

THE EVOLUTION OF THE GREENBERGIAN WORD ORDER CORRELATIONS

JEREMY COLLINS

University of Hong Kong, Pok Fu Lam Road, Hong Kong

Recurring traits across languages have been argued to relate directly to constraints imposed by the brain. A particular case is Greenberg's word order universals (Greenberg 1966), which have been argued to reflect the demands of language acquisition (Baker 2001, Chomsky 2010) or of processing sentences (Hawkins 1983, Kirby and Hurford 1997). This paper argues that typological universals are indeed relevant to the study of the biological basis of language, but for a different reason, that they are created by the process of grammaticalisation. Recent historical evidence for the grammaticalisation explanation will be presented, concluding that explanations invoking processing or acquisition are misguided. Grammaticalisation involves the use of metaphor and the 'chunking' of constructions similar to the chunking of motor sequences (Bybee 2002), and hence the origin of language universals is argued to be in these properties of the mind rather than those proposed by innatist or functionalist linguists.

1. Explaining Language Universals through Grammaticalisation

1.1 *Greenberg's Word Order Correlations*

Language universals refer to properties common to all languages, or more loosely referring to properties commonly found in unrelated languages.

Greenberg (1966) formulated word order universals such as SOV languages tending 'with overwhelmingly greater than chance frequency' to have noun-adposition and possessor-noun ordering, and for VSO languages to have adposition-noun and noun-possessor ordering. Patterns such as these require an explanation, and attempts to explain them have consequences for our view of what is adaptive in language. Kirby and Hurford (1997) summarises two views which are still widely held today: either these correlations are 'economical' in the grammar and make language acquisition easier (e.g. Pinker 1992, Baker 2001); or they are there because sentences made up of constructions obeying these correlations are easier to process (Hawkins 2004, Kirby and Hurford 1997). The former theory is found regularly in syntax textbooks and research (Roberts (2007), Chomsky (2010)) and the latter theory in prominent functionalist linguistics writing such as Evans and Levinson (2009) and several computer simulations in the evolutionary literature (see Van Everbroek 1999), showing that these hypotheses are still influential.

There is growing historical evidence from a number of language families that these hypotheses are redundant, and arguably incorrect. Instead, these patterns are remnants of a historical process, grammaticalisation, as argued for by Aristar (1991) and others. In brief, the argument is this: the ordering of adpositions and the nouns they modify often correlates with the ordering of genitive and noun simply because adpositions often derive etymologically from nouns. This

process is **grammaticalisation**, the process of grammatical elements developing from lexical words. Heine and Kuteva (2007) give numerous examples, such as Chinese *li* 'in' etymologically meaning 'interior' or 'village': hence *fangzi li* 'in the house' might be glossed more literally as 'the house's inside'. The ordering of elements is Noun (*fangzi*)-Postposition (*li*) simply because it derives from a construction which is Genitive (*fangzi*)-Noun (*li*). Similarly for many languages in Ghana such as Moore *nyingri* 'on' < 'sky', and numerous other examples in languages all over the world (Heine and Kuteva 2007). An alternative source of adpositions is verbs, such as Chinese *gen* 'with' from a verb meaning 'to follow'. *Gen* is a preposition because verbs precede objects in Chinese, and *gen* has inherited that ordering even while undergoing category change. Chinese hence has a mixture of prepositions and postpositions, the former systematically deriving from verbs and the latter from nouns. In each case the ordering of the adposition is inherited from a construction which it evolved from, because Chinese has verb-object order and genitive-noun order. Similar remarks apply to the Gbe languages in Ghana which also show typological mixing.

Grammaticalisation can also explain the ordering of verb and object correlating with genitive and noun ordering for at least some languages. Certain types of verb phrase derive from noun phrases made up of a nominalised verb and its object in a possessive case. An example is Ewe, which expresses 'I am seeing him' literally (or etymologically) as 'I am on his seeing': *Me-le e kpe dzi* (1Sg.-be.at his see on) (Claudi 1994).

Ewe is normally SVO but employs the genitive-noun ordering here 'his seeing', creating a construction which is SOV. Nominalisations of this kind are used cross-linguistically for expressing aspect (such as the continuous aspect in Ewe), for subordinate clauses (expressing 'I was surprised that he saw me' as 'I was surprised at his seeing of me' in Javanese, e.g. Adelaar and Himmelmann 2005:61) for voice marking (in Austronesian languages, *ibid.*), or for infinitives (Heine and Kuteva examples of all of these are given in the next section. These verb phrases can become the most frequently used and unmarked verb phrases in the languages, thus the basic verb-object order of a language can evolve from a genitive-noun construction, even if the nominal origins of the verb form are no longer transparent. If this process is sufficiently common across language families, this explains why the ordering of elements in the verb phrase reflects that of the ordering of elements in the noun phrase.

The correlations between constructions can thus be explained almost banally as the accidental product of a historical change. The arguments by Hawkins (1983) and Kirby and Hurford (1997) that these correlations might be easier to process are redundant, and also probably inaccurate in the context of grammaticalisation. The next two sections attempt to strengthen the evidence for this view, which has grown in the last few years of historical linguistic work, and answer possible objections. The final section attempts to make this view more than a negative result by discussing the cognitive basis of grammaticalisation.

1.2 Grammaticalisation Cross-Linguistically

How widespread are such examples of grammaticalisation, and to what extent can they explain the correlation? This was asked by Hawkins (1983), dismissing the historical explanation because it seemed to 'lack generality' (p. 131). Since the extensive work on grammaticalisation by Heine and Kuteva (2007) among others it is now no longer possible to argue that these patterns of change are exceptional: they seem to happen in every language family, to the point of being true universals of language.

The pattern of adpositions inheriting the ordering of the noun or verb they derive from is borne out in different families, such in many Oceanic languages (Lynch et al 2002:51) where adpositions are transparently nouns and reflect whatever ordering of genitive-noun the language has (hence can be either prepositional, as in Hawaiian, or postpositional as in Motu); Indo-European languages where reconstructible, such as English *across* < 13 c. Anglo-French *an cros* 'on cross' (Bordet et al. 2010:16); Japanese (e.g. *kara* 'from' < 'way', *si* restrictive particle < 'do', Frellesvig 2010); Australian languages in which adpositions are morphologically still nouns (Dixon 2002); Tibetan and Burmese (DeLancey 1997); and so on. Heine and Kuteva (2007:62) even remark that 'we are not aware of any language that has not undergone such a process'.

The pattern of verb phrases evolving from nominalised verbs with a possessor object