr. <i>maxi-</i> , <i>macro-</i> , <i>mega(lo)-</i>
Augmentative suffixes:	Sp. <i>-ón</i> , It. <i>-one</i> , Port. <i>-ao</i> etc.

Modern Greek:

Diminutive prefix:	μικρο-
Diminutive suffixes:	-άκι, -ούλι, -ίτσα etc.
Augmentative prefix:	μακρο-
Augmentative suffixes:	-άς, -άκλα, -άρα etc.

The data in (3) show that the same semantic instructions ‘small X’ and ‘big X’ can be expressed cross-linguistically either by suffixes, by prefixes (or circumfixes and infixes) or even by both types of affixes within the same language. Needless to say that this is an anti-economic and, consequently, typologically unusual situation.

At this stage we will restrict our observations to the Indo-European languages of Europe (henceforth simply European languages). The four evaluative semantic values (SMALL, BIG, GOOD, BAD) are spread similarly within all these languages (see the languages from Italian to Greek in Table B in the Appendix). In particular, we may notice some regularities: the meaning SMALL and the meaning BIG are (almost) always expressed both by prefixes and by suffixes in European languages, the meaning GOOD is (almost) always expressed by prefixes and the meaning BAD is always expressed (if it is expressed morphologically), by suffixes. The last issue, in particular, would certainly deserve to be investigated in more detail, and could provide matter for further work. The maps at the end of the paper summarize the situation.

It should be noticed, however, that the emergence and spread of evaluative prefixes in all European languages (not only Indo-European) is primarily due to the emergence of a pan-European cultural lexicon, by which almost all European languages borrowed a large number of Latin and Greek morphemes (such as *super-* or *hyper-*). This fact may actually cause some distortions in the observed data (for instance, French and English, two languages which possess very few evaluative suffixes have a wide set of evaluative prefixes). However, the fact that non-learned evaluative prefixes exist in

the majority of these languages (as they existed in Latin and Ancient Greek) suggests, in our opinion, that European languages do have a prefixal evaluation⁴.

For European languages, a first survey of the data suggests that what we called ‘prefix-suffix neutrality’ involves morphemes expressing the meaning SMALL, and partially involves morphemes expressing the meaning BIG (in particular for Romance languages, Slavic languages and Greek). It could be said, then, that it concerns the ‘quantitative’ side of evaluation (SMALL vs. BIG), but not the ‘qualitative side (GOOD vs. BAD).

Interestingly, this phenomenon is also attested – though to a smaller extent – in languages belonging to other families and spoken in other geo-linguistic areas:

(4) i. Ugro-Finnic languages

Finnish:

Diminutive prefix: *pikku-*

Diminutive suffix: *-nen*

ii. Afroasiatic languages

Berber:

Diminutive suffix: *-ush*

Diminutive circumfix: *t__t*

ii. Niger-Congo languages

Bantu languages:

Diminutive prefixes: usually class 12/13 (but also 2, 7, 8, 11, 14, 19, 20)

Diminutive suffix: *-ana*

Augmentative prefixes: class 3, 4, 5, 10, 21, 22 etc.

Augmentative suffix: *-hadi / -kati* etc.

Once again, only the “quantitative” side of evaluation is involved.

In the following sections of this paper, we aim to focus on this “prefix-suffix neutrality” in a typological perspective in order to understand how far this unusual situation is spread among languages other than Indo-European ones and, secondly, to understand if this phenomenon correlates with other typological features (or, in other words, if there are typological correlations that can favour or disfavour it). In this case, three parameters will be taken into account: the morphological type, the presence of prepositions vs. postpositions and the VO vs. OV word order.

3. Prefix-Suffix Neutrality in a Typological Perspective

The sample on which we tested the occurrence of “prefix-suffix neutrality” and the possible correlations with the previously mentioned typological parameters includes 55 languages, belonging to different families (Indo-European, Uralic, Altaic, Afro-Asiatic,

⁴ For a discussion on the status of prefixal vs. suffixal evaluation in Italian, cf. Grandi / Montermini (forthcoming).

Sino-Tibetan, Kam-Tai, Austric, Austronesian, Oceanic, Niger-Congo, Caucasian, Eskimo-Aleut, Amerind and Chukotko-Kamchatkan; also a few isolated languages were investigated). The data presented in this paper were collected from a questionnaire submitted to native speakers of the languages⁵. The whole list of languages is in the first column of tables A and B at the end of the paper. Before going into detail, we have to point out that the sample is unbalanced in favour of head-initial languages; as a consequence, it cannot be said to be really representative of the World's languages.⁶ It is important, then, to lay stress on the fact that our results are still partial and incomplete (this paper, in fact, represents just the first step of a wider project whose aim is to draw a 'typological map' of derivational morphology in languages). Consequently, our conclusions must not be read as proven, but as possible clues to general typological tendencies.

The languages of the sample have been analyzed in order to single out:

- (a) the morphological type;
- (b) the order of verb and direct object (VO vs. OV);
- (c) the presence of prepositions or postpositions;
- (d) a list of all the morphological strategies with evaluative meaning;
- (e) the formal (not only morphological) strategies used to express each evaluative meaning.

Tables A and B in the Appendix summarize the data we took into account. The information corresponding to points (a-d) above are located in the third, fourth, fifth and sixth columns of table A. The information corresponding to point (e) is displayed in table B.

When reading and analysing Table A, it should be kept in mind that it is often difficult – sometimes even impossible – to indicate some typological tendencies in a clear, univocal and, above all, concise way. This holds especially for morphological types or word order typology. So, the values in the third and fourth columns of Table A are to be considered as the expression of statistically relevant tendencies. If two or more values are present in the same slot, the one in capital letters corresponds to the prevailing tendency. Furthermore, question marks indicate typologically entangled situations or the presence of morphological items which cannot be classified in a clear and precise way.

The parameters to be related to the presence / absence of evaluative affixes have been chosen taking into account the results of the works by Greenberg, Hawkins / Gilligan and others on the 'suffixing preference'. First of all, we referred to the 18 implicational universals listed in Hawkins / Gilligan (1988), who relate the choice of prefixes or suffixes to express categories such as case, mood, etc. to the internal structure of the adpositional phrase (namely to the presence of prepositions or

⁵ All the informants have some 'metalinguistic competence'.

⁶ Languages of the sample can be grouped as follows: 38 VO / Pr languages; 12 OV / Po languages and 5 languages with other typological configurations.

postpositions) and of the verbal phrase (i.e. to the relative order of verb and direct object).

Among the languages of our sample, the distribution of evaluative affixes can be represented as follows:

(5)

	Evaluative morphology								
	Pref	Inf	Suf	Pref / Inf / Suf	Pref / Suf	Pref / / Inf	Inf / Suf	No eval. morph.	Other stra- tegies
VO & Pr (38)	4	6	5		18		2	3	2
VO & Pr/Po (2)					2				
VO & Po (1)			1						
OV & Po (12)	2		8		1				1
OV & Pr (2)			1					1	

N.B.: the total of the first row of the table is 40 instead of the expected 38 because Moroccan and Tunisian Arabic have been counted twice. In fact, these languages uses both infixes and suffixes, but the meaning of these strategies do not overlap: infixes have a diminutive meaning, suffixes have an augmentative meaning. So, Moroccan and Tunisian Arabic cannot be placed in the ‘Inf / Suf’ slot. Berber, which makes use of the diminutive circumfix *t__t* and of some diminutive suffixes has been inserted – maybe forcibly – into the ‘Pref / Suf’ slot.

The slots containing the values we consider as being the most relevant are marked in grey. It is useful to express these figures as percentages, in order to show the results of this survey in the clearest way. 52% of VO/Pr languages exhibit some kind of affixal neutrality (mostly between prefixes and suffixes, but also between infixes and suffixes). The remaining VO / Pr languages are equally divided into prefixal languages, infixal languages, suffixal languages, languages with no evaluative morphology and languages that use other strategies to form diminutives and augmentatives. So, the incidence of affixal neutrality among VO / Pr (or, in other words, among head-initial languages) seems to be high.

But if we turn to OV / Po languages, the situation is radically different: in our sample, the only language which displays a prefix-suffix neutrality is Hindi (an Indo-European language). As it could have been easily foreseen, suffixes are the most favoured strategy in evaluative morphology of OV / Po languages (they are attested in 66% of the languages in our sample).

Thus, the cross-linguistic distribution of ‘affixal neutrality’ also seems to be asymmetrical, favouring head-initial languages.

Table B gives a detailed picture of the situation: all the formal strategies used by each language of the sample to express the four main evaluative meanings (SMALL, BIG, GOOD, BAD) are grouped on the basis of their meaning. The grey areas correspond to the cases of neutrality between prefixes and suffixes (or between infixes and suffixes). The table contains 220 slots; 117 of them are empty. This means that the corresponding semantic value is not morphologically expressed. The great majority of empty slots refers to the ‘qualitative’ side of evaluative morphology, represented by the GOOD / BAD opposition. As to the filled slots, 38 out of 113 (about 33%) correspond to some kind of neutrality; but just 2 of them refer to the ‘qualitative’ side of evaluation:

(6)	SMALL	23 instances of affixal neutrality
	BIG	13 instances of affixal neutrality
	GOOD	1 instance of affixal neutrality
	BAD	1 instance of affixal neutrality

Therefore, data provide evidence for the hypothesis that prefix-suffix neutrality is a characteristic of the quantitative side of evaluative morphology. Moreover, very interestingly, about 92% of these slots (35 out of 38) correspond to VO languages (belonging above all to the Indo-European family, but also to the Afro-Asiatic family and to the Niger-Congo family).

Thus, data confirm that the absence of a ‘suffixing preference’ in evaluative morphology and the presence of the typologically unusual ‘prefix-suffix neutrality’ are widely (but not exclusively) attested in Indo-European VO languages.

There are probably two ways of interpreting these data. At first, one can wonder why prefix-suffix neutrality has emerged in many Indo-European VO languages and in a couple of head-initial non Indo-European languages.⁷ But, on the other hand, one can wonder why this neutrality has not taken place in the evaluative morphology of OV languages, although derivational prefixes are attested in many of them. In our opinion, this is the most puzzling aspect of the situation. In the last section of this paper, we will focus specifically on this issue, in order to explain the absence of prefix-suffix neutrality in head-final languages.

4. Conclusions

Although the sample taken into account is not balanced either genealogically or typologically, the data discussed in this paper convey, in our opinion, some promising hints and suggestions. The ‘prefix-suffix neutrality’⁸ seems to be a peculiarity of VO/Pr languages, possibly independently of their genetic affiliation (even if the degree of occurrence of the phenomenon we have investigated is particularly high in Indo-European languages). This unusual ‘affixal neutrality’ can thus be viewed as a typological characteristic of head-initial languages. In this picture, the problem is to understand why it is almost absent in head-final languages.

The hypothesis we suggest is that the possible explanation for this further asymmetry is to be looked for in the typological outline of morphological systems. It is well known that consistent OV languages tend to be agglutinative in their morphology (cf. Lehmann 1973: 47). Agglutinative languages tend to preserve a one-to-one correspondence between form and meaning. Prefix-suffix neutrality is a clear violation of this tendency: in this case, we have more formal items to express just one meaning. Such a situation would be very problematic for agglutinative languages. So, one can easily predict that languages with a low index of fusion tend to avoid these morphological strategies.

So, there are two possible answers to the question concerning the asymmetrical distribution of ‘prefix-suffix neutrality’ in evaluative morphology. In fact, we could state that the combination of VO word order and of a high index of fusion favours the development of ‘prefix-suffix neutrality’. But we could state also that the combination of OV word order and a low index

⁷ Some of these non Indo-European languages are directly influenced by some Indo-European languages (cf. Maltese).

⁸ We keep this expression as a broad term covering not only prefixes and suffixes but also discontinuous affixes.

of fusion disfavours the development of ‘prefix-suffix neutrality’. Which is the right path to follow? Bantu languages, which combine VO word order and agglutinative morphology, can play an important role in solving this problem. They can be a sort of ‘litmus paper’ to test our hypothesis. In fact, evaluative suffixes (which, to a certain extent, correspond to Indo-European evaluative prefixes) are a very recent innovation in Bantu languages. As a consequence, it is still not clear if their domain of application and that of evaluative prefixes do really overlap, at least partially. So, it will be very interesting to monitor the development of these suffixes in the coming years, in order to understand which is the strongest factor in conditioning prefix-suffix neutrality between VO word order and agglutinative morphology. Of course, if prefix-suffix neutrality widely takes place in Bantu languages such as in Indo-European languages, then VO word order (attested both in Bantu and Indo-European languages) should be considered as the prevailing factor in favouring prefix-suffix neutrality. On the contrary, if the spread of evaluative suffixes in Bantu languages is not wide enough to generate a real ‘prefix-suffix neutrality’, then agglutinative morphology (which distinguishes Bantu languages from Indo-European languages) should be considered as the strongest factor in disfavoursing prefix-suffix neutrality.

References

- Campbell, G. L. (2000), *Compendium of the world's languages*, London-New York, Routledge
- Comrie, B. (1990²) (ed.), *The world's major languages*, New York-Oxford, Oxford University Press.
- Cutler, A., J.A. Hawkins & Gilligan G. (1985), *The suffixing preference; a processing explanation*, “Linguistics” 23, 723–758.
- Grandi, N. (2002), *Morfologie in contatto. Le costruzioni valutative nelle lingue del Mediterraneo*, Milano, Franco Angeli.
- Grandi, N. & F. Montermini (forthcoming), *Valutativi suffissali e prefissali: un'unica categoria?*, in M. Grossmann & A.M. Thornton (eds.), *La formazione delle parole*, Atti del XXXVII Congresso Internazionale della SLI (L'Aquila, September 25–27 2003).
- Greenberg, J.H. (1957), *Order of affixing: a study in general linguistics*, in J.H. Greenberg, *Essays in Linguistics*, Chicago–London, University of Chicago Press, 86–94.
- Greenberg, J.H. (1963), *Some universals of language with particular reference to the order of meaningful elements*, in J.H. Greenberg (ed), *Universals of Language*, Cambridge (Mass.), MIT Press, 58–90.
- Hall, C.J. (1988), *Integrating diachronic and processing principles in explaining the suffixing preference*, in J.A. Hawkins (ed), *Explaining Language Universals*, Oxford, Basil Blackwell, 321–349.
- Hawkins, J.A. & G. Gilligan (1988), *Prefixing and suffixing universals in relation to basic word order*, “Lingua” 74, 219–259.
- Lehmann, W.P. (1973), *A structural principle of languages and its implications*, “Language” 49, 47–66.
- Mel'čuk I.A. (2000), *Formal processes*, in G. Booij et al. (eds), *Morphology. An International Handbook on Inflection and Word-Formation*, vol. I, Berlin–New York, de Gruyter, 523–535.
- Ruhlen, M. (1987), *A Guide to the World's Languages. Vol. 1 Classification*, Stanford, Stanford Univ. Press.
- Sapir, E. (1921), *Language. An Introduction to the Study of Speech*, New York, Harcourt, Brace & World Inc.

Appendix

TABLE A

Language	Classification ⁹	Morph. Type	VO/OV	Pr/Po	Evaluative morphology
Italian	Indo-European/Italic	Fusional	VO	Pr	SUFFIXES, Prefixes
French	Indo-European/Italic	Fusional	VO	Pr	SUFFIXES, Prefixes
Spanish	Indo-European/Italic	Fusional	VO	Pr	SUFFIXES, Prefixes
Catalan	Indo-European/Italic	Fusional	VO	Pr	SUFFIXES, Prefixes
Romanian	Indo-European/Italic	Fusional	VO	Pr	SUFFIXES, Prefixes
German	Indo-European/Germanic	Fusional	VO	Pr	SUFFIXES, Prefixes
Dutch	Indo-European/Germanic	Fusional	VO	Pr	SUFFIXES, prefixes
Swedish	Indo-European/Germanic	Fusional	VO	Pr	Prefixes
Danish	Indo-European/Germanic	Fusional	VO	Pr	Prefixes
Norwegian	Indo-European/Germanic	Fusional	VO	Pr	Prefixes
Icelandic	Indo-European/Germanic	Fusional	VO	Pr	Prefixes
Russian	Indo-European/Slavonic	Fusional	VO	Pr	SUFFIXES, Prefixes
Polish	Indo-European/Slavonic	Fusional	VO	Pr	SUFFIXES, Prefixes
Czech	Indo-European/Slavonic	Fusional	VO	Pr	SUFFIXES, Prefixes
Bulgarian	Indo-European/Slavonic	Fusional	VO	Pr	SUFFIXES, Prefixes
Croatian	Indo-European/Slavonic	Fusional	VO	Pr	SUFFIXES, Prefixes
Slovene	Indo-European/Slavonic	Fusional	VO	Pr	SUFFIXES, Prefixes
Irish	Indo-European/Celtic	Fusional	VO	Pr	Suffixes
Albanian	Indo-European/Albanian	Fusional	VO	Pr	Suffixes
Greek	Indo-European/Greek	Fusional	VO	Pr	SUFFIXES, Prefixes
Modern Stand. Armenian	Indo-European/Armenian	Agglutinative	VO	Pr/Po	SUFFIXES, Prefixes (only loan prefixes)
Hindi	Indo-European/ Indo-Iranian/Indic/Central	Fusional	OV	Po	Prefixes, SUFFIXES
Kâmv'iri	Indo-European /Indo-Iranian/Nuristani /Northern	Fusional	OV	Po	Suffixes
Hungarian	Uralic/Finn-Ugric/Ugric	Agglutinative	OV	Po	Suffixes
Finnish	Uralic/Finno-Ugric/Finnic	AGGLUTINATIVE/Fusional	VO	Pr/PO	Suffixes, Prefixes
Turkish	Altaic/Turkic/Southern	Agglutinative	OV	Po	Suffixes
Evenki	Altaic/Mongolian-Tungus/Tungus/Northern	Agglutinative	OV	Po	Suffixes,
Basque	(isolate)	Agglutinative	OV	Po	Suffixes
Modern Standard Arabic	Afro-Asiatic/Semitic/West/Central/Arabo-Canaanite/Arabic	Fusional	VO	Pr	Infixes
Moroccan Arabic	Afro-Asiatic/Semitic/West/Central/Arabo-Canaanite/Arabic	Fusional	VO	Pr	Suffixes, INFIXES
Tunisian Arabic	Afro-Asiatic/Semitic/West/Central/Arabo-Canaanite/Arabic	Fusional	VO	Pr	Suffixes, INFIXES
Libyan Arabic	Afro-Asiatic/Semitic/West/Central/Arabo-Canaanite/Arabic	Fusional	VO	Pr	Infixes
Egyptian Arabic	Afro-Asiatic/Semitic/West/Central/Arabo-Canaanite/Arabic	Fusional	VO	Pr	Infixes
Syrian Arabic	Afro-Asiatic/Semitic/West/Central/Arabo-Canaanite/Arabic	Fusional	VO	Pr	Infixes

⁹ Classification is from Ruhlen (1991²). The typological information in the third, fourth and fifth columns are from Comrie (1990²) and Campbell (2000).

Prefix-Suffix Neutrality In Evaluative Morphology

Maltese	Afro-Asiatic/Semitic/West/Central/Arabo-Canaanite/Arabic	Fusional	VO	Pr	Suffixes, INFIXES
Hebrew	Afro-Asiatic/Semitic/West/Central/Arabo-Canaanite/Canaanite	Fusional	VO	Pr	Suffixes, Infixes
Berber (Tamazight)	Afro-Asiatic/Berber/Northern/Atlas	Fusional	VO	Pr	CIRCUMFIX, subtractive morphology, Suffixes
Mandarin Chinese	Sino-Tibetan/Sinitic/Chinese	Isolating	VO	Pr	Pref/Suf
Korean	Korean-Japanese	Agglutinative	OV	Po	No (vowel and consonant alternations)
Japanese	Korean-Japanese	Agglutinative	OV	Po	Prefixes (consonant alterations)
Thai	Kam-Tai/Tai	Isolating	VO	Pr	Pre-nominal free (?) forms
Vietnamese	Austriac/Viet-Muong	Isolating	VO	Pr	Post-nominal free (?) forms
Malay	Austronesian/Western/Malayo-Polynesian/Malayic/Malayan	Agglutinative	VO	Pr	Pre-nominal (?) free forms
Makassarese	Austronesian/Western Malayo-Polynesian/Celebes/South Sulawesi	Agglutinative	VO	Pr	No (reduplication)
Samoa	Oceanic/Central Pacific/Polynesian/Samoic Outlier	Agglutinative	VO	Pr	No
Pileni	Oceanic/Central Pacific/Polynesian/Samoic Outlier	Agglutinative	VO	Pr	No
Swahili	Niger-Congo/Bantu/Central Bantu/Swahili	Agglutinative	VO	Pr	PREFIXES, Suffixes
Shona	Niger-Congo/Bantu/Central Bantu/Shona	Agglutinative	VO	Pr	PREFIXES, Suffixes
Yukaghir	Uralic-Yukaghir/Yukaghir	Agglutinative	OV	Po	Suffixes
Avar	Caucasian/North/Daghestanian	Agglutinative	OV	Po	Suffixes
Inuktitut (East. Canadian Inuit)	Eskimo-Aleut/Eskimo/Inuit	Polysynthetic	OV	Po	Suffixes
Mapudungun	Amerind/Andean/Southern	Polysynthetic	OV	Pr	No
Potawatomi	Amerind/Northern/Algonquian	Polysynthetic	OV	Pr	Suffixes
Nahuatl	Amerind/Uto-Aztecan/Aztec	Agglutinative	VO	Po	Suffixes
Chukchi	Chukotko-Kamchatkan/Northern/Chukchi	Incorporating	OV	Po	Suffixes

TABLE B¹⁰

Language	SMALL	BIG	GOOD	BAD
Italian	Pref / SUF	Pref / SUF	PREF / Suf (?)	Suf
French	Pref / SUF	Pref	Pref	Suf
Spanish	Pref / SUF	Pref / SUF	Pref	Suf
Catalan	Pref / SUF	Pref / SUF	Pref	
Rumanian	Pref / SUF	Pref / SUF	Pref	Suf
German	Pref / SUF	Pref		
Dutch	Pref / SUF	Pref		
Swedish	Pref	Pref		
Norwegian	Pref	Pref		
Danish	Pref	Pref	Pref	
Icelandic			Pref (?)	
Russian	Pref / SUF	Pref / SUF	Pref (?)	Suf
Polish	Pref / SUF	Pref / SUF	Suf	Suf
Czech	Pref / SUF	Pref / SUF	Pref	
Bulgarian	Pref / SUF	Pref / SUF	Pref (?)	Suf
Croatian	Pref / SUF	Pref / SUF		
Slovene	Pref / SUF	Pref	Pref	
Irish	Suf			
Albanian	Suf	Suf		
Greek	Pref / SUF	Pref / SUF		
Modern St. Armenian	Pref / SUF	Pref		
Hindi	Pref / SUF	Pref / SUF	Pref	Pref
Kâmv'iri	Suf			
Hungarian	Suf			
Finnish	Pref / Suf	Pref	Pref	
Turkish	Suf			
Evenki	Suf			
Basque	Suf			
Modern Stand. Arabic	Inf			
Moroccan Arabic	Inf	Suf		
Tunisian Arabic	Inf	Suf		
Libyan Arabic	Inf			
Egyptian Arabic	Inf			
Syrian Arabic	Inf			
Maltese	INF / Suf	Suf		Suf
Hebrew	Inf / Suf			
Berber (Tamazight)	CIRCUMFIX / Suf	subtractive morphology		
Mandarin Chinese	Pref / Suf		Suf	
Korean	vowel and consonant alternations			
Japanese	Pref		Pref	Pref / CONSONANT ALTERATION
Thai	pre-nominal free (?) forms	pre-nominal free (?) forms		

¹⁰ As to this table, we have decided not to make a distinction between native morphemes and borrowed morphemes. For example, Swedish prefixes are non-native, since they represent a consequence of the spread of the cultural pan-European lexicon we have mentioned in § 2. Of course, all the borrowed morphemes included in the table are fully integrated in the languages involved; in other words, their use is not limited to formal varieties of the languages.

Prefix-Suffix Neutrality In Evaluative Morphology

Vietnamese	post-nominal free (?) forms	post-nominal free (?) forms		
Malay	pre-nominal free (?) forms	pre-nominal free (?) forms		
Makassarese	reduplication			
Samoan				
Pileni				
Swahili	PREF / Suf	PREF / Suf		
Shona	PREF / Suf	PREF / Suf		
Yukaghir	Suf	Suf		
Avar	Suf			
Inuktitut (Eastern Canadian Inuit)	Suf	Suf	Suf	Suf
Mapudungun				
Potawatomi	Suf			
Nahuatl	Suf			
Chukchee	Suf	Suf		

N.B. In case of affixes with two possible meanings, only the primary one has been taken into account.

Pref = prefixes / Suf = suffixes / Inf = infixes

Appendix: Maps



SMALL



BIG



GOOD



BAD



Word Formation and Typology: Which Language Universals?

Livio Gaeta

Dipartimento di Linguistica, Università di Roma Tre

livio.gaeta@unito.it

1. Introduction

Typology was born with morphology, at least taking into consideration the old Humboldtian idea of comparing language types (isolating, inflectional, etc.) on the basis of the morphological properties of words (for a survey cf. Coseriu 1973). In this perspective, a central role was played by inflectional morphology, at least in its prototypical core. This is not surprising because of the “paradigmatic” nature of inflectional morphology, which allows one to identify categories and values in a rather precise way (cf. on this subject Ricca in press). This very nature is probably the reason why typological investigations on (some aspects of) inflectional morphology have been extremely fruitful both from a synchronic (just to mention a few, see for instance Blake 2001, Corbett 1991, 2000) and a diachronic perspective: consider in this latter respect the exemplary volumes by Bybee, Perkins & Pagliuca (1994), in which the source and the distribution of inflectional morphemes is investigated on the basis of a well-balanced language sample. Furthermore, the occurrence of well-profiled categories has favored the formulation of Greenbergian universals such as: “If a language has a trial, it also has a dual”.

Therefore, generalizations of a typological character involving inflectional morphology are numerous, even though in several cases still requiring an empirical validation. Much less so for derivational morphology and in general for word formation. In this respect, it is interesting to consult the Universals Archive (= UA) worked out by F. Plank and E. Filimonova at Konstanz University. The archive, which records about 2000 universals of various character occurring in typological literature, allows one to obtain an overview on what can be considered “received wisdom” in typology. Compared to approximately 170 universals concerning inflectional morphology, for derivational morphology the number of possible universals amounts to about 60, which illustrates how difficult it is to adopt a typological perspective when dealing with the latter. In the rest of the paper, I intend primarily to explore what is the state of the art for typology and derivational morphology, and in doing so I will rely on the Konstanzer archive.

Among the universals sampled in the archive, there are two basic types (in the rest of the paper, the archive universals will be identified by their archive number). The first type consists in unrestricted or non-conditional universals, which assert general properties of language:

- #919 As the number of contrastive segments in a language increases, the average length of a word will decrease.
- #662 There is no reduplication pattern which would not involve reference to lexical identity.

The properties asserted in such universals are conditions holding achronically in the first example as for the form of lexical morphemes, and in the second one as for the derivational meaning (“Wortbildungsbedeutung”) of a certain derivational process (reduplication). Furthermore, there are implicational universals, which relate the occurrence of two properties of language:¹

#892 OV languages tend to have suffixes and VO languages prefixes.

In this contribution I will attempt on the basis of the data provided by the UA a first exploration of the principal traits of word formation from a typological point of view, and I will attempt to establish a minimal set of prerequisites which a typology of derivational morphology must display to answer Bauer’s (2002) question on what the latter is definitely able to do. I must admit that this paper will only contribute to raise more questions instead of providing answers. Thus, it can be intended either as a *cahier de doléance* or as it really is, namely a research project. However, my conviction is that raising correct questions already offers the key to find right answers.

2. Delimiting the Field

The first problem to face is how to discriminate whether a morphological process is to be attributed to inflectional morphology, or rather to derivational morphology. In this respect the criteria usually assumed to distinguish between the two do not always provide reliable results. One could even ask whether the distinction is of a categorial (and so qualitative) or rather of a quantitative nature, and this in turn implies a theoretical model to which one may refer. I will not pursue this issue here, and rather refer the reader to the literature (cf. among others at least Scalise 1988, Dressler 1989, Anderson 1992).

Among the several approaches to the question, Haspelmath’s (1996) paper is very telling, since he attributes to inflectional rules the property of changing word class, a criterion usually held to be fundamental to define prototypical word formation. This might be true but, as the author admits, only under certain conditions, i.e. with non-prototypical inflection: participles, verbal adjectives, infinitives, and so on. In this uncertainty, light can be shed by proceeding in an empirical way taking into consideration single categories expressed by single morphological markers. This is for example the procedure followed by Bauer (2002). This procedure is not without contradictions. Among the derivational patterns of Kwakw’ala, Anderson (1985)

¹ As pointed out by Plank in the guidelines of UA, several allegedly implicational universals recorded have a para-conditional status rather than an implicational one. One such universal is the following one:

#219 In all languages: if there are non-root morphemes, then all such morphemes have a more limited inventory of phonemes than root-morphemes and the average length of non-root morphemes is not more than the length of root-morphemes.

In this case the condition can be paraphrased by “assuming that, given that”, and the positive formula does not coincide with the negation of the contrary: If p then q ≠ If ¬q then ¬p. Furthermore, the question arises whether the asserted correlation is relevant or not at a typological-structural level. In this sense, the following universal does not appear to highlight a relevant language property, since its domain is too narrow:

#1581 If there is a reflexive verb meaning 'to laugh' it is usually not derived from the transitive base meaning 'to make somebody laugh'.

mentions affixes (see the examples in (1) below) corresponding to what in many languages are inflectional categories: temporal (future, recent past, remote past, etc.), aspectual (inchoative, habitual, repetitive, etc.), voice, modality (optative, potential, exhortative, etc.), noun plural (simple plural, distributive, etc.):

- (1) $x^w ak^w \text{əna}$ ‘canoe’ → $x^w ak^w \text{ən}\lambda$ ‘canoe that will be, that will come into
existence’
 $x^w ak^w \text{əna}$ ‘canoe’ → $x^w ak^w \text{əna}\text{x}\text{di}$ ‘canoe that has been destroyed’

The main argument to support his approach is given as follows:

“[I]t is hard to find secure criteria for classifying these elements as derivational or inflectional: we take it to be significant for the derivational status of at least the temporal, aspectual and plural groups that they are (a) optional, and present only where necessary for emphasis or disambiguation; and (b) equally applicable to words of any syntactic function or word class ... These forms involve the same suffixes as those appearing with verbs to mark the same categories, and this is general across all members of these classes” (Anderson 1985: 30)

The argument in favor the derivational status of these tense markers is of a distributional nature: the affixes occur with different word classes. Notice that this argument allows one to interpret the affixes as operating a transcategorization, changing word class. This makes these affixes in a way similar to the Dutch example of bracketing paradox reported in (2), for which Booij (2002: 161) assumes a conversion from noun to verb, which is however only contextually conditioned:

- (2) *breedgeschouderd* ‘broad-shouldered’ [A [[ge- [N]_v -d]_v]_A]_A

The theoretical justification is again of a distributional nature, since the conversion pattern is independently well-established in Dutch. The bracketing paradox is solved once that one expresses “that certain independently established word formation patterns co-occur: the use of one pattern implies the use of the other” (Booij 2002: 161). We would clearly ascribe neither the Dutch affix nor the category to which it belongs to derivation. We would rather speak of conversion, or of zero derivation, depending on the theoretical persuasion, assuming an abstract derivational level. However, in the light of the Kwakw’ala verbalizing suffixes, nothing prevents us from considering the affix derivational!

These uncertainties require a very careful approach, which not only looks at the individual patterns, but more in general considers the whole morphological structure of a language. Adopting Bauer’s procedure condemns us to replicate his negative results, as for instance for derivation producing nouns:

“There are no implicational scales observable here. There are languages which appear to allow abstract nouns derived from adjectives without allowing abstract nouns derived from verbs (though the latter could be counted as inflectional, it must be recalled)” (Bauer 2002: 40)

Given the difficulty of discriminating between inflection and derivation, it seems to me a better alternative to check if implications come out, when morphology as a

whole is considered. In my opinion, in order to verify if there are interesting connections in form-function relations, it is first necessary to ask what is morphologized, and then look for more fine-graded distinctions in terms of inflection/derivation (see for instance Noonan 1997).

Related to this question is the problem of determining which and how many are the word classes in a given language (see on the question Comrie & Vogel 2000). In fact, the debate on whether at least nouns and verbs must be considered universal categories is still open (compare for instance Sasse 1993 vs. Mithun 2000). Establishing the nature and the kind of word classes also allows one to specify the selection domain of word formation rules, even though Plag (1999: 144) has recently pointed out that “one could even come up with the strong hypothesis that with any given productive affix, the syntactic category of potential base words is only a by-product of the semantics of the process” (Plag 1999: 144). In this view, the role played by word classes in morphology is strongly diminished.

Moreover, the problem arises of verifying whether there exist derivational categories that can be considered “universal” similarly to those assumed for inflectional morphology, in order to look on this basis for possible implicational universals (see in this regard the scepticism of Bauer 2002), as the following one:

#1945 If a language has denominal derivation, it has nominal derivation (= derivation of something else to nouns).

Again, here the approach must be probably broader, and more “functionally”-based, in the sense of first looking at general strategies adopted by languages to carry over specific functions such as nominalization, verbalization, modification in the verbal (i.e. adverbs) and in the nominal (i.e. adjectives) domain, intensification / evaluation, etc. Only after this scrutiny might fine-graded morphological investigation really start.

Connected with word classes is the question of derivational categories or types. Also on this subject opinions are diverging (see Scalise 1999 for a discussion), but it can be generally agreed that derivational types are rather neglected within the theoretical debate, the only exception being Zwanenburg’s (1980, 1984) framework. Should derivational categories be assumed as general viewpoints, under which to look for generalizations, or are they simply to be discarded since constituting a mere abstraction, meaningless for a “morpheme-based” approach? In the archive a universal responding to such a question regards (perhaps expectedly...) evaluatives:

#2015 Suggested hierarchy of base types for diminutivisation and augmentativisation:
Noun > Adjective, Verb > Adverb, Numeral, Pronoun, Interjection > Determiner.

Finally, one has to ask: how are the morphological techniques connected with the functions they perform? This question is in a way specular to what has just been said above because it is related to the expression side of morphology. For instance, it must be investigated whether there are strict relations between morphological techniques and lexical classes, as claimed by the following universal:

#1746 There is more prefixing on verb than on noun. If a language has any prefixes on noun, it will also have prefixes on verb with considerably more than chance frequency.

Moreover, it is not without interest to ascertain whether with respect to certain morphological techniques it is possible to sustain generalizations such as for instance “In a given language if composition expresses action nouns, then it also expresses agent nouns”. In this sense consider the following universal concerning reduplication:

#1868 IF reduplication is used for grammatical purposes in any other word class, THEN it is also used (for whatever purpose: gradation, superlative, intensification, distributivity, diminution, etc.) for adjectives or adjective-derived nouns.

In this vein, it would be interesting to know what is the range of possible derivational meanings expressed by affixation with respect to composition, or the relation between endocentric and exocentric composition, and so on.

3. Universals and Word Formation

After having touched upon these questions of a general character, which have rather been hinted at than answered, let us now consider the archive in search of morphological universals and more in particular of universals concerning word formation. As reported in the archive, the sum of universals connected with morphology amounts to about 230, among which a first set concerns morphology as such, independent of the morphological categories involved. A second set is of a transmorphological character since it concerns the connection of morphology with other language components, especially phonology and syntax. A third set of universal is specified for inflectional morphology; they will not be discussed here (cf. Ricca in press): morphological categories such as tense, aspect and mood for verbs, case and number for nouns, etc., have been excluded from the analysis, even though the caveats hinted at in the preceding section must be kept in mind. Finally, about 60 universals are devoted to word formation proper, which must be again grouped either into categorial universals which relate to a specific word formation category, or into transcategorial ones, if two different (not necessarily both derivational) categories are taken into consideration.

3.1 Structural Universals

Let us start from more general universals relating to morphology as such, in its form – meaning dimension, as in the following one:

#362 The extent of “material” articulation, pertaining in particular to (a) the elaboration of sound systems, (b) the complexity of syllable structures, (c) word length, (d) accentual differentiation (as opposed to not-so-articulated tonal modification), correlates with the extent of “formal” articulation, pertaining in particular to (a) the differentiation of parts of speech, (b) the

elaboration of inflectional and derivational systems, (c) analytic syntax (as opposed to not-so-articulated polysynthesis).

A similar structural dimension is shared by the two following “classical” Greenbergian universals, which describe the relation between inflection and derivation in distributional terms:

#508 If a language has inflection, it always has derivation.

#507 If both the derivation and inflection follow the root, or they both precede the root, the derivation is always between the root and the inflection.

A handful of universals concern conditions on possible morphemes, as for instance the following ones which constrain the distribution of affixation, always implied by less “diagrammatic” techniques, to use a term of Natural Morphology (cf. Dressler 1985):

#505 If a language has discontinuous affixes, it always has either prefixing or suffixing or both.

#1946 The use of all other processes of nominal derivation and inflection (namely internal modification, suprasegmental processes, subtraction, conversion, suppletion), with the exception of total reduplication, implies the use of affixation.

Further conditions define the limits of allomorphy, as in the following case:

#908 Allomorphy cannot be conditioned across (grammatical) word boundaries.

Finally, about ten universals deal with reduplication, which has been the object of several investigations from a typological viewpoint (above all, cf. Moravcsik 1978). This is not surprising given the pretty well defined nature of reduplication as a morphological technique (although much less so as for the range of its derivational meanings), and its limited distribution. Among others, the following two respectively refer to the form and to the content of the reduplication rules:

#663 There is no reduplication pattern that would involve reference to phonological properties other than syllable number, consonantality-vowelhood, and absolute linear position.

#268 If in a language reduplication (full or partial) exists as a productive grammatical means of word- and formbuilding, then, included in the meanings expressed by means of reduplication, we find the meaning “change of quantity or degree”.

3.2 *Transmorphological Universals*

The second group of universals concerns what I call transmorphological relations, namely the relation of morphology with other language components, basically phonology and syntax. The first subset of about 20 universals touches phonology and morphology, and displays the following range of topics, each exemplified by a couple of universals:

- *Conditions on Segmental Structure*

#1963 IF there are consonant clusters C_iC_j , THEN there are also stems of the form C_iVC_j .

#1965 IF a phonotactic constraint holds for a syllable-edge, THEN it also holds for a corresponding word edge, but not vice versa.

- *Conditions on Suprasegmental Structure*

#374 There is a positive correlation between higher syllable-per-sentence and syllable-per-word ratios, simpler (or shorter) syllables, agglutinative morphology, and (S)OV basic word order on the one hand and between lower syllable-per-sentence and syllable-per-word ratios, more complex (or longer) syllables, flective (or no) morphology, and (S)V(S)O basic word order on the other.

#713 IF morphology is agglutinative, THEN there is vowel harmony. IF morphology is flexive, THEN there is stress accent.

- *Conditions on Stress*

#711 IF morphology is agglutinative, THEN (stress) accent will be demarcating, falling on word edges (either on initial or final syllables), rather than be free and centralizing, and there consequently will not be much phonological reduction of initial or final syllables.

#1964 If a language has the basic word-form stem+suffix, the accent will fall on a non-final syllable. If a language has the basic word-form that coincides with the stem, the accent will fall on the final syllable.

Among the about fifteen transcategorial universals relating to morphology and syntax, a large majority regards the connection between morphological properties and the basic word order, as in the “classical” Greenbergian universal:

#506 If a language is exclusively suffixing, it is postpositional; if it is exclusively prefixing, it is prepositional.

Other morphosyntactic universals are related to verb argument structure (cf. #608), and there is also a morpho-lexicological universal as #1201:

#608 If a language has a derivational morpheme whose distributional characterization makes reference to objects, it will also make reference to intransitive subjects but not to transitive one.

#1201 IF a language is (more or less) analytic, THEN it has a (more or less) regular phraseological system.

3.3 *Categorial Universals*

The last relevant group of universals deals with single derivation categories or types. I use the term “category” in a rather broad sense here, meaning both what some linguist would call supercategory (for instance, evaluation, comprising both diminutives and augmentatives), and single instantiations of categories such as causatives and reflexives, which could be theoretically subsumed under a supercategory “valence-changing operations”. They are quite limited in number, and can be subdivided into a first subset of intracategorial universals, which comprises the following categories illustrated in the usual way:

- *Causatives*

#1583 If there are causative affixes in a language which serve to form causative verbs from transitives, then this language also has causative affixes which serve to form causative verbs from intransitives.

- *Numerals*

#536 When a number is expressed by subtraction, or when a subtraction occurs as a constituent of a complex expression, the subtrahend is never larger than the remainder.

- *Honorifics*

#657 Honorific affixes to pronouns are more common than pejorative ones; if a language has pejorative pronominal affixes, it also has honorific ones.

- *Reflexives*

#1579 If there are any reflexive verb derived from intransitives by adding reflexive marker and an affix (or a predicative adjective, etc.), a great number of them is likely to imply intensity of action and resultant state.

- *Evaluatives*

#1932 There is an apparently universal iconic tendency in diminutives and augmentatives: diminutives tend to contain high front vowels, whereas augmentatives tend to contain high back vowels.

A second subset of transcategorial universals connects two different derivational categories as in the following cases where participles and deverbal nouns, or reflexives and causatives are put into relation:

- #396 If a language has a morphological means to indicate verbal modification (i.e., if a language has participles), then it has a morphological means to indicate verbal reference (i.e., a language has nominalized verb forms).
- #1582 If both reflexive marker and the causative marker in a language are affixes, both are: (a) either prefixes (cf. Abkhaz, Amharic, Klamath), or (b) suffixes (Yakut, Kechua, Aymara) or (c) reflexive marker is a prefix and the causative marker is a suffix (cf. Georgian, Ainu, Nivkh, Luganda, Shoshone); it is unlikely for the reflexive marker to be a suffix and the causative marker, a prefix.

The quantitative extension of the universals is, however, strongly different for the single categories. The most represented category are numerals, both because of the substantial investigation by Greenberg (1978), and because of their “paradigmatic” nature, which makes them a less prototypical instance of word formation. Similarly, the comparatively high number of universals relating to causatives is partly due to the study of Nedjalkov & Sil'nickij (1973), and partly to the fact they are often expressed by means of inflectional techniques. Recapitulating all data in the following table, the result for the other categories is rather miserable:

(3)

<i>Categories</i>	<i>Universals</i>
Numerals	45
Causatives	12
Reflexives	4
Evaluatives	2
Honorifics	1
Nominalizations	1
<i>Tot.</i>	65

These figures make it evident, that word formation is almost an unexplored continent for typology. In what respect this may be done must be verified with the help of specific investigations, which will take into consideration in a systematic way the derivational categories, the techniques employed to express them, and the possible connections with other properties of grammar.

4. Substantial and Formal Universals

Besides these substantial universals, which inductively result from investigations on more or less balanced language samples, within theoretical morphology several proposals of universal generalizations are current, which pertain to the shape of grammar, such as for instance Aronoff's Unitary Base Hypothesis. The compatibility of these latter universals with the traditional typological approach outlined above is not

straightforward, if only because they are not of a merely descriptive character, but presuppose a certain theoretical model including a number of (theory-dependent) constructs. For instance, the Unitary Base Hypothesis crucially relies on a clearcut notion of word class. However, it is not excluded that such “formal” universals might be of interest from a typological point of view, also because the latter perspective is often faced with similar definitory problems as discussed in § 2 above. With respect to “formal” universals, in my view it is possible to identify (at least) three types. The first type can be labeled as constraints on the form of the grammar, in other words as universal conditions modeling the grammar of a single language, and is exemplified by the Lexical Integrity Principle (cf. among others Bresnan & Mchombo 1995, and Gaeta 2003) or by the following more specific condition (cf. Menn & MacWhinney 1984):

- (4) Repeated Morph Constraint: *XY, where X and Y are adjacent surface strings such that both could be interpreted as manifesting the same underlying morpheme through regular phonological rules, and where either (a) X and Y are both affixes, or (b) either X or Y is an affix, and the other is a (proper subpart of a) stem.

This condition can be seen in action for instance in Italian to exclude that verb stems ending with an affricate be further derived by suffixes containing affricates, as in the following verbs:

- | | | | |
|-----|--|----------------------------------|-------------------------|
| (5) | *[[...tsV] _V -zione] _N | <i>avvizzi-re</i> ‘to wither’ | → * <i>avvizzione</i> |
| | | <i>tappezza-re</i> ‘to paper’ | → * <i>tappezzione</i> |
| | *[[...dʒV] _V -aggio] _N | <i>arrangia-re</i> ‘to arrange’ | → * <i>arrangiaggio</i> |
| | | <i>scheggia-re</i> ‘to splinter’ | → * <i>scheggiaggio</i> |

Formal universals of this kind have good chances of holding as general (only restrictive?) conditions for morphology, and therefore of being put on a pair with the substantial universals providing a picture of the types of possible complex morphemes throughout the world languages. A second type of formal universals is more specific since it pertains to the form of morphological rules, which as such crucially relies on the theoretical model adopted. For instance, the following conditions are both claimed to be universal, although giving opposite predictions:

- (6) Adjacency Condition: No WFR can involve X and Y, unless Y is uniquely contained in the cycle adjacent to X. (Siegel 1978, Allen 1979)
- (7) Atom Condition: A restriction on the attachment of af_x to Y can only refer to features realized on Y. (Williams 1981)

Both conditions involve a number of notions (such as for instance the idea of a derivational cycle or of percolation) which are related to a certain model of grammar. Moreover, they claim the morphology to function in a certain way, and accordingly make precise predictions on how complex morphemes should be. For instance, the Adjacency Condition claims that an affix may only have access to features realized on the previous derivational cycle. Accordingly, it correctly predicts that the Italian suffix

-*aggine* only selects adjectives bearing a negative semantics, which is provided by the prefix *in-* in the base *insensato* ‘senseless’:²

- | | | | | |
|-----|----|--|----|--|
| (8) | a. | [[<u>in[sensat]</u>] <i>aggine</i>] | b. | [[<u>in[sensat]</u>] _Y <i>aggine</i>] |
| | c. | <i>maturità</i> ‘maturity’
<i>sicurezza</i> ‘certainty’
<i>efficacia</i> ‘efficacy’
<i>precisione</i> ‘precision’
<i>cautela</i> ‘caution’ | d. | <i>immaturità</i>
<i>insicurezza</i>
<i>inefficacia</i>
<i>imprecisione</i>
<i>incautela</i> |
| | e. | [[<u>in[sicur]</u>] <i>ezza</i>] | f. | [[<u>in[sicur]</u>] _Y <i>ezza</i>] |

On the contrary, the Atom Condition is not able to predict the correct form, because the suffix may only have access to the lexical head, onto which the negative semantics of the prefix cannot percolate in Italian as shown in (8b), because prefixes are not heads. On the other hand, the Atom Condition correctly predicts that prefixed adjectives as in (8d) select the same suffixes as their bases in (8c), whereas the Adjacency Condition cannot express this regularity, because of the blocking effect of the intervening derivational cycle as shown in (8e). As can be seen, both conditions present shortcomings in accordance with the set of examples considered. Because of this restricted validity, and for the reasons mentioned above, this second type of universals related to the rule format cannot in my opinion be easily generalized about. Even more idiosyncratic, in the sense of theory-internal, is the third type of formal universals, which is related to the grammar format. One example of this kind of universals is in my opinion the Right-Hand Head Rule (cf. Williams 1981), which assigns only to suffixes the property of being heads, and accordingly of inducing feature percolation. Notice that this property is the basis for the Atom Condition seen above. Similarly theory-internal is the so-called Mirror-Principle of Baker (1985):

- (9) Mirror-Principle: The order of morphological operations, as revealed by the order of affixation, is always identical to that of syntactic operations.

The empirical testability of this kind of universals is highly problematic (cf. Carstairs-McCarthy 1992: 119–130 for a discussion), especially because of the high number of abstract levels requested, which makes the cross-linguistic comparison difficult (and in several cases vacuous).

Independently of the nature and the validity of the single theoretical constructs, the conditions that in my view allow formal universals to be put together with substantial universals are firstly the extent to which they are able to predict a large amount of data, i.e. they grasp universal tendencies. This is in my opinion the case for the notion head, even though with all possible caveats (see in this respect Bauer 1990 and Haspelmath 1992). Second, and more importantly, they should not be theory-

² Notice that, accordingly, **sensataggine* does not occur, and the positive base *sensato* ‘sensible’ selects *-ezza*: *sensatezza* ‘good sense’.

internal in the sense of requiring theoretical constructs which cannot be exported cross-linguistically.

5. Conclusion

To sum up, I hope to have made clear what is already shared knowledge among typologists as to which universal generalizations occur for morphology and in particular word formation. The result is rather miserable: no homogeneous picture either regarding the derivational categories investigated or the morphological techniques involved seems to emerge. On the other hand, an approach based on formal properties of derivational morphology has produced till now few concrete results to be used as guidelines for typological (or even only theoretical) research in a satisfactory way. This does not mean, however, that word formation should not be seen as an adequate research field to explore. On the contrary, my conviction is that also the latter should become a main research object for typological research on a well-balanced language sample. And this both from an achronic perspective, such as the one proper of typology, and from a diachronic viewpoint of “system ontogenesis”, as in the perspective adopted by grammaticalization theory.

As a final word, let me end by quoting Anderson (1992: 335), and fully subscribe to his programmatic point of view:

“[T]here is no substantial difference between typology and theory when correctly viewed. Of course, if it turns out that the correct descriptive framework admits of only a very few dimensions of variation for languages, with few possible values on each, some will say that we have discovered a typological framework while others will say that we have found the right set of parameters for Universal Grammar. There is no reason to think that what would make the one set happy should not make the others happy too”.

References

- Allen, M. (1979), *Morphological investigations*, PhD. Diss., University of Connecticut.
- Anderson, S. (1985), *Typological distinction in word formation*, in T. Shopen (ed.), *Language typology and syntactic description*, vol 3, Cambridge, Cambridge University Press, 3–56.
- Anderson, S. (1992), *A-morphous Morphology*, Cambridge, Cambridge University Press.
- Baker, M. (1985), *The Mirror Principle and morphosyntactic explanation*, “Linguistic Inquiry”, 16, 373–415.
- Bauer, L. (1990), *Be-heading the word*, “Journal of Linguistics”, 26, 1–31.
- Bauer, L. (1997), *Evaluative morphology: In search of universals*, “Studies in Language” 21, 533–575.
- Bauer, L. (2002), *What you can do with derivational morphology*, in S. Bendjaballah et al. (eds.), *Morphology 2000*, Amsterdam & Philadelphia, John Benjamins, 37–48.
- Blake, B. J. (2001), *Case. Second Edition*, Cambridge, Cambridge University Press.
- Booij, G. (2002), *The Morphology of Dutch*, Oxford, Oxford University Press.
- Bresnan, J.W. & S.M. Mchombo (1995), *The Lexical Integrity Principle: Evidence from Bantu*, “Natural Language and Linguistic Theory” 13, 181–254.
- Bybee, J., R. Perkins & W. Pagliuca (1994), *The Evolution of Grammar. Tense, Aspect and Modality in the Languages of the World*, Chicago, The University of Chicago Press.

Word Formation and Typology: Which Language Universals?

- Carstairs-McCarthy, A. (1992), *Current morphology*, London, Routledge.
- Comrie, B. & P.M. Vogel (2000) (eds.), *Approaches to the Typology of Word Classes*, Berlin & New York, Mouton de Gruyter.
- Corbett, G. (1991), *Gender*, Cambridge, Cambridge University Press.
- Corbett, G. (2000), *Number*, Cambridge, Cambridge University Press.
- Coseriu, E. (1973), *Sulla tipologia linguistica di W. von Humboldt. Contributo alla critica della tradizione linguistica*, "Lingua e stile", 8/2, 235–266.
- Dressler, W.U. (1985), *On the predictiveness of Natural Morphology*, "Journal of Linguistics", 21, 321–337.
- Dressler, W.U. (1989), *Prototypical Differences between Inflection and Derivation*, "Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung" 42, 3–10.
- Gaeta, L. (2003), *Ai limiti della morfologia basata sulle parole*, in A. Bisetto et al. (eds.), *Scritti di morfologia. In onore di Sergio Scalise in occasione del suo 60° compleanno*, Roma, Caissa, 47–59.
- Greenberg, J. (1978), *Generalizations about numeral systems*, in J. Greenberg (ed.), *Universals of Human Language*, vol. 3, Stanford, Stanford University Press, 250–295.
- Greenberg, J. (ed.), *Universals of Human Language*, 4 vols. Stanford, Stanford University Press, 1978.
- Haspelmath, M. (1992), *Grammaticization theory and heads in morphology*, in M. Aronoff (ed.), *Morphology now*, Albany, NY, SUNY Press, 69–82.
- Haspelmath, M. (1996), *Word-class-changing inflection and morphological theory*, in G. Booij & J. van Marle (eds.), "Yearbook of Morphology" 1995, Dordrecht, Kluwer, 43–66.
- Lehmann, Ch. & E.A. Moravcsik (2000), *Noun*, in G. Booij et al. (eds.), *Morphology: An International Handbook zur Flexion und Wortbildung*, Berlin, de Gruyter.
- Menn, L. & B. MacWhinney (1984), *The Repeated Morph Constraint: toward an explanation*, "Language", 60, 519–541.
- Mithun, M. (2000), *Noun and verb in Iroquoian languages: Multicategorisation from multiple criteria*, in B. Comrie & P.M. Vogel (eds.), *Approaches to the Typology of Word Classes*, Berlin & New York, Mouton de Gruyter, 397–420.
- Moravcsik, E. (1978), *Reduplicative constructions*, in J. Greenberg (ed.), *Universals of Human Language*, vol. 3, Stanford, Stanford University Press, 297–334.
- Nedjalkov, V.P. & G. Sil'nickij (1973), *The typology of morphological and lexical causatives*, in F. Kiefer (ed.), *Trends in Soviet Theoretical Linguistics*, Dordrecht, Reidel, 1–32.
- Noonan, M. (1997), *Versatile Nominalizations*, in J. Bybee et al. (eds.), *Essays on Language Function and Language Type. Dedicated to T. Givón*, Amsterdam & Philadelphia, John Benjamins, 373–394.
- Plag, I. (1999), *Morphological productivity*, Berlin & New York, Mouton de Gruyter.
- Ricca, D. (in press), *Appunti per una tipologia degli universali morfologici*, in N. Grandi (ed.), *Morfologia e dintorni. Studi di linguistica tipologica e acquisizionale*, Milano, Franco Angeli.
- Sasse, H.-J. (1993), *Das Nomen – eine universale Kategorie?*, "Sprachtypologie und Universalienforschung", 46/3, 187–221.
- Scalise, S. (1988), *Inflection and derivation*, "Linguistics" 26, 561–581.
- Scalise, S. (1999), *Rappresentazione degli affissi*, in P. Benincà et al. (eds.), *Fonologia e morfologia dell'italiano e dei dialetti d'Italia*, Roma, Bulzoni, 453–481.
- Siegel, D. (1978), *The Adjacency Constraint and the theory of morphology*, "North-Eastern Linguistic Society" 8, 189–197.
- UA = *Universals Archive*, by F. Plank & E. Filimonova, accessible at the website: <http://ling.uni-konstanz.de/pages/proj/sprachbau.htm>
- Williams, E. (1981), *On the notions 'lexically-related' and 'head of a word'*, "Linguistic Inquiry", 12, 245–274.
- Zwanenburgh, W. (1980), *Form and Meaning in Morphology*, "Quaderni di Semantica", 1, 327–337.
- Zwanenburgh, W. (1984), *Word Formation and Meaning*, "Quaderni di Semantica", 5, 130–142.

Rule Counting vs. Rule Ordering: Universal Principles of Rule Interaction in Gender Assignment

Tore Nessel
CASTL, Tromsø
tore.nessel@hum.uit.no

Gender (agreement class) represents a perfect testing ground for hypotheses about rule interaction given that conflicts between rules are frequently attested. Enger (2002) draws a distinction between two types of analysis: the classical rule ordering approach and a rule counting approach (Doleschal 2000 and Steinmetz 1986, 2002). In the present paper I shall propose that an adequate theory must invoke both ordering and counting of rules. My point of departure will be the framework proposed by Steinmetz (1986) and further developed by Rice (2003). While Enger (2002) characterizes this framework as a rule counting approach, we shall see that it also invokes what I shall call formal principles of rule ordering. However, going beyond Steinmetz (1986) and Rice (2003), I shall argue that in addition to formal principles of rule ordering, we also need *substantial* ordering principles. To this end I shall advance what I refer to as the “Core Semantic Override Principle”.

Section 1 discusses Steinmetz’ (1986) principle “Gender Tally”, according to which a noun is assigned the gender suggested by the majority of assignment rules. As a research strategy, I shall pursue the idea that assignment rules are not ordered unless a ranking is imposed by universal principles of rule ordering. One such principle is Kiparsky’s (1982) Elsewhere Condition explored in section 2. It is shown that the use of rule ordering in Steinmetz’ theory in part falls out as a consequence of this principle. In section 3 it is suggested that default hierarchies (Steinmetz 1986, Rice 2003) are relevant for gender assignment. However, at the same time it is pointed out that hierarchies postulated for individual languages must receive support from independent evidence in order to be more than *ad hoc* solutions. Section 4 argues that Corbett’s (1991:68f.) hypothesis that semantic rules take precedence in gender assignment may be too strong. As an alternative, I advance the Core Semantic Override Principle, whereby semantic rules referring to biological sex take precedence in gender assignment.

1. Rule Counting: Gender Tally

The reason why Enger (2002) characterizes the theory of gender assignment advocated in Steinmetz (1986, 2002) as a “rule counting approach” is that it invokes a principle that Steinmetz (1986) calls Gender Tally. It can be expressed as the following instruction:

- (1) Gender Tally:
Count the number of times each gender is assigned and assign the noun the gender with the highest value. (Steinmetz 1986:193)

In order to see how this works, consider the assignment of gender to German nouns like *Gemüse* ‘vegetable’ and *Gebäude* ‘building’, for which Steinmetz (2002:4) assumes the following rules to be relevant:

- (2) a. German nouns ending in *-e* are feminine (e.g. *die Treppe* ‘staircase’)
- b. German nouns with the prefix *ge-* are neuter (e.g. *das Geräusch* ‘noise’)
- c. Superordinate nouns in German are neuter (e.g. *das Möbel* ‘furniture’)

The term “superordinate” in (2c) may require clarification. *Möbel* is a superordinate noun in the sense that it is a cover term for the semantic field comprising chairs, sofas, tables etc. In the same way, *Gemüse* is a cover term for various vegetables and *Gebäude* for various types of building. According to Steinmetz’ rule (2c) such superordinate terms are neuter in German. In the case of *Gemüse* and *Gebäude* two rules point towards neuter (i.e. 2b-c) and one towards feminine gender (i.e. 2a). Hence, Gender Tally predicts neuter gender, a prediction that is borne out by the facts.

Another illustration of the effect of Gender Tally in German involves words like *Gefängnis* ‘prison’ and *Gedächtnis* ‘memory’ (Steinmetz 1986:200f.). In addition to the *ge-* prefix referred to in rule (2b) above, the suffix *-nis* is relevant for the gender of this type. While the suffix is compatible with both feminine and neuter gender as witnessed by feminines like *Finsternis* ‘darkness’ and *Erlaubnis* ‘permission’ and neuters like *Zeugnis* ‘testimony’, there are no masculine nouns in *-nis*. One way to represent this, is to let two gender assignment rules refer to the suffix:

- (3) a. German nouns ending in *-nis* are feminine (e.g. *die Finsternis* ‘darkness’)
- b. German nouns ending in *-nis* are neuter. (e.g. *das Zeugnis* ‘testimony’)

These rules facilitate an account of the assignment of gender to *Gefängnis* ‘prison’ and *Gedächtnis* ‘memory’ in terms of Gender Tally. Two rules – (2b) and (3b) – indicate neuter gender, while only one – (3a) – points towards the feminine. Since the majority suggests neuter gender, this gender is correctly assigned.

Gender Tally receives support from connectionist processing (cf. e.g. McClelland and Elman 1986). When a target (in our case a noun) activates certain units in a network (assignment rules in our case), one of the factors facilitating the selection of a certain unit is the amount of conceptual overlap. The higher the degree of overlap, the greater are the chances that a certain unit is selected. This is analogous to Gender Tally. When a majority of rules competes with a minority, the majority represents the higher degree of conceptual overlap. In other words, Gender Tally assigns gender on the basis of conceptual overlap. While this does not indicate that one has to believe in connectionism in order to adopt Gender Tally, the parallelism is nevertheless interesting.

After these brief illustrations of Gender Tally and the rule counting approach, the question arises as to what an alternative analysis in terms of rule ordering would look like. The rules in (2) above are illustrative. Examples like *Gemüse* and *Gebäude* suggest that either (2b) or (2c) or both must outrank (2a), because otherwise feminine

gender would be assigned to these words. Now, in German there are superordinate nouns in *-e* lacking the *ge-* prefix, e.g. *Waffe* ‘weapon’ and *Pflanze* ‘plant’. Since these nouns are feminine, we are forced to order (2a) before (2c). Thus, we arrive at the ranking in (4a) below where the symbol \gg reads “outranks”. Consider now the feminine noun *Gemeinde* ‘community, congregation’. Since this noun posits both the *-e* suffix and the *ge-* prefix, rules (2a-b) are relevant. However, *Gemeinde* is not a superordinate, so (2c) does not apply. In order to predict that *Gemeinde* is feminine, (2a) must be ordered before (2b), as summarized in (4b). As pointed out by Rice (2003), these rankings are incompatible. Rule (2a) cannot be ordered both before and after (2b).

(4) Ordering paradox (after Rice 2003):

- a. (2b) \gg (2a) \gg (2c) (motivated by *das Gemüse* ‘vegetable’)
- b. (2a) \gg (2b) (motivated by *die Gemeinde* ‘community’)

This suggests that the adoption of ordered gender assignment rules produces ordering paradoxes. In view of this, a possible reaction would be to dismiss the rules in (2) and (3) altogether. Notably, however, they seem to represent fairly well established generalizations about German, and in any case the onus of proof would be on those who would want to present an alternative to these rules.

The strategy I shall explore in the following is to adopt the null hypothesis that gender assignment rules are not ranked. However, while I dismiss free ranking of individual rules, I shall assume that Gender Tally interacts with certain universal principles of rule ordering. In the model I propose, therefore, rules are ranked only when universal ranking principles force them to be so. Only when universal principles have been carefully investigated, a need for stipulated, language-specific rankings can possibly be established. The nature of universal principles and their interaction with Gender Tally is the topic of the remainder of this study. The first principle to be discussed is the Elsewhere Condition.

2. Rule Ordering: Elsewhere Condition

Kiparsky’s (1982) Elsewhere Condition regulates the order of application of rules of different degrees of specificity. If rule A refers to a proper subset of the nouns referred to by rule B, A takes precedence over B.¹ (This takes place no matter whether A belongs to a majority of rules favoring a certain outcome.) The notion of “default” has been widely used in studies of gender assignment (cf. e.g. Fraser and Corbett 1997), so there is every reason to believe the Elsewhere Condition to bear on gender assignment. Consider, as a simple example, the case of so-called indeclinable nouns in Ukrainian, i.e. nouns taking a zero ending throughout their inflectional paradigm. In Ukrainian, nouns of this type tend to belong to the neuter gender, e.g. *sari* ‘sari’ and *frykase* ‘fricassée’. However, indeclinable nouns denoting animates are masculine, e.g. *flamingo*

¹ The generalization that specific information takes precedence is also known as “Proper Inclusion Precedence” (Koutsoudas et al. 1974) and “Panini’s Principle” (Prince and Smolensky 1993).

‘flamingo’ and *poni* ‘pony’ (Pugh and Press 1999:56f.). The following two rules capture these generalizations.²

- (5) a. Ukrainian indeclinable nouns are neuter (e.g. *frykase* ‘fricassée’).
b. Ukrainian indeclinable nouns denoting animates are masculine (e.g. *flamingo* ‘flamingo’).

Since indeclinable nouns denoting animates constitute a proper subset of indeclinable nouns, rule (5b) takes precedence over (5a) by the Elsewhere Condition and masculine gender is correctly assigned to nouns like *flamingo* and *poni*.

A somewhat more complex example comes from Old Norse, as analyzed in Trosterud (2003):

- (6) a. Old Norse nouns are neuter.
b. Old Norse nouns for concepts related to time are masculine (e.g. *timi* ‘time’).
c. Old Norse nouns for concepts related to the annual cycle are neuter (e.g. *sumar* ‘summer’).
d. Old Norse nouns related to winter are masculine (e.g. *vetr* ‘winter’).

These rules constitute a nested structure where the nouns referred to in (6d) form a subset of those in (6c), which in turn are a subset of the nouns invoked by (6b). Rule (6a) is least specific – it is a global default rule stating that Old Norse nouns are neuter as long as other rules do not apply. Given the subset relationships between the rules, the Elsewhere Condition predicts a hierarchy where (6d) receives the highest ranking and (6a) the lowest. This prediction is borne out by the facts. As pointed out by Trosterud, *vetr* ‘winter’ is masculine because of (6d) although (6a) and (6c) point towards the neuter. The names of the other three seasons, *sumar* ‘summer’, *haust* ‘fall’ and *vár* ‘spring’ are neuter since (6c) overrides the conflicting (6b). Nouns for time-related concepts not covered by (6c-d), e.g. *aptann* ‘evening’ and *timi* ‘time’ are masculine in view of (6b), which takes precedence over the default rule (6a).

While Steinmetz (1986) does not refer to the Elsewhere Condition, it is in fact implicitly acknowledged in his framework. Consider, again, the interaction of rules (6a) and (6b). A pure rule counting approach would run into problems with nouns like *aptann* ‘evening’ and *timi* ‘time’. Here, one rule – (6b) – suggests masculine and one – (6a) – neuter. We are in other words facing a tie, and Gender Tally would therefore not

² Notice in passing that there are some systematic exceptions to these rules. For instance, according to Pugh and Press (1999: 57) indeclinable nouns denoting languages tend to be feminine, e.g. *bengali* ‘Bengali’ and *urdu* ‘Urdu’. Indeclinable common nouns like *madam* ‘madame’ and *ledi* ‘lady’, as well as a few indeclinable female first names like *Esfir* are feminine contra rule (4b) because Ukrainian has a general rule assigning feminine gender to nouns denoting females. Since the cases mentioned in this footnote do not bear on any conclusion to be drawn in the present study, they will not be discussed in the following.

be able to decide which gender to assign. The move made by Steinmetz (1986) is to assume that global default rules like (6a) only come into play when more specific rules tie. This seems tantamount to saying that default rules are ranked below specific rules, and this is in fact made explicit in Rice's (2003) Optimality Theory account of Steinmetz' framework, where default rules are ranked below specific rules. Interestingly enough, however, neither Steinmetz nor Rice attempts at justifying the ranking by invoking the Elsewhere Condition. Nevertheless, Steinmetz' move does not involve a merely stipulated ordering of rules, but rather a ranking that follows automatically from a well-established principle of rule ordering.

The upshot of this discussion is that Gender Tally must be supplemented by the Elsewhere Condition. In view of this, Enger's (2002) characterization of Steinmetz' model as a rule counting approach is to some extent misleading. While rule counting (Gender Tally) is pivotal in Steinmetz' framework, his model also involves rule ordering. This becomes even clearer when we consider Steinmetz' notion of default hierarchies, the topic to which we turn in the following section.

3. Rule Ordering: Default Hierarchies

The discussion of the Elsewhere Condition illustrates the relevance of defaults in gender assignment. In the following, I shall explore an extended use of defaults originating in the work of Steinmetz (1986) and Rice (2003). According to this theory, all languages contain global default rules for each gender, and these rules are mutually ranked. To see how this works, consider German nouns like *Waffe* 'weapon' and *Pflanze* 'plant'. According to Steinmetz (1986), two rules are relevant. In section 1, they were given as (2a) and (2c), but for the convenience of the reader I repeat them here:

- (7) a. German nouns ending in *-e* are feminine (e.g. *die Treppe* 'staircase')
- b. Superordinate nouns in German are neuter (e.g. *das Möbel* 'furniture')

Since one rule points towards the feminine and one towards the neuter, we are facing a tie, and Gender Tally does not enable us to select the right gender for *Waffe* and *Pflanze*. Furthermore, the Elsewhere Condition is of no help, because the rules in (7) do not stand in a subset relation to each other. In order to be able to handle cases of this type, Steinmetz (1986) assumes the German genders to form the hierarchy masculine >> feminine >> neuter. This means that the masculine is the default gender in German, while the feminine outranks the neuter. As is clear from Rice (2003), Steinmetz' gender hierarchies can be expressed in terms of rule interaction if one assumes default rules of the following type:

- (8) a. German nouns are masculine
- b. German nouns are feminine
- c. German nouns are neuter

In order to reflect Steinmetz' gender hierarchy, (8a) must be ranked above (8b), which in turn must outrank (8c). Given the Elsewhere Condition, the specific rules in (7) outrank the default rules in (8). Thus, we arrive at the hierarchy in (9).

(9) (7a), (7b) >> (8a) >> (8b) >> (8c)

Given the complexity of the matter, it may be fruitful to illustrate the interaction of the rules by means of the Optimality Theory tableau in (10), which is adapted from Rice (2003). (Rice states the default rules as negative restrictions, but the question of whether constraints are to be stated in negative or positive terms in Optimality Theory, does not bear on the question under scrutiny here.)

(10) Gender assignment to German *Waffe* 'weapon'
(tableau adapted from Rice 2003)

	-e=F (7a)	Sup=N (7b)	BeMasc (8a)	BeFem (8b)	BeNeut (8c)
<i>der Waffe</i> (masc.)	*!	*!		*	*
<i>die Waffe</i> (fem.)		*	*		*
<i>das Waffe</i> (neut.)	*		*	*!	

As can be seen from the tableau, (7a) and (7b) make it clear that *Waffe* and *Pflanze* cannot be masculine, but do not enable us to choose between the feminine and the neuter. Therefore, we must proceed to the lower-ranked default rules. Rule (8a) is indecisive, but the second default rule, (8b), enables us to assign the feminine gender.

In section 1 I introduced Gender Tally and argued against free ordering of individual gender assignment rules. Is the notion of "default hierarchy" explored above compatible with this research paradigm? As we have seen, Steinmetzian hierarchies involve ordered rules. A priori, there is no reason to preclude free ranking of them. At least, general restrictions on the ranking of default rules have not been discussed in the literature. However, free ranking of default rules is quite different from free ranking of *all* gender assignment rules. The number of default rules is limited and in most cases small since it equals the number of genders in a given language. Hence, the number of possible rankings is limited and the overall restrictiveness of the framework is not jeopardized. Moreover, the notion of "default hierarchy" raises interesting questions for further research. For instance, are there languages with feminine as the global default? What conditions changes in a default hierarchy over time? In view of the fact that default hierarchies provide a restricted framework that yields implications for further research, I propose to include it in a general theory of gender assignment. Let me hasten to add, however, that a default hierarchy postulated for any given language should be corroborated by independent evidence in order to be more than an *ad hoc* solution. We shall return to this point in the next section.

4. Rule Ordering: The Core Semantic Override Principle

So far I have argued that a rule counting approach should be supplemented by principles of rule ordering such as the Elsewhere Condition and default hierarchies. Since these

principles concern the logical relationship between rules, they may be referred to as “formal ordering principles”. In the following, I shall go further than Steinmetz and Rice and suggest that we also need *substantial* ordering principles, i.e. principles favoring rules invoking certain types of information. To this end I propose what I call the Core Semantic Override Principle.

4.1 *The Generalization*

The problem we shall consider concerns the assignment of grammatical gender to nouns denoting biological males and nouns denoting biological females. By way of illustration, consider Russian *djadja* ‘uncle’. In Russian, nouns ending in *-a* belonging to the second declension are generally feminine. Nevertheless, *djadja* and other second declension nouns denoting male persons are masculine. Seemingly, then, the semantics takes precedence over the declension for the purposes of gender assignment. The case of Russian *djadja* is not isolated as witnessed by the examples in (11) from otherwise quite different languages:

(11) Examples:

- Russian *djadja* ‘uncle’ is masculine although second declension nouns ending in *-a* are generally feminine (cf. Corbett 1982 and 1991).
- Norwegian *gubbe* ‘old man’ is masculine although nouns in *-e* tend to be feminine (Trosterud 2001).
- Arapesh *nakor* ‘husband’s father’ belongs to gender VII although nouns in /r/ belonging to declension 18 are normally in gender X (Fraser and Corbett 1997).³
- Old Norse *bruðr* ‘bride’ is feminine although nouns in /r/ are generally masculine (Trosterud 2003).
- Latvian *puika* ‘boy’ is masculine although declension four nouns in *-a* are normally masculine (Mathiassen 1997: 40).
- Lithuanian *sesuõ* ‘sister’ is feminine although declension five nouns in *-uõ* are normally masculine (Mathiassen 1996: 37).
- Lithuanian *de≥, de≥* ‘uncle’ is masculine although second declension nouns in *-a* or *-e≥* are normally feminine (Mathiassen 1996: 39).

There is solid typological evidence in favor of a privileged position of gender assignment rules based on biological sex. According to Dahl (2000: 101f.), who has investigated a large language sample including all languages discussed in Corbett (1991), sex is the “major criterion” for the assignment of gender in languages with more than one gender for animates. While Dahl’s term “major criterion” may seem opaque, it is clear from his discussion that it implies that sex-based gender assignment tends to take precedence. Notice that the provision “tends to” does not indicate that we are dealing with a mere statistical generalization. Rather, the set of cases where sex-based

³ Arapesh is a Torricelli language spoken on the north coast of Papua New Guinea. The gender system of Arapesh is discussed in Aronoff (1994), Fraser Corbett (1997), Corbett and Fraser (2000) and Dobrin (1997 and 1999). Data are from Fortune (1942/1977), but Dobrin has also carried out fieldwork on Papua New Guinea.

rules are overridden is limited and well defined. Dahl (2000: 103) isolates the following:⁴

- (12) a. Special morphological rules may take precedence for augmentative and diminutive derivations.
- b. Special semantic rules may take precedence for nouns denoting young or small animates.
- c. Special semantic rules may take precedence for certain kinds of animals.
- d. The “wrong” gender may be used in order to obtain special rhetorical effects (“downgrading” and “upgrading”).

German diminutives in *-chen* and *-lein* are well known examples of (12a). As an illustration of special treatment for nouns denoting young or small animates in (12b), Dahl (2000: 103) mentions the assignment of neuter gender to unmarried women in certain Polish dialects (see also Corbett 1991: 100). As for (12c), in the Australian language Ngangikurrunggurr nouns denoting animals hunted for meat are relegated to a special gender (Dahl 2000: 105). Finally, the special effects obtained by the use of *s/he* about inanimate objects and *it* about humans in American English serve to illustrate downgrading and upgrading in (12d) (Dahl 2000: 105). A detailed discussion of cases of these types is beyond the scope of the present study. Suffice it to say that Dahl’s typological evidence strongly suggests that sex-based rules take precedence universally in gender assignment, with the exception of the four well-defined cases in (12). For explicitness, I suggest formulating the following principle on the basis of Dahl’s evidence:⁵

- (13) The Core Semantic Override Principle:
Rules referring to biological sex take precedence in gender assignment.

I refer to (13) as the “Core Semantic Override Principle” because biological sex may be considered the semantic core of the category of gender.

4.2 *Is a Stronger Hypothesis Possible – Do Semantic Rules Take Precedence?*

Could the principle in (13) have been stated more inclusively so as to embrace *all* semantic rules, not only those involving biological sex? Corbett and Fraser have adopted this position:

⁴ Dahl also mentions arbitrary exceptions, but I have not included that in the list in (7) since we are interested in the systematic properties of gender systems.

⁵ Curt Rice (p.c.) suggests that exceptions of the type found in (12) can be ranked higher than the rules for biological males and females by the Elsewhere Condition. For instance, in order to account for examples like *das Weib* in German one might assume a rule Downgraded female → N. This rule will take precedence over Female → F by the Elsewhere Condition since “downgraded females” constitute a subset of females.

- (14) a. “If there are conflicting factors at work, semantic factors usually take precedence”. (Corbett 1991: 68f.)
- b. “As is universally the case, the formal gender assignment rules [...] are dominated by the semantic gender assignment rules.”
(Corbett and Fraser 2000a: 321)

This seems correct for languages like Russian and Arapesh discussed by Corbett and Fraser, since in these languages all the semantic rules refer to biological sex.⁶ In both languages there is a strong correlation between declension and gender. For most nouns in these languages the gender can be established on the basis of the noun’s membership in a certain declension class. The main exception is nouns denoting male or female beings, which are assigned gender according to biological sex even if this conflicts with the declension class. However, languages with a less strong correlation between declension and gender appear to be problematic for the claims in (14). Examples include Germanic languages like German (Köpcke and Zubin 1984, 1995 and references therein), Old Norse (Trosterud 2003) and Norwegian (Trosterud 2001, Enger 2002). Since in languages of this type the morphological rules cover a smaller portion of the vocabulary, researchers have postulated numerous semantic rules, not all of which refer to biological sex. For instance, Trosterud (2001) assumes 28 semantic rules for Norwegian. It seems fair to say that at present the interaction of semantic and other rules in complex systems of this type is not well understood. Trosterud explicitly avoids making strong claims about rule interaction on the grounds that the rules themselves are not sufficiently well understood.

As counterexamples to Corbett and Fraser’s position, let us, for instance, consider the interaction of the following three assignment rules for German (after Steinmetz 1986: 190), two of which have been discussed above:⁷

- (15) a. Superordinate nouns in German are neuter (e.g. *das Möbel* ‘furniture’)
- b. German nouns ending in *-e* are feminine (e.g. *die Treppe* ‘staircase’)
- c. German nouns in /uxt/ are feminine (e.g. *die Bucht* ‘bay’)

From Corbett and Fraser’s position we would expect the semantic rule (15a) to override the morphological (15b) and the phonological (15c). However, despite this superordinate nouns like *die Waffe* ‘weapon’ and *die Pflanze* ‘plant’ are feminine in accordance with (15b) and *die Frucht* ‘fruit’ in accordance with (15c). Clearly, we cannot draw strong conclusions on the basis of such examples without being sure that the rules in (15) are actually correctly stated. Furthermore, we cannot know whether the rules are correct before we have considered the German gender system in its entirety. Nevertheless, the German examples suggest that Corbett and Fraser’s proposal may be too strong, and that it may be wise to adopt a somewhat more cautious position. Until a

⁶ Russian has a rule whereby indeclinable nouns referring to animates are masculine. Thus nouns like *kenguru* ‘kangaroo’ are masculine (cf. Corbett 1991: 40). Notice, however, that this is not a purely semantic rule since it refers to the morphological property of indeclinability in addition to animacy.

⁷ Further counterexamples from various languages are discussed in Rice (2003).

fuller understanding of rule interaction in languages like German is arrived at, I suggest adopting the Core Semantic Override Principle in (13).

4.3 Theoretical Status

Even if we accept the Core Semantic Override Principle as a valid descriptive generalization, it does not follow that it should be granted the status of an independent principle in a general theory of gender assignment. If it can be shown that it follows from other, independently motivated principles of the theory, the Core Semantic Override Principle is nothing more than a descriptive generalization. Now, it seems quite clear that the Elsewhere Condition does not subsume the Core Semantic Override Principle. In the case of Russian *djadja* discussed in section 4.1, for instance, we have a conflict between a semantic rule invoking biological sex and a morphological rule referring to a declension class, and these rules are clearly not in a subset relation. Furthermore, the example of *djadja* indicates that Gender Tally does not make the Core Semantic Override Principle redundant, since we are dealing with a tie between two rules suggesting different genders.

However, an account of nouns like *djadja* in terms of Gender Tally and the Elsewhere Condition in conjunction with default hierarchies may be viable without reference to the Core Semantic Override Principle. The tableau in (16) illustrates this (cf. Rice 2003):

- (16) Assignment of gender to Russian *djadja* ‘uncle’
(tableau adapted from Rice 2003)

	Male=M	-a=F	BeMasc	BeFem	BeNeut
☞ <i>djadja</i> (masc.)		*		*	*
<i>djadja</i> (fem.)	*		*!		*
<i>djadja</i> (neut.)	*	*!	*	*!	

As can be seen from the tableau, the two competing rules assigning masculine gender to males and feminine gender to second declension nouns ending in *-a* suffice to rule out the neuter. Masculine gender is selected because the most highly ranked default rule militates against the feminine.

The success of this analysis hinges on the assumption that masculine outranks feminine in the default hierarchy of Russian. The question therefore arises as to whether there is any independent evidence in support of this ranking. Without such evidence, the analysis only states that the masculine takes precedence over the feminine, which is exactly what one wants to explain. Steinmetz (1986) does not discuss general criteria for establishing default hierarchies, but in Steinmetz (2002) he invokes statistics on type frequency in order to support the Russian default hierarchy. According to these data, the Russian masculine has slightly more members than the feminine: 21,516 masculines vs. 21,067 feminines. More data pointing in the same direction are given in Corbett and Fraser (2000b). They also assume the masculine to be the default gender for Russian nouns, although they do not invoke a hierarchy of default rules.

An evaluation of this argument for Russian will not be attempted here. However, on a more general level, it has interesting implications. It shows that if it is possible for all cases where sex-based rules take precedence to establish adequate default hierarchies

and come up with independent evidence in favor of them, the “Core Semantic Override Principle” might turn out to be an epiphenomenon. In the present context, the prediction would be that genders encompassing nouns denoting biological males or females universally have more members than competing genders. However, even if this prediction turned out to be true for all gender systems of the world’s languages, I find it somewhat hard to believe that the universally consistent override of sex-based rules in gender assignment conflicts is a mere coincidence conditioned by the relative sizes of genders. In any case, the onus of proof is on those who would want to argue this. Hence, at our present level of knowledge about gender assignment, it appears premature to conclude that default hierarchies offer a general solution to the problem of why sex-based rules take precedence in gender.

A typological argument also goes against a solution in terms of default hierarchies. It is not difficult to conceive of a language that is identical to Russian except that the feminine outranks the masculine in the default hierarchy. In such a quasi-Russian, *djadja* would belong to the feminine rather than the masculine by the logic of the default-based approach. A default-based analysis would even be compatible with a language where semantic rules were never decisive for the assignment of gender, since there is no principle ensuring that semantic rules take precedence. This is highly problematic because such languages are not attested:

- (17) “In a sense all gender systems are semantic in that there is always a semantic core to the assignment system” (Corbett 1991: 8 based on Aksenov 1984: 17f.).

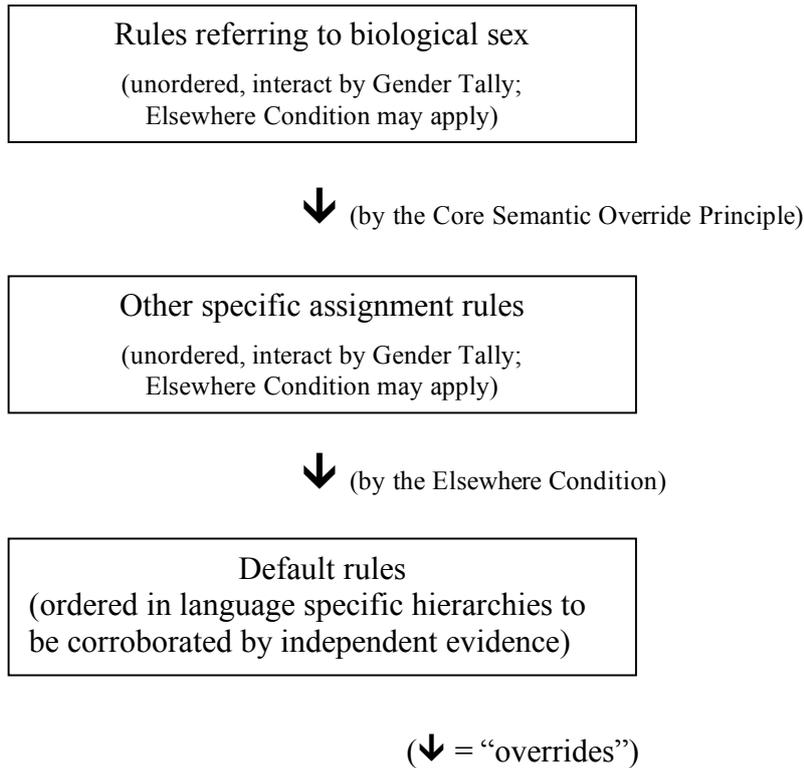
Thus, a pure default-based approach yields dubious typological predictions, in that it is compatible with unattested gender systems without a semantic core. In view of the evidence provided, I propose including the Core Semantic Override Principle in the general theory of gender assignment as an independent rule ordering device. We have seen in sections 1 through 3 that the so-called rule counting approach advocated by Steinmetz (1986, 2002) and Rice (2003) involves a certain amount of rule ordering. Supplementing it with the Core Semantic Override Principle moves the theory one step closer towards a rule ordering approach.

5. Conclusion

In this paper I have explored four general principles of a general theory of rule interaction in gender assignment: Gender Tally, the Elsewhere Condition, default hierarchies and the Core Semantic Override Principle. The contribution of the paper can be summed up in four points – one for each principle. First of all, I have suggested as a working hypothesis that assignment rules are not ordered unless universal principles force them to be so. As long as such principles do not apply, rule conflicts are resolved by Steinmetz’ Gender Tally, whereby a noun receives the gender indicated by the majority of the assignment rules. Secondly, we have seen that the Elsewhere Condition plays an important part in gender assignment. In part, Steinmetz’ framework falls out as an automatic consequence of the Elsewhere Condition, an observation that has not been made explicit in the literature. Thirdly, it has been suggested that Steinmetz’ notion of “default hierarchy” bears on gender assignment, although it has been pointed out that hierarchies postulated for individual languages must be corroborated by independent

evidence in order to be more than *ad hoc* solutions. Fourthly, I have argued that formal principles of rule ordering like the Elsewhere Condition must be supplemented by substantial ordering principles. I have suggested that Corbett’s hypothesis that semantic rules take precedence may be too strong. As an alternative, I have proposed the Core Semantic Override Principle. The impact of the four principles is summarized in (18) where the arrow represents override:

(18) Overview of the proposed model:



The present study does not offer a complete theory of rule interaction in gender assignment. The principles explored are likely to require revision, and further principles may have to be added. These qualifications notwithstanding, the principles I have explored in this study would seem to form a fruitful starting point for further investigation of the properties of the gender systems of the world’s languages.

References

i
 Aronoff, M. (1994), *Morphology by Itself*, Cambridge (Mass.), The MIT Press.
 Corbett, G.G. (1991), *Gender*, Cambridge, Cambridge University Press.
 Corbett, G.G. & N.M. Fraser (1997), *Defaults in Arapesh*, “Lingua”, 103, 25–57.
 Corbett, G.G. & N.M. Fraser (2000a), *Gender Assignment: A Typology and a Model*, in G. Senft (ed.), *Systems of Nominal Classification*, Cambridge, Cambridge University Press, 293–325.
 Corbett, G.G. & N.M. Fraser (2000b), *Default Genders*, in B. Unterbeck & M. Rissanen (eds.), *Gender in Grammar and Cognition*, Berlin & New York, Mouton de Gruyter, 55–97.

- Dahl, Ö. (2000), *Animacy and the Notion of Semantic Gender*, in B. Unterbeck & M. Rissanen (eds.), *Gender in Grammar and Cognition*, Berlin & New York, Mouton de Gruyter, 99–115.
- Dobrin, L.M. (1997), *The Morphosyntactic Reality of Phonological Form*, in G. Booij & J. van Marle (eds.) “Yearbook of Morphology” 1997, 59–81.
- Dobrin, L.M. (1999), *Phonological Form, Morphological Class, and Syntactic Gender: The Noun Class Systems of Papua New Guinea Arapeshan*, doctoral dissertation, University of Chicago.
- Doleschal, U. (2000), *Gender Assignment Revisited*, in B. Unterbeck & M. Rissanen (eds.), *Gender in Grammar and Cognition*, Berlin & New York, Mouton de Gruyter, 117–167.
- Enger, H.-O. (2002), *Stundom er ein sigar berre ein sigar*, “Maal og Minne” 2, 135–151.
- Fortune, R. (1942 [1977]), *Arapesh*, New York, J.J. Augustin Publisher. (Reprinted 1977 by AMS Press, New York)
- Fraser, N.M. & G.G. Corbett (1997), *Defaults in Arapesh*, “Lingua”, 103, 25–57.
- Kiparsky, P. (1982), *Explanation in Phonology*, Dordrecht, Foris.
- Köpcke, K.-M. & D. Zubin (1984), *Sechs Prinzipien für die Genuszuweisung im Deutschen: Ein Beitrag zur natürlichen Klassifikation*, “Linguistische Berichte” 93, 26–51.
- Köpcke, K.-M. & D. Zubin (1995), *Prinzipien für die Genuszuweisung im Deutschen*, in E. Lang & G. Zifonum (eds.), *Deutsch-typologisch*, Berlin & New York, Walter de Gruyter, 473–491.
- Koutsoudas, A., G. Sanders & C. Noll (1974), *The Application of Phonological Rules*, “Language” 50.1, 1–28.
- Mathiassen, T. (1996), *A Short Grammar of Lithuanian*, Columbus OH, Slavica Publishers.
- Mathiassen, T. (1997), *A Short Grammar of Latvian*, Columbus OH, Slavica Publishers.
- McClelland, J.L. & J.L. Elman (1986), *The TRACE Model of Speech Perception*, “Cognitive Psychology” 18, 1–86.
- Prince, A. & P. Smolensky (1993), *Optimality Theory. Constraint Interaction in Generative Grammar*, Report no. RuCCS-TR-2, New Brunswick, Rutgers University.
- Pugh, S.M. & I. Press (1999), *Ukrainian. A Comprehensive Grammar*, London & New York, Routledge.
- Rice, C. (2003), *Optimizing Gender*, ms., University of Tromsø.
- Steinmetz, D. (1986), *Two Principles and Some Rules for Gender in German: Inanimate Nouns*, “Word”, 37, 189–217.
- Steinmetz, D. (2002), *Gender Shifts in Germanic and Slavic*, paper presented at the conference “The Grammar of Gender”, Oslo, November 28–29 2002.
- Trosterud, T. (2001), *Genustilordning i norsk er regelstyrt*, “Norsk Lingvistisk Tidsskrift” 19, 29–57.
- Trosterud, T. (2003), *Gender Assignment in Old Norse*, ms., University of Tromsø.

The Morphological Typology of Change of State Event Encoding

Andrew Koontz-Garboden & Beth Levin

Stanford University

andrewkg@csl.stanford.edu , beth.levin@stanford.edu

Words denoting non-causative and causative change of state (COS) predicates often are morphologically related to words denoting the related state predicates, though the relationship sometimes differs for different types of states. For the state of ‘brokenness’, for example, in English the word denoting the state in (1c) is derived from the words denoting the change of state. In contrast, the word denoting the state of ‘looseness’ in (2c) is morphologically basic, with the words denoting the changes of state being derived from it.

- | | | | |
|-----|----|-------------------------|---------------------------------------|
| (1) | a. | The cup broke | (non-causative change of state) |
| | b. | Sandy broke the cup | (causative change of state) |
| | c. | The cup is broken | (state predicate is deverbal) |
| (2) | a. | The knot loosened | (non-causative change of state) |
| | b. | Sandy loosened the knot | (causative change of state) |
| | c. | The knot is loose | (state predicate is simple adjective) |

This paper reports on preliminary research aimed at clarifying the morphological and lexical semantic relationship between states such as those highlighted above and their causative and non-causative COS counterparts.

The morphological typology of words denoting non-causative (e.g. (1a), (2a)) and causative (e.g. (1b), (2b)) COS predicates has been relatively well studied (Nedjalkov 1969; Nedjalkov and Silnitsky 1973; Haspelmath 1993), with one important finding being that for certain types of COS events, languages tend to have morphologically simple words denoting the causative predicates, morphologically deriving the corresponding word denoting the non-causative COS predicate. For other types of events, the opposite direction of derivation is favored. This pattern of behavior is observed in Tongan (Polynesian), as shown in (3) and (4).

- | | | | |
|-----|-----------|---------------------|---------------------------------|
| (3) | Tongan | | |
| | pelu | ‘cause become bent’ | (causative change of state) |
| | ma-pelu | ‘become bent’ | (non-causative change of state) |
| (4) | Tongan | | |
| | lahi | ‘become big’ | (non-causative change of state) |
| | faka-lahi | ‘cause become big’ | (causative change of state) |

While certain types of events are lexicalized with the causative as the morphologically basic form, deriving the word denoting the non-causative change of state, as in (3) for the word for ‘bend’, other events have the non-causative change of state lexicalized as the morphologically basic form, deriving the word denoting the causative change of state as in (4) for the word for ‘big’. Haspelmath (1993) argues that the direction of

morphological derivation correlates with the likelihood that the event can occur spontaneously – events more likely to occur spontaneously are lexicalized in their morphologically basic form as words denoting non-causative COS predicates (e.g. *melt*), while those less likely to occur spontaneously are lexicalized in their morphologically basic form as words denoting causatives (e.g. *break*). The leading idea behind his research program is that the morphological direction of derivation, within and across languages, is suggestive of how non-causative and causative COS predicates are conceptually related to one another.

We take Nedjalkov and Silnitsky's and Haspelmath's ideas further by bringing states into the picture, examining how the non-causative and causative COS predicates are related to their associated states. Specifically, for a given state such as 'broken' or 'wide', there has been no systematic investigation of the morphological relationship between words denoting the state, a non-causative change into the state, and a causative change into the state. In this paper we take the first steps in such an investigation. We begin by laying out what we believe to be some of the more important questions in this domain. We follow this with discussion of some suggestive data culled from reference grammars and native speakers of relevant languages.

1. Three Questions about Change of State Encoding

1.1 *How Are Words Denoting States and Changes of State Morphologically Related to One Another?*

The question of how words denoting states are related to their non-causative and causative COS counterparts is prefigured in the work of Hale and Keyser (2002) and Baker (2003), whose theories predict a very specific type of relationship between states and their causative and non-causative COS counterparts. Namely, causative and non-causative COS predicates are predicted to be derived from their state counterparts.

Hale and Keyser, especially, give suggestive data supporting the idea that words denoting non-causative and causative COS predicates are morphologically derived from words denoting the corresponding state.

- (5) O'odham (Hale and Keyser 1998: 92, (31))
- a. (s-)moik 'be soft'
 - b. moik-a 'become soft'
 - c. moik-a-(ji)d 'cause to become soft'
- (6) Warlpiri (Hale and Keyser 1998: 92, (31))
- a. wiri 'be big'
 - b. wiri-jarri- 'become big'
 - c. wiri-ma- 'cause to become big'

In O'odham in (5) the word denoting the causative is derived from the word denoting the non-causative, which is in turn derived from the word denoting the state. In Warlpiri in (6), on the other hand, the words denoting the causative and the non-causative COS predicates are derived from the word denoting the state. In both cases the state is morphologically basic, an observation Hale and Keyser use to argue for the derivation

of the changes of state from the state itself. Though it is clear that this sort of relationship holds sometimes, work by Dixon (1982) makes us wonder whether it can be taken for granted that the relationships between states and changes of state are identically encoded for all types of languages and for all types of states.

1.2 Is the Relationship the Same for All Ontological Types of States?

In contrast to what is suggested by the theories of Hale and Keyser (2002) and Baker (2003), Dixon shows that "... certain states, naturally described by adjectives, contrast with states that are the result of some action" (1982: 50), for example, they differ in their morphological encoding. Dixon refers to the class of states naturally described by adjectives – in languages that have that lexical category – as *property concepts* (e.g. predicates denoting states related to speed, age, dimension, color, value, etc. and that presuppose no prior change). Contrasting with the class of property concepts is the class of states "that are the result of some action", *result states*, which are morphologically derived from verbs in many languages. This contrast shows up even in English, which otherwise does not have much verbal morphology.

- (7) English
a. The road is *wide*
b. The machine is *brok+en*

While the word denoting a property concept in (7a) is morphologically basic, the word denoting the result state in (7b) is morphologically derived from its corresponding change of state verb. Just as Hale and Keyser (1998: 100), Haspelmath (1993) and others argue that morphological makeup is an indication of semantic composition for non-causative and causative COS verbs in the causative alternation, so we believe that morphological makeup should be considered in understanding the semantic nature of states, and their relationship to related COS predicates.

1.3 What Effect Does a Language's Lexical Category Inventory Have on this Relationship?

An additional relevant question in this domain of study is what effect a language's lexical category inventory has on the relationship between words denoting states and words denoting their associated changes of state. It is well-known that not all languages have adjectives. Property concepts show up as nouns in some of these languages, and as verbs in others (Dixon 1982). Given that derivational morphology is often sensitive to lexical categoryhood, it seems quite possible that crosslinguistic variation in lexical category inventory might contribute to different types of relationships between words denoting states and their related changes of state. So far as we are aware, this is a question that has not been asked before.

2. Some Suggestive Data

Having now laid out several questions regarding the relationship between states and changes of state, we turn to some preliminary data suggesting answers and further areas

for research related to these questions. We begin by addressing the questions in §1.1 and §1.2 and then move on to the question raised in §1.3.

2.1 *Are All States Conceptually and Morphologically Basic?*

Data from a variety of languages, such as English and Quechua, suggest that in contrast to what is suggested by theories such as those of Hale and Keyser (2002) and Baker (2003), not all states are conceptually and morphologically basic. In the following sections we give data supporting this observation.

2.1.1 *English*

A major finding of Dixon's (1982) study is that the morphological complexity of a word denoting a state depends on the nature of the state: words denoting property concepts are morphologically simple in their stative denotation, while words denoting states that presuppose some change (i.e. result states) are often morphologically complex. The data in (8) and (9) illustrate this point.

Words whose denotation includes a property concept are morphologically basic in their stative denotation, as shown in (8) for *loose*, where the words denoting the changes of state are derived from the word denoting the property concept state with the *-en* suffix.

- (8) a. The knot is loose
b. The knot loosened
c. Kim loosened the knot

The same sort of relationship between states and changes of state holds for other adjectives in English, such as *bright*, *broad*, *cheap*, *coarse*, *damp*, *dark*, *deep*, *fat*, *flat*, *fresh* and others (Levin 1993). In other instances, the word denoting the change of state and the associated state are morphologically identical, but we assume that the COS predicates are again derived, as represented by the category change. We attribute the absence of the affix to a failure to meet the phonological conditions governing its appearance (Jespersen 1939).

This contrasts with the situation for words whose denotation includes a result state – for these types of words in English, the word denoting the state tends to be the one that forms English past participles, derived with the *-en* suffix (and its allomorphs) from the word denoting the changes of state, as shown in (9).

- (9) a. The machine is broken
b. The machine broke
c. John broke the machine

The same sort of relationship holds for other verbs denoting an action giving rise to a result state, such as *bend*, *crease*, *crinkle*, *crumple*, *fold*, *rumple*, *wrinkle*, *break*, *chip*, *crack*, *crash*, *crush*, *fracture*, *rip*, *shatter*, *smash*, *snap*, *splinter*, *split*, *tear*, and others (Levin 1993).

2.1.2 *Cuzco Quechua*

It is not only in English that this asymmetry between property concepts and result states is observed. In Quechua it is also the case that words whose denotation includes a property concept have a morphologically underived form that denotes a state. This is illustrated by the data in (10) from the Cuzco dialect.

- (10) a. wasi-qa *hatun-mi* (ka-sha-n)
 house-TOP big-EVIDENTIAL be-PROG-3P
 ‘The house is big’ (Martina Faller, p.c.)
- b. *hatun-ya-y*
 big-TRANSFORMATIVE-INF
 ‘become big’ (cf. Sp. *agrandarse*) (Cusihuaman 1976: 195)
- c. wasi-ta *hatun-ya-chi-rqa-n*
 house-ACC big-TRANSFORMATIVE-CAUS-PAST-3P
 ‘(s)he made the house big’ (Martina Faller, p.c.)

Other words denoting property concepts seem to behave similarly. According to Weber, describing the related Huallaga dialect, *-ya:* is an inchoative marker and “... seems to be completely productive ...” occurring with property concept words with meanings such as ‘big’, ‘crazy’, ‘white’, ‘rich’, ‘red’, ‘sick-ness/sick person’, ‘curly’, ‘hard’, ‘deaf’, etc. (Weber 1989: 30–31). Words denoting causative changes of state can then be derived from the *-ya:* marked non-causative changes of state with the *-chi* causative suffix (Weber (1989: 166), Cusihuaman (1976: 194), Martina Faller, p.c.); compare (10b) to (10c).

This direction of derivation from state to non-causative change of state to causative change of state contrasts with the direction of derivation for states that presuppose a change. In these cases, the word denoting the state is a participle derived from a verb (Weber 1989: 282–283; Cusihuaman 1976: 225). The word denoting the non-causative change of state, for its part, is derived from the word denoting the causative change of state with some some sort of reflexive marker. This is illustrated by the data in (11).

- (11) a. Tela *qhasu-sqa* ka-sha-n.
 cloth tear-PAST.PART be-PROG-3P
 ‘The cloth is torn’ (Martina Faller, p.c.)
- b. tela *qhasu-ku-n*.
 cloth tear-REFL.-3P
 ‘The shirt tore/got torN’ (Martina Faller, p.c.)
- c. tela-ta *qhasu-sha-n*.
 cloth-ACC tear-PROG-3P
 ‘She/he tore the shirt / She tears/is tearing the cloth’ (Martina Faller, p.c.)

In both English and Quechua, then, while the direction of derivation for words whose denotation includes a property concept meaning appears to be state to change of state,

the direction of derivation for words whose denotation includes a result state is the reverse – from change of state to (result) state.¹

2.2 *Which States Are Morphologically Derived, and Which Are Basic?*

Given the asymmetry observed above for both English and Quechua, one wonders if there is any sort of generalization holding across languages. These data, taken alongside Dixon's study of languages without adjectives, suggest that property concepts are denoted by morphologically simple words. They may be lexicalized as either stative verbs, nouns, or adjectives, depending on the language, but are morphologically simple words whatever their lexical category encoding. This generalization is stated in (12).

(12) Generalization 1:

If X is a property concept meaning, then the word Y denoting X is morphologically simple.

Given (12), if there is any overt derivational relationship between words denoting states, non-causative and causative changes of state, then, the words denoting the changes of state will be derived from the word denoting the state, as illustrated in (8) for English and in (10) for Quechua. The generalization also holds in other languages we have looked at.²

The lexicalization of result states and COS predicates related to them requires further research, as some languages such as Lakhota (Boas and Deloria 1939; Foley and Van Valin 1984) and Tagalog (Foley and Van Valin 1984) seem to lexicalize result states as morphologically simple forms, with words denoting the non-causative and causative changes of state built on top of them. What is noteworthy, though, is that in all languages we have examined, the paradigms involving result states are morphosyntactically distinct from those involving property concepts. For example, based on data in Boas and Deloria (1939), it seems that only roots with property concept meanings can be used without additional affixes in Lakhota, while roots with result state meanings must combine with certain affixes to be used with a stative meaning. Further, the two kinds of roots take different affixes to form non-causative and causative changes of state. Data like these and those discussed above support the idea that property concepts and result states are two fundamentally different types of states, down to the level of morphological encoding.

2.3 *What Is the Impact of Crosslinguistic Variation in Lexical Category Inventory?*

Dixon's observation regarding the diversity in lexical category encoding of property concepts was discussed above. This diversity turns out to have an interesting impact on the relationship of words denoting property concept states to words denoting their

¹ More research is needed on the possible morphological relations between words denoting causative and non-causative changes of state. Haspelmath's (1993) work on this question is suggestive, but unfortunately his survey preponderantly involves words whose denotation includes result states, so that it only presents a partial picture.

² This empirical generalization is predicted if the construction of word meaning is monotonic, as proposed e.g. in Olsen (1996) and Rappaport Hovav and Levin (1998). See Koontz-Garboden (2004) for discussion related to these facts specifically and for a proposal to derive (12) from monotonicity.

associated non-causative changes of state. We have observed two types of languages so far as this relationship is concerned. The more familiar kind of language is exemplified by O’odham, Spanish, and Warlpiri in (13)-(15). These are languages in which the word denoting the non-causative change of state is derived from the word denoting the property concept through some sort of morpholexical process overtly marked by morphology. In O’odham, as shown in (13), where property concepts are said to be lexicalized as adjectives, the addition of a suffix derives a non-causative change of state from the property concept state, and the causative change of state is, in turn, derived from the non-causative change of state. In Spanish, as shown in (14), where property concepts are also lexicalized as adjectives, this is done by some combination of prefixes and suffixes. Warlpiri, as shown in (15), where property concepts are lexicalized as nouns, derives words denoting non-causative changes of state from the word denoting the state with a suffix. Words denoting causative changes of state are also derived from the state-denoting word, but with a different suffix.

(13) O’odham (Hale and Keyser 1998: 92)

	<u>Adjective</u>	<u>Non-causative COS</u>	<u>Causative COS</u>	
a.	(s-)wegi	weg-i	weg-i-(ji)d	‘red’
b.	(s-)moik	moik-a	moik-a-(ji)d	‘soft’
c.	(s-)'oam	'oam-a	'oam-a-(ji)d	‘yellow’

(14) Spanish

	<u>Adjective</u>	<u>Non-causative COS</u>	<u>Causative COS</u>	
a.	rojo	en-roje-cer	en-roje-cer	‘red’
b.	duro	en-dure-cer se	en-dure-cer	‘hard’

(15) Warlpiri (Hale and Keyser 1998: 93)

	<u>Noun</u>	<u>Non-causative COS</u>	<u>Causative COS</u>	
a.	wiri	wiri-jarri-	wiri-ma-	‘big’
b.	maju	maju-jarri-	maju-ma-	‘bad’

This situation contrasts with that observed in certain other languages, such as Tongan (Polynesian). In this language property concepts are lexicalized as verbs and the same word is polysemous between a state and a non-causative COS denotation, as shown by the data in (16). Words denoting causative changes of state are derived from the state denoting words with a distinct morpheme, *faka-*, as shown in (16c).

(16) Tongan (Koontz-Garboden, field notes)

a.	Ko	e	hala	'oku	<i>lahi</i>	
	PRSTNL	the	road	PRES	wide	
	‘The road is wide’					
b.	Hili	pe	'uluaki	fo'i'akau',	kuo	<i>lahi</i> ia.
	after	only	first	medicine,	PERF	big him
	‘After only one pill, he became big’					

- c. Na'e *faka-lahi* e he puleanga 'a e hala
 PAST CAUSE-wide ERG the government ABS the road
 'The government widened the road'

Though there is no derivational morpheme signaling the difference between the state and the non-causative COS denotation in (16a,b) above, there is a difference in aspect marking – while the use of the continuous marker *'oku* correlates with an ongoing state denotation, use of the perfect marker *kuo* correlates with a non-causative COS denotation.³ This polysemy is not unusual as it has been observed in the literature on the typology of aspect marking that perfective marking of a stative verb often yields a change of state interpretation (Comrie 1976: 19–20; Bybee *et al.* 1994: 75–76; Chung and Timberlake 1985: 217). Further, similar facts have been observed for other languages in which property concepts are lexicalized as verbs, such as Fongbe (Lefebvre and Brousseau 2002: 88), Thai (Prasithratsint 2000: 262), Lao (Enfield 2003: 6), Mokilese (Chung and Timberlake 1985: 238), and Mandarin, as illustrated in (17), for example.

- (17) Mandarin
 a. tā gāo
 'he is tall'
 b. tā gāo-le
 (Pfv. [perfective]) 'he became tall, has become tall'
 (Comrie 1976: 19–20)

It seems that this sort of polysemy arises only in languages where property concepts are lexicalized as verbs; in languages where they are lexicalized as nouns or adjectives, we observe no such polysemy. This leads us to a second generalization, stated as in (18).

- (18) Generalization 2:
 Only in languages where property concepts are lexicalized as verbs can a single word be polysemous, denoting a property concept state and its associated non-causative change of state.

The typological generalization, then, is that there seem to be two types of languages as far as the derivation of non-causative changes of state from property concept states is concerned, and that the type of derivation a language uses is in part correlated with how it lexicalizes property concept notions. There can only be polysemy to derive non-causative changes of state from states where the latter are lexicalized as verbs. The explanation for this lies in the nature of the mapping between lexical semantics and morphosyntax. While words of many different lexical categories can denote states, only verbs can denote changes of state. Because of this, the same word

³ Here we are actually simplifying significantly due to space considerations. It is actually the case that a COS meaning can arise with *'oku* marked states in the presence of an adverb requiring such a meaning, though the default interpretation of *'oku* constructions is a stative one. This suggests that what determines whether a property concept word has a state or a COS reading goes beyond grammatical aspect marking. Which reading arises depends on the sentential context, which can lead to the coercion of one meaning or another (Zucchi 1998). These issues are discussed extensively in Koontz-Garboden (2004).

can denote both states and changes of state only in a language where states are lexicalized as verbs. In languages where property concepts are lexicalized as nouns or as adjectives, these cannot be polysemous between a state and change of state reading, since only verbs can denote meanings of the latter type.⁴ The facts we have seen above support this claim. Indeed, Spanish, Warlpiri, and O’odham are all languages where property concepts are said to be lexicalized as either nouns (Warlpiri) or as adjectives (Spanish and O’odham). In this way, these languages contrast with Tongan and Mandarin, where property concepts are said to be lexicalized as verbs.

3. Conclusion

Though the research we have reported is still in its preliminary stages, several important empirical generalizations have already emerged. First, we find that property concepts and result states are lexicalized as words with different morphological makeups. While property concepts are lexicalized as morphologically simple words, this is not always the case for result states. Secondly, we find that some languages fail to have morpholexical non-causative changes of state derived from the associated property concept state. Rather than having a morpholexical derivation of a change of state from a property concept state, in these languages, one morphologically simple word is polysemous between a property concept state and a non-causative COS meaning. We find, then, that there exist two types of languages – those with non-causative changes of state derived morpholexically and those with polysemy. Due to a constraint on the mapping between lexical semantic and morphosyntactic categories that only verbs can denote changes of state, polysemy arises only in languages where property concepts are lexicalized as verbs.

From a theoretical perspective, we believe that our observations suggest that theories of event structure that give homogeneous representations to all COS predicates (e.g. Hale & Keyser 2002; Baker 2003) need to be revisited. There seems to be a contrast in the behavior of property concept states and result states, and in how non-causative changes of state are derived from property concepts, depending on other morphosyntactic properties of different languages. Theories of event structure should capture these asymmetries.

Acknowledgements

We would like to thank the participants of the Fourth Mediterranean Morphology Meeting for their questions and comments. We are also grateful to Sisilia Lutui for help with Tongan and to Martina Faller for help with Quechua. This research was supported in part by Graduate Research Funds from the Department of Linguistics at Stanford to Koontz-Garboden and by NSF Grant BCS-0004437 to Levin.

⁴ This claim is fleshed out in Koontz-Garboden (2004), where it is shown that when properly formulated, potential counterexamples such as *birth*, *conception*, etc. actually support the theory.

References

- Baker, M.C. (2003), *Lexical categories: Verbs, nouns, and adjectives*, Cambridge, Cambridge University Press.
- Boas, F., & E. Deloria (1939), *Dakota grammar*, *Memoirs of the National Academy of Sciences*, volume 23, number 2.
- Bybee, J., R. Perkins & W. Pagliuca (1994), *The evolution of grammar: Tense, aspect, and modality in the languages of the world*, Chicago, University of Chicago Press.
- Chung, S. & A. Timberlake (1985), *Tense, aspect, and mood*, in T. Shopen (ed.), *Language typology and syntactic description*, vol. 3, Cambridge, Cambridge University Press, 202–258.
- Comrie, B. (1976), *Aspect*, Cambridge, Cambridge University Press.
- Cusihuaman, A. (1976), *Gramática Quechua: Cuzco-Collao*, Lima, Perú, Ministerio de Educación, Instituto de Estudios Peruanos.
- Dixon, R.M.W. (1982), *Where have all the adjectives gone? and other essays in semantics and syntax*, The Hague, Mouton.
- Enfield, N.J. (2003), *Adjectives in Lao*, in R.M.W. Dixon & A.Y. Aikhenvald (ed.), *Adjective classes: a cross-linguistic typology*, Oxford, Oxford University Press.
- Foley, W.A. & R.D. van Valin (1984), *Functional syntax and universal grammar*, Cambridge, Cambridge University Press.
- Hale, K.L. & S.J. Keyser (1998), *The basic elements of argument structure*, in *Papers from the UPenn/MIT roundtable on argument structure and aspect*, volume 32, MITWPL, 73–118.
- Hale, K.L. & S.J. Keyser (2002), *Prolegomenon to a theory of argument structure*, Cambridge (Mass.), MIT Press.
- Haspelmath, M. (1993), *More on the typology of inchoative/causative verb alternations*, in B. Comrie & M. Polinsky (eds.), *Causatives and transitivity*, Amsterdam, John Benjamins, 87–120.
- Jespersen, O. (1939), *The history of a suffix*, “*International Journal of Structural Linguistics*” 1, 48–56.
- Koontz-Garboden, A. (2004), *Tongan and the typology of change of state predicates*, Stanford University, ms. Available on WWW: <http://www-csli.stanford.edu/~andrewkg/>.
- Lefebvre, C. & A.-M. Brousseau (2002), *A grammar of Fongbe*, Berlin, Mouton de Gruyter.
- Levin, B. (1993), *English verb classes and alternations: a preliminary investigation*, Chicago, University of Chicago Press.
- Nedjalkov, V.P. (1969), *Nekotorye verojatnostnye universalii v glagol'nom slovoobrazovanii*, in I.F. Vardul (ed.), *Jazykovye universalii i lingvisticheskaja tipologija*, Moscow, Nauka, 106–114.
- Nedjalkov, V.P. & G.G. Silnitsky (1973), *The typology of morphological and lexical causatives*, in F. Kiefer (ed.), *Trends in Soviet theoretical linguistics*, Dordrecht, D. Reidel Publishing, 1–32.
- Olsen, M.B. (1996), *A semantic and pragmatic model of lexical and grammatical aspect*, Northwestern University dissertation, published 1997 by Garland.
- Prasithrathsint, A. (2000), *Adjectives as verbs in Thai*, “*Linguistic Typology*” 4, 251–271.
- Rappaport Hovav, M. & B. Levin (1998), *Building verb meanings*, in M. Butt & W. Geuder (eds.), *The projection of arguments: Lexical and compositional factors*, Stanford, CSLI Publications, 97–134.
- Weber, D. (1989), *A grammar of Huallaga (Huánuco) Quechua*, Berkeley, University of California Press.
- Zucchi, S. (1998), *Aspect shift*, in S. Rothstein (ed.), *Events and grammar*, Dordrecht, Kluwer, 349–370.

Morphemes and Lexemes versus “Morphemes or Lexemes ?”

François Nemo
Univ. Orleans
fnemo@univ-orleans.fr

More than a century after the first linguistic definition of the notion of morpheme by Baudoin de Courtenay (1895) and Sweet (1876), an ever-lasting debate – which I shall refer here as the “Morpheme or Lexeme” (M or L) debate – on the nature of linguistic bricks is still going on.

Since a by-product of this debate is terminological confusion in the use of the four notions of *morpheme*, *lexeme*, *word* and *item*, I shall start by describing very briefly the conflicting uses of these terms and then show that to accounting for the generation of the lexicon, i.e. of both new senses and new lexical units, requires accepting the co-existence of two distinct semantic stocks and explaining how morphemes which belong to the first stock may become lexemes or be involved in lexeme-formation processes.

1. Uses of the Notions of *Morpheme*, *Word*, *Lexeme* and *Item*

1.1 Uses of the Notion of “*Morpheme*”

There are basically four uses of the noun *morpheme*:

- *morpheme* may be used, following Baudoin de Courtenay (1895), in order to refer to the smallest meaningful linguistic unit, a minimal sign identified as a semantic atom through a process of decomposition, regardless of its syntactic autonomy. Within such a methodology, the definition includes both a unit like the English *milk*, which cannot be decomposed in smaller elements and is syntactically autonomous, and a unit like the French *-spir-* which is the result of the decomposition of the words *re-spir-er* (to breathe), *in-spir-er* (to breathe in), *ex-spir-er* (to breathe out) but is not syntactically autonomous;
- *morpheme* is commonly used to refer to infra-lexical semantic units, typically affixes, which are not syntactically autonomous and hence are not words. Following Corbin, I shall refer to such bound morphemes as *infra-lexical units*, leaving out of this category bound morphemes which are used for flexion, since such bound morphemes do not belong to the lexicon.
- *morpheme* may be used to refer to grammatical bound morphemes only. In such a case, the notion of morpheme is associated with the notion of flexion and therefore morphemes are not lexical units at all;
- *morpheme* has been used in contemporary linguistic semantics to refer to a form/signification pair which can be isolated by considering all the uses of a single semantic unit, including categorically distinct ones. For instance, semanticists will

speak of a single morpheme *but* in English, encoding stable semantic indications, so as to account for all the uses of *but* in English, for instance its uses with the non connective meanings of *almost*, *without*, *except*, *only* and with the connective meanings of *but*, *rather*, etc. . This semantic definition differs from the classical one by its using a distributional methodology – i.e. considering all the uses of a given unit, so as to isolate encoded meaning – instead of the decompositional methodology advocated by the structuralists.

The aim of this paper will be to show that this last definition is of crucial importance for any morphological theory.

1.2 *Uses of the Notion of “Word”*

I shall not detail here all the conflicting definitions of the notion of *word* which have been proposed so far (for an overview of the problem, see Di Sciullo & Williams 1987; Dixon & Aikhenvald 2002), and I shall limit myself to the few issues directly at stake in the “M or L” debate:

- it has been repeatedly asserted that *words* and not *morphemes* are the minimal signs of a language (Aronoff 1976; Anderson 1992). The usual justification of such a view is that all the combinatory rules, whether grammatical or morphological, are word-based and not morpheme-based and that ordinary speakers have a semantic intuition about the meaning of words, but no intuition at all about the meaning of morphemes;
- it has been repeatedly asserted that infra-lexical units such as affixes have no stable meaning (Aronoff 1976) and hence are not signs in the Saussurean sense, but processes.
- it has been assumed (Di Sciullo & Williams 1987) that there are two class of words, listemes on the one hand (which have to be learnt one by one and are either semantic atoms or unpredictable complex units) and generated words on the other hand (which are the outputs of regular word-formation processes);
- it has been argued within a constructional approach to morphology that such infra-lexical units possess a meaning indeed, but that this meaning is instructional/procedural and not conceptual;
- it has been repeatedly asserted in linguistic semantics that data-based observation of the uses of “words” shows all too clearly that the actual uses of words are semantically distinct from our intuition about their meanings (e.g. that possibly 90% of the lexicalised uses of the French verb *balayer* are not predictable from intuition), and that doing semantics implies forbidding the use of intuition, adopting a clear distinction between *signification* and *sense* (Benveniste 1954; Ducrot 1987) and admitting that only form/signification pairs are signs in the Saussurean sense, and that the form/sense pairs provided by ordinary dictionaries are not linguistic signs but only local interpretations of these signs and of the constructions in which they are inserted;

1.3 Uses of the Notion of “Lexeme”

The notion of *lexeme* is frequently used as a technical and less ambiguous equivalent of the notion of *word*. It has the advantage of allowing the integration in the lexicon of many lexical units which are not *words* in the ordinary sense (for instance because they are formed of smaller units which are also *words*). *Lexemes* may thus be a cover term so as to refer to all the semantic units stored in the lexicon, and may thus refer to lexical units (e.g. *milk*), infra-lexical units (e.g. units as *micro-*, affixes – if we admit they have a signification – and bound bases) and supra-lexical units which behave like syntactic atoms (e.g. *pomme de terre – potato*, literally *apple of ground* – or phrases like *tout à fait*, lexicalised idioms).

1.4 Uses of the Notion of “Item”

Given the “L or M” debate, and the typological differences between agglutinative and polysynthetic languages and isolating ones for instance, the term *item* has often been used to avoid the more controversial terms *morpheme* or *word*. *Item* is thus compatible with both a concatenative and a non-concatenative view of morphology, and also with a sign-based and process-based view of morphology. Semanticists do not use it at all.

2. The “Morpheme or Lexeme” Debate.

From structural linguistics to contemporary morphology, it has been widely assumed that a choice had to be made between morpheme-based models and lexeme – (i.e. words) based ones. In order to understand why we should rather consider morphemes and lexemes as two kinds of linguistic and semantic units which co-exist (and are complementary) and hence shouldn't be opposed, what must be remarked is that according to the classical definition within structural linguistics morphemes were simultaneously: i) the basic **semantic units** of a language; ii) the basic **combinatorial units** of a language. Within such a view the basic **semantic units** and the basic **combinatorial units** of a language were assumed to be the same thing. For instance the unit *table* is considered at the same time as an atomic semantic unit and a noun, i.e. as a syntactically defined unit.

The problem with such a view is that it leaves no choice but to **list** a huge part of the lexicon (e.g. words like *rétablir*, *tabler*, *tableur*) and to postulate endless sense enumerative lexicons (with as many entries for *table* as senses that the noun may have in its different uses, and with as many different units *table* as needed to explain the existence of bound bases in words like *se rétablir*, *rétablir*, *tableur*¹).

Fortunately, since accounting for the diversity of uses (and hence senses) of a semantic unit is precisely the aim of semantics – according to the linguistic semantics framework shared by a large part of contemporary semantics (Benveniste 1954; Ducrot

¹ One needs to add to the questions of polysemy, “semantic drift” or “semantic bleaching” the question of polycategoriality: since a unit like *timap* in Palikur (Arawakan) “means” simultaneously *to hear*, *to shout*, *echo*, *loudly*, etc., depending on the way it is used, we can either adopt mere degrouping homonymy and have as many lexical units *timap* as there are ways to use it, or refuse this “solution” (Pustejovsky 1995) and admit that a distinction must be made between a non-categorial semantic unit *timap* and the categorial (and contextual) interpretations it receives in each of its uses.

1987, Bouchard 1995, Pustejovsky 1995, Nemo 1998, 1999, 2002a, to appear) – recent developments have enabled a clear understanding of the nature of the problem and, more importantly, of the nature of its solution, advocating for the co-existence of morphemes and lexemes in a language.

2.1 The Signification / Sense Distinction.

As mentioned earlier, the distinction between (lexical) senses and (encoded) signification is the founding postulate of all Linguistic Semantics models and descriptions. It is associated with the ideas that:

- (encoded) linguistic signification is accessible to the linguist only by considering the variety of uses of a semantic item;
- (encoded) linguistic signification must explain these uses just as rules explain sentences;
- (encoded) linguistic signification is usually not intuitive.
- significations are to senses what the equations of functions are to the points created by these equations;
- significations and senses are different in nature (Ducrot 1987; Nemo 2001c) since signification is neither some kind of a very abstract sense, nor the common denominator of these senses ;

2.2 The Lexicon as a Memory of Interpretations

Within such a view, which has proved to be extremely efficient to account for polysemy and polycategoriality, our understanding of the nature of the lexicon itself is deeply transformed. For most linguists outside of semantics, the received idea has long been that languages were formed of a lexical stock of combinatorial units on the one hand and of a set of combinatorial rules on the other, and that sentences (or words) are the outputs of a generative process whose inputs are the lexicon and the combinatorial rules.

Within linguistic semantics on the contrary, it is more and more widely acknowledged that the role of semantics is to account both for the generation of new meanings (and thus for polysemy) and for the generation of new lexical units (and thus of the lexicon itself). Within such a view, it is not legitimate to take the existence of the lexicon for granted, and the lexicon itself is what has to be explained and is therefore the **output** of a process which has to be described and whose inputs, as we shall see, are morphemes and constructions.

Understanding the co-existence of morphemes and lexemes in that perspective requires only to understand that morphemes, which encode significations, are the inputs of a process in which:

- each time a morpheme is used, it is inserted in a construction and in a context and it receives a constructional and contextual local interpretation;

- if the same use is repeated, i.e. if the same morpheme is used in the same construction and the same context, the interpretation process is not repeated and the interpretation becomes memorised, a process called conventionalisation or lexicalisation.

Thus, within linguistic semantics, the lexicon is only a memory of the interpretations of morphemes in their different uses. As a result, lexemes are not the basic semantic units and all languages have two stocks of semantic units, a stock of linguistic units on the one hand (that we shall call morphemes from now on) and a stock of lexical units on the other hand (called lexemes and including lexical, supra-lexical and infra-lexical units).

And finally the whole picture consists in a triple (and parallel) distinction between: i) signification and sense; ii) morphemes which encode signification and lexemes which have senses; iii) the linguistic stock consisting of morphemes and the lexical stock consisting of lexemes.

2.3 The Signification / Sense Distinction and Morphology

As for morphology, the distinction between signification and senses has far-reaching consequences:

- the notorious instability of the form/sense relationship, which has led many linguists to consider meaning as irrelevant for morphological theory (Aronoff 1976) has misled them about the semantics of morphemes. Morphemes have a very stable meaning in all their uses, and are also very stable in diachrony (i.e. much more stable than grammatical or morphological structures);
- the importance of listemes in the lexicon, possibly 40% of the word-forms found in corpus-based studies and probably up to 80–90 % of the senses of apparently well-formed words, can be accounted for only by using the morpheme/lexeme distinction.

2.3.1 Accounting for Semantic Instability

Within contemporary linguistic semantics, i.e. by adopting the signification/sense distinction, it has become possible to show that the diversity of uses of a semantic unit was compatible with the fact that this unit encodes a very stable signification.

So that for instance, the English semantic unit *but* does not encode a connective and pragmatic sense (whose equivalents would be the German *aber* or *sondern*, the Spanish *pero* or *sino*, or the French *mais*) on the one hand and have unpredictable non connective and non pragmatic uses on the other (with the meanings of *almost*, *without*, *only*, *except*, etc.). Instead, what we have is:

- a single morpheme *but*, encoding the indication that “something had (could have, should have, etc.) been stopped”, and which may be inserted in different constructions and positions (for instance in connective and non-connective positions) where it receives a local (constructional and contextual) interpretation;

- different lexemes *but* (with the lexicalised meanings of *aber*, *sondern*, *almost*, *without*, *only*, *except*, etc.) with their own polysemy.

Within such a view (Nemo 2002a, to appear), it becomes indeed possible to account for the various interpretations of *but* in the three following utterances: (1) *The price is interesting but I have no money*; (2) *But for Peter, I would be dead*; (3) *This specie has but disappeared*; since despite constructional differences, it describes in the three cases the fact that a process is not completed, with “having no money” being the blocking factor in (1), Peter's intervention being the blocking factor in (2), and the not fully completed disappearance accounting for the “almost” interpretation in (3).

Within such a view: i) morphemes are semantic units (i.e. form meaning pairs) but not syntactic units (they provide no combinatory information); ii) lexemes are syntactic-semantic units; iii) morphemic meaning can be identified only by taking into consideration all the uses of the morpheme regardless of its syntactic status; iv) morphemic meaning is indicational and lexical meaning is a conventionalization of a morphemic/constructional/contextual complex; v) morphemic meaning is encoded, lexical meaning is memorized; vi) lexical meaning is not the starting point of semantic analysis but an intermediate level.

A semantic account which can be formalised (Gasiglia, Nemo & Cadiot 2001) by saying that the senses *s* of both lexemes and non memorised uses are only **the results** of a function *f*, which may be described as:

$$f(\text{morpheme, construction, context}) = s$$

having as a result the necessity for the linguist to admit the existence of three semantic stocks, namely:

- morphemes, which are form/signification pairs that exist independently of the construction and context in which they are inserted;
- constructions which are form/interpretation pairs that exist independently of the morphemes used (Goldberg 1995);
- lexemes which minimally are morpheme/construction pairs and are associated with lexicalised meanings (i.e. senses),

and to avoid any methodology taking for granted that lexemes are the inputs of linguistic processes. Otherwise, as we shall see now, one has no other choice but to **list** whatever is not predictable from these units and their intuitive senses, i.e. to list most of the lexicon.

2.3.2 *The Origin of Listemes*

Within contemporary morphology, words are indeed to be listed if they are *somehow* irregular, i.e. whenever:

- the “input” is problematic;
- there are no rules to account for the observed pattern;

- the semantic output is unpredictable from the semantic input.

As a result, most of the lexicon has to be listed, even though:

- whenever the “input” is problematic it may be observed that i) the frequency of problematic inputs may be as high (see French *en-*) as to be considered “regular”; ii) problematic inputs often produce unproblematic “outputs”, with meanings very similar to those of non-listed words (e.g. *s’enticher* and *s’enamourer*); iii) bound bases are simply not studied before listing is decided (e.g. *re-tali-ation*).
- whenever there are no rules allowing prediction, the existence of categorially problematic listemes is as high as to be considered “regular” (e.g. *buteur*, *footballeur*), the listed “outputs” often share the same global meaning as non-listed one (*pétrolier*, *chimiquier*), the possibility of exocentric derivation is not tested.
- whenever *semantic drift* or semantic incoherence is postulated, its existence is based on the hypothesis of the existence of a primary meaning directly accessible through intuition and familiarity, no study of polysemy is ever made and the “input”/“output” semantic relationship is believed to be stable.

On the contrary, within the Linguistic Semantic framework presented here, instead of the classical view according to which lexemes are the inputs of morphology:

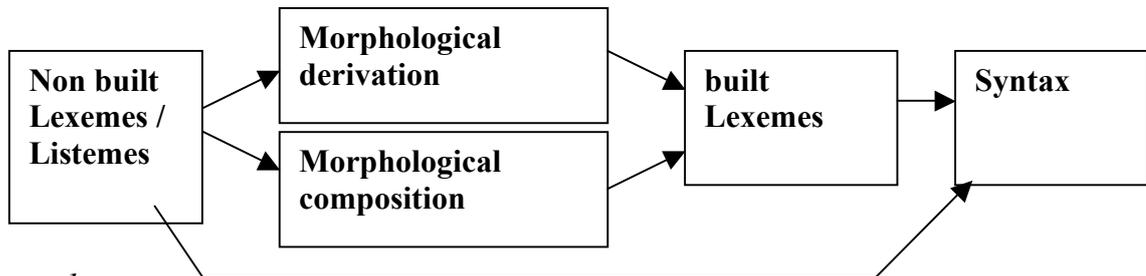


Figure 1

what we have is a morpheme/lexeme/construction distinction:

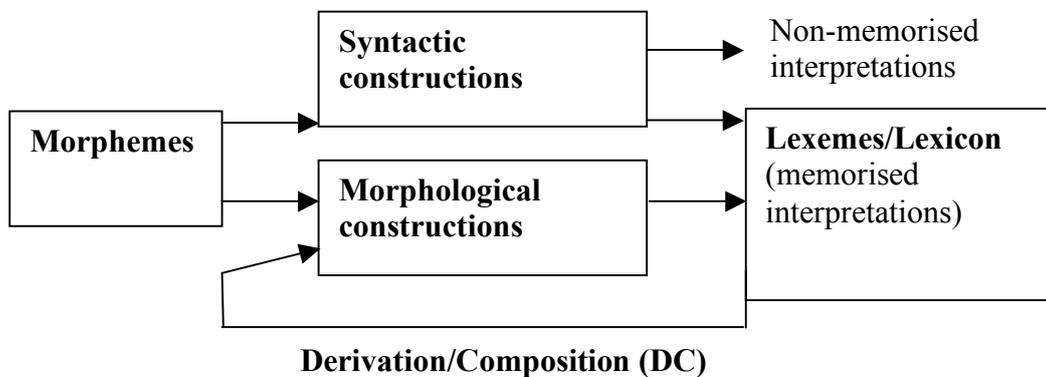


Figure 2

according to which morphemes and constructions are the inputs of the process which must be described and the lexicon its output.

2.3.2 *Listing and Falsification*

The main difference between the two models contrasted here consists in the predictions they make about the existence or inexistence of listemes: since listemes (which are not basic words) are those lexical units whose existence or meaning cannot be predicted as a result of lexeme-based word-formation process, it is easy to see that what figure 1 describes is in fact the DC re-entering arrow of figure 2, and that whatever is directly generated by inserting morphemes into constructions will have to be listed.

One of the main problems associated with the view presented in figure 1 is that even though much time has been dedicated to a theorisation of what listemes are (Di Sciullo & Williams 1987), no quantification of listing has ever been made on real data. So that even if criteria have been defined so as to decide whether a given word (or meaning of a word) has to be listed, the fact that using these criteria often leads to the listing of a large number of word-forms and most word meanings is not acknowledged at all. Tenants of the classical view never indeed acknowledge (or simply mention) that applying the criteria defined in 2.3.1. to the diversity of uses of morphemes like *table* or *coll* leads to the listing of most words-forms in which these morphemes are inserted – namely words such as *collecte*, *collection*, *collision*, *collusion*, *accolade* – and also to the listing of a large part of the meanings of the remaining word-forms, such as the *taking off* sense of *décoller* for a plane, . It should on the contrary be remarked that given the semantic methodology adopted by tenants of the classical view, which combines intuition and prototypicality to isolate the supposedly basic/true meaning of a word, almost all the lexicon should be listed, since the average frequency of these intuitive meanings is never higher than 10–20 % of the uses of a word (see Gasiglia, Nemo & Cadiot for an illustration of this about the French word *balayer*).

Tenants of figure 2 on the other hand, who are today's leading semanticians, do not accept the idea of such a global frequency (80–90%) of “semantic drift”, “semantic bleaching” and “idiosyncrasy”. They also refuse the underlying methodology and the consequence it has on our understanding of the generation of meanings. For instance, if we consider the description of the meaning of *but* proposed by B. Fraser (1998) as an illustration of the shortcomings of such an intuitive and unexplanatory methodology, it is clear that the linguist has no choice but to pick out a prototypical meaning (on a “trust me” basis), to declare it basic (“*the core meaning of but is to signal simple contrast, nothing more, and the speaker will select it when intending to highlight a contrast*”), to declare this description unfalsifiable even when it is directly falsified (e.g. saying that “*even if one cannot find two specific areas of contrast between the direct S2 and S1 messages, the messages may nevertheless be contrasted in one of several ways*” in order to account for uncontrastive uses such as “*Paul is brilliant but so is John*”), then to add that defining definition is impossible (“*I can offer no precise definition of what qualifies as a Contrastive Discourse Marker*) and finally, concerning other uses of *but* (i.e. concerning the so-called semantic drift), to declare that “*I am not treating other uses of but such as found in: « All but one left today », « There was no doubt but that he won », « it has not sooner started but it shopped », « He was but a poor man », « I may be wrong but I think you are beautiful ».* *Whether or not they could be included under my analysis is left open*”. The important point is of course to understand that what Fraser is saying here about *but*, whose morphemic account was presented above, is due to the

same attitude adopted by many morphologists (especially in the generative framework), according to which: i) semantics should be intuitive and not explanatory; ii) semantics should not take more time than five minutes; iii) whatever might require more than 5 minutes should be left over and forgotten.

It is indeed quite clear that the same thing holds in morphology when we have to account for words like *décoller* (to take off, for a plane), *collecte*, *collection*, *collision*, *collusion*, *accolade*: the fact that only a minority of these words and/or a minority of the uses of these words are predictable from the ordinary meanings of the words *colle* (glue) or *coller* (stick) is not a trace of the fact that “*the lexicon is like a prison – it contains only the lawless, and the only thing that its inmates have in common is lawlessness*” nor of the fact that since it “*is simply a collection of the lawless, there neither can nor should be a theory directly about it*” (Di Sciullo & Williams 1987), but a trace of the fact **that the ordinary meanings** of the words *colle* and *coller* are **only local interpretations of the indications encoded by the morpheme [coll]**, interpretations which, since they are not coded, are completely unable to block the generation of new interpretations in new contexts and the generation of new lexemes in new constructional positions. Listing all these uses and words in the lexicon because they cannot be predicted from these local interpretations is thus the equivalent of describing the lexicon as an endless sense enumerative lexicon criticised by Bouchard (1995) and Pustejovsky (1995), a list of leaves unrelated one to another by any branch or tree.

Thus, adopting the “Morpheme and Lexeme Hypothesis” is a way to avoid adopting the Generative Morphology’s hypothesis of an “Ungenerated Lexicon”:

- if the only thing morphology can say about words such as *tabler* (to bank on), *rétablir* (to restore, to re-establish, to reinstate), *se rétablir* (to recover, to return), *tableur* (spreadsheet) *tableau* (board, chart, table, instrument panel, dashboard), or idioms like *se mettre à table* (to tell everything), *dresser un tableau de la situation* (to paint the picture of the situation) – which are not compositionally predictable from the meaning of the lexeme *table* (as a *piece of furniture*) they seem to include – is that these words (and/or meanings) are listed, unrelated and have to be learnt one by one;
- if the only thing morphology can tell us about all the uses of the noun *table* (dining table, changing table, arithmetic charts, book contents, editing bench, etc.) is to describe them in terms of semantic drift, semantic bleaching or homonymic degrouping;
- if the only words predictable from the DC arrow are the words *tablée* and *s'attabler*, whose interpretation clearly presuppose the lexicalised meaning of *table* as a *piece of furniture*.

then it would have to be acknowledged that morphology and common sense have exactly the same (un)explanatory power, morphology being unable to account for anything more than what immediate intuition would.

Consequently, it seems clear that instead of assuming that “*there neither can nor should be a theory*” about listemes, linguists should understand that (most) listemes are a direct **empirical falsification** of the classical view, and that they should be considered

as such. It may be the case that listemes are the nightmare of combinatorial/categorical morphology, but it may also be the case that listemes are in fact an open window into the reality of word-formation processes:

Limiting morphology to the DC arrow implies that describing the generation of the French noun *re-spir-ation* from the verb *re-spir-er* is possible, but that accounting for *re-spir-er* itself is impossible. It also implies that it is impossible:

- to account for the production of words/lexemes like French *rot-ation*, or *obstruction* which are listemes since their input is problematic (*roter* is not a French verb, *obst* has no syntactic autonomy);
- to account for words/lexemes like French *dé-coll-er* (to take off, for a plane) which are also listemes since their semantic output is not predictable from the sense of an existing lexeme,
- to account for the existence of listemes such as *re-cycl-er*, *but-eur* or *chimiqu-ier* which cannot be predicted by any combinatorial WFR but are the result of the general existence in French of constructions associated with a pattern of exocentric interpretation;

All of which can be predicted within the “Morpheme and Lexeme Hypothesis” (see Nemo 2001a) presented here, if a correct description of the signification of the morpheme and of the variety of possible morphological constructions is provided, i.e. if polymorphy and morphological flexibility are considered.

Polymorphy, i.e. the fact that *morph* and *form*, *coul-er* and *dé-goul-in-er*, *rot-ation* and *tor-dre* are two forms of the same morpheme, allows to account for a large part of problematic bases and for the inexistence of semantic drift: for instance even though the meaning of the word *amorphe* (inactive) in French is not compositional in a DC sense, it may be directly predicted from a lexicalised interpretation of *forme* in *être en forme*, *avoir la forme*. This leads to the conclusion that instead of the systematic postulation of semantic drift, we should rather consider the reality of the kinds of formal drift involved in polymorphy, namely that the same signification (and sometimes meaning) can be associated with various (related) “signifiants”, and study such polymorphy as a regular phenomenon².

Flexibility of morphological constructions is another issue, directly related to the interpretability constraint: comparing words like *chimiqu-ier* (chemical tanker), *but-eur* (striker) and *re-cycl-er* which are all listemes because their base is syntactically distinct from what we would expect (respectively an adjective and not a noun, a noun and not a verb, a noun and not a verb), with predictable words such as *pétrolier* (tanker), *tu-eur* (killer) or *re-pousser* (to push back), allows us to understand (and therefore to predict) the possibility of generating such “listemes”, either because:

- French systematically admits the possibility of an exocentric interpretation of the base of affixed words; so that instead of requiring a forcefully nominal or verbal

² Systematic polymorphy in French consists mainly in: i) alternating non-voiced and voiced consonants (p/b, k/g, t/d) as in *coul/goul*; ii) permutation/metathesis, as in *uple, plu, pul, supplément* and *plus*; iii) expansion, as in *-able* and *habile*; iv) alternating au/al, ou/ol, etc. as in *autre, alterner, haut, altitude*; v) combining any of the former, as in *obst/stop*.

base, it allows the nominal or verbal head to remain implicit and one of its arguments to become the base, thus replacing the “[*pétrole*]_N]_{NP} (oil) transported in *pétrolier* (tanker)” interpretation of *pétrolier*, by the exocentric “[*produits*]_N]_{[*chimiques*]_A]_{NP} (chemicals) transported in *chimiquier*” interpretation, or replacing the [*tuer*]_V]_{VP} interpretation of *tu-eur* by the [*marquer*]_V]_{[*des buts*]_O]_{VP} interpretation of *but-eur*.}}

or because:

- the meaning of the noun *cycle* actually unifies/coincides with the indications encoded by the morpheme *re*, thus allowing *recycle* to be semantically interpretable, and hence well-formed.

So that in both cases, it is possible to show that the criteria proposed in order to decide what had to be accounted for and what does not have to be accounted for, lead morphologists to overlook the existence and diversity of word-formation processes, and to ignore the fact that their model is heavily falsified.

2.3.2 *Word-formation ≠ Derivation/Composition*

Much more important in a certain way, the distinction proposed in figure 2, allows the linguist to draw a clear line between true **derivational and compositional processes** (represented in figure 2 by the re-entering arrow DC), which really take *lexemes* with their lexicalised meaning *s* as inputs for word-formation, and **insertional processes** which take morphemes – and the **f(m, cstr, ctxt)** functions they encode – as inputs and force a new (contextual and constructional) interpretation of the morpheme (which may become lexicalised if the use becomes a usage).

In other words, it is of considerable importance for the linguist to be able to distinguish between an horizontal relationship between two words (or meanings), such as the relation between two leaves of the same branch (polysemy or polycategoriality), and a vertical/derivational relationship, in which an indisputable transfer of meaning occurs.

As we have seen there is only a horizontal relationship between the different lexemes *but* in English, and none of the *almost, without, only*, etc. meanings associated with these lexemes may be said to be derived from a “basic” connective meaning nor be the result of any bleaching of this supposedly basic meaning. The same thing holds in morphology when we have to account for words like *décoller* (to take off, for a plane), *collecte, collection, collision, collusion, accolade*: the fact that only a minority of these words and/or a minority of the uses of these words are predictable from the ordinary meanings of the words *colle* (glue) or *coller* (stick) is not a trace of the fact that “*the lexicon is like a prison – it contains only the lawless, and the only thing that its inmates have in common is lawlessness*” nor of the fact that since it “*is simply a collection of the lawless, there neither can nor should be a theory directly about it*”, but a trace of the fact **that the ordinary meanings** of the words *colle* and *coller* are **only local interpretations of the indications encoded by the morpheme [coll]**, interpretations which, since they are not coded, are completely unable to block the generation of new interpretations in new contexts and the generation of new lexemes in new positions. Listing all these uses and words in the lexicon because they cannot be predicted from

these local interpretations is thus the equivalent of describing the lexicon as an endless sense enumerative lexicon (see Bouchard 1995; Pustejovsky 1995).

All this will lead us to a single conclusion: if we are to account for the existence of so-called listemes, we need to understand that: i) it is always possible for a speaker to use a morpheme in a new construction or a new context, thus creating new interpretations and freeing him/her from conforming with the senses associated with previous uses; ii) the well-formedness of a new lexeme is not a matter of applying existing rules to an existing lexical stock, as in figure 1 above, but mainly a matter of **interpretability**.

If such is the case, and if indeed most listemes (such as *multiple*, *rotation*, *rétablir*) are semantically and constructionally interpretable despite their not being formed by the kind of combinatorial mechanisms (WFRs) morphologists were looking for (i.e. the DC arrow of figure 2), then it means that understanding what interpretation is about, how contextual unification works and what the relationship between non-categorial morphemes and categorially defined lexemes is, should be a central issue in morphology.

A word like *re-tali-ation* in English is not well-formed because it can be produced by general combinatorial rules, but only, as the French word *re-cycle*, because the signification of *re-* is to indicate the existence of two anti-oriented processes p1 and p2, and because the meaning of the base (*Talion*), as opaque as it may seem, does unify with these indications (losing an eye as a p2 punishment for the p1 crime of making somebody lose an eye, etc.). Word-formation and word-construction, it seems, is thus cemented by interpretation.

This conclusion directly falsifies one of the founding postulates of the Chomskian approach to linguistics, according to which: (i) a linguistic theory should describe the combinatorial mechanisms which allow the generation of new sentences or new words; (ii) there is no way semantic considerations could help us explain the combinations that are acceptable and the ones that are not. It seems quite clear on the opposite, that it is impossible to account for the generation of the lexicon, i.e. of a large part of the first task, without taking into account the fact that, ultimately, the cement of word-formation is interpretation, i.e. without dropping the second assumption.

What we need hence in order to be able to account for the generation of the lexicon, to avoid listing most of it and to integrate the repeated demonstration, within Linguistic Semantics, of the fact that no combinatorial information is attached to (encoded by) the basic semantic units of a language (i.e. morphemes), is to understand that **well-formedness** in morphology is **not a matter of grammaticality but a matter of interpretability**. And thus that what we need is a theory of interpretation consistent with the empirical observations of data-based studies, and a methodology which strictly forbids the use of introspection and intuition in the definition of what has to be accounted for and of how to account for it. Ultimately, the choice is not between doing morphology with or without semantics, as Chomsky seemed to suggest, but between doing it with bad or good semantics. Any semantic model whose ambition (and result) is not to account for the generation of the lexicon, i.e. of new senses and new lexemes, should be abandoned in morphology if morphology wants to be something else than a formalisation of the shortcomings of common sense.

References

- Anderson, S.R. (1992), *A-Morphous Morphology*, Cambridge, Cambridge University Press.
- Anshen, F. & M. Aronoff (1988), *Producing Morphologically Complex Words*, “Linguistics” 26, 641–655.
- Aronoff, M. (1976), *Word Formation in Generative Grammar*, Cambridge (Mass.), The MIT Press.
- Aronoff, M. (1993), *Morphology by Itself: Stems and Inflectional Classes*, Cambridge (Mass.), The MIT Press.
- Aronoff, M. & F. Anshen (1998), *Morphology and the Lexicon*, in A. Spencer & A. Zwicky (eds.), *The Handbook of Morphology*, Oxford, Blackwell, 248–271.
- Baayen, H. (1991), *Quantitative Aspects of Morphological Productivity*, in G. Booij & J. van Marle (eds.), “Yearbook of Morphology” 1991, Dordrecht, Kluwer, 109–149.
- Baayen, Harald (1993), *On Frequency, Transparency and Productivity*, in G. Booij & J. van Marle (eds.), “Yearbook of Morphology” 1992, Dordrecht, Kluwer, 181–208.
- Baudouin de Courtenay, J. (1895), *Versuch einer Theorie phonetischer Alternationen: Ein Kapitel aus der Psychophonetik*, Strassburg/Crakow.
- Benveniste, E. (1954), *Problèmes de reconstruction en sémantique*, in *Problèmes de linguistique générale*, 2, Paris, Gallimard.
- Bouchard, D. (1995), *The Semantics of Syntax*, Chicago, Chicago University Press.
- Broschart, J. (1997), *Why Tongan does it differently?: categorial distinctions in a language without nouns and verbs*, “Linguistic typology” 1, 123–165.
- Bybee, J.L. (1988), *Morphology as Lexical Organization*, in M. Hammond & M. Noonan (eds.), *Theoretical Morphology: Approaches in Modern Linguistics*, San Diego, Academic Press.
- Bybee, J.L. & P. Hopper (2001)(eds.), *Frequency and the Emergence of Linguistic Structure*, Amsterdam, John Benjamins Publisher.
- Corbin, D. (1987), *Morphologie dérivationnelle et structuration du lexique*, Max Niemeyer Verlag, Tübingen.
- Di Sciullo, A.-M & E. Williams (1987), *On the Definition of Word*, Cambridge (Mass.), The MIT Press.
- Dixon, R & A. Aikhenvald (2002), *Word: A Cross-linguistic Typology*, Cambridge, Cambridge University Press.
- Ducrot, O. (1987), *L'interprétation comme point de départ imaginaire de la sémantique*, in *Dire et ne pas dire* ([1972], 1991), Paris, Hermann.
- Fradin, B. (1993), *La théorie morphologique face à ses choix*, “Cahier de Lexicologie”, 1993/2, 63, 5–42.
- Fradin, B. (1999), *Syntaxe et morphologie*, “Histoire Epistémologie et Langage” XXI, 2, 7–44.
- Gasiglia, N., F. Nemo & P. Cadiot (2001), *Meaning and the Generation of Reference*, in P. Bouillon (ed.), *Generative Approaches to the Lexicon*, Université de Genève.
- Fraser, B. (1998), *Contrastive Discourse Markers in English*, in A.H. Jucker & Y. Ziv (eds.), *Discourse Markers: Descriptions and Theory*, Amsterdam, John Benjamins.
- Gil, D. (2000), *Syntactic categories, cross-linguistic variation and universal grammar*, in P. Vogel & B. Comrie (eds), 173–216.
- Goldberg, A. (1995), *Constructions*, Chicago, University of Chicago Press.
- Jalenques, P. (2000), *Contribution à l'étude du préfixe RE en français contemporain: pour une analyse compositionnelle du verbe regarder*, PhD Thesis, Paris 7.
- Levin, B. & M. Rappaport Hovav (1998), *Morphology and Lexical Semantics*, in A. Spencer & A. Zwicky (eds.), *The Handbook of Morphology*, Oxford, Blackwell, 248–271
- Lieber, R. (1992), *Deconstructing Morphology: Word Formation in Syntactic Theory*, Chicago, University of Chicago Press.

- Matthews, P.H. (1991), *Morphology: An Introduction to the Theory of Word-Structure*, Cambridge, Cambridge University Press.
- Nemo, F. (1998), Enfin, encore, toujours entre indexicalité et emplois, in *Actes du Colloque de Bruxelles. Sens et fonctions*, XIV, Tübingen, Max Niemeyer Verlag.
- Nemo, F. (1999), *The Pragmatics of Signs, The Semantics of Relevance, and The Semantic/Pragmatic Interface*, in *The Semantics-Pragmatics Interface from Different points of View*, CRiSPI Series, Amsterdam, Elsevier Science, 343–417.
- Nemo, F. (2001a), *Morpheme Semantics and the Autonomy of Morphology. The Stable Semantics of (Apparently) Unstable Constructions*, in M. Andronis et al. (eds.), *CLS 37: The Panels. Papers from the 37th Meeting of the Chicago Linguistic Society*, vol. 2. Chicago, Chicago Linguistic Society
- Nemo, F. (2001c), *Pour une approche indexicale (et non procédurale) des instructions sémantiques*, “Revue de Sémantique et de Pragmatique” 9/10, 195–218
- Nemo, F. (2002a), *But (and mais) as morpheme(s)*, “Delta” (Sao Paulo), n°18–2.
- Nemo F, (to appear), *Discourse Particles as Morphemes and as Constructions*, in K. Fischer (ed.), *Approaches to Discourse Particles*, Amsterdam, Elsevier.
- Plag, I. (1998), *The Polysemy of -ize Derivatives. The Role of Semantics in Word Formation*, in G. Booij & J. van Marle (eds.), “Yearbook of Morphology” 1997, Dordrecht, Kluwer, 219–242.
- Pustejovsky, J. (1995), *The Generative Lexicon*, Cambridge (Mass.), The MIT press.
- Spencer, A. (1991), *Morphological Theory: An Introduction to Word Structure in Generative Grammar*, Blackwell Publisher.
- Sweet, H. (1876), *Words, logic and grammar*, Transactions of the Philological Society, in W.F. Bolton & D. Crystal (eds.)(1969), *The English language: essays by linguists and men of letters*, vol. 2, Cambridge, Cambridge University Press
- Vogel, P. & B. Comrie (eds)(2000), *Approaches to the Typology of Word Classes*, Berlin-New York, Mouton de Gruyter.

Typology and Boundaries: The Acquisition of a New Morphological Boundary by Modern Hebrew*

Irit Meir

Department of Hebrew Language, University of Haifa, Israel
imeir@univ.haifa.ac.il

Abstract

Despite the many changes that Modern Hebrew (MH) has undergone since its revival, its morphology seems to have remained largely intact, in that the Biblical Hebrew root-and-pattern word structure is dominant in MH as well. However, it has been pointed out (by e.g., Bolotsky 1978, Schwarzwald 2002 and many others) that even morphology is not immune from changes. Borrowed suffixes such as *-ist*, *-nik* and *-çik* have found their way into MH word formation. The extensive use of prefixes has also been regarded as foreign influence. In this paper I argue that the morphology of MH shows yet another deviation from the Biblical Hebrew structure, by acquiring not a new affix, but rather a new morphological boundary and a new level for suffixation, the # (word level) boundary. This boundary applies to words lacking the canonical root-and-pattern structure (that is, borrowings, acronyms, names and compounds). The affixation of the # suffixes to these words does not cause stress shift to the suffix (these suffixes were stress-attracting in earlier stages of the language, even in borrowed forms). This accounts for the distinct stress pattern exhibited by some inflected and derived forms of non-canonical words. One consequence of this change is that MH developed several default suffixes, (in the sense of Kiparsky 1973, Aronoff 1976), e.g., in the plural and feminine forms. Another consequence of this change is the emergence of two distinct gender systems in MH, one that does not constitute an inflectional class (in the sense of Aronoff 1994), and one that does. The suggested analysis also ties together several observations and analyses concerning plural formation and stress assignment in the nominal system of MH, which previously were not regarded as related.

1. Plural Affixation in Hebrew

Nouns in Hebrew fall into two gender classes, masculine and feminine. There is a rather strong correlation between the phonological form of a noun and its gender. The feminine is the marked gender, feminine nouns typically ending with *-a* (e.g., *simxa* ‘happiness’) or *-ut/-it/-et/-at* (*xanut* ‘shop’, *xavit*, ‘barrel’, *rakevet* ‘train’, *tsalaxat* ‘plate’). Masculine nouns are unmarked: nouns lacking a feminine ending are masculine. However, this correlation is not entirely consistent. Some masculine-sounding nouns, that is nouns which do not have a feminine ending, are nonetheless feminine (e.g., *zeven* ‘stone’, *zerets* ‘country/land’, *tsipor* ‘bird’), and a smaller number of nouns ending with *-a* or *-it/-et* are masculine (*layla* ‘night’, *tsevet* ‘crew’, *ramit* ‘colleague’).

Hebrew has two nominal plural suffixes: *-im* and *-ot*. The latter has several allomorphs: *-iyot/uyot*, and *-a?ot*. Masculine nouns usually take the *-im* suffix, and feminine nouns the *-ot* suffix.¹ Once again, the correlation is not entirely consistent.

* Thanks to Mark Aronoff, Edit Doron and Wendy Sandler for very helpful comments and discussion. I would also like to thank the participants of the MMM4 conference.

¹When the feminine plural suffix *-ot* attaches to words ending with *-a*, it replaces the vowel in word final position: *?agala* – *?agalot* (‘wagon’).

Aronoff (1994) notes that there are about 80 masculine nouns in current use taking the *-ot* suffix, and 30 or so feminine nouns taking the *-im* suffix. Thus the choice of plural suffix cannot be inferred from the gender of the noun. Furthermore, it cannot be reliably inferred from the phonological form of the noun: feminine-sounding nouns may take the *-im* suffix², and some masculine-sounding nouns take the *-ot* suffix. Hence, although "...the morphological structure along with gender marking are the main causes for the choice of the plural suffix" (Schwarzwald 1991: 596), neither the gender nor the phonological structure of the base can fully predict the choice of the plural suffix (as illustrated in table 1 below). The specific phonological form and the choice of plural suffix have to be stated for each noun *independently* (Aronoff 1994: 78), which means that there are no noun paradigms in the language. Therefore, gender in Hebrew is not an inflectional class (in the sense of Aronoff 1994, that is the set of lexemes whose members each select the same set of inflectional realizations).³

Noun gender	Regular		Irregular	
Masculine	<i>xof – xofim</i>	‘beach’	<i>kol – kolot</i>	‘voice’
Feminine	<i>?erets – ?aratsot</i>	‘country/land’	<i>?even – ?avanim</i>	‘stone’
Phonological form				
Masculine sounding	<i>maxfev – maxfevim</i> (m.)	‘computer’	<i>mafteax – maftexot</i> (m.)	‘key’
Feminine sounding	<i>layla – leylot</i> (m.)	‘night’	<i>nemala – nemalim</i>	‘ant’

Table 1: The unpredictability of plural formation in Hebrew.

Plural formation in Hebrew is yet irregular in another way. Plural affixation usually shifts the stress to the suffix. This stress shift may result in additional phonological changes to the base. Though the Mishkal (pattern) of the singular form is a good predictor of these phonological changes (Berent et. al 1999), their occurrence is nonetheless not always predictable. For example, in *gamad – gamadim* (‘dwarf’) plural inflection does not alter the base, but in the phonologically similar *gamal – gmalim* (‘camel’), suffixation causes the deletion of a vowel in the stem. Similarly, in *xanit – xanitot* (‘spear’), suffixation does not change the base, whereas in *mapit – mapiyot* (‘napkin’), suffixation results in the deletion of the feminine suffix (*-it*) of the base (Schwarzwald 1991: 601). Thus, plural formation in Hebrew is irregular in two ways: both the choice of the plural suffix (*-im* or *-ot*) and the phonological changes caused by

² In Schwarzwald’s (1991:595) dictionary count, she found that out of 3926 nouns with a feminine ending, 69 took the *-im* suffix.

³ The gender of Hebrew nouns is reliably revealed only by agreement. Agreeing adjectives, verbs and participles agree in gender with the noun. Thus, an adjective modifying a feminine noun is morphologically marked as feminine, whether or not the noun is phonologically marked as feminine (e.g., *?even levan-a* ‘a white (fem.) stone (fem.)’). Similarly, the choice of the plural suffix in adjectives is entirely predictable from the gender of the head noun: adjectives modifying masculine plural nouns take the *-im* suffix, and adjectives accompanying feminine plural nouns take the *-ot* suffix. The predictability of plural marking in adjectives led Schwarzwald (1991) to suggest that adjectival pluralization takes place in the grammar, while nominal pluralization takes place in the lexicon.

suffixation are not reliably predictable from the phonological form or the gender of the base.⁴

2. Plural Formation and Stress

Plural affixation generally shifts the stress to the suffix, e.g.: *sipur* – *sipurim* (‘story’ m.); *rakevet* – *rakavot* (‘train’ f.). However, there is a class of nouns in which the stress does not shift to the plural affix. This class includes words which are outside of the canonical root-and-pattern word formation structure,⁵ (referred to by Berent et. al. 1999 as words lacking a canonical root). It consists of the following sub-classes:

Borrowings:	<i>student</i> – <i>studentim</i>	‘student’,
	<i>banana</i> – <i>bananot</i>	‘banana’
Words containing a borrowed affix:	<i>kibutsnik</i> – <i>kibutsnikim</i>	‘a Kibutz member’ (m.)
	<i>kibutsnikit</i> – <i>kibutsnikiyot</i>	‘a Kibutz member’ (f.)
Acronyms:	<i>rabat</i> – <i>rabatim</i>	(<i>rav</i> – <i>turai</i> , ‘corporal’)
	<i>tatsa</i> – <i>tatsot</i>	(<i>tatslumei</i> – <i>?avir</i> ‘aerial photographs’)
Nouns used as proper names:	<i>?afik</i> – <i>?afikim</i>	‘the Afik family’
	<i>dina</i> – <i>dinot</i>	‘the Dina’s’
Some blends:	<i>midrexov</i> – <i>midrexovim</i>	‘pedestrian walkway’
Some highly lexicalized compounds:	<i>kadursal</i> – <i>kadursalim</i>	‘basketball’

When suffixation does not result in stress shift, there are also no accompanying phonological changes in the base. Thus, the plural of the noun *barak* (‘lightening’) is *brakim*, exhibiting the expected vowel change. But when used as a family name, its plural form is *Barakim*, with no stress shift and no vowel change (Berent et. al 1999: 31)⁶. Plural suffixation, then, applies very differently to canonical vs. non-canonical

⁴ As pointed out in fn. 1, pluralization of adjectives is much more regular than that of nouns, in that the choice of the plural suffix can be fully inferred from the gender of the head noun. However, even in adjectives, the phonological changes to the based caused by suffixation are not fully predictable, as in the following examples: *gadol* – *gdolim* ‘big’ vs. *varod* – *vrudim* ‘pink’; *salit* – *salitim* ‘reigning’ vs. *favir* – *fvirim* ‘fragile’.

⁵ In addition to nouns constructed by the root and pattern combination, canonical words in Hebrew include also most nouns formed by stem+Hebrew suffix (as opposed to borrowed suffixes), whether the stem is of Hebrew origin or not, e.g., *traktoron* – *traktoranim* (‘Dune buggy’). Some foreign stems, though, exhibit non-canonical behavior even when they combine with a Hebrew suffix, e.g., *politikai* – *politikaim* ‘politician’. See Schwarzwald 2002, for further discussion.

⁶ The only possible phonological change to the base is stress shift. When a stressless suffix attaches to a base with stressed antepenult, stress often shifts to the penult in the suffixed form, as in *?otobus* – *?otobusim* (‘bus’), *telefon* – *telefonim* (‘telephone’). This stress shift occurs in some forms but not in

words. In the former, the plural suffix is stress-attracting, and suffixation may result in phonological changes to the base. In the latter, the plural suffix does not attract stress, and suffixation does not cause phonological changes to the base.

The plural suffixes are not the only suffixes in the languages exhibiting such a dual behavior. There are few other suffixes characterized by dual behavior when attached to canonical vs. non-canonical words (Schwarzwald 2002):

The suffix	Canonical word	Non-canonical word
Feminine inflection <i>-it</i>	<i>rakdan</i> – <i>rakdanit</i> 'dancer'	<i>rabat</i> – <i>rabatit</i> 'corporal'
Adjectivizing suffix <i>-i</i>	<i>šemeš</i> – <i>šimšī</i> 'sunny'	<i>tel-ʔaviv</i> – <i>tel-ʔavivi</i> 'Tel-Aviv' – 'Tel-Avivian'
A derivational suffix forming abstract nouns <i>-(iy)ut</i>	<i>yeled</i> – <i>yaldut</i> 'child' – 'childhood'	<i>šlumper</i> – <i>šlumperiyut</i> 'slob' – 'slobbishness' <i>diva</i> – <i>divaiyut</i> 'diva' – 'diva-ness'

Table 2: Dual-behavior suffixes

However, not all suffixes exhibit such dual behavior. Some suffixes are consistently stress-attracting, even when affixed to non-canonical words. (Bat-El 1993, Schwarzwald 2002).

The suffix	Canonical word	Non-canonical word
<i>-an</i>	<i>sefer</i> – <i>safran</i> 'book' – 'librarian'	<i>solo</i> – <i>solan</i> 'solo' – 'solist'
<i>-iya</i>	<i>sefer</i> – <i>sifriya</i> 'book' – 'library'	<i>djunk</i> – <i>djunkiya</i> 'junk yard'
<i>-ai</i>	<i>ʔiton</i> – <i>ʔitonai</i> 'journal' – 'journalist'	<i>bank</i> – <i>bankai</i> 'bank' – 'banker'
<i>-on</i>	<i>yeled</i> – <i>yaldon</i> 'boy' – 'small child'	<i>traktor</i> – <i>traktoron</i> 'tractor' – 'dune buggy'

Table 3: Uni-behavior suffixes

Of special interest is the construct state masculine plural *-ei*. Though morphologically related to the plural suffix *-im* (Berman 1978: 75), it does not exhibit the dual behavior of *-im*. Rather, it consistently attracts stress. Thus, in non-canonical words construct state plurals and free state plurals show different stress patterns:

- (1) a. *milyon* – *milyonim* ('million') *kurs* – *kursim* ('course')
- b. *milyonei* 'anašim ('millions of people'), *kursei-mavo* ('introductory courses')

others, and varies among speakers (Bat-El 1993). It can also be attested in some adjectives derived from penult bases (*london* – *londoni* 'a Londoner').

The above facts indicate that stress shift or the lack of it is not a property of bases or of suffixes by themselves. The same base may either retain its stress in suffixation or not, depending on the suffix (as in 1.a-b). Conversely, the same suffix may or may not attract stress, depending on the base (as illustrated in table 2). Hence the occurrence or non-occurrence of stress shift is determined by the combination of a base and a suffix. Stress fails to shift to the suffix only when a dual-behavior suffix is attached to a non-canonical base. In all other combinations, stress shifts to the suffix.

3. Semantic and Distributional Correlates of Dual-behavior Suffixation

The two distinct phonological patterns exhibited by the dual-behavior suffixes correlate neatly with a cluster of properties. Stress-neutral suffixation is more regular and coherent than stress-shifting suffixation.

(a) *Semantics*: Stress-shifting suffixation is less coherent semantically, in that the meaning of the suffixed form is not always componential. Some plural forms have idiosyncratic meanings. For example, *ferutim* (*ferut-im*, ‘services’) has the additional meaning of ‘WC’. Others are pluralia tantum (e.g., *panim* ‘face’, *raxamim* ‘compassion’, *xayim* ‘life’, *?atikot* ‘antiquity’, *sonot* ‘miscellany’, Schwarzwald 1991,593). And there are at least two nouns which are morphologically plural, but are syntactically singular: *behemot* ‘behemoth /hippopotamus’ and *be?alim* ‘possessor/owner’. These nouns are homophonous with the regular plural forms *behemot* (‘beasts’) and *be?alim* (‘husbands’). In contrast, stress-neutral plural suffixes are semantically coherent: the meaning of the complex forms is a compositional function of the meaning of its parts.

(b) *Morphology*. Stress-shifting suffixes are sensitive to the internal morphological structure of the words to which they attach. They attach to forms constructed by the root and pattern combination, or to forms ending with a Hebrew suffix (see fn. 5). Stress-neutral suffixes attach across the board to all nouns and adjectives to which there is no lexically specified form.

(c) *Distribution*: The distribution of stress-shifting suffixes is not entirely regular. There are nouns which do not take the plural suffix, for no apparent semantic or phonological reasons (see Schwarzwald 1991 for an extensive discussion of such nouns). Additionally, there are a few nouns which can take both suffixes, e.g., *?eser* ‘ten’ – *?esrim* ‘twenty’ – *?asarot* ‘decades’, and *yom* ‘day’ – *yamim* ‘days’ – *yemot-* ‘times of’ (Schwarzwald 1991: 588). Stress-neutral suffixation, on the other hand, is fully productive. The plural suffixes can be affixed to any count noun, regardless of its phonological or morphological forms.⁷ Finally, while the choice of the plural suffix is not predictable when the suffix is stress-attracting, it is fully predictable when the suffix is stress-neutral: nouns ending with *-a* take the *-ot* suffix (*viola* – *violot* ‘viola’, *?ameba*

⁷ Schwarzwald’s list of nouns which do not pluralize includes some non-core nouns as well, including professional areas of studies such as *filologya* ‘philology’, *geometrya* ‘geometry’, *?akustika* ‘akustics’. I disagree with her judgments here. Such nouns can be pluralized in appropriate contexts.

– *ʔamebot* ‘ameba’, *pitsa* – *pitsot* ‘pizza’), all other nouns take the *-im* suffix (*avokado* – *avokadoim* ‘avocados’, *kontsert* – *kontsertim* ‘concert’, *kartiv* – *kartivim* ‘popsicle’, *guru* – *guruim* ‘guru’).⁸ I am aware of one exception to this generalization: when a family name ends with *-a*, the plural (denoting the members of the family) is formed by the *-im* suffix rather than the *-ot* (e.g., *ha-moria-im* ‘the Moria family’, **ha-moriyot*).

4. Default Plural Marker

A different aspect of plural formation in MH has been investigated by Berent, Pinker and Shimron (1999). They raise the question of whether MH has a default plural marker, that is, regular inflection that applies by the ‘elsewhere condition’ to any target that fails to trigger a more specific process (in the sense of Kiparsky 1973). Berent et. al. hypothesize that although plural formation is irregular, native speakers use the *-im* suffix as the default plural marker for all masculine-sounding words outside of the canonical root-and-pattern morphology, e.g., borrowings, acronyms and names. In a series of experiments, they presented native speakers with masculine sounding non-words that are highly dissimilar from existing Hebrew words, as well as masculine sounding words identical in form to existing Hebrew words, but used as borrowings or names (e.g., the word *kir* ‘wall’) was presented as a French drink or a family name). The subjects were asked to provide the plural forms for these invented words. Subjects invariably chose the *-im* suffix, although many of the homonymous Hebrew words are pluralized by *-ot*. Hence Berent et. al. conclude that *-im* indeed functions as a general default plural marker in MH.

What has gone unnoticed so far is that the Berent et. al. study is directly related to the dual behavior of plural suffixation described above, in that the class of words that takes the default plural marker is precisely the class that does not allow stress shift in plural formation. The experiments in the Berent et. al. study were conducted in writing, hence the stress pattern of the target words was not documented (Hebrew orthography does not encode stress).⁹ However, had they done the experiment orally, it would become clear that the default suffix does not attract stress. In other words, the plural marker, when functioning as a default marker, is stressless. This correlation calls for an explanation.

5. Suggested Analysis

One possible explanation is to assume that Hebrew has acquired a number of stressless suffixes. Hebrew has indeed borrowed a few stressless derivational suffixes, e.g., *-nik* (*kibutsnik* – *kibutsnikim* ‘a Kibutz member’), and the diminutive *-çik* (*katañçik* ‘very small, minute’). These suffixes, though stressless, are not stress-neutral: they require the preceding syllable to be stressed. The suffixes analyzed in this paper, in contrast, are

⁸ As was pointed out to me by Edit Doron, the plural form of nouns ending with *-i* is *-im* rather than the expected *-iim* (e.g., *sini* – *sinim* ‘Chinese persons’). In adjectives, however, plural forms often retain both vowels: *siniim* ‘Chinese (adj)’.

⁹ Berent et. al. do mention that default suffixation is stressless. However, their experiments were designed to examine the choice of the plural marker (*-im* or *-ot*), and did not take stress into consideration.

both stressless and stress-neutral. If we assume that these suffixes are borrowed as well, it would be difficult to explain why all these suffixes have homophonous stressed counterparts. It also fails to explain the semantic and distributional correlates of the two types of suffixation.

The approach I wish to pursue here is that Hebrew has acquired a new way of combining a suffix to a base, that is, that Hebrew acquired a different boundary, or a new level for suffixation. This approach accounts straightforwardly for the cluster of properties associated with each type of suffixation, and for the development of default forms as well.

As has long been observed (e.g., by Sapir 1925¹⁰), suffixes attach to bases in two different ways. These have been formalized in terms of two different boundaries: + and # (Chomsky & Halle 1968, Aronoff 1976), which correspond to two different levels of affixation: stem level and word level respectively (Kiparsky 1982, 2000, Aronoff & Sridhar 1987).¹¹ Stem level suffixes typically trigger and may undergo phonological changes, may cause stress shift in the base, are less coherent semantically and less productive. Word level suffixes cause no phonological changes to the base, they are stress neutral, and are much more regular, both semantically and distributionally.

Hebrew nominal suffixes (both inflectional and derivational), are basically stem level suffixes. They attract stress, and may alter the phonological structure of the base. They are also semantically less coherent, and their distribution is not completely regular. A few suffixes, however, behave like word level suffixes when attached to non-canonical bases: they are stress-neutral, do not cause any phonological changes to the base, are semantically coherent and their distribution is completely regular. In other words, the dual behavior of certain suffixes can be captured in terms of different levels of suffixation: these suffixes behave as stem-level suffixes when attached to bases with canonical roots, and as word-level suffixes when attached to non-canonical bases.¹² The cluster of properties characterizing each type of suffixation follow straightforwardly from the assumption that they apply at different morphological levels, as summarized in table 4:

¹⁰ Sapir (1925, fn. 6) attributes to L. Bloomfield the observation that "the agentive *-er* contrasts with the comparative *-er*, which allows the adjective to keep its radical form in *-ŋg-* (e.g., *long* with *-ŋ*: *longer* with *-ŋg-*)." Consequently, Sapir analyzes the agentive *-er* as an affix that attaches to a word, while the comparative *-er* is affixed to stems. I thank Mark Aronoff for bringing this reference to my attention.

¹¹ Kiparsky maintains that the levels are ordered with respect to each other, while Aronoff & Sridhar explicitly argue against level ordering. The analysis presented here does not have any bearings on the issue.

¹² Hebrew is not unique in having homonymous word vs. stem level suffixes. Aronoff (1976) and Aronoff & Sridhar (1987) discuss such suffixes in English and Kannada, showing that the morphological differences are accompanied by the expected semantic and distributional differences.

Stem Level Suffixes (+boundary)	Word Level Suffixes (#boundary)
<ul style="list-style-type: none"> • Trigger phonological changes to the base (<i>tof</i> – <i>tupim</i> ‘drum’) • Attract stress (<i>gir</i> – <i>girim</i> ‘chalk’) • Less coherent semantically (<i>jerutim</i> ‘service+pl., =WC’) • Less productive: do not apply to some words (<i>behemot</i> ‘hippopotamus’) • Irregular distribution: choice of plural suffix cannot be determined by the form or gender of the singular. 	<ul style="list-style-type: none"> • Cause no phonological changes to the base (<i>avokado</i> – <i>avokadoim</i>) • Stress neutral: (<i>gir</i> – <i>girim</i> ‘gear’) • Semantically coherent • Fully productive: can attach to words of any phonological structure, even words ending with a vowel (<i>homo</i> – <i>homoim</i> ‘homosexual’) • Regular distribution: determined by the form of the singular: words ending with <i>-a</i> take the <i>-ot</i> suffix. All other words take the <i>-im</i> suffix.

Table 4: Two different types of suffixation in Modern Hebrew

This analysis has the following advantages: first, the default nature of these suffixes is accounted for. Word level affixes are much more regular and productive than stem level affixes, in that they apply across the board to an entire class of words. Hence only word level affixes can function as default marker in this case. Second, it explains the fact that all stressless suffixes have stressed counterparts: the suffixes themselves are not new, only the way they combine with the bases. Third, it accounts for the specific nature of the bases which take stem-level suffixes. These words lie outside the canonical word-formation processes of the language, and hence fail to trigger any more specific affixational rules.

According to this analysis, the diachronic change that Hebrew is undergoing is the activation of a new level for suffixation, the word level. In earlier stages of Hebrew, all nominal suffixation processes took place at stem level. In Modern Hebrew, suffixation takes place at two levels, depending on the nature of the base and the nature of the suffix. The core lexicon still exhibits the same pattern found in earlier stages of Hebrew: suffixation is restricted to stem level. The non-core lexicon, in contrast, introduces the change: some suffixation processes take place at stem level, while others occur at word level. The word level suffixes are the most productive and regular suffixes in the language: the plural and feminine inflectional suffixes, and the *-i* and *-iyut* derivational suffixes. All other suffixes are stem level¹³.

¹³ The stem level suffixes include all derivational suffixes, as well as two inflectional suffixes: the masculine plural construct state suffix *-ei*, and the possessive suffixes. These suffixes, though inflectional, are non-obligatory, since they have synthetic paraphrases, and in fact they become quite rare in current language use.

These diachronic changes are quite recent. In earlier stages of the language, plural suffixes were always stress-attracting, even when attached to borrowed words, e.g.: *teʔatron* – *teʔatraʔot* (‘theatre’, of Greek origin), *maʃkanta* – *maʃkantaʔot* (‘mortgage’ of Aramaic origin), *ʔadrixal* – *ʔadrixalim* (‘architect’, of Akkadian origin, via Aramaic), and even the more recent *ʔuniversita* – *ʔuniversitaʔot* (‘university’).

<p><u>Stem Level :</u> All nominal suffixation (inflectional and derivational)</p>	<table border="1"> <tr> <td data-bbox="756 481 1043 651"> <p><u>Stem Level :</u> Core Lexicon – All nominal suffixation</p> </td> <td data-bbox="1043 481 1353 651"> <p><u>Stem Level :</u> Non-core lexicon Non-regular (mainly derivational) suffixes</p> </td> </tr> <tr> <td colspan="2" data-bbox="756 651 1353 813"> <p><u>Word Level:</u> Non-core lexicon: Regular (default) suffixes: inflection (pl., fem.), derivation (<i>-i</i>, <i>-iyut</i>).</p> </td> </tr> </table>		<p><u>Stem Level :</u> Core Lexicon – All nominal suffixation</p>	<p><u>Stem Level :</u> Non-core lexicon Non-regular (mainly derivational) suffixes</p>	<p><u>Word Level:</u> Non-core lexicon: Regular (default) suffixes: inflection (pl., fem.), derivation (<i>-i</i>, <i>-iyut</i>).</p>	
<p><u>Stem Level :</u> Core Lexicon – All nominal suffixation</p>	<p><u>Stem Level :</u> Non-core lexicon Non-regular (mainly derivational) suffixes</p>					
<p><u>Word Level:</u> Non-core lexicon: Regular (default) suffixes: inflection (pl., fem.), derivation (<i>-i</i>, <i>-iyut</i>).</p>						
<p>Earlier stages of Hebrew</p>	<p>Recent Modern Hebrew</p>					

Table 5: Levels of suffixation in Hebrew

The bifurcation of suffixation in MH results in another change in its morphological system: the emergence of two distinct gender systems in the language. In the core lexicon, gender assignment is unpredictable, and therefore has to be assigned lexically. In addition, gender is not an inflectional class, as there are no inflectional paradigms in the gender system. In the non-core lexicon gender assignment is completely predictable by the phonological form of the word (as has been pointed out by Schwarzwald 2002), and gender is an inflectional class, since the form of the plural is predictable from the phonological form of the singular: if the singular ends in *-a*, it is feminine, and the plural suffix is *#ot*; otherwise, it is masculine, with *#im*. (e.g. *viola* is feminine, but *çelo* is masculine; plural *violot* and *çeloim*). Hence the novel development in Hebrew – the activation of the word level – results in two significant changes in Hebrew word formation: the development of default inflectional markers and a split in the inflectional category of gender.

The model suggested above makes the following predictions:

1. If a word takes a word-level suffix it is a non-canonical word.
2. If a dual-behavior suffix exhibits stem-level behavior, then the base it attaches to is a canonical word.

To the best of my knowledge, there are no counterexamples to the first prediction. Only non-canonical words take word-level suffixes. As for the second prediction, there are two types of possible counterexamples. First, old borrowings take only stem level suffixes. As pointed out above, word-level suffixation is a new phenomenon in the language. In that respect, old borrowings behave as canonical words. Thus the suffixation pattern of a foreign word is an indicator of the point in

which it entered the language: if a foreign word exhibits only stem level suffixation, it has entered the language in earlier stages.¹⁴

The second type of counterexamples consists of non-canonical words which share the vocalic pattern of canonical words. Typically, these are disyllabic stress-final words, with 3-5 consonants. Thus, *mankal* ('C.E.O.' acronym), *salat* ('salad', borrowing), *smartaf* ('babysitter', blend) are perceived by speakers as having a canonical pattern (on a par with the canonical *malʔax* 'angel', *tabax* 'cook' and *klavlav* 'a little dog/puppy'), and consequently are restricted by some speakers to stem-level suffixation.¹⁵ These two types of counterexamples indicate that the diachronic change Hebrew is undergoing is still very dynamic, being shaped by forces such as the relative youth of a word in the language, and the resemblance of newly formed or borrowed words to canonical forms.

6. Against a Phonological Analysis

Bat-El (1993) and Becker (2003) offer a phonological account of the stress behavior of suffixed forms in MH. According to Bat-El (1993), Hebrew has a class of words that are inherently marked for stress ('accented formatives'), and consequently do not allow the stress to shift to the affixes. Thus, in *traktor* – *traktorim* ('tractor'), stress does not shift to the plural suffix since the base is lexically accented. In order to account for the stress shift in some suffixes (such as *-an*, as in *traktor^{an}* 'tractor driver'), she further distinguished between cyclic and non-cyclic affixes. Cyclic suffixes always precede non-cyclic suffixes, and they trigger the Stress Erasure Convention; that is, cyclic suffixes remove any metrical structure previously assigned. Suffixes such as *-an* are cyclic, hence they remove the lexically assigned accent of the base. In contrast, the non-cyclic plural suffixes respect previously assigned metrical structure.

Bat-El's analysis is similar to the one suggested here in assuming different classes of bases (formatives) and different classes of suffixes. Stress assignment is the result of attaching a specific type of suffix to a specific base. It differs from the analysis suggested here in that the bases and the suffixes are categorized only according to their phonological structure, without making reference to their morphological status.

Becker (2003) further suggests that all the items that have no underlying stress (which he refers to as 'words with mobile stress') are subject to a disyllabic maximum constraint. That is, stress shift to the suffix is restricted to words whose roots are maximally disyllabic. Thus, *psanter* ('piano') has mobile stress (*pasnterim*), since it is disyllabic, while *diktator* ('dictator') has fixed stress (*diktatorim*) since it is tri-syllabic. This analysis faces some empirical problems, in that there are a few tri-syllabic words with mobile stress in Hebrew, such as *livyatan* – *livyatanim* ('whale'), *pilegef* – *pilag^{sim}* ('concubine'), *'akavif* – *'akavif^{sim}* ('spider'), *tsiporen* – *tsipornim* ('carnations'), *taklitor* – *taklitorim* ('CD'), *kadureglan* – *kadureglanim* ('a soccer

¹⁴ When, precisely, the change took place is unclear. However, I think it is reasonable to assume that this diachronic change is closely related to the revival of Hebrew as a spoken language, in the end of the 19th century and the first decades of the 20th century.

¹⁵ Blends ending with *-or* seem to constitute another type of counterexamples. For most speakers, they are pluralized at stem level, though they do not have a canonical vocalic pattern: *migdator* – *migdatorim* ('lighthouse'), *taklitor* – *taklitorim* ('CD'). I have no explanation for that.

player’). In addition, the old loans mentioned above exhibit mobile stress, whether or not their root is maximally disyllabic.

The main problem, however, for a strict phonological analysis, is its failure to account for the specific nature of the class of words with fixed stress (Bat-El’s ‘accented formatives’). Under Bat-El’s analysis, whether a word has fixed or mobile stress is an idiosyncratic property of each word. In Becker’s analysis, this falls out from its syllabic structure. Indeed many foreign words and acronyms have stems consisting of more than two syllables, but there are also numerous monosyllabic or disyllabic borrowings in the language. Whether a mono/disyllabic word has fixed or mobile stress must be stipulated in Becker’s model.

The behavior of nouns used as names is also incompatible with a strict phonological account, as pointed out by Berent et. al. (1999: 32). Names having phonological forms identical to existing canonical nouns, nonetheless have different plural forms (e.g., *barak* – *brakim* ‘lightening’ vs. *Barakim* ‘The Barak family’). This difference cannot be explained without referring to the morphological make-up of these forms, specifically to ‘rootlessness’ of names.

Finally, a phonological analysis cannot account for the semantic and distributional correlates of the two types of suffixation. These arguments strengthen the conclusion reached by Berent et. al., namely that an analysis which views suffixation as a morphological process is more explanatory and adequate than a strict phonological analysis.

7. Conclusions

The dual behavior of certain suffixes in Modern Hebrew with respect to stress-assignment has been accounted for in terms of a new morphological level for nominal suffixation in the language. This level is the site for concatenation of regular suffixes to non-canonical nominals. Irregular suffixation and suffixation of canonical nouns take place at the stem-level, which was the only level available for nominal suffixation in earlier stages of the language. This morphological change brought about two additional modifications to the system: the development of true default markers and the emergence of two distinct gender systems in the language.

Aronoff & Sridhar (1987: 19) point out that English is considered odd in having two levels of affixation, and that this oddity is often attributed to the mixed ancestry of the language – “bastard child of Germanic out of Romance”. Kannada (also discussed in Aronoff & Sridhar), a Dravidian language heavily Sanskritized, is another example of such a language. While modern Hebrew retained much of the morphological system of Biblical Hebrew, in particular the root-and-pattern non-concatenative morphology, it might be that the flux of foreign borrowings and foreign word formation processes (such as prefixation and blends) have led to a similar change in its morphological structure. If levels of affixation contribute to the morphological typology of languages, then it seems that MH is undergoing a change in its typological characterization, by adding word-level to its stem-level nominal suffixation.

References

- Aronoff, M. (1976), *Word formation in generative grammar*, LI Monographs, Cambridge, Mass., MIT Press.
- Aronoff, M. (1994), *Morphology by itself*, LI Monographs, Cambridge, Mass., MIT Press.
- Aronoff, M. & S. Sridhar (1987), *Morphological levels in English and Kannada*, in E. Gussmann (ed.), *Phonomorphology*, Lublin, Redakcja Wydawnictw Katolickiego Uniwersytetu Luelskiego.
- Bat-El, O. (1993), *Parasitic metrification in the Modern Hebrew stress system*, "The Linguistic Review" 10, 189–210.
- Becker, M. (2003), *Lexical stratification of Hebrew – the disyllabic maximum*, talk presented at the 19 Annual meeting of the Israeli Association for Theoretical Linguistics (IATL), Beer-Sheva.
- Berent, I, S. Pinker & J. Shimron (1999), *Default nominal inflection in Hebrew: evidence for mental variables*, "Cognition" 72, 1–44.
- Bloomfield, L. (1933), *Language*, Holt, New York.
- Bolozky, S. (1978), *Some aspects of Modern Hebrew phonology*, in R.A. Berman (ed.), *Modern Hebrew structure*, Tel-Aviv, Universities Publishing Projects, 11–67.
- Chomsky, N. & M. Halle (1968), *The sound pattern of English*, New York, Harper and Row.
- Kiparsky, P. (1973), *'Elsewhere' in phonology*, in S. Anderson & P. Kiparsky (eds.), *A festschrift for Morris Halle*, New York, Holt, Rinehart and Winston, 93–106.
- Kiparsky, P. (1982), *Lexical morphology and phonology*, in I.-S. Yang (ed.), *Linguistics in the morning calm*, Seoul, Hanshin, 3–91.
- Kiparsky, P. (2000), *Opacity and cyclicity*, "The Linguistic Review" 17, 351–365.
- Pinker, S. (2000), *Words and Rules*, New York, Perennial.
- Sapir, E. (1925), *Sound patterns in language*, "Language" 1, 37–51.
- Schwarzwald (Rodrigue), O. (1991), *Grammatical vs. Lexical plural formation in Hebrew*, "Folia Linguistica" XXV/3–4, 577–608.
- Schwarzwald (Rodrigue), O. (2002), *Studies in Hebrew Morphology*, Tel-Aviv, The Open University of Israel.

Modifier-Head Person Concord¹

Irina Nikolaeva
University of Konstanz
irina_a_nikolaeva@yahoo.com

1. Introduction

Several studies on the typology of grammatical agreement have stated that agreement features depend on the syntactic domain where the agreement relation holds. This has been one of the primary motivations for dividing agreement into two relations resulting from different grammatical processes: NP-internal agreement (modifier-head concord) and NP-external agreement (argument-predicate agreement). In typology this idea goes back to Lehmann (1982, 1988), who draws a critical distinction between these two types of agreement based on how the features are transmitted from the controller to the target.

According to Lehmann, NP-external agreement is pronominal and referential in nature. Its purpose is hypothesized to be the tracking of referents in the discourse by recording pronominal features on the target, hence it involves a pronominal Person feature. In contrast, for NP-internal agreement, a modifier does not contain a pronominal indication to the controller, because the target and the controller are constituents of the same NP. Therefore the modifier need not agree in Person. On the other hand, modifier-head agreement involves Case, which is semantically and syntactically a category of the NP. Speaking informally, the modifier agrees in Case with the NP rather than with the head noun. Therefore adnominal modification may exhibit Case agreement, while Person agreement is prohibited, and it is predicted that no target can agree in both Case and Person (Lehmann 1988: 58). These ideas are further confirmed by diachronic facts: according to Lehmann (1983), the markers of internal agreement sometimes come from deictic demonstratives, whereas the markers of external agreement normally go back to personal pronouns.

Lapointe (1988: 71) also observes that Person agreement on adjectival modifiers is unavailable, while Plank (1994) confirms this observation on the basis of data from a 45 language sample and formulates several universals on features involved in modifier-head concord. According to Plank, if a modifier agrees in one feature it will most likely be Number. If there is agreement in two features, they are most likely to be Number and Gender, other permissible combinations being Number and Case or Gender and Case. Lastly, if NP-internal constituents agree in more than two categories, the maximum being four, those will include Number and Gender, very likely also Case, and finally Definiteness, but Person, consistent with the claims of Lehmann, does not occur in this type of agreement.

These assumptions concerning the relevance of features for agreement relations have received the most explicit formal accounts within GPSG and HPSG, where the modifier-head concord is determined by feature compatibility between the head and its projection. According to Gazdar et al. (1985: 83–94), NP-internal agreement involves Case, Number and Gender. Case and Number belong to the category of HEAD features

¹ I am grateful to Doug Arnold, Greville Corbett, Berthold Crysmann, Paul Kiparsky, Louisa Sadler, and especially Farrell Ackerman for discussions and comments on previous versions of this paper.

which if assigned to the NP are transmitted to its head noun. The feature Gender is a lexical property of a noun and is duplicated on the NP by the Head Feature Convention. These features are copied on the dependants via the Control Agreement Principle, which specifies possible controllers and targets. Anderson (1992) provides a similar account within the A-morphous Morphology framework, except that he eliminates the Control Agreement Principle and introduces the category of DEPENDENT features whose value is assigned to the phrase and transmitted to all its daughters.

HPSG explicitly encodes the notion that different principles and features are involved in NP-internal and NP-external agreement. For Pollard & Sag (1994: 60–99) agreement with the verb is a matter of the referential INDEX of the nominal that triggers it. Indices are part of the value of the CONTENT feature structure and therefore part of the semantic contribution of nouns. They are associated with referential expressions and have to be anchored to real world entities via anchoring conditions. Indices involve Person, Number and Gender. In contrast, Case is not an attribute of referential indices, but a purely syntactic property. It arises from language-specific constraints requiring structure sharing between a noun's Case value and that of a noun's dependent.

This particular account makes no explicit claims as to whether NP-internal concord can involve features other than Case. In particular, it does not make any predictions about Person. This has been modified in more recent HPSG accounts by Kathol (1999) and Wechsler & Zlatić (2000, 2003). Following Lehmann's conjectures, they exclude Person from modifier-head concord by explicitly specifying the allowable features for different agreement relations. In Kathol's proposal, NP-internal agreement information is expressed under a feature called AGR, represented as part of the HEAD specification. Modifier-head concord results from structure sharing with the noun's AGR specification. Person never plays a role in NP-internal agreement, because of the assumption that NPs in general do not have a Person attribute in their AGR. Instead, Person information is recorded in the noun's or pronoun's INDEX. Unlike modifier-head concord, subject-verb agreement typically refers to INDEX and therefore can include Person.

The most significant evidence for separating INDEX and AGR comes from the fact that a noun can trigger different features on two classes of agreement targets. This has been richly exemplified in the recent book by Wechsler & Zlatić (2003), who argue that INDEX agreement is more semantically driven than NP-internal concord (in their terminology, CONCORD), because it is a morphosyntactic reflex of anchoring conditions and plays an important role in the semantic interpretation. INDEX features are grammaticalizations of the constraints on anchoring in a discourse and include Person, Number and Gender. In contrast, the CONCORD relation is simply a sharing of morphosyntactic features between certain designated elements. For example, adjective-noun concord follows from the fact that subcategorization of a noun specifies that its modifier's features must match its own features. CONCORD features are Case, Number and Gender. Person is not involved because it is not dependent on local syntactic relations, but has a purely pronominal motivation. Consequently, the analysis reflects the belief there are no languages that list Person under their CONCORD features.²

The primary goal of this paper is to challenge some of these assumptions concerning the distributions of particular features across different types of agreement

²A possible exception is provided by a rather restricted Swahili example where the quantifier 'all' shows agreement with the 1st and 2nd Plural pronouns. However, it is unclear what kind of syntactic relation holds between the two.

relations. I will demonstrate that Tundra Nenets (Samoyed branch of Uralic) exhibits fairly regular, albeit optional, Person concord between an adjectival modifier and its head. However, this occurs in a special context: the Person feature comes from the possessor which is recorded on the head noun via a suffixal head marking strategy. I will argue that this kind of modifier-head concord is in fact expected in some languages that have head marked possessives, if we assume Wechsler & Zlatić's theory of agreement. Plank (1994) explains the absence of adjective-noun Person agreement by the simple fact that all nouns are 3rd Person. The situation in Nenets is more complex because possessed nouns are marked for two Person features simultaneously: they are 3rd Person by virtue of being a noun and additionally carry Person/Number features that come from their possessor. Crucially, I will show, the latter are encoded as part of their CONCORD specification and therefore copied on the adjectival modifier via modifier-head concord. This provides an additional argument for separating morphosyntactic features of a noun into two sets, along the lines suggested by Kathol and Wechsler & Zlatić.

In the next section I cite the basic data on Tundra Nenets agreement. Section 3 presents my analysis, and section 4 provides conclusions.

2. Internal Agreement and the Tundra Nenets NP

2.1. Possessive Agreement

The basic NP in Nenets is head-final. Within nominal possessive constructions a pronominal possessor triggers Person/Number marking on the head noun. Although an independent pronoun is optional, when it is overt it stands in the Nominative (1a). A lexical possessor, in contrast, stands in the Genitive and normally shows no Person/Number agreement on the head (1b).

- (1)³
- a. (pidø^o) te-r^o
 you.SG.NOM reindeer-2SG
 'your (SG) reindeer'
- b. Wata-h ti
 Wata-GEN reindeer
 'Wata's reindeer'

The possessive affixes simultaneously express the Person and Number of the possessor and therefore I will refer to them as Person/Number affixes. They are shown below for a Singular possessor and Nominative possessed noun.

(2)

	SG	DU	PL
1	-w ^o /-myi	-myih	-waq
2	-r ^o	-ryih	-raq
3	-da	-dyih	-doh

³ The Nenets data comes from my own fieldwork supported by an ELDP grant. I use the transcription of Salminen (1997). The glossing for the Nominative will be omitted in further examples.

In non-Nominative cases and with non-Singular possessed nouns, affixes cumulatively express several features: the Case and Number of the possessed and the Person/Number of the possessor. I will not cite the relevant paradigms here for reasons of space.

After Kathol (2001), I will assume that possessed nouns are formed by means of a lexical rule that maps a lexeme to a word inflected for possessive Person/Number.⁴ The possessed head noun can be viewed as selecting for a possessor argument. It corresponds to a two-place relation \mathfrak{R} whose specifier is identified with the possessor. The possessive affix is associated via identically numbered tags with the specifier and therefore with the possessor. A representation for *ter*^o ‘your reindeer’ below follows Kathol (2001).

$$(3) \left(\begin{array}{l} \text{PHON} \quad F_{\text{poss}} \langle te_{[1]}, r^o_{[4]} \rangle \\ \text{ARG-ST} \quad \langle [6] \text{ NP } [nom] [4] : [5] \rangle \\ \\ \text{SEM | CONT} \quad \left[\begin{array}{l} \text{INDEX } [2] \\ \text{RESTR } [3] \end{array} \right] \cup \left\langle \left[\begin{array}{l} \text{RELATION } \mathfrak{R} \\ \text{POSSESSOR } [4] \\ \text{POSSESSED } [2] \end{array} \right] \right\rangle \\ \text{SPR } [3] \langle ([6]) \rangle \end{array} \right)$$

Where [5] *pro* if [3] = $\langle \rangle$

In (3) the specifier requirement is optional, as indicated by parentheses. In the absence of the overt possessor phrase the possessor is interpreted pronominally. *F_{poss}* is a morphological spell-out function that specifies the exponence for particular values associated with Person/Number features. The possessive affix is the realization of the features associated with the INDEX of the possessor argument.

One point that remains unclear from Kathol’s analysis of head-marked possessives is the distribution of features. According to the representation in (3), the possessed noun has its own INDEX represented as [2]. It is further passed to the phrasal category, due to the Semantic Inheritance Principle (Sag & Wasow 1999: 116). This principle ensures that the INDEX value of the NP is identical to that of its head daughter. For example, the word *ter*^o has the 3rd Person INDEX feature and triggers the 3rd Person agreement on the main verb.

$$(4) \quad te-r^o \quad x\text{o}ya \quad / *x\text{o}ya-n^o$$

reindeer-2SG leave.3SG / leave-2SG
‘Your reindeer left’

The subject-verb agreement in this language may refer to INDEX since, first, it is pronominal in nature, and second, it allows semantically motivated feature mismatches, as is typical of INDEX agreement. The pronominality of subject agreement is seen from the fact that overt subjects are not required and in fact overt pronouns in the subject function are very rare. Semantically motivated feature mismatches are illustrated in (5).

⁴ See Ackerman and Nikolaeva (forthc.) for a detailed exposition of Tundra Nenets possessive constructions.

- (5) nyax^or serako/*serako-q ti / *ti -q to^o-q / to^o
 three white/white-PL reindeer/reindeer-PL come-3PL/come.3SG
 ‘Three white reindeer came’

Nouns quantified by numerals must be in the Singular, although they refer to Plural entities. As shown in (5) such nouns must trigger Singular agreement on NP-internal modifiers. Accordingly, they have the Singular CONCORD feature. On the other hand, their INDEX specification includes the Plural feature, which reflects a true semantic property of the expression’s referent. Unlike modifier-head concord, the subject-verb agreement refers either to CONCORD or INDEX, as follows from the variations shown in (5). In the former case the Singular agreement on the verb is a pure reflection of morphosyntactic features of the subject. In latter case the Plural agreement is more semantically motivated.

So, the INDEX of the possessive NP comes from the INDEX of the possessed head noun. On the other hand, in Kathol’s account, the INDEX of the possessor is identical to the INDEX of the specifier and is realized as a Person/Number affix by the morphological spell-out function. In representation (3) these features are not shown. Given the binary typology of Wechsler & Zlatić, the question is then whether they are INDEX or CONCORD features.

Notice that by either analysis we end up with two conflicting values of the same feature. If possessive features are specified in the head noun’s INDEX, then the word *ter^o* has two conflicting values for the attribute Person: the 3rd Person from the possessed nominal and the 2nd Person from the possessor. On the contrary, if possessive features are CONCORD, the possessed noun may have conflicting values of the Number feature. This is shown in (6).

- (6) (pidø^o) serako-q tí-d^o
 you.SG white-PL reindeer-PL.2SG
 ‘your (SG) white reindeer (PL)’

In (6) the Plural head noun triggers Plural agreement on its modifier via CONCORD. But it is also marked as 2nd Person Singular by virtue of being a possessed noun in the possessive relation where the possessor is the 2nd Person Singular. The CONCORD Plural feature and the possessive Person/Number features have a cumulative exponence as the suffix *-d^o*. If possessive features are registered in the CONCORD attribute of the head, this suffix expresses two conflicting values of the CONCORD feature Number: Singular and Plural.

In the next section I will show that this second alternative is in fact correct, that is, possessive Person/Number belongs to the CONCORD specification of the head noun.

2.2 *Agreement on Modifiers*

Nenets shows modifier-head concord in Number and Case. Modifiers include adjectives, modifying nouns and participial relative clauses, but I will only concentrate on adjectives in this paper. (7) shows the attributive concord in Number and Case in non-possessive NPs.

- (7) serako-xøt° te-xøt°
 white-ABL.PL reindeer-ABL.PL
 ‘from white reindeer (PL)’

Number concord is obligatory, while Case concord is highly optional and in fact infrequent.

Crucially, possessive NPs where the possessed noun bears a possessive suffix show another type of NP-internal feature matching: the adjective may take the same possessive affix as the head. Unlike the regular Number concord which exists in all varieties of Tundra Nenets, possessive agreement on adjectives seems to be limited to the Eastern dialectal area. Although it is mostly typical of the archaic language of folklore, it may occasionally occur in everyday speech, and the speakers have clear intuitions on the grammaticality of such constructions. As indicated in (8), possessive agreement is optional.

- (8) a. (møny) serako(-myi) te-myi
 I white-1SG reindeer-1SG
 ‘my white reindeer’
- b. (pidør°) serako(-r°) te-r°
 you.SG white-2SG reindeer-2SG
 ‘your white reindeer’
- c. (pidør°) serako-q / serako-d° tí-d°
 you.SG white-PL / white-PL.2SG reindeer-PL.2SG
 ‘your (SG) white reindeer (PL)’

These examples demonstrate that the head noun and its modifier exhibit matching Person/Number features. Example (8c) also demonstrates an important behavior pattern concerning number agreement, namely, that when the possessed head noun is Plural, the modifier must also show Plural agreement. Additionally it can show possessive agreement in Person and Number, and all these features are expressed in (8c) with the cumulative affix *-d°*. As illustrated in (9), possessive agreement can also accompany Case concord.

- (9) (pidør°) serako-m-t° te-m-t°
 you.SG white-ACC-2SG reindeer-ACC-2SG
 ‘your white reindeer (ACC)’

(9) violates the universal statement mentioned above that disallows agreement in Case and Person on the same target.

Although the data reviewed thus far suggests the existence of an agreement relation between a modifier and its head, the fact that the Person/Number of the possessor participates in this relation raises the question as to whether the agreement is actually between a syntactically independent, albeit optional, possessor and modifier. In other words, what controls possessive agreement on the adjective? The following evidence definitively shows that we are dealing with the true modifier-head concord

here, by demonstrating that the Person/Number features on the adjective are not interpretable as simply reflecting the features of a syntactically expressed possessor.

Consider possessive NPs where the possessor corresponds to a lexical noun. As was shown in the previous subsection, the lexical possessor does not normally trigger possessive agreement. However, a discourse marked lexical possessor can in fact be cross-referenced by a 3rd Person possessive affix on the head. The notion of discourse markedness will be explained later in the paper. At this stage it is important to indicate the contrast between example (1b), without possessive agreement, and example (10), with possessive agreement.

- (10) Wata-h te-da
 Wata-GEN reindeer-3SG
 ‘Wata’s reindeer’

With lexical possessors possessive affixes on the adjective are only possible in the presence of possessive agreement on the head. This is illustrated below. When the adjective bears no possessive marking, the head noun either takes the 3rd Person possessive affix or not (11a). However, when the adjective is marked for Person/Number, the possessive affix is obligatorily present on the head (11b).

- (11) a. Wata-h serako ti / te-da
 Wate-GEN white reindeer / reindeer-3SG
 ‘Wata’s white reindeer’
- b. Wata-h serako-da te-da /* ti
 Wate-GEN white-3SG reindeer-3SG /reindeer
 ‘Wata’s white reindeer’

Thus, when the possessor is lexical the possessive marking on the head is optional. Crucially, adjectival possessive marking is only available in the presence of nominal possessive marking, as in (11b). This indicates that the relationship of feature matching obtains between the adjective and the head noun rather than between the adjective and the possessor. Therefore it is an instance of true modifier-head concord.

Since we can conclude that attributive concord in Tundra Nenets involves Person, it provides a counterexample to proposals that exclude Person from this kind of agreement. It also presents a challenge to representation (3) because, as was discussed at the end of the previous subsection, the feature structure of the head noun accommodates two conflicting values for the same feature.

3. An Analysis

3.1. Pronominality of Person/Number Affixes

In Kathol’s analysis of Luiseño possessive constructions, as presented previously, Person/Number affixes are pronominal, if an independent pronominal possessor is not overt. This is represented as a disjunction on the value of the possessor argument: the possessor either corresponds to an overt specifier NP or is expressed as a Person/Number

affix with a pronominal interpretation. Basically the same situation can be assumed for Nenets as well, as was represented in (3). As in Luiseño, possessive affixes are interpreted pronominally in the absence of the possessor, but in Nenets this can also hold even when the possessor is overt. The claim of this subsection is that the modifier-head possessive concord obtains when possessive affixes on the head are pronominal.

First, I will demonstrate that the Nenets NP has two structural positions for the possessor. The regular possessor is presumably a specifier of the possessive phrase, but there is another possessor position located at its very left periphery. I will refer to this kind of possessor as the peripheral possessor. We have seen in the previous section that a lexical possessor optionally triggers possessive agreement. Agreement correlates with the position of the possessor: while the regular possessor does not trigger agreement, a peripheral one does. The evidence for this claim comes from the position of the possessor with respect to a determiner.⁵ Cf. (12a) and (12b).

- (12) a. tyuku° Wata-h ti / *te-da
 this Wata-GEN reindeer / reindeer-3SG
 ‘this reindeer of Wata’
- b. Wata-htyuku° te-da / *ti
 Wata-GEN this reindeer-3SG / reindeer
 ‘this reindeer of Wata’

If the possessor follows the determiner as in (13a), agreement on the head is impossible. In contrast, when the possessor precedes the determiner as in (13b), it must trigger possessive agreement. A pronominal possessor triggers agreement independently of its position, i.e. whether it precedes or follows the determiner, so it is impossible to determine its position based on the surface form alone.

There is additional syntactic evidence for two types of loci for possessors. What I referred to as the peripheral possessor seems to have some effect on the clausal syntax, although it remains NP-internal. In particular, it participates in switch-reference. Nenets has a so-called Modal Gerund which is used in same-subject adverbial manner clauses. However, as shown by example (13), in the presence of a peripheral possessor subject coreferentiality may be violated. As can be seen, the Gerund is controlled not by the main clause subject (*ngæwada*) but by the peripheral possessor (*Watah*) that triggers possessive agreement. The regular possessor that does not trigger agreement on the head cannot control the Modal Gerund.

- (13) [Ø_i tol°-h tyax°na ngamtyo°] Wata-h_i (*yetryi) ngæwa-da/*ngæwa ye°
 table-GEN at sit.GER Wata-GEN always head-3SG/head hurt.3SG
 ‘When he sits at the table, Wata’s head (always) hurts’

This example also demonstrates that the peripheral possessor remains a subconstituent of the NP. While in some cases it can be fully extracted out of the host phrase, this is not necessarily so. In (13) the possessor cannot be separated from the rest of the NP by other clausal constituents, for example, the adverbial ‘always’. Other constituency tests, such as questioning and coordination, also point towards its NP-internal position.

⁵ The so-called demonstrative pronouns in this language function as determiners.

So the peripheral possessor differs from the regular possessor in that it triggers possessive agreement when lexical, can control switch-reference and precedes the determiner. This indicates that the NP has an additional possessor position located “higher” than the regular possessor. The status and the syntax of this position is the matter of a separate discussion, which is outside the scope of this paper.⁶ What is important is that the peripheral possessor is in non-local, or at least “less” local configuration with the head noun on which it triggers agreement.

This suggests that possessive agreement between the peripheral possessor and the head is anaphoric in the sense of Bresnan & Mchombo (1987) and Bresnan (2000). In their theory, grammatical agreement obtains with elements selected by an argument-taking predicate. Such arguments must be expressed by syntactically independent elements within the phrase structure headed by the predicate or be marked on the predicate itself, so grammatical agreement is structurally local. If the latter situation obtains, the agreement marker itself can satisfy the selectional requirement of the head, functioning as an incorporated pronoun. When an overt antecedent is independently expressed as well, a feature matching relation between the antecedent and the incorporated pronoun is referred to by Bresnan and Mchombo as anaphoric agreement. This relation can occur outside a local domain, because there is no requirement for non-arguments to be local. So, non-local agreement is unambiguously anaphoric and acts in tandem with pronominal incorporation.

An additional argument for the anaphoric nature of agreement between the head and the peripheral possessor comes from the clausal function attributed to the latter. In Bresnan & Mchombo’s original analysis of Chicheŵa the antecedent of an incorporated pronominal has the discourse function of topic and is generated as some kind of adjunct. This is argued to follow from the independent assumption within Bresnan’s Lexical Functional Grammar that only a single argument can serve to satisfy each of the selectional demands of a predicator (the LFG’s principle of Functional Uniqueness). Since the incorporated pronominal satisfies the demands of the predicate, the overt independent element cannot do this as well. So if possessive affixes are pronominal, an overt co-referring peripheral possessor is predicted to fail to satisfy the selectional requirement of the head. This prediction turns out to be true.

The Nenets peripheral possessor normally functions as topic, as demonstrated by the next example.

- (14) a. What about this girl?
b. tyuku^o nye ngøcyeki-h bantø-da / *bant^o ngarka
this woman child-GEN ribbon-3SG / ribbon big
‘This girl’s ribbon is big’

After Gundel (1988) and others, I assume here that the context ‘what about X?’ establishes the topical role of the element X in the answer. As can be seen from (14b), the topical possessor must trigger possessive agreement and therefore is characterized as peripheral. Consider now (15).

⁶ Under the DP analysis this position can correspond to the Spec DP, as in fact was suggested by Szabolsci (1987, 1994, and other works) for Hungarian, a language distantly related to Nenets, where a similar, though not identical, situation is observed. Alternatively, it may be associated with a functional projection on its own or adjoined to a minimal NP.

- (15) a. Whose ribbon is big?
 b. tyuku° nye ngöcyeki-h bantø / *bant°-da ngarka
 this woman child-GEN ribbon / ribbon-3SG big
 ‘This girl’s ribbon is big’

The context (15a) ensures that the possessor in (15b) cannot be interpreted topically. In fact, it has a focus function. In this situation possessive agreement and therefore the peripheral possessor are ungrammatical. So when the possessor is peripheral, it has some kind of discourse marked function comparable to topic, rather than an argument possessor function.⁷ This is expected if the possessive affix is pronominal. The relationship between the two can be characterized as a non-local anaphoric agreement.

Crucially, it is exactly in this situation when the modifier-head Person concord can occur. First, we have seen in (11) that with lexical possessors a possessive affix on the modifier is available when there is a possessive affix on the head. As I have just argued, the agreeing lexical possessor is peripheral. Second, agreement does not disambiguate between the regular and peripheral pronominal possessors. However, (16) demonstrates that possessive concord depends on the position of the possessor.

- (16) a. pidø° tyuku° serako / serako-r° te-r°
 you this white / white-2SG reindeer-2SG
 b. tyuku° pidø° serako / *serako-r° te-r°
 this you white / white-2SG reindeer-2SG
 ‘this white reindeer of yours’

In (16a) the possessor precedes the determiner and so is peripheral. The pronominal possessive affix stands in a non-local configuration with its antecedent. In this situation the possessive concord on the modifier is available. In contrast, in (16b) the regular possessor follows the determiner and therefore must be in the local specifier position. It satisfies the argument requirement, while the possessive affix is simply a grammatical agreement marker. Possessive concord is here ungrammatical.

This data shows that possessive concord correlates with the pronominal interpretation of possessive affixes on the head noun, which satisfy its possessor requirement. The possessor is either absent or structurally non-local to the head and has a non-governable discourse function. So if possessive affixes are present both on the possessed noun and its modifier, their status is different. In the former case they are incorporated pronouns, while in the latter case they are simply affixes of grammatical concord.

⁷ It should be noted that in some cases the NP-internal peripheral possessor is an unlikely clausal topic. Instead it is interpreted as an element prominent in the interpretation of the respective NP. That is, its discourse status is still marked compared to the regular possessor. I will not address such cases here (more discussion on this can be found in Nikolaeva, *forthc.*), but they seem to demonstrate that the inventory of discourse functions is larger than was originally thought, cf. more recent LFG publications, for example, Butt & King (1996).

3.2 Index-to-Concord Principle

If possessive affixes on the head are analyzed as incorporated pronouns, we are dealing with a kind of mismatch between morphology and function. On the one hand, possessive affixes are pronouns and therefore have referential indices. For example, the incorporated pronoun *-r^o* in the word *ter^o* in (4) has the features [PERS 2, NUM *sg*] in its INDEX anchored to the addressee of the respective utterance. On the other hand, they are bound morphemes. The lexical rule of possessive formation creates a complex morphological object where two entities each with its own set of INDEX features are combined within one morphological word. A noun associated with a referent cannot have multiple INDEX values, since referential indices are reflections of the anchoring conditions. As we have seen, the INDEX of the NP comes from the INDEX of its head noun rather than the incorporated pronoun.

I therefore suggest that the INDEX features of incorporated pronouns are specified in the head noun's CONCORD attribute. CONCORD is a pure result of structure sharing and has little, if any, semantic motivation, so stacking several CONCORD features does not lead to a collapse of the semantic interpretation. In principle, this situation should arise each time a single lexical head contains multiple values for distinct arguments, e.g. when a verb agrees with two or more arguments. NP-internally a similar situation is demonstrated by double case constructions, as represented in some languages of Australia. The idea that a single noun can have two or more conflicting values of the CONCORD Case feature has been formalized in Malouf (2000). He suggests a Case Concord Principle that ensures that a dependent NP copies the Case of the head, so that its Case value consists minimally of the Case value of its head and another locally assigned Case. The Case Realization Principle then maps the morphosyntactic Case feature onto a morphological realization and the resulting word can take more than one Case affix.

The situation in Nenets is partly reminiscent of this in the sense that a noun can carry two conflicting Number features, and they both belong to the attribute CONCORD. This is because CONCORD includes the INDEX of the incorporated possessive pronoun in addition to the Number feature that comes from the head. Since both CONCORD and INDEX make reference to Number, the possessed head noun may have two conflicting values of the Number feature.⁸ For example, the word *tíd^o* in (6) bears the Plural and the 2nd Person Singular CONCORD features. The Plural comes from the CONCORD value of the head, while the 2nd Person Plural comes from the INDEX of the incorporated pronominal. As was shown in (8c), both Number features participate in the NP-internal concord and can be copied on the modifier.

This situation can be represented as a constraint on heads. I will refer to it as the Index-to-Concord Principle and represent it as follows.⁹

$$(17) \left[\begin{array}{c|c} \text{HEAD} & \text{CONCORD [1]} \\ \text{SPR} & \textit{pro} [2] \langle \rangle \end{array} \right] \rightarrow \left[\text{HEAD} \mid \text{CONCORD} \langle [1] \oplus [2] \rangle \right]$$

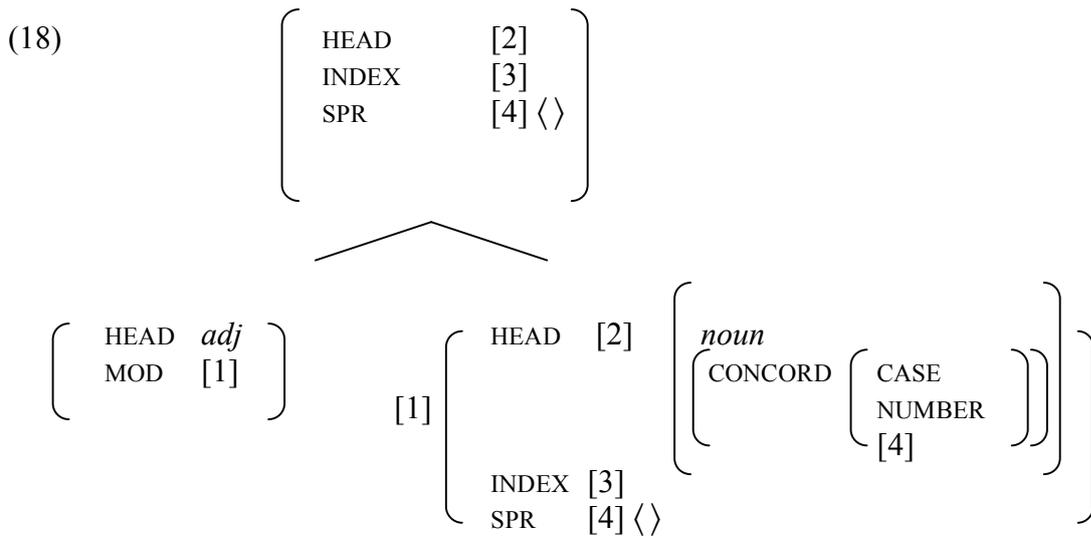
⁸ I assume after Kathol (1999) that non-possessed nouns do not have Person as part of their CONCORD specification.

⁹ I use the list addition sign \oplus to indicate that the value of the CONCORD feature is a list of features: CONCORD values of the head are added to the INDEX values of the dependent, which results in multiple values for the same feature.

This principle ensures that the CONCORD value of the possessed nominal consists of the CONCORD value of the head noun with the addition of INDEX features associated with the specifier. This has two consequences: first, the attribute CONCORD has multiple values for the feature Number; second, it includes the feature Person.

Additionally, the Index-to-Concord principle indicates that the specifier is interpreted pronominally. The pronominal specifier is realized as a bound possessive affix on the head by the morphological spell-out function, as shown in (3). The question of the morphological expression of the stacked CONCORD features is, strictly speaking, independent of the analysis of agreement patterns and therefore is left outside the scope of this paper. I simply assume a list of realizational relationships that obtain between the morphosyntactic characteristics of the head and their cumulative morphological exponence, as described in Salminen (1997). For instance, the combination of the Plural, the 2nd Person Plural and the Nominative Case is realized as the suffix *-d°*.¹⁰

On the proposed account, attributive concord is ensured via the usual mechanism within HPSG. The combination of a noun and its adjectival modifier into a well-formed constituent structure is licensed by the Head-Adjunct Schema which specifies structure-sharing between the head daughter and the MOD value of the adjunct daughter (Pollard & Sag 1994: 56). For Nenets possessed nouns where the adjective shows Person/Number concord with the head it is represented by the following structure.



As indicated in this representation, there is no phrasal specifier. By the lexical rule introduced in (3) the specifier is associated with the possessor argument and is realized as a pronominal affix on the head. The specifier's INDEX is registered in the CONCORD attribute of the head together with the other CONCORD features present in the noun's HEAD field, due to the Index-to-Concord principle (17). As a result, the adjectival modifier shares the value of the stacked CONCORD features of its head.

As follows from this analysis, the Index-to-Concord Principle is applicable to those languages that have head-marked possessives and attributive concord, and where possessive affixes on the head are interpreted pronominally. This combination of

¹⁰ This realization perspective is further developed in Ackerman and Nikolaeva (forthc.).

properties does not seem to be widespread, which may explain why previous research has excluded the possibility of modifier-head Person concord across languages.¹¹

4. Conclusion

The purpose of this paper was to contribute to the cross-linguistic profile of attributive concord. Tundra Nenets provides a counterexample to previous claims that Person never participates in this type of agreement. This is important for two reasons. First, this bears on the more general question of whether agreement can be split into two different relations based on the syntactic domain in which it holds. NP-external and NP-internal (modifier-head) agreement have been said to involve different features: the former cannot be based on Case, while the latter cannot involve Person. However, there are examples of NP-external Case agreement (e.g. Comrie 1997), and the Nenets data cited in this paper shows that NP-internal Person concord is also available. This means that at least with respect to the relevant features no principled difference exists between NP-internal and NP-external agreement.

A noun can bear different (sometimes conflicting) sets of agreement features which participate in different agreement processes referred to as INDEX and CONCORD in recent HPSG publications. Subdividing agreement into these two relations is orthogonal to the question of domains, because at least CONCORD can hold both within an NP and NP-externally. The present treatment has also shown that, contrary to the conventional claims implemented most recently in Weschler & Zlatić (2003), these two types of agreement do not necessarily involve different features: while Case is excluded from the INDEX relation, nothing prevents Person from participating in CONCORD. Thus, syntactic domain, morphosyntactic feature inventories, and the grammatical processes that ensure agreement appear to be independent parameters, although we might be able to talk about some frequent cross-linguistic correlations between them. This conclusion argues for a gradient approach to agreement where the notion of domain plays no essential role (cf. Corbett *forthc.* a, b).

Second, the paper has touched on pronominal incorporation. The modifier-head concord in Nenets involves some features that come from the referential index associated with incorporated pronouns. That is, at first glance incorporated pronouns are fully functionally identical to free standing pronouns in that they seem to be able to function as agreement controllers, in violation of lexicalist assumptions. However, the paper has introduced the Index-to-Concord Principle, which suggests that the referential features of incorporated pronouns are “passed” to the host word and can participate in the concord relation triggered by it.¹² In other words, incorporated pronouns do not control pronominal index agreement, unlike their free standing counterparts.

¹¹ However, Tundra Nenets is not unique. Modifier-head Person concord exists in the related Samoyed languages Nganasan and Enets, but evidence about them is scarce. Outside Samoyed it is attested in Evenki (Tungus), but in this language it is only available on relative clauses. This has some interesting consequences for the analysis, but I leave them for another occasion.

¹² This principle may have a wider application, not necessarily NP-internally. Nenets seems to provide another example: it has a class of adverbials which match in features the pronominal subject agreement affixes on the verb even in the absence of an overt subject.

Abbreviations

ACC – Accusative, DU – Dual, GEN – Genitive, GER – Gerund, NOM – Nominative, PL – Plural, PRET – Preterit, SG – Singular.

References

- Ackerman, F. & I. Nikolaeva (forthc.), *Comparative grammar and grammatical theory: A construction-based study of morphosyntax*, Stanford, CSLI.
- Anderson, S. (1992), *A-morphous morphology*, Cambridge, Cambridge University Press.
- Barlow, M. & Ch. Ferguson (eds.) (1988), *Agreement in natural languages: Approaches, theories, description*, Stanford, Stanford University.
- Bresnan, J. (2001), *Lexical-functional syntax*, Malden-Massachusetts, Blackwell.
- Bresnan, J. & S. Mchombo (1987), *Topic, pronoun, and agreement in Chicheŵa*, “Language” 63, 741–83.
- Butt, M. & T.H. King (1996), *Structural Topic and Focus without Movement*, in M. Butt & T.H. King (eds.), *Online Proceedings of LFG 1996*. <http://csli-publications.stanford.edu/LFG/1/lfg1.html>.
- Comrie, B. (1997), *The typology of predicate case marking*, in J. Bybee, J. Haiman & S.A. Thompson (eds.), *Essays on language function and language type. Dedicated to T. Givón*, Amsterdam-Philadelphia, John Benjamins, 39–50.
- Corbett, G. (forthc. a), *Agreement: Canonical instances and the extent of the phenomenon*, in J. De Cesaris et al. (eds.), *Proceedings of the Third Mediterranean Morphology Meeting*, Barcelona, September 2001.
- Corbett, G. (forthc. b), *Agreement: terms and boundaries*, in W. Griffin (ed.), *The role of agreement in natural language: Proceedings of the 2001 Texas Linguistic Society Conference*, Austin, Texas, 2–4 March 2001.
- Gazdar, G. et al. (1985), *Generalized phrase structure grammar*, Oxford, Blackwell.
- Gundel, J. (1988), *The role of topic and comment in linguistic theory*, New York-London, Garland.
- Lapointe, S. (1988), *Toward a unified theory of agreement*, in M. Barlow & Ch. Ferguson (eds.), 67–87.
- Lehmann, C. (1982), *Universal and typological aspects of agreement*, in H. Seiler & F.J. Stachowiak (eds.), *Apprehension: Das sprachliche Erfassen von Gegenständen*, Vol 2, *Die Techniken und ihr Zusammenhang in Einzelsprachen*, Tübingen, Gunter Narr Verlag, 201–267.
- Lehmann, C. (1983), *Rektion und syntaktische Relationen*, “Folia linguistica” 17, 339–78.
- Lehmann, C. (1988), *On the function of agreement*, in M. Barlow & Ch. Ferguson (eds.), 55–65.
- Nikolaeva, I. (forthc.), *Agreement and linguistic construal*, in J. Fernandez-Vest (ed.), *Uralic languages today*, Paris.
- Plank, F. (1994), *What agrees with what in what, generally speaking?*, “EUROTYP Working Papers”, VII-23, 39–58.
- Pollard, C. & I. Sag (1994), *Head-driven phrase structure grammar*, Chicago, University of Chicago Press.
- Sag, I. & Th. Wasow (1999), *Syntactic theory: a formal introduction*, Stanford, CSLI.
- Szabolcsi, A. (1994), *The Noun phrase*, in F. Kiefer & K.É. Kiss (eds.), *The syntactic structure of Hungarian. Syntax and Semantics 27*, San Diego-New York-Boston-London-Sydney-Tokyo-Toronto, Academic Press, 179–274.
- Wechsler, S. & L. Zlatić (2000), *A theory of agreement and its application to Serbo-Croatian*, “Language”, 76, 799–832.
- Wechsler, S. & L. Zlatić (2003), *The many faces of agreement*, Stanford, CSLI.

Unexpressed Features of Verb Agreement in Signed Languages

Christian Rathmann
The Ohio State University
rathmann.1@osu.edu

Gaurav Mathur
Haskins Laboratories, Yale University
mathur@haskins.yale.edu

1. Introduction

The goals of this paper are two-fold. First, it examines the extent of cross-linguistic variation in the expression of person and number features through verb agreement in three signed languages: American Sign Language (ASL), German Sign Language (DGS) and Japanese Sign Language (Nihon Shuwa). Second, it discusses how two different morphological approaches handle the cross-linguistic phenomena revealed in the study.

Background: The working definition of verb agreement adopted here is a syntactic relationship between a verb and its arguments that is encoded by a morphological process expressing the features of the arguments.

In all the signed languages documented to date, verbs fall into one of three inflectional classes depending on their argument structure. The first inflectional class includes all verbs that have two animate arguments as part of their argument structure. The second inflectional class involves verbs of motion and location, while the third inflectional class contains the rest of verbs: intransitives, and transitives that have one animate argument along with other inanimate arguments. These inflectional classes correspond roughly to Padden's (1983, 1990) classes of agreement, spatial and plain verbs, which are based on morphological criteria rather than argument structure.

The paper focuses on the first inflectional class, since only verbs in this class show agreement with their arguments in person and number.¹ Verbs in the second inflectional class agree with their arguments in different features, which require separate treatment. The verbs in the last inflectional class do not exhibit any agreement. While some of these verbs may be modulated for aspectual inflection like continuative, iterative and habitual, this inflection is distinct from agreement with an animate argument. Moreover, some verbs have more than one meaning; each meaning may be associated with a different argument structure so that a verb may appear in more than one inflectional class. For example, the ASL verb TEACH can select for two animate arguments (as in *I teach him*) and appear in the first inflectional class, or it can select for one animate argument and an inanimate argument (e.g. *I teach mathematics*) and appear in the third inflectional class. The paper focuses on those senses that fit the argument structure of verbs in the first inflectional class.

¹ Gender and other possible agreement features do not seem to play a role in verb agreement in the signed languages researched to date. This is also true for Nihon Shuwa, even though it has "gender morphemes" that appear throughout its lexicon (Supalla and Osugi 1996, Fischer 1996).

Roadmap: The paper starts with a description of verb agreement in the three signed languages. Specifically, it shows how person and number features are expressed (section 2). Next, it turns to cases where the expression of person and number features is blocked for some reason and introduces the notion of unexpressed features (section 3). The next section clarifies how these unexpressed features constitute special cases of syncretism and points out unique features of these cases (section 4). To account for the case of syncretism, two approaches are introduced and compared: an inferential-realizational approach and a lexical-realizational approach (section 5).

2. Person and Number Features

2.1 Person

The person feature may be theoretically decomposed into two subfeatures, [+/- 1] and [+/- 2] (Noyer 1992, Halle 1997 and Frampton 2002). The combinations of these subfeatures yield the following values: [+1], [+2] = first person inclusive; [+1], [-2] = first person exclusive; [-1], [+2] = second person; and [-1], [-2] = third person.

In the case of signed languages, it is not necessary to use the subfeature [+/- 2] for two reasons. First, there is no grammatical distinction between second and third person (Meier 1990). For example, the pronoun for second person is identical to the pronoun for third person; the distinction is seen only at the pragmatic level. Second, there seems to be no distinction between inclusive and exclusive first person at the grammatical level; rather, the distinction is made at the pragmatic level. There are no pronouns that are just inclusive nor are there pronouns that are purely exclusive (Cormier 2002). If there is no formal distinction between second and third person, and if there is no linguistic distinction between exclusive and inclusive first person, it is sufficient to use just the [+/- 1] subfeature for signed languages.

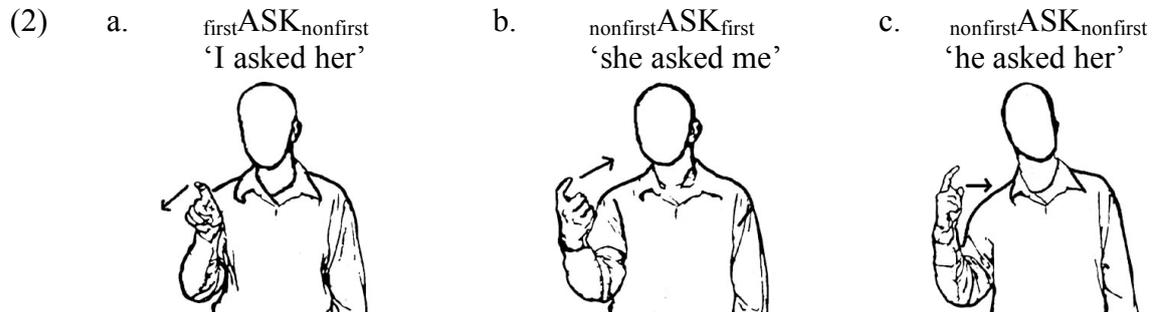
- (1) Person features for signed languages
 - [+1] = first person
 - [-1] = nonfirst person

All the signed languages mark agreement with these features in the same way. Some verbs mark the person feature of the object only (called ‘single agreement’) while other verbs mark the person feature of both the object and the subject (called ‘double agreement’). Agreement is manifested through a change in the direction of movement and/or orientation of the verb so that the hand points toward the location of the object referent (and away from the location of the subject referent).

The location for first person referent is the center of the signer’s chest. The location for a nonfirst person referent corresponds to one’s conceptualization of it within signing space, defined roughly as the empty area in front of the signer’s body. (Rathmann and Mathur 2002, see also Aronoff, Meir and Sandler 2000, Lillo-Martin 2002 and Liddell 2003).

Thus, for a first person subject and a nonfirst person object, the verb moves from the center of the chest to the location of the nonfirst person referent. At the same time, the palm of the hand faces the location of the nonfirst person referent. This is illustrated

below with the ASL sign ASK in (2a). When the person features of the subject and object are reversed (i.e. a nonfirst person subject and a first person object), so is the direction of the movement and the orientation of the palm (see 2b).



The third illustration on the right (2c) shows the form for a nonfirst person subject and a nonfirst person object. In this case, the palm of the hand faces the location of the object referent and the hand moves from the location of the subject referent to that of the object referent.

Below are examples of verbs in each signed language that undergo agreement with the object (and the subject) in its person feature. Some of the verbs change only in direction of movement, while others change only in orientation of the hand, while yet others change in both. It does not matter which specific change occurs, as long as some change occurs to mark the person feature of the object and subject.

(3) Examples of verbs showing person agreement in ASL, DGS and Nihon Shuwa

ASL	DGS	Nihon Shuwa
ASK	BESUCHEN 'visit'	DAMASU 'deceive'
BOTHER	ENTLASSEN 'fire'	HIHAN-SURU 'criticize'
FILM	IGNORIEREN 'ignore'	KOTAERU 'answer'
JOIN	SCHIMPFEN 'bawl-out'	OKORU 'be angry at'
SAY-NO	VERSPOTTEN 'tease'	RENRAKU-SURU 'contact'

There are no differences across the signed languages with respect to the expression of person features.

2.2 *Number*

The signed language literature assumes that there are four possible values for the number feature: singular, dual, exhaustive and multiple (Klima and Bellugi 1979, Padden 1983).

Here, this paper assumes just two values for the feature of number: singular and multiple. It does not count 'exhaustive' as a possible value for the feature of number, because it is assumed for now that the exhaustive form results from several instances of singular agreement, one for each conjoined noun phrase. This is consistent with the

meaning of the exhaustive form that events are distributed over different individuals.² The conjoined agreement forms may be then phonologically reduced. The ‘dual’ form is also not included, since it is taken to be a subcase of the ‘exhaustive’ form, i.e. it consists of two instances of singular agreement, one for each of the noun phrases.

For the purpose of this paper, the two values of the number feature are defined in terms of the binary feature [+/- pl]. The number feature is defined in terms of the plural feature rather than the singular feature, because, as seen below, the plural feature is marked by a morphological process, whereas the singular feature is not marked.

(4) Number features for signed languages

[-pl] = singular

[+pl] = plural

DGS and ASL mark agreement with these features in the same way. Verbs mark the [-pl] feature of a subject or an object through zero marking. All of the examples above show zero marking for number and thus show agreement with a singular subject and a singular object.

Verbs mark the [+pl] feature of an object through the insertion of a horizontal arc into the movement of the verb stem. The overall result is that the hand makes a sweeping motion roughly in the location of the object referent. The [+pl] feature may be marked for a nonfirst person object (see 5a) or for a first person object (see 5b). Note that the plural marking is produced simultaneously as the marking for person, which is manifested through a change in the orientation of the palm.

(5) a. $\text{first}(\text{sg})\text{ASK}_{\text{nonfirst}(\text{pl})}$ ‘I asked them’



b. $\text{nonfirst}(\text{sg})\text{ASK}_{\text{first}(\text{pl})}$ ‘he asked us’



For the [+pl] feature of an subject, there is zero marking (in other words, the ‘multiple’ form is not available for a subject, Padden 1983). Thus, marking for the number feature of the subject is ambiguous between singular and plural in the absence of context.

Not all verbs allow the plural marking for the object. For example, the ASL sign STAB means to stab a person in the back with a knife. It is not possible to stab many people at once. The ‘multiple’ form then cannot be used with verbs that require distributed events for a plural entity. (In such cases, the ‘exhaustive’ form may be used.) Here are examples of verbs that allow plural marking for the object in ASL and DGS.

² Padden (1983) distinguishes a similar form that is done more slowly and clearly for each participant’s location. Here, this difference is taken to be one of specified vs. unspecified individuals, but both still involve singular agreement for each conjoined noun phrase.

(6) Examples of verbs showing number agreement in ASL and DGS

ASL	DGS
ASK	FRAGEN ‘ask’
BAWL	HELFEN ‘help’
FILM	INFORMIEREN ‘inform’
GIVE	VERBESSERN ‘correct’
SAY-NO	VERTEIDIGEN ‘defend’

Nihon Shuwa, unlike ASL and DGS, does not seem to use the ‘multiple’ form regularly. Rather, it uses the singular form for both singular and plural noun phrases. No examples are thus provided from this language.

In sum, there are six possible combinations of features that an agreeing verb in ASL and DGS can show overt marking for:

(7) Combination of features that can be overtly marked on the verb in signed languages:

- a. **First person singular subject and nonfirst person singular object**
(e.g. *I to you*)
- b. **Nonfirst person singular subject and first person singular object**
(e.g. *you to me*)
- c. Nonfirst person singular subject and nonfirst person singular object
(e.g. *you to him*)
- d. **First person singular subject and nonfirst person plural object**
(e.g. *I to y’all*)
- e. **Nonfirst person singular subject and first person plural object**
(e.g. *you to us*)
- f. Nonfirst person singular subject and nonfirst person plural object
(e.g. *you to them*)

Nihon Shuwa, which does not mark plural, shows overt marking only for the combination of features in (7a) through (7c).

Since verbs marking (7b) always mark (7c), and likewise those marking (7e) always mark (7f), the (7b) and (7c) forms are collapsed together, and the (7e) and (7f) forms together. The rest of the paper thus focuses only on four of these combinations (7a, b, d, e). These forms are schematically represented below.

(8)

(7a)	(7b)	(7d)	(7e)
addressee ○ ↑ ○ signer	addressee ○ ↓ ○ signer	addressee ○ ↘ ○ signer	addressee ○ ↘ ○ signer
first sg subj, nonfirst sg obj	nonfirst sg subj, first sg obj	first sg subj, nonfirst pl obj	nonfirst sg subj, first pl obj
‘I to you’	‘you to me’	‘I to y’all’	‘you to us’

Since the expression of person features is the same across the three signed languages, and since the number feature is expressed in the same way in ASL and DGS and not expressed in Nihon Shuwa, ASL data will be used for illustration for ease of exposition, and where relevant, differences from other signed languages will be pointed out.

3. Unexpressed Features

There are verbs that *should* show agreement with a subject and object in person and number yet do not manifest all of the marking overtly. Four examples from ASL illustrate this point. The first three examples (FLATTER, FLIRT and ANALYZE) show that the lack of marking can be due to some phonological reason; the last example (TEST) shows that the lack of marking can be due to the fact that it has not yet been grammaticized as a verb showing agreement.

The first example is the ASL sign FLATTER. It can be modulated to show person. That is, it can mark agreement with first person subject singular and nonfirst person object singular (see 9a) and with nonfirst person subject singular and first person object singular. It, however, cannot be modulated to show plural number, whether this feature is combined with first or nonfirst person. For example, to agree with a nonfirst person plural object, a horizontal arc movement must be inserted into the sign (see 9b). It is not possible to produce this movement simultaneously with the lexical movement of the sign, because they use the same joints of the arm differently. It is also not possible to produce the arc movement after the lexical movement due to a principle of phonological well-formedness that constrains movement in a sign to a complex one. In such cases, ASL forgoes the marking for the plural feature on the verb.

(9) a. $_{\text{first}(\text{sg})}\text{FLATTER}_{\text{nonfirst}(\text{sg})}$



b. $^*_{\text{first}(\text{sg})}\text{FLATTER}_{\text{nonfirst}(\text{pl})}$



Another example is the ASL sign FLIRT, which requires contact between the thumbs of the two hands, as shown in (10a). While the sign can show agreement with nonfirst person singular and plural object, it cannot agree with a first person object because this form violates principles of phonological well-formedness. For instance, one way is to twist the arms inwards, while preserving contact between the thumbs, so that the fingers point to the chest, as depicted in (10b). While this option is articulatorily feasible, it is not possible because the side-by-side relation between the hands is a lexical property that must be preserved. Given that such options are not available, ASL does not express the first person feature on the verb.

(10) a. $\text{first(sg)FLIRT}_{\text{nonfirst(sg)}}$



b. * $\text{nonfirst(sg)FLIRT}_{\text{first(sg)}}$



Yet another example is the ASL sign ANALYZE. It can show agreement with first and nonfirst person singular noun phrases (see 11a); it can also show agreement with a nonfirst person plural object. Yet it cannot show agreement with a first person plural object. The reason is again phonological. This sign involves both hands in an upright posture. To agree with a first person plural object, the arms must be twisted so that the palms face the signer's body; in addition, the arms must move in an horizontal arc (see 11b). This places the nondominant arm in an articulatorily awkward configuration. To avoid this configuration, the language marks only the first person feature, leaving the plural feature unexpressed for a first person object.

(11) a. $\text{first(sg)ANALYZE}_{\text{nonfirst(sg)}}$



b. * $\text{nonfirst(sg)ANALYZE}_{\text{first(pl)}}$



Phonetic/phonological constraints are not the only reason that a verb can fail to mark all the features of a noun phrase. Another reason may be that it takes time for some verbs to become grammaticized as verbs that show agreement. For example, older generations of ASL signers do not express the first person feature of an object on TEST, because the sign has no direction of movement that could be changed under agreement, as seen in (12a). In contrast, a sign like HELP involves path movement, whose direction is readily changed under agreement to show first person object agreement. It is only over time that a change in orientation becomes sufficient for showing agreement on verbs like TEST (see 12b).

(12) a. $\text{first(sg)TEST}_{\text{nonfirst(sg)}}$



b. $\text{nonfirst(sg)TEST}_{\text{first(sg)}}$



Verbs vary in how far they travel along the path of grammaticization from not showing any features to showing features. Variation also appears across generations of

are determined by looking across the paradigms of verbs. If a form is substituted for another form in some but not all paradigms of a verb, syncretism is unidirectional. For example, in the preterite paradigms but not in the other paradigms, a Bulgarian verb's 2nd person singular forms are the same as the third person singular forms (see table 2.3 in Stump 2001: 39).

If the first form is substituted for the second form for some verbs and if it happens the other way around for other verbs, syncretism is bidirectional. For instance, in Rumanian, for verbs in some conjugations, the first person singular form is the same as the third person plural form. For verbs in other conjugations, it is the other way around (see table 7.1 in Stump 2001: 213).

Unstipulated syncretism occurs when there are never distinctive forms for two feature sets in a certain context; if the two feature sets form a natural class, it is sufficient to posit one form for this natural class. In the same example from Rumanian, the third person singular form and the third person plural form are always the same in the present tense for verbs in one conjugation.

If the two feature sets do not form a natural class, yet if there is a systematic syncretism between these sets across paradigms and across verbs, this syncretism is called symmetric. In Rumanian, the first person singular form is the same as the first person plural form in the imperfect tense for all verbs (see table 7.2 in Stump 2001: 215).

The verb agreement patterns in signed languages illustrate two different cases of syncretism. One case is unidirectional while the other case is unstipulated. The first case of syncretism occurs for verbs in the first four columns of the table in (13). Let us go over each column. The first column contains verbs that do not express any features. The forms are syncretized as follows:

(14) Syncretism for verbs that do not express any feature (e.g. KNOW)

(7b) → (7a)

(7d) → (7a)

(7e) → (7a)

This syncretism is unidirectional, because forms for (7b) through (7e), which are distinctive on other verbs, are substituted by the same, singular form. The second column lists verbs that do not express plural features. In such cases, the forms syncretize to singular forms:

(15) Syncretism for verbs that do not express plural features (e.g. FLATTER):

(7d) → (7a)

(7e) → (7b)

This syncretism is unidirectional because there are distinctive forms for (7c) and (7d) on other verbs, yet on the particular verbs above, these forms syncretize to the corresponding singular forms and never the other way around. The next column of verbs do not express the first person feature for an object. In such cases, the forms syncretize to one form, the nonfirst person singular form:

- (16) Syncretism for verbs that do not express first person feature for object (e.g. FLIRT):

(7b) → (7a)

(7e) → (7a)

Note that (7e) could theoretically syncretize to (7b) which preserves the number feature for the object, but this is not what happens. If the person feature syncretizes from first person to nonfirst, so does the number feature from plural to singular. Otherwise, this syncretism is still unidirectional, since there are distinctive forms for (7b) and (7e) (as seen on other verbs). The fourth set of verbs do not express the feature for a first person plural object. In these cases, the form is syncretized to the corresponding singular form.

- (17) Syncretism for verbs that do not express first person plural feature for object

(7e) → (7b)

(e.g. ANALYZE):

Note that this syncretism appears in the second set of verbs. The second set of verbs are actually a subset of the verbs here. As shown above, this syncretism is unidirectional.

All these cases of syncretism are unidirectional. Another case of syncretism is of a different type. This syncretism is unstipulated and occurs in the plural forms for the subject. These forms are always syncretized to the singular form, i.e. there is no distinctive form for the plural feature for a subject. This is true for all verbs in all the signed languages.

- (18) Syncretism in subject number
subject plural → subject singular

In all the cases of syncretism seen so far, there are two features that particularly stand out. First, all of the forms syncretize to the marking for the unmarked feature. Nonfirst person and singular number are both unmarked, so nearly all of the forms syncretize to forms expressing these features. This differs from the usual cases of syncretism seen in spoken languages, which may occur between two marked forms.

The other feature of the syncretism seen above is that it is mostly driven by phonetic-phonological reasons, in contrast to cases of syncretism in spoken languages that can be purely morphological. For example, in the English present-tense paradigm, there is syncretism to a form with a zero affix, i.e. an affix with no phonological content. This contrasts with the affix for the third person singular form, *-s*, which has phonological content. There is no such contrast in the signed languages. Rather, the contrast is between, on the one hand, forms that mark all features overtly and on the other hand, forms that do not mark all of them and that syncretize to forms expressing an unmarked feature, and this contrast is driven by phonetic-phonological factors.

While all three signed languages behave the same way with regard to the above patterns, other signed languages than ASL offer an additional option for expressing the features in case they cannot be expressed on the verb. DGS and Nihon Shuwa may express the features on an auxiliary-like element (called Person Agreement Marker, PAM) (Rathmann 2000). In DGS, PAM may mark singular features (see 19a) or plural features (see 19b), although the latter form is not frequently used. In Nihon Shuwa, PAM has a different phonological form (see 19c) and can mark singular features. It is not clear whether

This approach posits three affixes for subject and object agreement: (i) ‘pl’ which inserts an arc movement onto the verb stem; (ii) ‘1’ which stands for the fixed location of first person, i.e. the chest of the signer; and (iii) ‘0’ which is a placeholder for the location of the non-first person argument (location is to be later matched with content from spatio-temporal conceptual structure at output). At the output of the paradigm function, a process applies that changes the direction of the verb stem according to the locations specified by the affixes. The paradigm holds for both ASL and DGS. The paradigm for Nihon Shuwa is similar except that there are no columns for the plural feature on the object. For DGS and Nihon Shuwa, the features may be expressed on PAM instead of the verb.

The paradigm accommodates all the forms of a verb that shows all features, like ASL ASK. The approach is also able to handle the two types of syncretism seen above. First, the unstipulated syncretism that the verb cannot mark a plural subject is built into the paradigm. In the rows for the subject features, there are no rows that differentiate between singular and plural.

Second, the various cases of unidirectional syncretism are handled by rules of referral (Zwicky 1985) that specify which cells syncretize in which contexts. Here, the relevant contexts are the sets of verbs. Recall that there are four sets of verbs that exhibit varying degrees of syncretism. A rule of referral will be needed for each set of verb:

(21) Rules of referral

- a. If a verb is of the first set (e.g. KNOW), all forms are realized as (7a).
- b. If a verb is of the second set (e.g. FLATTER), the plural form is realized as the singular form.
- c. If a verb is of the third set (e.g. FLIRT), the first person object form is realized as (7a).
- d. If a verb is of the fourth set (e.g. ANALYZE), the first person plural object form is realized as (7b).

Applying these rules of referral to the above paradigm results in the following paradigms, one for each set of verbs.

(22) Paradigms resulting from the application of rules of referral in (16)

1st Set (KNOW)		object			
		first person		non-first person	
		sg	pl	sg	pl
subject	first person			$_1V_0$	
	non-first person				

2nd Set (FLATTER)		object			
		first person		non-first person	
		sg	pl	sg	pl
subject	first person			$_1V_0$	
	non-first person	$_0V_1$		$_0V_0$	

3rd Set (FLIRT)		object			
		first person		non-first person	
		sg	pl	sg	pl
subject	first person			$_1V_{0pl}$	
	non-first person	$_1V_0$			

4th Set (ANALYZE)		object			
		first person		non-first person	
		sg	pl	sg	pl
subject	first person			$_1V_0$	$_1V_{0pl}$
	non-first person	$_0V_1$		$_0V_0$	$_0V_{0pl}$

It is possible that some of the forms are still phonetically unpronounceable, depending on the location used for the non-first person argument. In that case, the verb switches to another paradigm with unmarked forms.

The inferential-realizational approach then relies on rules of referral to handle unexpressed features. The rules of referral are stated in the context of a particular set of verbs; thus it must be stipulated which set a verb belongs to, even though a verb may switch between different paradigms. Apart from the context, the rules of referral are quite similar in that they result in syncretism to the same unmarked forms.

This approach does not make any specific predictions about which direction the development of the agreement system can go in. The paradigms can either become simpler (Carstairs-McCarthy 1991) through increased syncretism to unmarked forms or can become complete through an increased number of distinctive forms for each set of features.

5.2 *Lexical-Realizational Approach*

The lexical-realizational approach assumes that the notion of a paradigm is not required in the grammar. Rather, this approach is based on lists of morphemes, rules for using them and multiple derivations to generate the set of forms. The framework of Distributed Morphology (Halle and Marantz 1993) is used for illustration.

Every word is the result of a series of derivations. In the initial derivation, person and number features of the subject and the object are copied onto the verb. These elements are sent to a ‘morphology’ component that rewrites features through impoverishment rules (Bonet 1991) and spells out the features according to a list of disjunctively ordered morphemes before being submitted to further phonological

processes. If this derivation does not crash due to a violation of a phonetic constraint, the agreement forms are pronounced.

If the derivation crashes, a new derivation is attempted in which the features on the verb are not expressed. In languages with PAM, another option is available: an Agreement Phrase is projected, which is manifested by PAM and the features from the verb; the features are then spelled out on PAM.

After the features for the subject and for the object are copied onto the features of the verb as part of agreement, the features are subject to an impoverishment rule which deletes the plural feature for a subject. This has the effect that a verb (or PAM) never marks a plural subject.

- (23) Impoverishment
[Pl] → ∅ / ___ [subject]

Next, the person and number features are spelled out separately for object and subject agreement respectively. The location for non-first person is left blank, which is to be filled by a location that matches content from spatio-temporal conceptual structure at output.

- (24) Vocabulary items for person agreement
[-1] ↔ location: _____
Else ↔ location: center of chest

Vocabulary items for number agreement
[Pl] ↔ insert movement in horizontal arc convex outwards
Else ↔ ∅

These spell-outs are then subject to the rule that changes the direction of the verb stem (including the morpheme for a plural feature) according to the locations of subject and object agreement affixes. In case the surface form violates some phonetic constraint(s), the derivation crashes, and another derivation is attempted in which the features of the subject and the object are not copied to the verb. The result is that there is no feature to be spelled out on the verb.

The analysis for DGS is the same as that for ASL, with one difference. If the initial derivation crashes, the next derivation can copy the features to PAM. The features are then spelled out just as if they were on a verb. The analysis for Nihon Shuwa is similar, with one difference. Since the ‘multiple’ is rarely used in the language, the morpheme [Pl] is assumed not to be available in the list of items for number agreement. Given just one item, which is a zero morpheme, number agreement becomes vacuous and may be assumed to be absent altogether.

In sum, the lexical-realizational approach assumes a new derivation each time there is a crash at the phonetic-interface, and each successive derivation expresses fewer features in order to converge. The phonetic constraints are then sufficient for determining whether the features are expressed, so that it is not necessary to stipulate which set a verb belongs to.

Finally, the approach makes a specific prediction regarding the development of the morphological system: due to the principle of economy, the number of crashes at the phonetic interface should be minimized over time; this would push more features to be expressed over time, i.e. there would be less syncretism. Various studies are consistent with this prediction. Verbs that do not express any features gradually express them during

language change (Engberg-Pedersen 1991); children acquire verbs that do not express features before verbs that do (Meier 1982); and verbs that do not express features are used more frequently than verbs that do express them in language innovation (Senghas 1995, Nicaraguan Sign Language; Stack 1999, acquisition of Signed Exact English; Abu-Shara Sign Language, Aronoff, Meir, Padden and Sandler 2003).

6. Summary

This paper has made three points. First, person features are expressed through a change in the direction/orientation of the verb, while number features are expressed through the insertion of a horizontal arc into the movement of the verb.

The second point concerns the extent of cross-linguistic variation. There is no cross-linguistic difference in the expression of the person features. That is, ASL, DGS and Nihon Shuwa express person features in the same way. There is, however, cross-linguistic variation in whether the number feature is expressed or not. It is expressed in ASL and DGS but not in Nihon Shuwa. In ASL and DGS, the number feature is expressed in the same way.

There is cross-linguistic pattern with respect to another property. In all of the signed languages, the person (and number for DGS and ASL) features may be unexpressed. Signed languages vary in whether these features remain unexpressed on the verb (ASL) or whether they get expressed on another element like PAM (DGS and Nihon Shuwa).

The last point is that the unexpressed features constitute cases of syncretism. One kind (singular marking for both singular and plural subjects) is unstipulated, while all the other cases are unidirectional. There are two cross-linguistic generalizations regarding the unidirectional cases of syncretism:

- (25) a. If a verb cannot express the first person feature, it uses the form marking the nonfirst person feature with the other features held constant.
- b. If a verb cannot express the plural feature, it uses the form marking the singular feature with the other features held constant.

These patterns support the assumption that first person and plural number constitute marked features. Moreover the number feature is dependent on the person feature as seen in one case of syncretism.

Two realizational approaches handle these unexpressed features in different ways. Under the inferential-realization approach, unexpressed features are encoded within paradigms that are artificially similar to one another. Under the lexical-realizational approach, they are predicted by crashes at the phonetic interface. Minimizing these crashes is also sufficient to predict the path that verbs take in language change, acquisition and innovation: verbs go from not expressing features to expressing features over time.

References

- Aronoff, M., I. Meir & W. Sandler (2000), *Universal and particular aspects of Sign Language morphology*, in K. Grohmann & C. Struijke (eds.), "University of Maryland Working Papers in Linguistics" 10, 1–34.
- Aronoff, M., I. Meir, C. Padden & W. Sandler (2003), *Morphological universals and the sign language type*, paper presented at the Fourth Mediterranean Morphology Meeting (MMM4) in Catania, Sicily.
- Bonet, E. (1991), *Morphology after syntax: pronominal clitics in Romance languages*, Ph.D. dissertation, Massachusetts Institute of Technology.
- Carstairs-McCarthy, A. (1991), *Inflectional classes: two questions with one answer*, in F. Plank (ed.), *Paradigms: the economy of inflection*, The Hague, Mouton de Gruyter, 213–253.
- Cormier, K. (2002), *Grammaticization of indexical signs: how American Sign Language expresses numerosity*, Ph.D. dissertation, The University of Texas at Austin.
- Engberg-Pedersen, E. (1993), *Space in Danish Sign Language*, Hamburg, Signum-Verlag.
- Fischer, S. (1996), *The role of agreement and auxiliaries in sign language*, "Lingua" 98: 103–120.
- Frampton, J. (2002), *Syncretism, impoverishment and the structure of person features*, in *Papers from the 2002 Meeting of the Chicago Linguistic Society*.
- Halle, M. (1997), *Distributed Morphology: impoverishment and fission*, "MIT Working Papers in Linguistics" 30, 425–449.
- Halle, M. & A. Marantz (1993), *Distributed Morphology and the pieces of inflection*, in K. Hale & S.J. Keyser (eds.), *The view from Building 20*, Cambridge, Mass., MIT Press, 111–176.
- Klima, E. & U. Bellugi (1979), *Signs of language*, Cambridge, Mass., Harvard University Press.
- Liddell, S. (2003), *Grammar, gesture and meaning in American Sign Language*, Cambridge, Cambridge University Press.
- Lillo-Martin, D. (2002), *Where are all the modality effects?*, in R. Meier, K. Cormier & D. Quinto-Pozos (eds.), *Modality and structure in signed and spoken languages*, Cambridge, Cambridge University Press, 241–262.
- Meier, R. (1982), *Icons, analogues and morpheme: the acquisition of verb agreement in American Sign Language*, Ph.D. dissertation, University of California, San Diego.
- Meier, R. (1990), *Person deixis in American Sign Language*, in S. Fischer & P. Siple (eds.), *Theoretical issues in sign language research*, Chicago, The University of Chicago Press, 175–190.
- Noyer, R. (1992), *Features, positions and affixes in autonomous morphological structure*, Ph.D. dissertation, Massachusetts Institute of Technology.
- Padden, C. (1983), *Interaction of morphology and syntax in American Sign Language*, Ph.D. dissertation, University of California, San Diego.
- Padden, C. (1990), *The relation between space and grammar in ASL verb morphology*, in C. Lucas (ed.), *Sign language research: theoretical issues*, Washington, D.C., Gallaudet University Press, 118–132.
- Rathmann, C. (2000), *The optionality of Agreement Phrase: evidence from signed languages*, Masters report, The University of Texas at Austin.
- Rathmann, C. & G. Mathur (2002), *Is verb agreement the same cross-modally?*, in R. Meier, K. Cormier & D. Quinto-Pozos (eds.), *Modality and structure in signed and spoken languages*, Cambridge, Cambridge University Press, 370–404.
- Senghas, A. (1995), *Children's contribution to the birth of Nicaraguan Sign Language*, Ph.D. dissertation, Massachusetts Institute of Technology.
- Stack, K. (1999), *Innovation by a child acquiring Signed Exact English II*, Ph.D. dissertation, University of California in Los Angeles.
- Stump, G. (2001), *Inflectional morphology*, Cambridge, Cambridge University Press.
- Supalla, T. & Y. Osugi (1996), *Structural analysis of gender handshapes in Nihon Shuwa*, Paper presented at the 5th Internatl. Conference on Theoretical Issues in Sign Language Research, Montreal.
- Zwicky, A. (1985), *How to describe inflection*, "Berkeley Linguistics Society" 11, 372–386.

The Morphosemantics of Transnumeral Nouns

Paolo Acquaviva
University College Dublin
paolo.acquaviva@ucd.ie

0. Introduction

This paper studies the interaction between number morphology and semantic interpretation on nouns that are semantically neither singular nor plural. After exemplifying the notion of transnumeral nouns in section 1, it will be shown in section 2 that a transnumeral interpretation has morphological reflexes also on nouns which have morphological number; in particular, nouns where singular or plural marking does not straightforwardly correlate with singular or plural semantics tend to be morphologically irregular along similar ways. On the semantic level, section 3 will argue that these common morphological patterns define a semantic class of nouns more precisely characterized as “weakly individuated concepts”. On the morphological level, it will be argued in sections 4 and 5 that the various idiosyncrasies of these nouns have a lot in common, which can be traced back to the fact that number is *not* assigned to the noun by a syntactic [Number] head distinct from [N] (as is normally the case):

A noun may be transnumeral only if it is not assigned number from a separate [Number] head.

This subsumes apparently singular “numberless” nouns, inherent plurals, and even *pluralia tantum* like *scissors*. Beside offering a semantically unified approach to the morphology of *pluralia tantum*, irregular plurals, classifiers and collectives, this analysis also explains under what conditions transnumeral semantics can be compatible with number morphology, and why this cannot happen when number is fused with gender.

1. Point of Departure: Transnumeral Nouns

There are different ways in which a noun may be said to transcend the number opposition.¹ In the clearest case, a noun not formally marked for any number value occurs in a construction that makes it problematic, or impossible, to decide which number it is. Such examples of morphosyntactic transnumerality must be distinguished from the simple property of lacking a number exponent: the English *pen*, for example, has no singular marking, but it is not transnumeral because all and only the occurrences of the noun in the form *pen* are unambiguously singular (both syntactically and interpretively). In certain languages and in certain constructions, however, the lack of explicit number marking correlates with an interpretation that is neither clearly singular nor plural.

¹ Of course, there is no single number opposition, as the comprehensive survey of Corbett (2000) makes clear. What I have to say here applies to nouns that neutralize a number opposition elsewhere present in their respective language, very often falling in what Corbett calls “general number”.

1.1 *Complements to Classifiers*

Classifiers are overt markers of countability, which express a unit of the referent of their complement noun, like *blade* in *a blade of grass* (cf. Greenberg 1974). Although such unit expressions can semantically be analyzed as classifiers even in languages like English (cf. Chierchia 1998), both the unit noun *blade* and its complement *grass* are full lexical nouns: they have autonomous meaning, they can occur without a complement mass noun, and they can be either singular or plural. This last property has particular significance, because it discriminates unit nouns with a classifier semantics from classifiers proper, which are grammaticalized expressions of countability. The English *head* in *three head of cattle*, which lacks the expression of plural otherwise mandatory for nouns in this context, is closer to being a classifier in the morphosyntactic sense.²

The distinctive trait of classifier constructions in the strict sense, however, lies not so much in the classifier itself as in the complement noun. In English, unit nouns like *blade* and the quasi-classifier *head* are followed by mass nouns that are unambiguously singular or, more rarely, plural like *cattle* (we will consider exceptions below). Classifier languages differ in two respects: *all* nouns occur as complement to classifiers in counting contexts (except measures and unit-nouns, which are by themselves expressions of countability), which gives the impression that all nouns are mass; and they are morphosyntactically neither singular nor plural. The languages of South-East Asia, here exemplified by Mandarin Chinese, are the best-known instantiation of this type: morphology just does not provide a number opposition for nouns (apart from a “collective” marker *-men* for animate nouns or pronouns), and in contexts that entail countability (not only after numerals), all nouns must be preceded by a classifier. As Cheng and Sybesma explain (1999: 514–515), some classifiers “create a unit of measure” over a mass like ‘rice’ (*mass-classifiers*), while others apply to conceptually bounded referents like ‘pen’, and “simply name the unit of natural semantic partitioning” (*count-classifiers*):

- | | | |
|-----|--------------------------|--|
| (1) | <i>mass-classifier</i> : | <i>count-classifier</i> : |
| | san ba mi | san zhi bi |
| | 3 hand(ful) rice | 3 CL pen |
| | | (Mandarin Chinese: Cheng and Sybesma 1999) |

Löbel (2000) and Bisang (1999) show further that in a language like Vietnamese the same lexical item can have the function of lexical noun and of classifier:

- | | | |
|-----|---------------|---|
| (2) | hai cái bao | hai bao cam |
| | two thing bag | two bag(fuls) orange (Vietnamese: Löbel 2000) |

Clearly, the noun governed by a classifier is not just morphologically unmarked for number (which could in principle be an accident of the inflectional morphology of these languages), but lacks any syntactic or even semantic characterization as either singular or plural. Such “concept nouns” (Rijkhoff 1991), which Chierchia (1999) analyzes as

² Multipliers like *dozen* or *hundred* can also appear as invariable singulars (*three dozen / hundred students*), but they differ from classifiers in that their complement noun must be independently countable. In English, this correlates with the lack of preposition *of* before the head noun.

kind-referring expressions, do not designate one or multiple entities: as such, they are transnumeral.

1.2 *Formally* [SG] *Nouns After Numerals* >2

Transnumerality emerges in a different fashion in languages that, unlike those of South-East Asia, have a well-established number opposition in nominal morphology and syntax. A typical case involves the use of formally singular nouns in a semantically plural context. In agglutinating languages where a plural suffix is attached to the base singular form, numerals often govern what is morphologically the singular form:

- (3) két kocsi (Hungarian; Uralic languages generally)
2 car.SG
- (4) iki ev (Turkish; Turkic languages generally)
2 house.SG

As Corbett (2000: 211) notes, the absence of plural marking on semantically plural nouns is typologically most common for nouns governed by numerals, which is unsurprising because formal marking is redundant where plurality is semantically implied. But this does not explain why this is much more common in morphologically agglutinating languages than in inflecting / fusional ones. In fact, the use of singular after semantically plural numerals is but a facet of a more general pattern: where the plural is morphologically an extension of the singular (typically arising from suffixation of a non-suffixed singular), the latter form can typically be used with an interpretation as kind, or as group:

- (5) a. a bálma a legnagyobb emlosállat (Hungarian: Rounds 2001: 91)
the whale.SG A largest mammal.SG
'whales are the largest mammals'
- b. az alma a sarokban, a költe a porcon van
the apple A corner.LOC, the pear A shelf.LOC are
'the apples are in the corner, the pears are on the shelf'
- (6) polis 'the police, the policeman' (Turkish: Lewis 1967: 26)
bir polis 'a policeman'

Viewed in this context, the "singular" after plural numbers is not really a singular at all, but a base form morphologically and semantically unspecified for number. Unlike the previous case, transnumeral nouns in such agglutinating languages are formally members of a regular number opposition (hence their traditional label of "singulars"); but the wide availability of a semantically non-singular interpretation shows that the number opposition is more aptly analyzed as "base vs. plural" than as "singular vs. plural".

Russian seems to provide a counterexample to the claim that a "singular" noun form after plural numbers is in fact a numberless base form. As is well known, the numbers 2-3-4 seem to govern a singular form (in the genitive case) which is *not* a bare stem on which plural is affixed:

- (7) dva žurnal-a (PL.NOM žurnal-y, PL.GEN žurnal-ov) (Russian)
 2 journal.SG.GEN

In fact, there are independent reasons to view this as an apparent counterexample. First, the singular is only mandated if the noun phrase appears in the nominative case (and accusative when the two are identical); second, an adjective modifying the putative genitive singular noun is *plural* (with nominative or genitive case); third, the “genitive singular” form used after ‘two’ carries in some nouns a different stress from that of the regular genitive singular.

As Corbett (1993) has expressly argued, this is enough evidence to consider *zurnala* in (7) a special form of the noun mandated by the governing ‘two’, identical with the genitive singular form but synchronically distinct from it, in particular not marked [singular] for agreement purposes.

1.3 Base to Singulative Affixation

Singulative affixes derive nouns interpreted as single individuals (objects or events). Given this discretizing function, the singulative derivation therefore presupposes a class of nouns with transnumeral interpretation, in so far as they derive individual referents from bases that, regardless of their grammatical number, must be interpretively distinct from both singular individuals and plural aggregates. The Arabic derivations known as “unit noun” (*ism l-wahda*) and “instance noun” (*ism l-marra*) provide the clearest and best-known example of a morphological process that derives an individual entity or event from a base noun interpreted as a mass, as an activity predicate, or as a pure property:

- (8) a. baqar^{unn} ‘cattle’ – baqarat^{unn} ‘cow’
 b. hadiid^{unn} ‘iron’ – hadiidat^{unn} ‘piece of iron’ (classical Arabic)
- (9) ‘akil ‘food’ – ‘akla ‘a meal’ (Gulf Arabic; Qafisheh 1977)
- (10) boos ‘kissing’ – boose ‘a kiss’ (Syrian Arabic; Cowell 1964)

The tight relation between the interpretation of nouns that serve as bases for singulative derivation and that of complements to classifiers comes to the fore in the Omani dialect, where Greenberg (1974) has documented the simultaneous existence of both constructions:

- (11) a. baqar ‘cattle’ – baqra (fem) ‘cow’
 b. thalaath baqraat ‘3 cows’ (3 + N.FEM.PL)
 c. thalaathit rwaas baqar ‘3 cows’ (3 + CL + N) (Omani Arabic; Greenberg 1974)

As can be seen, the discretization into individuals, required by the numerical construction, can be achieved either by resorting to a singulative like *baqrat*, or by having the uncountable base-form *baqra* governed by an individualizing classifier.

The distribution of singulatives in Breton sheds further light on the transnumeral interpretation of the nominal bases which singulatives are derived from. The singulative suffix *-enn* turns into feminine nouns with individual referents bases with various interpretations:

- (12) a. *collectives:* (Breton: Trépos 1957)
plouz ‘straw’ → eur blouzenn ‘a straw’
stered ‘stars’ → eur steredenn ‘a star’
- b. *plurals:*
bran ‘crow’, brini ‘crows’ → brinienn ‘a crow’
- c. *singulatives:*
lod ‘part’ → lodenn ‘part’

In the examples in (12a), the input to singulative derivation is a mass noun, whether grammatically singular like *plouz* or plural like *stered* (cf. the English *clothing* and *clothes*, neither of which is countable). The transnumeral interpretation of the input is less obvious in (12b), where the singulative is formed by suffixation of a plural which, unlike *stered*, has its own unsuffixed singular. Apparently, a plural like *brini* is liable to being interpreted as a collective mass (like *cattle*), which the singulative suffix makes countable. The most surprising case is (12c), where the singulative attaches to a base which, judging by the gloss, is already every bit as countable as the output. The explanation by Trépos (1957: 268) is enlightening: ‘le suffixe *-enn* rend l’objet plus proche, plus materiel, plus tangible; c’est ainsi que *lod* désigne plutôt la part lorsque le partage n’est pas encore fait: *peb hini* ‘no e *lod* ‘chacun aura sa part’, et *lodenn* la part que chacun reçoit: *brasoc’h eo e lodenn* ‘sa part est plus grande’. The unsuffixed basis, then, refers to an abstract equivalence class rather than an actual individual object. *Lod* does not refer to a mass or a kind, or to a referent conceptualized as plural without being an aggregate of salient individuals (such as *brini*); still, it can feed singulative derivation. This suggests a connection between the interpretation as an equivalence class and the interpretations of referents that are neither singular nor plural (typically mass or kind), and this connection leads us to an empirical domain traditionally disregarded in the analysis of transnumeral nouns.

2. The Irregularity of Number on Unit Nouns

That measure nouns often show irregular morphology is well known. But their morphological idiosyncrasies should be seen in the context of the morphology and semantics of transnumeral nouns. The examples overviewed in this section will show that a host of unit concepts, not just measure nouns, display a *certain* kind of irregularity which is strongly reminiscent of the transnumeral status of classifiers, although in these cases we are dealing with nouns with morphological number.

2.1 Exceptionally Singular Measure Terms in Germanic

In English (especially in its European dialects), many units of measurement are irregular with respect to morphosyntactic number: they can, or sometimes must, appear as singular nouns in a context that would mandate the plural for all other nouns.

Expressions that are part of the counting system (“large numbers”: *dozen, score, hundred, thousand, million*) would appear to be nouns, in so far as they can all appear as single complements of the singular indefinite article and all can be suffixed by the plural *-s*. Distributionally, however, they resemble classifiers more than lexical nouns, because they can be followed immediately by a head noun, without an intervening preposition (*one hundred pens*). The crucial observation is that they all can appear in the singular after semantically plural determiners (numbers above 1 or count determiners like *a few*):

(13) three dozen / score / hundred / thousand / million (pens)

Note that the lexical noun is not obligatory, and its presence has no bearing on the morphological number of these numerical expression. Together with the fact that the plural form is generally available (although usage varies), this shows that we are indeed in the presence of a morphosyntactic irregularity: these units of counting *can* behave just like any other noun, but the expression of the plural is liable to being suspended.

The same occurs with units of measurement that are unambiguously nouns: semantically, they define a dimension (space for *fathom*, weight for *pound*, otherwise monetary value) in addition to a quantification; syntactically they cannot be immediately followed by a noun.

(14) three bob / quid / pound / cent / Euro / fathom

Indeed, the plural is morphologically ill-formed for *bob* and *quid*.

This irregular singular in a plural context should not be confused with the singular of phrases like *three foot long*, where the measurement appears as a pre-nominal or pre-adjectival modifier. The singular in this latter construction is generalized to *all* nouns provided they can have a unit interpretation (*a three-page document, three year old*).

The irregular singular for measure terms is even more prominent in German. The “large numbers” 100 and 1,000 are full-fledged nouns (with regular plural) if and only if they refer to sets of individuals (*Hundert sind gestorben* ‘hundreds died’); otherwise, they are invariable and orthographically attached to the governing number (*dreihundert Leute* ‘three hundred people’). Units of quantity (monetary or otherwise) are instead obligatorily singular:

(15) drei Mark / Pfund / Kilo / Gramm / Mann / Fuss / Faden
‘3 mark.SG / pound.SG / kilo.SG / gram.SG / man.SG / foot.SG / fathom.SG’

I have included *Mann* ‘man’, as a unit measuring the numerical strength of groups (often in a military context). German also allows, with a number of unit nouns, the construction that English restricts to *head* in *three head of cattle*; the classifier function

of such unit nouns is in German further enhanced by the lack of a preposition in front of the lexical noun:

- (16) drei Sack Kohle drei Glas Wein drei Korb Kartoffeln
 3 sack.SG coal 3 glass.SG wine 3 basket.SG potatoes

Usage varies greatly, and speakers disagree on the set of nouns that can be thus employed (partly, this has cultural reasons: measuring commodities by traditional containers is much less common today than fifty years ago). However, variation does not obscure the irregularity of unit nouns with respect to morphological number.

2.2 *Exceptional Plurals in Irish and Italian*

Irish and Italian provide two more genetically unconnected examples of the way irregularities in morphological number affect a class of nouns that centres on units of measurement but, crucially, extends beyond this class.

The Irish data concern a class of exceptions to the general pattern of morphosyntactic number in numerically quantified noun phrases: a noun governed by 3-10 is generally singular, but some nouns appear in the plural. Abstracting away from considerable dialectal variation and the complications of numerical quantification in Celtic (cf. Ó Siadhail 1982, Acquaviva 2004), the irregular use of plural after 3-10 is characterized by two main features: first, morphologically, there are some nouns that have a special plural form only employed after numerals 3-10; second, the nouns that exceptionally appear in the plural (whether the regular plural or a special form) after 3-10 in all dialects comprise units of measurement, plus concepts like ‘instance’, ‘item’, ‘year’, ‘week’ and, in single dialect groups, notions like ‘egg’ (Connacht) or ‘boat’ and ‘man’ (Munster). For reasons of space, only the less dialectally characterized nouns are reproduced here:

- (17) Some nouns that take the plural after 3-10 (GGBC 1999: 70)

<i>singular</i>		<i>plural</i>
ceann	‘head (as a unit), one’	cinn
cloigeann	‘head (counting persons)’	cloigne
orlach	‘inch’	orlaí
slat	‘rod (measure), yard’	slata

- (18) Some nouns that take a special plural form after 3-10 (*ibidem*)

<i>singular</i>		<i>plural</i>	<i>plural after 3-10</i>
bliain	‘year’	blianta	bliana
fiche	‘twenty’	fichidí	fichid
pingin	‘penny’	pinginí	pingine
uair	‘time, occasion’	uaireanta	uaire

In the context of our previous observations, this selection raises three questions:

- (i) why do the Irish irregular nouns resemble so much a list of classifiers and unit nouns?

- (ii) why are normal nouns singular and the exceptions plural rather than the other way around?
- (iii) why a special plural form?

Related questions are raised by irregular plurals in Italian. In this case, unlike in Irish, the irregularity resides in the morphology of the nouns, and is not restricted to numerically quantified contexts. The nouns in this class (a group comprising between 10 and 20 items, depending on usage) are all masculine and their singular ends in *-o*; their plural, however, ends in *-a*, which is nowhere else in Italian an exponent for plurality, and is feminine for the purposes of syntactic agreement. To compound the irregularity, many of these nouns also have a regular masculine plural in *-i*, giving rise to a series of plural doublets:

- (19) Some Italian irregular plurals in *-a* (Acquaviva 2002)

<i>singular (masc.)</i>	<i>regular plural (masc.)</i>	<i>irregular plural (fem.)</i>
cervello 'brain'	cervelli 'brains' (organs)	cervella 'brains' (mass)
fondamento 'ground'	fondamenti 'grounds'	fondamenta 'foundations'
dito 'finger'		dita 'fingers'
centinaio 'hundred'		centinaia 'hundreds'
uovo 'egg'		uova 'eggs'

Leaving aside the non-trivial complexities of these plurals, let us focus on the concepts associated with this morphologically irregular class. The lexical choice comprises units of measurement (*miglia* 'miles', *centinaia* 'hundreds', *migliaia* 'thousands'), of quantity (*staia* 'bushels', *paia* 'pairs', obsolete *carra* 'cartloads'), members of cohesive aggregates (*braccia* 'arms', *corna* 'horns'), complexes of non-individual parts (*budella* 'entrails', *mura* 'city walls'), and objects perceived as indistinguishable (*uova* 'eggs'; note the parallel with Irish *uibhe* 'eggs'). The association between units of measure and irregular number is once more confirmed; comparing the Irish and Italian lists, however, we see that a host of other concepts is involved.

In the face of these facts, one possibility is to deny the existence of a common semantic basis underlying the irregularity of all these nouns, beyond the central core of measure nouns. I want instead to argue that the morphological idiosyncrasies considered in this section (for languages in which nouns are fully integrated in the number opposition) should be considered on a par with those reviewed in the preceding section, where nouns were shown to be beyond the number opposition, only interpretively or morphologically as well. The next section will clarify the semantic connection between classifiers, unit nouns, measurements, "collectives", abstract notions (Breton *lod* 'part') and concepts like 'eggs'; this afford a deeper understanding of the morphology-semantics connection in transnumeral nouns.

3. Semantic Generalization

The complements of classifiers, the classifiers themselves, the bases for singulative affixation, *and* the irregular nouns reviewed in the preceding section all involve a

natural semantic class: they are associated with concepts without individual properties, as schematically set out in (20):

(20) *Concepts without individual properties*

homogeneous masses	NON-DISCRETE
collective masses (e.g. <i>furniture</i>)	
activity predicates	
abstract nouns	
abstract units (including Breton <i>lod</i> ‘part’)	EQUIVALENCE CLASSES
measures of quantity and amounts	
members of cohesive collections	WEAKLY INDIVIDUATED
objects without salient distinctive properties (e.g. <i>eggs, times</i>)	

Those nouns that require classifiers to establish a criterion of countability are like mass nouns for grammatical purposes (although the mass-count distinction is preserved semantically, even in languages like Chinese: cf. Cheng and Sybesma 1999). This is the category which most clearly transcends the semantic opposition between singular and plural: masses conceived as atomless (e.g. *water*, assuming it has no smallest parts for linguistic purposes), as well as mass nouns interpreted as aggregates (e.g. *furniture, clothes, embers*) cannot be said to be “many” because they lack an intrinsic criterion to define “one”. Semantically, they are all transnumeral, whether or not they carry grammatical number (as in English) or not (as in Chinese). Nouns that denote activity predicates, like Arabic *boos* ‘kissing’ (cf. (10) above), are also semantically transnumeral, as are abstract nouns (unless they are made countable by some other interpretive means, like the abstract *beauty* when it is turned into the concrete *beauty-beauties*). In all these cases, the noun’s domain of reference is non-discrete.

Unit nouns, encompassing classifiers, measurements and all other expressions of quantity, are instead discrete; indeed, their interpretation amounts to a criterion for segmenting a domain into units. But they are all equivalence classes: a litre, a sack-ful, or even just a “part” have no individual properties that could set them apart from another litre, sack-ful or “part”. In so far as these nouns express different criteria for segmentability, they refer to ways to discretize a domain, *not* to individuals or amount of matters. Of course these nouns are countable (that is their function), but they too are beyond the singular-plural semantic opposition, because a phrase like *three litres* does not refer to a plurality of litres as opposed to *one litre*: *three litres* refers to an amount of matter three times big as that referred to by *one litre*. I think this is the reason why measure nouns, and less consistently nouns used as criteria for standard sizes, tend to be irregular in the expression of number: because morphological number on them is not related to the interpretive distinction between one and more than one instance of an entity – and this is because they do not refer to entities.

What this second class has in common with the class of non-discrete concepts is the lack of distinctive individual properties for their referents: non-discrete concepts define no units, and unit nouns define no individuals. It is this crucial semantic trait that explains why, in a variety of languages, concepts in the third group, such as ‘egg’ or ‘finger’, may pattern with unit nouns. These concepts are discrete and refer to actual entities, but these entities are conceptualized as interchangeable, or weakly individual.

A noun like ‘time, circumstance’ (Irish *uair*, French *fois*, Italian *volta*) cannot identify an individual time interval unless it is deictically anchored. In some cases, the lack of distinctive individuality has a basis in the low perceptual salience of the objects involved (cf. phrases like *alike as two peas*). In other cases, it depends on the cohesiveness of aggregates: in the singular, a concept like ‘finger’ or ‘star’ clearly refers to an individual entity, but the plural of such concepts is easily conceptualized as a cohesive aggregate, a larger structure in which each part presupposes the others. And obviously, the greater the cohesion of parts in a whole, the lesser their individuality. Nouns in this third class, then, are not transnumeral in the sense that their interpretation precludes a semantic contrast between one and many, but in the sense that their plurals forms mean something different from just a plurality of singulars.

4. Morphological Generalizations

Now that we have a semantic basis for viewing in a unified fashion all the dissociations we have considered between morphological and semantic number, we can focus on the morphological generalizations that emerge.

4.1 Germanic Irregular “Singulars” as Bare Stems

Section 1.2 above featured the use of apparently singular nouns with plural sense in (some) agglutinating languages. As was pointed out, this singular is better seen as a numberless stem (an approach that seems confirmed by descriptive grammarians). It is at least a coincidence that English irregularly singular measure terms (cf. 2.1 above) also appear in a form that has no exponent for number. As explained in (1), the mere absence of number marking on a noun like *pen* is no ground for regarding it as transnumeral, because that form systematically appears in a context that is interpretively and morphosyntactically singular. Things are different with measure nouns like *quid*, however, which never have a competing form **quids*; and also for *dozen* (or *head*), which is semantically neither singular nor plural when used as a unit of measurement. One can, of course, regard these cases as zero-plurals, akin to *sheep* or *aircraft* in *these sheep are grazing* or *these aircraft have landed*. But, aside from the fact that zero-plurals are *always* suffixless and not just in quantified contexts (unlike the nouns in (13)), this move would treat as accidental the concomitance of transnumeral interpretation and numberless form. This is especially unlikely when viewed from a comparative perspective: there is a definite tendency, as we saw in 1.2 above, for nouns to have “singular” form but plural sense after numbers when the “singular” has no number marking, and vice-versa, languages where a noun is always formally marked for number (as in Russian) tend to shun such semantics-morphology mismatches.

German allows us to test and refine the idea that irregularly singular measure nouns are formally numberless. *Mark* and *Gramm* are invariable, as is *Faden* ‘fathom’ (in fact much less than a fathom). *Kilo* is just like English: its plural is *Kilos*. These cases are all consistent with the hypothesis of numberless stems used as counting units, either because there is no competing plural, or because the plural is an agglutinative suffix attached to a form without a number marker (*Kilo-s*). The remaining nouns considered, *Fuss*, *Glas*, *Korb*, *Mann*, *Pfund* and *Sack*, are more complicated cases. Their plurals all involve the addition of a suffix: *Füsse*, *Gläser*, *Körbe*, *Männer*, *Pfunde*,

Säcke. If the marker of plurality was only a suffix, pasted on a bare stem identical with the singular, we could simply extend to German the analysis of English (and Turkish and Hungarian). But, except for *Pfunde*, pluralization also involves umlaut of the root vowel, so that at least on the surface the stems *fus*, *glas*, *korb*, *man*, *sak* contrast with the plural stems *füs-*, *gläs-*, *körb-*, *män-*, *säk-*; and because of this contrast, *fus*, *glas*, *korb*, *man*, *sak* appear as singular, not numberless.

However, root revowelling can also be seen as a secondary reflex of suffixation (cf. Carstairs 1987, Noyer 1997 for such an approach in terms of primary vs. secondary exponence). This means that a form like *Männer* can still be regarded as arising from suffixation to a bare stem which corresponds with the singular form: *man* + *er* (UMLAUT). Therefore, all of the German unit nouns above considered conform to the pattern SINGULAR = BARE NUMBERLESS STEM. The hypothesis that, even in German, what appear as irregular singulars are in fact numberless is straightforwardly compatible with this state of affairs. What is more, it predicts that no German unit noun can appear as an irregular singular if it is morphologically marked as singular. This is, in my opinion, the basis for the systematic exclusion of feminine unit nouns from this “quasi-classifier” construction:

- (21) *drei Flasche Wein *drei Tasse Wasser *drei Elle Stoff
 ‘3 bottle.SG wine’ ‘3 cup.SG water’ ‘3 cubit.SG cloth’

Unlike nouns like *Mann* or *Sack*, feminines like *Flasche* encode information about number through the final schwa, which is systematically connected with the singular number for feminine nouns (as opposed to masculines). What is more, a speaker of German would also be able to infer that a feminine noun ending in *-e* in the singular will end in *-en* in the plural, and that a feminine singular adjective will always end in *-e* (in the direct cases), which means that final *-e* has a morphological significance in the German nominal morphology as an exponent of the properties [feminine, singular]. This does not mean that *-e* spells out only these features in German, of course; but it does mean that a word form like *Flasche*, unlike *Fuss*, contains morphological information on singular number (for a feminine noun) and therefore cannot be said to be a bare numberless stem. My contention is that this explains the systematic lack of unit nouns as in (21).

4.2 Italian and Irish Irregular Plurals Have no Canonical Plural Suffixes

Germanic unit nouns are irregular because they appear as singulars with plural interpretation; I have argued that they are morphologically not singular, and that their interpretation is neither singular nor plural. The Irish and Italian exceptions of 2.2 comprise nouns of the same semantic category as the Germanic exceptions (weakly individualized concepts), but they are irregular for the opposite reason: they are plural where the language would normally mandate a singular (Irish), or their plural form is irregular (Italian, partly Irish). On closer inspection, the formal irregularity of Italian and Irish special plurals turns out to systematically involve lack of a specifically plural suffix.

The point is straightforward for Italian. Not only, as mentioned above, is a plural ending *-a* a complete *unicum* in Italian morphology; when an irregular plural in *-a* is combined with an evaluative suffix such as *-ino/a*, the resulting form has the

inflectional ending determined by the suffix, but it crucially retains the (exceptional) feminine gender of the irregular plural: *dita* ‘fingers’ → *dit-ine* (fem. pl.). This means that the feminine gender is a feature of the base itself, which is retained even when the final *-a* is deleted. Therefore, *dita* does not inherit its [fem., pl.] features from the ending *-a*. (Cf. Acquaviva 2002 for several arguments to the effect that *dita* is an inherently plural lexeme.)

The Irish facts are more complex, but the crucial point for present purposes is that the irregular plurals systematically make use of palatalisation of the last consonant and addition of a neutral vowel (or a combination of the two). Both processes find wide application in Irish morphology outside of the function as plural markers (Ó Siadhail 1989: 135–140, 159–161). Regular plurals, on the other hand, feature specifically plural suffixes in addition to vowel extension and palatalisation:

- (22) *Regular plurals:*
- specifically plural suffixes (bus-anna, tamall-acha, blian-ta, scór-tha, seachtain-í ...)
 - suffix with stem extension (uibh-each-aí, uair-ean-ta ...)
 - palatalisation (fear / fir, punt / puint, bord / boird ...)
 - vowel extension (lámh / lámha, ceann / ceanna ...)
- Irregular plurals:*
- palatalisation (ceann / cinn, scór / scóir)
 - vowel extension (uair / uair-e, pingin / pingin-e, bliain / blian-a, seachtain / seachtain-e ...)
 - vowel extension + palatalisation (ubh / uibh-e)

The restriction to palatalisation and vowel extension typically means that irregular plurals are shorter than regular ones, a fact recognized by the traditional label of “short plurals”. The systematic restriction of irregular plurals to stem extensions that are not specifically plural suggests that short plurals are in fact morphologically anomalous among noun plurals. This is confirmed by the observation that specifically plural suffixes almost always attach to both direct and genitive case forms (“strong” plurals), while the form of short plurals fails to generalize to both case forms:

(23)

	Strong plural: <i>bliain</i> ‘year’		Weak plural: <i>muc</i> ‘pig’	
	<i>singular</i>	<i>plural</i>	<i>singular</i>	<i>plural</i>
<i>Nominative</i>	bliain	blian-ta	muc	muc-a
<i>Genitive</i>	blian-a	blian-ta	muic-e	muc

The conclusion I wish to draw from these observation is that Irish special plurals are irregular in a specific sense: their morphological structure is never STEM + PLURAL AFFIX. This is the same conclusion that arises from an examination of Italian irregular plurals, and it is reminiscent of the conclusion reached in connection with Germanic irregular singulars, which are never STEM + SINGULAR AFFIX. The underlying semantic uniformity of nouns with weakly individualized referents is thus matched by a morphological uniformity: when nouns with a transnumeral interpretation are morphologically irregular, they are either bare numberless stems (as in Turkish or

Germanic), or intrinsically plural stems, or *lexical* plurals (Italian or Irish). This latter category also includes Arabic broken plurals (cf. McCarthy and Prince 1990), and can be further exemplified by the English *pence*, which differs from *pennies* precisely in not being decomposable into STEM + PLURAL AFFIX. ‘Pence’, which refers to an abstract monetary value rather than to a plurality of penny-coins, also falls in the semantic class of transnumerals. I claim this match of form and meaning is systematic.

5. Conclusions

A simple statement about the abstract structure of transnumeral nouns encompasses all of the facts so far considered:

- (24) A noun may be transnumeral (and fall in the semantic class in (20)) only if it is not assigned number from a separate [Number] head.

Assuming that the abstract number features of a noun phrase are expressed not on N itself, but on a separate Number head, the morphological resources of a language can spell out in different ways an input schematically like in (25):

- (25) [DP D [NumberP Num [NP N]]]

(24) states that, if a noun has a transnumeral interpretation, its morphological form will be affected by the fact that it will not be “assigned number from a separate Number head”.

In the simplest case, [Num] is either absent, or in any case does not encode number features. Classifier languages typically feature a marker of countability (classifier) in place of [Num]; both N and the classifier itself are semantically transnumeral and fall under (24).

In languages with an established number opposition, N normally raises to Num, but here too nouns can remain Num-less: the bare “singulars” of Turkic and Uralic languages are N stems spelled out without Num, which is null (but syntactically present to provide the DP with number agreement features). Bases for singulative derivation, and more generally bare stems which do not enter into a number opposition, are amenable to the same kind of analysis as bare N without association with Num (in so far as they are not countable and display no number marking).

The disassociation between N and Num is especially common when N is governed by a numeral. Why this is so depends on the syntax of numerically modified DPs in the respective languages, a vast topic I have neither the ability nor the space to explore here. In general, basic numerals (2-10), which semantically force a count interpretation, require a marker of countability in the DP, which can either be the head Num itself or a classifier-like unit noun expressing the criterion for countability:

- (26) Numeral [NumberP [Num / Class] [NP N]]

English and German bare-stem unit nouns are in [Num/Class], if they are followed by a N (English *three million people*, German *drei Sack Kohle*); nouns that express a unit but

are not followed by another noun (as in *three quid*) can be seen as bare Ns raised to double up as criterion of countability:

(27) Numeral [NumberP [N_i] [NP t_i]]

The full significance of (24) emerges with lexical plurals, like the Italian and Irish examples of 4.2. These nouns are indeed plural, morphologically as well as syntactically; but we have seen that they are not constructed with the usual plural affixes of their respective language. This means that their plural formatives are really part of N itself, not spell-outs of Num. I think this is the crucial connection between semantics and morphology: Italian and Irish irregular plurals have a common semantic basis in the notion of weakly individualized concepts, and they are morphologically similar in not being decomposable into STEM + PLURAL AFFIX. Setting N = STEM and Num = PLURAL AFFIX, (24) provides the beginning of an explanation for this match: a N with that interpretation may be a bare stem (only apparently singular), or an internal plural (without a plural suffix that spells out Num).

In fact, (24) leaves open just one possibility for a “synthetic” plural (STEM + PL. AFFIX) to have a transnumeral interpretation. Consider a N which is inherently plural, regardless of the syntactic context. On some such *pluralia tantum* the morphological expression of plural is indeed fused with the stem: *pence* or *cattle* provide two English examples (differing in countability). But nouns like *scissors* are also inherently plural, even though they are clearly segmentable as STEM + PLURAL AFFIX. So, *scissors* is morphologically made up of N + Num, but the value [plural] on Num is part and parcel of the morphosyntactic characterization of this N. In this single case, I suggest, regular “synthetic” plurals can be transnumeral: indeed, *pluralia tantum* like *scissors* or *clothes* are uncountable and semantically transnumeral, despite their morphological number.

This unified perspective on the morphosemantics of transnumeral nouns affords some interesting typological consequences. Suppose a N is ill-formed without a gender, and gender and number are fused in that language. Then, number *must* have an exponent (the fusional [gender, number] affix). Hence, fusional languages like Latin, Russian or Italian are *predicted* to have no morphologically transnumeral nouns; that is, no “bare stems” comparable to Turkic or Germanic (cf. 1.2, 2.1). This explains on a principled basis why the pattern ‘Numeral + N.SG’ is especially common in agglutinating languages without gender. That would also explain why English (which has no morphological gender on its nouns), but not German nor Romance, may have transnumeral constructions like *twenty police / faculty / personnel*. These nouns are compatible with a singular or plural reading, and the reason I am proposing is that they are morphologically numberless in such constructions (but not in e.g. *three faculties*). But they can be morphologically numberless because they are genderless; English allows this, German does not.

Finally, I have claimed that a noun may be morphologically marked for number, but semantically transnumeral, only if the number feature is a property of the stem itself, as in *pluralia tantum* like *blues* or *scissors* or in internal plurals like *pence* or the Italian and Irish irregular plurals. In all other cases, a transnumeral interpretation demands a bare, Num-less N stem. But this last avenue is precluded for strongly fusional languages like Latin or Russian, in which every N must have gender and number in each of its word forms. This means that inherent number is the only way in which these languages can express transnumerality on nouns (apart from kind-readings,

as in *homo hominis lupus* ‘man [is] man’s wolf’). If correct, we would expect *pluralia* and *singularia tantum* to be particularly frequent in such languages, more than in languages that can express this reading via a bare stem. And, although this is no more than an educated guess, I submit it is correct.

References

- Acquaviva, P. (2002), *Il plurale in -a come derivazione lessicale*, “Lingue e linguaggio” 2, 295–326.
- Acquaviva, P. (2004), *Goidelic inherent plurals and the morphosemantics of number*, ms., University College Dublin.
- Bisang, W. (1999), *Classifiers in East and Southeast Asian Languages: Counting and Beyond*, in J. Gvozdanović (ed.), *Numeral Types and Changes Worldwide*, Berlin – New York, Mouton de Gruyter, 113–185.
- Carstairs, A. (1987), *Allomorphy in Inflection*, London, Croom Helm.
- Cheng, L. & R. Sybesma (1999), *Bare and Not-So-bare Nouns and the Structure of NP*, “Linguistic Inquiry” 30, 509–542.
- Chierchia, G. (1998), *Plurality of Mass Nouns and the Notion of Semantic Parameter*, in S. Rothstein (ed.), *Events and Grammar*, Dordrecht, Kluwer, 53–103.
- Corbett, G. (1993), *The Head of Russian Numeral Expressions*, in G. Corbett, N. Fraser & S. McGlashan (eds.), *Heads in Grammatical Theory*, Cambridge, Cambridge University Press, 11–35.
- Corbett, G. (2000), *Number*, Cambridge, Cambridge University Press.
- Cowell, M. (1964), *A Reference Grammar of Syrian Arabic*, Washington, DC., Georgetown University Press.
- GGBC = *Graiméar Gaeilge na mBráithre Criostaí*, (1999), L. Ó hAnluain (ed.), Dublin, an Gúm.
- Greenberg, J. (1974), *Numeral Classifiers and Substantival Number: Problems in the Genesis of a Linguistic Type*, in L. Heilmann (ed.), *Proceedings of the Eleventh International Congress of Linguists*, Bologna, il Mulino, 17–37.
- Löbel, E. (2000), *Classifiers versus genders and noun classes: A case study in Vietnamese*, in B. Unterbeck, M. Rissanen, T. Nevalainen and M. Saari (eds.), *Gender in Grammar and Cognition*, Berlin – New York, Mouton de Gruyter, 259–319.
- McCarthy, J. & A. Prince (1990), *Foot and word in prosodic morphology: the Arabic broken plural*, “Natural Language and Linguistic Theory” 8, 209–283.
- Noyer, R. (1997), *Features, Positions and Affixes in Autonomous Morphological Structure*, New York, Garland.
- Ó Siadhail, M. (1982), *Cardinal numbers in Modern Irish*, “Ériu” 23, 101–107.
- Ó Siadhail, M. (1989), *Modern Irish. Grammatical Structure and Dialectal Variation*, Cambridge, Cambridge University Press.
- Qafisheh, H. (1977), *A short reference grammar of Gulf Arabic*, Tucson, University of Arizona Press.
- Rijkhoff, J. (1991), *Nominal Aspect*, “Journal of Semantics” 8, 291–309.
- Rounds, C. (2001), *Hungarian*, London, Routledge.
- Trépos, P. (1957), *Le pluriel breton*, Rennes, Imprimeries Réunies.

Loan Words and Declension Classes in Czech

Darya Kavitskaya
Yale University
darya.kavitskaya@yale.edu

1. Introduction

Czech, a West Slavic language with a rich system of noun inflection, provides two general ways of treating borrowed nouns. They either get assigned to a morphological class thus joining an inflectional paradigm or remain indeclinable, lacking the inflectional paradigm altogether. In this paper, we will look at the regularities of the assignment of borrowings to inflectional classes in Czech. In particular, some borrowed nouns are supplied with the *-a* ending, even in cases when these borrowings do not violate native phonotactics. Word-final *-a* is a marker of feminine gender in Czech, however, while inanimate borrowings with non-etymological final *-a* are treated as belonging to feminine gender, animate borrowings which acquire this ending are assigned to a small class of *a*-final masculines.

2. Indeclinables

The most discussed example of borrowings in Slavic comes from Russian. (1) shows that whenever borrowed nouns remain uninflected in Russian, they are indeclinable which amounts to saying that they do not have any separate case forms, surfacing as in (1) in the six cases of Russian in both singular and plural. Aronoff (1994: 126) proposes that “borrowings that do not fit the phonological pattern of any noun class are likely to be indeclinable” (see also Corbett 1991). Note, however, that the words in the first column in (1) which end in *-o* are problematic if the definition of the indeclinable class is to remain strictly phonological. These examples are of the form of Russian neuter declinable nouns, such as [okno] ‘window’, and they are borrowed as neuters. This problem is resolved if we adopt Repetti’s (to appear) proposal that borrowed nouns are likely to be analyzed as stems; for now, it will suffice to say that the final vowel is not treated as a morphological ending in the examples in (1), and thus these nouns remain uninflected.

(1) Indeclinable borrowings in Russian¹

[paɫʲto]	‘coat’	[pensne]	‘pince-nez’
[metro]	‘metro’	[kaʃne]	‘scarf’
[saɫʲto]	‘somersault’	[kafe]	‘café’
[flamingo]	‘flamingo’	[tabu]	‘taboo’
[ura]	‘hurrah’	[viski]	‘whiskey’

In Czech, as in Russian, there is a fairly large group of indeclinable nouns, as shown in (2). These nouns are mostly vowel-final, with a number of exceptions such as

¹ In colloquial Russian, *e*- and *o*-final borrowed nouns are declined as neuters.

tangens, *blues*, etc. Interestingly, while in Russian most indeclinables are assigned neuter gender (with the exception of words like *kofe* ‘coffee’ and *viski* ‘whiskey’ which are variably masculine or neuter, at least in colloquial Russian), in Czech indeclinable nouns come in all three genders.²

(2) Indeclinable nouns in Czech (Grepł *et al* 1995: 280–281)

- a. Masculine³
V: atašé, abbé (adjectival declension is possible, e.g. abbého *gen.sg.*),
sou [su:], penny, leu [lei]
-u: emu, zebu, kakadu
-ns: tangens, kotangens, sekans
- b. Feminine
-i/-u/-e: brandy, rallye [reli], whisky, jury [žiri], revue; Lori, Noemi,
Kaliopi, Bety; Nike
C: Ingrid, Marylin, Dolores, Mercedes, Iris, Ruth
- c. Neuter
-V: aroma, malaga, agáve, aloe, entrée, filé, alibi, Tbilisi, zoo, šodo,
tabu
-C: blues, Buenos Aires, copyright⁴, rekviem, Cannes, Los Angeles,
Port au Prince

The words in (2) remain uninflected since they do not fit the phonological pattern of any declension class in Czech. For example, if *aroma* were to be borrowed as a feminine noun and analyzed as having a morphological ending *-a*, it would decline according to the feminine declension. However, it is neuter and thus indeclinable since no neuter noun in Czech can end in *-a*.

In a paper on morphology and phonology of English borrowings into Italian, Repetti (to appear) proposes two constraints whose interaction accounts for indeclinable borrowings. The fact that borrowed nouns remain unchanged can be accounted for by a principle in (3a) which requires speakers to analyze borrowed words as morphologically simple, thus not interpreting final vowels which could be treated as inflectional endings as such. This analysis was developed for Italian but extends easily to Czech and other languages, as in (3b).

- (3) *Principle of Morphological Analysis of Borrowed Nouns* (Repetti to appear)
- a. foreign noun = Italian stem
 - b. foreign noun = native stem

² For assignment of gender to loan words see Corbett (1991), Fisiak (1995), Poplack, Pousada & Sankoff (1982), Rabeno & Repetti (1997), Thornton (2001), among others.

³ The assignment of *u*-final nouns to masculine, feminine or neuter, or *i*-final nouns to feminine or neuter is idiosyncratic. For example, *ralley* [reli] is variably feminine or neuter, and *bronz*, *esej*, *kredenc* are variably masculine or feminine (Grepł *et al.* 1995: 233).

⁴ More frequently masculine (Grepł *et al.* 1995: 281).

A further constraint in (4) is responsible for the fact that no additional morphological material is added to such a stem, that is, the right edge of the stem is aligned with the right edge of the prosodic word.

- (4) Repetti (to appear):
 “If possible, no additional morphological material (i.e., inflectional morphemes) should be added to the noun.”
 Align-R (Stem, PrWd)
 i.e., do not add an inflectional morpheme

The constraints in (3) and (4) allow us to account for the examples in (2): borrowed nouns are analyzed as stems and no additional inflectional material is supplied. The most harmonic stems do not fit the phonological pattern of any noun class available in Czech and thus are assigned to the uninflected class.

3. Inflectional Classes in Czech

While there is a sizeable class of indeclinables, many borrowed nouns in literary Czech are declined, including recent loans.⁵ Traditionally, the division into inflectional classes in Slavic languages including Czech is based on their gender and the ending. Within a given gender and final vowel, a further subdivision into types and subtypes is made (Grepl *et al.* 1995). If we consider a class of animate masculine nouns, further subdivision to declension classes is dependent on the last segment (usually, a consonant) of the nominal stem.

(5) shows examples of declension for animate masculine nouns in seven cases of Czech both in singular and plural. The division into subtypes ‘mister’ and ‘husband’ depends on the phonological properties of the stem-final consonant: the first subtype specifies stems which end in a ‘hard’ consonant, while the second includes stems which end on a ‘soft’ consonant.

- (5) Animate masculine nouns (Grepl *et al.* 1995: 244)

	<i>Singular</i>	<i>Plural</i>	<i>Singular</i>	<i>Plural</i>
N	pán-Ø ⁶	pán-i/ové ‘mister’	muž-Ø	muž-i/ové ‘husband’
A	pán-a	pán-y	muž-e	muž-e
G	pán-a	pán-ů	muž-e	muž-ů
D	pán-ovi/u	pán-ům	muž-i/ovi	muž-ům
L	pán-ovi/u	pán-ech	muž-i/ovi	muž-ích
I	pán-em	pán-y	muž-em	muž-i
V	pan-e	pán-i/ové	muž-i	muž-i/ové

(6) shows the declension paradigms of masculine and feminine nouns in *-a* which will be relevant for the analysis of borrowings proposed below; note that there is a mismatch

⁵ In colloquial Czech most nouns are declined. For a description of colloquial Czech, see Townsend (1990).

⁶ Diacritics here signify vowel length (I use traditional Czech spelling in the following examples).

between the masculine gender of a noun ('chairman' in our example) and the ending *-a* which usually marks feminine gender.

(6)

	a. Masculine in <i>-a</i>		b. Feminine in <i>-a</i>	
	Singular	Plural	Singular	Plural
N	předsed-a	předsed-ové 'chairman'	žen-a	žen-y 'wife'
A	předsed-u	předsed-y	žen-u	žen-y
G	předsed-y	předsed-ů	žen-y	žen-Ø
D	předsed-ovi	předsed-ům	žen-ě	žen-ám
L	předsed-ovi	předsed-ech	žen-ě	žen-ách
I	předsed-ou	předsed-y	žen-ou	žen-ami
V	předsed-o	předsed-ové	žen-o	žen-y

In (7), there is a table of inflectional classes of singular nouns in Czech⁷, constructed on the basis of Aronoff's (1994) definition of an inflection class as a group of nouns which share the same set of inflectional generalizations, that is, the same set of endings for a given paradigm. Ignoring the further division into phonological subtypes, Czech has roughly six general classes of declinable nouns and a class of uninflected nouns.⁸ The classification in (7) is very general, and there are many exceptions to the patterns which have to be separately listed.

(7) Inflectional classes in Czech

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Nominative	Ø	a	Ø	a	Ø	o	Uninflected
Accusative	a/e ⁹	u	Ø	u	Ø	o	
Genitive	a/e	y/i	u/a/i	y/i	i	a	
Dative	ovi/u/i	ovi	u/i	e	i	u	
Locative	ovi/u/i	ovi	u/e/i	e	i	e/u	
Instrumental	em	ou	em	ou	í	em	
Vocative	e/u/i	o	e/u/i	o	i	o	

The examples of nouns belonging to each declension class are shown in (8):

(8) Class 1: masculine animate in *-C*
pán 'mister', *muž* 'husband'

Class 2: masculine animate in *-a*
předseda 'chairman'

Class 3: masculine inanimate
hrad 'castle', *stroj* 'mechanism'

⁷ Aronoff (1994) and Corbett (1991) present accounts of Russian noun classes and their relation to gender, see also Zaliznjak (1977) for the fullest proposed system of Russian declension classes and Harris (1985, 1991, 1992) for the account of inflectional classes in Spanish;

⁸ Note that Class 2 and Class 4 share inflectional markers in all cases except Dative/Locative.

⁹ The allomorphy is phonologically conditioned.

- Class 4: feminine in *-a*
žena ‘wife’, *ruka* ‘hand’
- Class 5: feminine in *-C*
kost ‘bone’, *řeč* ‘speech’
- Class 6: neuter
město ‘town’, *jablko* ‘apple’
- Class 7: indeclinables
whisky, *zoo*

In the following discussion we will be primarily concerned with declinable classes 1, 2 and 4 as well as with the class of indeclinables.

4. How Do Loanwords Get Assigned to Declension Classes?

In this section we discuss how loanwords get assigned to declinable noun classes. As was mentioned in the previous sections, morphologically most borrowings into Czech are inflected. Phonologically, there are two possible strategies of loan adaptation: borrowed nouns either remain unchanged¹⁰ or, if consonant-final, supplied with the final *-a*. This loan adaptation process results in masculine animate or feminine inanimate nouns.

For borrowed words whose phonological form remains unaltered in Czech, the assignment to noun classes depends on the phonological shape and the inherent gender of the word in question. (9) illustrates this type borrowings: (9a) shows masculine nouns ending in a consonant or a consonant cluster (assigned to declension class 1), (9b) gives examples of feminine nouns in *-a* (declension class 4), and (9c) lists examples of neuter nouns in *-o* (declension class 6).

- (9) a. Masculine nouns in *-C*
- | | |
|-----------------|--------------------------------------|
| -ent: | asistent, aspirant, imigrant |
| -CC: | adept, architekt, elf |
| -r, -m, -n, -l: | agresor, agronom, dominikán, admirál |
| -ang: | bumerang |
| -ik: | akademik |
| -p: | biskup, filantrop |
| -log: | dialektolog |
| -krat: | advokát, byrokrat |
| -at: | diplomat, homeopat |
- b. Feminine nouns
- | | |
|-----|------------|
| -a: | láva |
| | ekliptika |
| | ropa ‘oil’ |
| | charisma |

¹⁰ That is, no morphological ending is supplied. Of course, borrowed nouns undergo phonological changes, e.g. stress shift, in compliance with the phonotactics of Czech. Stress in Czech is word-initial with no exceptions.

- c. Neuter nouns
-o: pončo
rádio
auto
tango
bendžo

The animate nouns listed in (9a) not only end in a consonant (which is expected from a masculine noun in Czech) in a source language, but also have inherently masculine semantics interpretable precisely because of their animacy.

In section 2, it was mentioned that one of the constraints responsible for the fact that borrowed nouns remain unchanged was a requirement that the right edge of the stem should be aligned with the right edge of the prosodic word (Repetti to appear):

- (10) Align-R (Stem, PrWd)
i.e., do not add an inflectional morpheme

The constraint in (10) is only operative if the borrowed stems can be assigned to the existing inflectional classes of the language. If a stem cannot be assigned to a morphological class, it either join the class of indeclinables or a vowel suffix is added. This is the usual situation described by Repetti for Italian. In (11), there are examples of consonant-final borrowings in Standard Italian which retain their segmental structure and join the indeclinable class of feminine or masculine nouns (Repetti's class VI).

- (11) Standard Italian (class VI)

French:	bazar	[baddzar]	mas.
	boutique	[butik]	fem.
English:	computer	[kompju:ter]	mas.
	jeep	[ʤip]	fem.

If an Italian noun cannot be assigned to a morphological class without the addition of an inflectional morpheme, a vowel suffix is added. The constraint responsible for this is given in (12):

- (12) Align-R (Stem, σ) (Repetti to appear)
i.e., if a suffix must be added, keep it prosodically distinct from the stem

(13) shows the integration of loans into North American varieties of Italian: a suffix (*o*, *a*, *e*) is added and then the noun is assigned to the declension class I, II, or III, according to its final vowel.

- (13) North American varieties of Italian (class I, II, III) (Repetti to appear)

- a. Noun becomes type I noun (mas.)

lock	['lɔkk+o]
suit	['sutt+o]

b. Noun becomes type II noun (fem.)

brush	['brɔʃ+a]
tape	['tepp+a]

c. Noun becomes type III noun (mas. or fem.)

business	[bisi'niss+e]	mas.
home	['ɔmm+e]	fem.

As in Italian, the non-etymological vowel suffix appears in certain borrowings in Czech, as shown in (14).

(14) Borrowed feminine nouns in *-a*¹¹

a.	fakulta	'department'	from Latin <i>facultas</i>
	syntéza	'synthesis'	from Greek <i>synthesis</i>
	kapitula	'chapter'	from Old Latin <i>capitulum</i>
	modalita	'modality'	from Latin <i>modalitas</i>
b.	apokalypsa	'apocalypses'	from Greek <i>apokálypsis</i>
	komuna	'commune'	from German <i>Kommune</i>
	šablona	'template'	from German <i>Schablone</i>
c.	replika	'rejoinder'	from German <i>Replik</i>
	disketa	'floppy disk'	from English <i>diskette</i>
	karanténa	'isolation'	from French <i>quarantaine</i> 'forty days'
	kapota	'hood'	from French <i>capote</i>

Consonant-final inanimate nouns in (14) are not phonotactically acceptable, and the available strategy for loan integration is to supply the *-a* ending.¹² These nouns are thus assigned feminine gender and belong to the declension class 4.

However, there is a handful of borrowed masculine animate nouns in which non-etymological *-a* is supplied word finally, as in (15a). Note that without the final *-a* these words do not violate Czech phonotactics.¹³

(15) Masculine animates in *-a*

a.	asketa	'ascetic'	from Greek <i>askētēs</i>
	despota	'tyran'	from Greek <i>despótēs</i>
	bandita	'bandit'	from Italian <i>bandito</i>
	hoplita	'hoplite'	
	chetita	'hittite'	
	invalida	'invalid'	from French <i>invalide</i>

¹¹ The data here come from Klimes (2002), Pech (1948). Etymological information is from Gebauer (1903), Lyer (1978), Rejzek (2001).

¹² Note that in some cases the consonant of the source noun is lost and the final vowel is changed to *-a*.

¹³ Most animate nouns which acquire a non-etymological *-a* are [+human]; however, there is an animate non-human example such as *doga* 'mastiff' from English 'dog'.

b.	poeta	‘poet’	from Latin <i>poeta</i>
	patriarcha	‘patriarch’	from Late Latin <i>patriarcha</i>
	kolega	‘colleague’	from Latin <i>collega</i>

Examples of borrowings with the etymological *-a* are given in (15b) for comparison. The nouns in (15b) are analyzed as having internal morphological structure (*-a* is treated as a morphological ending). These *a*-final nouns stay in declension class 2 since they have inherently masculine semantics.

(16) gives examples of masculine animate borrowings which exhibit final *-C/-a* variation. Note that even though these nouns are consonant-final in the source languages, the form with the final *-a* in Czech can be a more or a less common variant.

(16) Masculine animates: *C/-a* variation

<i>More common</i>	<i>Less common</i>
archimandrita	archimandrit
akolyta	akolyt
despota	despot
anachoret	anachoreta

Finally, (17) shows examples from a large class of masculine animate borrowings ending in *-ista/-asta*. This suffix has the semantics of ‘belonging to a profession’ or ‘participating in an activity on a regular basis’. The suffix was borrowed into Czech through several sources (e.g. from Latin *baptista* ‘baptist’, from French *cycliste* ‘bicyclist’), and subsequently nativized, so the coining of such new words as *bohemista* ‘a scholar specializing in Czech language’ became possible.

(17) Masculine animates in *-ista/-asta*

a.	arabista	‘arabist’	
	cellista	‘cello player’	
	expresionista	‘expressionist’	
	fatalista	‘fatalist’	
	artista	‘artist’	from French <i>artiste</i>
b.	fantasta	‘fantasy writer’	
	dynasta		
	chiliasta		

So, as opposed to Italian, in Czech the vowel /a/ is added to phonotactically acceptable stems resulting in masculine nouns. The puzzle is thus twofold: what is the reason for the addition of the final *-a* to the consonant-final inanimate borrowings, and why they remain masculine given that *-a* signifies feminine gender elsewhere in the language. The fact that statistically, feminine nouns in *-a* (Class 2) are the most common in Czech, and masculine animates in *-a* are quite rare also makes it surprising that borrowed masculine nouns are frequently assigned to this class and supplied with a final *-a*.

To solve this puzzle it is important to pay attention to two regularities of Czech declension paradigms. First, we need to notice that declension classes 2 (feminine nouns

in *-a*) and 4 (masculine animate nouns in *-a*) have identical endings except in the dative and locative cases, as shown in (18) (in the plural, however, the set of endings for *a*-final masculines is identical to the consonant-final masculines). The endings are predictably different depending on gender, so classes 2 and 4 are collapsible. The new class is statistically the largest.

(18)

	<i>Singular</i>		<i>Plural</i>	
	Fem. <i>-a</i> /masc. <i>-a</i>	Masc. <i>-C</i>	Fem. <i>-a</i>	Masc. <i>-a</i> / masc. <i>-C</i>
Nominative	a	Ø	y	i/ové
Accusative	u	a/e	y	y
Genitive	y/i	a/e	Ø	ů
Dative	e ovi	ovi/u/i	ám	ům
Locative	e ovi	ovi/u/i	ách	ech
Instrumental	ou	em	ami	y
Vocative	o	e	y	i/ové

Yet another important observation is the frequency of the suffix *-ista/-asta* which denotes professions and occupations. It is worth noting that most masculine borrowings which acquire final *-a* are always *t*-final (with one exception ending in *-d* as in *invalida* and the noun *doga* which etymology and the time of borrowing is uncertain) in the source language. The existence of a large class of *-ista/-asta* nouns belonging to the declension class 2, together with the high frequency of the *a*-final nouns in general, makes it possible to generalize the *a*-final borrowings to a class of masculine animates. The fact that variability still exists for certain nouns of this type shows that the analogy is still incomplete.

Conclusion

In this paper, we provided an account of loan word adaptation in Czech. In particular, we concentrated on declinable masculine animate nouns which surprisingly acquire a non-etymological ending while the source form does not violate the phonotactics of Czech. We argued that the solution for this puzzle is connected with the high frequency of *a*-final nouns in Czech, together with the existence of the *-ista* suffix denoting professions and occupations and surfacing in masculine nouns.

References

- Anderson, S. (1992), *A-Morphous Morphology*, Cambridge, Cambridge University Press.
 Aronoff, M. (1994), *Morphology by Itself*, Cambridge, Mass., MIT Press.
 Corbett, G. (1991), *Gender*, Cambridge, Cambridge University Press.
 Fisiak, J. (ed.) (1980), *Historical Morphology*, The Hague, Mouton.
 Gebauer, J. (1903–1913), *Slovník Staročeský*, Praha, Unie.
 Grepl, M. et al. (1995), *Příruční Mluvnice Češtiny*, Brno, Lidové Noviny.
 Harris, J. (1985), *Spanish word markers*, in F.H. Neussel, Jr. (ed.), *Current Issues in Spanish Phonology and Morphology*, Bloomington, Indiana University Linguistics Club, 34–54.
 Harris, J. (1991), *The exponence of gender in Spanish*, “Linguistic Inquiry” 22, 27–62.

- Harris, J. (1992), *The form classes of Spanish substantives*, in G. Booij & J. van Marle (eds.), "Yearbook of Morphology" 1991, Dordrecht, Kluwer, 65–88.
- Klimeš, L. (2002), *Slovník cizích slov*, Praha, SPN.
- Pech, V. (1948), *Velký Slovník Cizích Slov*, Praha, Nakladatelé Kvasnicka a Hampl.
- Poplack, S., A. Pousada & D. Sankoff (1982), *Competing influences on gender assignment*, "Lingua" 57, 1–28.
- Rejzek, J. (2001), *Česky Etymologický Slovník*, Leda.
- Rabeno, A. & L. Repetti (1997), *Gender assignment of English loan words in American varieties of Italian*, "American Speech" 72, 373–380.
- Repetti, Lori (to appear), *How English nouns become Italian nouns: The morphology and phonology of borrowings*, "Probus".
- Thornton, A. (2001), *Some reflections on gender and inflectional class assignment in Italian*, in C. Schaner-Wolles et al. (eds.), *Naturally! Linguistic Studies in Honour of Wolfgang Ulrich Dressler Presented on the Occasion of his 60th Birthday*, Torino, Rosenberg & Sellier, 479–487.
- Townsend, C. (1990), *A Description of Spoken Prague Czech*, Slavica Publishers, Inc.
- Zalizniak, A. (1977), *Grammaticheskii slovar' russkogo iazyka: slovoizmenenie*, Moskva, Russkii Iazyk.

The Challenge of Typologically Unusual Structures¹

Alice C. Harris
SUNY Stony Brook
acharris@notes.cc.sunysb.edu

1. Introduction

Among the types of explanation that have been offered for typologically unusual structures are claims that the structure is rare because

- our innate endowment discourages this structure (perhaps as part of a more general feature)
- this structure does not function well
- this structure cannot be acquired easily by children
- this structure is not easily processed.

All of these proposed explanations share several problems. (i) In some cases there is no direct evidence to indicate what information our innate endowment provides about the structure at issue. In some cases evidence that the specific structure functions poorly, or is difficult to acquire, or is difficult to process is also lacking. (ii) In many instances, the reasoning that supports the proposed generalization is circular: This structure is rare because it does not function well (or is difficult to acquire, or is difficult to process, or is not part of our innate endowment), and we know this because the structure is rare. (iii) In many instances, including those discussed below, the unusual structure has existed for a very long time. If it is not easily acquired (or not easily processed, or not part of our innate endowment, or dysfunctional) how do we explain its longevity? (iv) None of the explanations summarized above explains why a few languages do have the structure or feature at issue. If it is not easily processed (or is not innate, or does not function well, or is difficult to acquire) how and why do some languages manage to have this feature or structure? If one or all of the explanations above are correct, we must still explain under what circumstances a dispreferred structure or feature may exist and under what circumstances it may not.

In this paper I argue that in many instances there is a different kind of explanation for typologically unusual features or structures. In many cases such a structure is the result of a complex series of very ordinary diachronic changes. I am suggesting that there is nothing unusual in any of the changes; the only thing unusual is the fact that all occur together here, and in a manner and order that produce this system.

¹ A different version of this paper was presented at a workshop, "Explaining Linguistic Universals: Historical Convergence and Universal Grammar", held at the University of California, Berkeley in March 2003, and a more complete version of it will be published with the papers from that conference.

The research reported here was supported in part by the National Science Foundation under grant BCS 0215523; gathering and analysis of data were supported by earlier grants, including a National Science Foundation National Needs Postdoctoral Fellowship (1978–79), the American Council of Learned Societies' exchange with the Academy of Science of the USSR (administered by the International Research and Exchanges Board, 1981, 1989), and National Science Foundation grants BNS-7923452, BNS-8217355, and SRB-9710085.

Diachronically, biclausal focus structures are often reanalyzed as monoclausal (Harris and Campbell 1995: Chapter 7), and this very common change occurred also in Udi, yielding the structure shown in (11).

- (11) [FOCC -PM ... Verb ...]
FINITE

The PM in (11) is derived from the pronoun indicated as ‘that;’ in (10). In the first person singular, for example, the independent pronoun is *zu* and the PM is *zu*. The second person forms have undergone some changes, and the third person forms are not yet well understood. Third person forms occurring in sentences with the structure of (10) may have been *t’e* ‘that’ or **no*.⁴ The independent pronouns of (10) cliticized to the focused constituent in a way that is known to occur; for example, in Somali, subject pronouns cliticized to a focus marker in the formation of the focus construction (Harris and Campbell 1995 and sources cited there).

While the structure in (11) is attested in 19th century Udi, it has been replaced almost entirely with the structure in (12), where the focused constituent occurs immediately before the verb.

- (12) [... FOCC -PM Verb ...]
FINITE

This is a common position for a focused constituent, occurring, for example, in Hungarian, Korean, and Armenian (Kiss 1995, Lambrecht and Polinsky 1997: 197).

For some combinations, the structure in (12) was reanalyzed as a lexicalized phrase, and this in turn was reanalyzed as a complex verb. The structure in (12) was not itself lost and continued to exist beside the reanalyzed structure. Lexicalized phrases and complex verbs consisting of noun-verb or adjective-verb are very common in the Lezgian subgroup, to which Udi belongs, and indeed in the family as a whole. For example, one or both of these constructions are found in the following other Lezgian languages: Lezgi, Tabasaran, Rutul, Tsaxur, Budukh, Khinalug, and Archi. In Udi, the lexicalized phrases were formed from (12), with the focused constituent becoming the incorporated element (IncE), and the verb becoming a light verb (LV) in many instances, as in (13).

- (13) [INCE-PM-LV]_v

During the process of univerbation, or consolidation of a verbal element and incorporated element, the PM, enclitic to the incorporated element, was trapped between these two lexical elements. A similar process in Indo-European has been discussed by Jeffers and Zwicky (1980), Klavans (1979), and Watkins (1963, 1964), among others, and this is discussed as a general process in language in Yu (2003). (13) represents the structure of the verb in (5) above, one of the unusual structures we are trying to explain.

⁴ The form *t’e* in the modern language occurs only before a noun, e.g. *t’e išu* ‘that man’; **no* occurs in the modern language only as parts of deictic pronouns – *meno* ‘this one’, *kano* ‘that one (close by)’, *šeno* ‘that one (distant)’.

The last structure to explain is that illustrated in (6), in which a monomorphemic verb root is divided by a PM. This developed, at least in part, through analogy to the structure in (13). All of the light verbs in Udi, except *-bak-* ‘be, become’, consist of a single consonant. Most of the time, then, the PM in (13) occurs between the incorporated element and a consonant, followed by the tense-aspect-mood suffix.

(14) [INCE-PM-C-SUF]_v

The structure of monomorphemic verbs can be analyzed on this pattern:

(15) [INCE- PM- C-SUF] _v	[CV- PM- C -SUF] _v
ci- ne- p- e	bɛ- ne- γ- e
down-3SG-LV-AORII	see ₁ - 3SG-see ₂ -AORII
‘she poured down’	‘she saw’

Speakers can analyze the structure on the left in (15), exemplified by the example on the left, as the structure on the right (that is, in terms of sound segments instead of morphs) and apply this analysis to the example on the right. The difference is that in the example on the left, the incorporated element and the light verb (usually of the form *-C-*), are different morphs in the stem, whereas in the structure on the right the CV- and *-C-* are in a single morpheme.

One way of taking stock of why these structures in Udi are typologically uncommon is to examine why the same thing did not happen in its sister languages. Proto-North-East-Caucasian had gender agreement, but not person agreement. Udi and Tabasaran are the only two languages in the family that have (independently of one another) created complete new person agreement systems from pronouns, though some of the other languages have some more limited innovative person agreement. The agreement markers in the other languages are affixes, not clitics as in Udi. It appears in structures of some of the other languages that agreement affixes there have also been trapped, but because they are affixes, this same process in the other languages of the family has created infixes, not endoclitics. So it is the combination of the fact that Udi created new agreement marking from pronouns, the fact that these markers are clitics, and the fact that the language has undergone extensive univerbation that has led to its being unique in its family in having the structure in (13), illustrated by (5). Note that it is the retention of the structure in (12), illustrated by (4), together with certain other structures, that keep these PM clitics from becoming affixes.

Since the Romance languages have well known person-number clitics that some analyze as marking agreement, another way of taking stock of why these structures in Udi are typologically uncommon is to examine why the same thing did not happen in the Romance languages. The simple answer is that although the formation of complex verbs is quite common, it has not occurred in the recent history of the Romance languages, and thus there has been nothing to trap the clitics.

Although analogy is known to be a very common diachronic process, the application of it described above may seem unusual, but that is only because few languages have the structure on the left in (15). Without this key analogue, it is clear that this particular use of analogy cannot be applied.

We can summarize this discussion by listing the changes that led to the intermorphemic clitic in (13) and (5).

- (16) Changes involved in the development of the intermorphemic clitic:
- a. development of focus cleft
 - b. loss of copula
 - c. use of pronoun to introduce the embedded clause
 - d. loss of the inherited agreement system
 - e. development of person-number clitics out of independent pronouns
 - f. univerbation
 - g. maintenance of structures such as (12), which prevent the clitics from being reanalyzed as affixes.

Thus, it appears that a complex sequence of common changes is responsible for the development of this structure in Udi. While each change is common, the combination appears to be uncommon and does not occur elsewhere in the family.

But the fact that Udi underwent such a complex development does not prove that this is the only route to developing endoclitics. Part of explaining why endoclitics are typologically unusual involves examining whether there are other possible historical routes to this same structure. Probably there are. However, to maintain agreement markers as clitics entails, by most definitions of clitics, that the markers occur in some instances in some other position, as in (1–4) here, to prevent them from being reanalyzed as affixes. For example, when we compare endoclitics with infixes, we see that the occurrence of the former in other positions is the only characteristic that distinguishes them. The complex origin described above (and in more detail in Harris 2002) accounts for the occurrence of Udi clitics in these various positions, while a simpler history would not. In other languages, it is most likely that only innovative agreement markers would be clitics, for eventually clitics are usually reanalyzed as affixes. The only known source of endoclitics is entrapment in the course of univerbation or some similar process. It may be possible for another change to have the same outcome, but there is no reason to believe that it would be a simpler process than entrapment in Udi.⁵ Thus it seems that at least (16e–g), or substitutes for them, are likely to be present most of the time, and other changes parallel to (16a–d) would most likely be required to set the stage for these, including getting the elements into the order required. Thus, while the changes summarized here are probably not the only possible route to the formation of endoclitics, it is unlikely that any other route would be significantly simpler.

If our innate endowment discourages endoclitics, the only evidence of this we have is that they are uncommon among languages of the world. As an explanation of their infrequency, this is circular. There is no specific evidence that this structure does not function well or is difficult to acquire or difficult to process, since these issues have not been researched. On the other hand, there is good reason to believe it is primarily

⁵ Yu (2003) proposes four mechanisms for the creation of infixes, and one might assume that any one of these might in principle create endoclitics as well. His four are entrapment, metathesis, reduplication mutation, and prosodic stem association.

the complexity of the history of Udi clitics that has insured that they would occur in a variety of positions and in this way has prevented their being reanalyzed as infixes. The complex history thus explains the typological rarity of this structure; it also explains why endoclitics do occur, in spite of their rarity. Other accounts cannot accomplish this.

3. Other Unusual Structures

While infixes and circumfixes are not as unusual as endoclitics, they are less common than either prefixes or suffixes. On the approach taken here, the reason is clear. Existence of a prefix or suffix requires only the creation of that affix – one historical step. In contrast, an infix would seem to require two steps – creation of a prefix or suffix, together with some mechanism for getting that affix inside the word (see note 5); some of the processes described by Yu (2003), however, are considerably more complex than this. A circumfix in most instances requires three steps – creation of a prefix, creation of a suffix, and the linking of these two morphemes into one. This is probably not the only way in which a circumfix can be created, but it is likely that any route to formation of a circumfix will be more complex than formation of a simple prefix or suffix.

In the paper cited in note 1, I have shown that the infrequency of a very different kind of structure is likewise best explained in terms of the many changes required to create it. This is the case system in Georgian, where three different tense-aspect-mood characteristics of verbs are associated with three different case patterns for their arguments. Again, while there may be alternative routes to such systems, it is unlikely that the creation of a system with three distinct case-marking patterns will ever be simple.

I am by no means suggesting that the relative frequencies of all structures are determined by the complexity of the processes that create them. For example, we assume that the creation of prefixes and suffixes are parallel processes, yet suffixes are believed to be more common. Historical complexity cannot account for this and other facts. Yet it seems that in a number of instances, infrequent structures are infrequent simply because their creation requires more steps than that of more common structures.

References

- Harris, A.C. (2000), *Where in the word is the Udi clitic?*, “Language” 76, 593–616.
- Harris, A.C. (2002), *Endoclitics and the origins of Udi morphosyntax*, Oxford, Oxford University Press.
- Harris, A.C. & L. Campbell (1995), *Historical syntax in cross-linguistic perspective*, Cambridge, Cambridge University Press.
- Jeffers, R.J. & A.M. Zwicky (1980), *The evolution of clitics*, in E. Closs Traugott, R. LaBrum & S. Shephard (eds.), *Papers from the 4th International Conference on Historical Linguistics*, Amsterdam, John Benjamins, 221–232.
- Kazenin, K.I. (1994), *Focus constructions in Daghestanian languages and the typology of focus constructions*, unpublished ms., Moscow State University.
- Kazenin, K.I. (1995), *Focus constructions in North Caucasian languages*, unpublished ms., Moscow State University.

- Kazenin, K.I. (2002), *Focus in Daghestanian and word order typology*, “Linguistic Typology” 6, 289–316.
- Kiss, K.É. (ed.) (1995a), *Discourse configurational languages*, Oxford, Oxford University Press.
- Klavans, J.L. (1979), *On clitics as words*, Chicago Linguistic Society Parasession volume, 68–80.
- Lambrecht, K. & M. Polinsky (1997), *Typological variation in sentence-focus constructions*, in K. Singer, R. Eggert & G. Anderson (eds.), *CLS 33: The panels*, Chicago, Chicago Linguistic Society, 189–206.
- Watkins, C. (1963), *Preliminaries to a historical and comparative analysis of the syntax of the Old Irish verb*, “Celtica” 6, 1–49.
- Watkins, C. (1964), *Preliminaries to the reconstruction of Indo-European sentence structure*, in H.G. Lunt (ed.), *Proceedings of the ninth international congress of linguists*, The Hague, Mouton, 1035–45.
- Yu, A.C. (2003), *The morphology and phonology of infixation*, Ph.D. Dissertation, University of California, Berkeley.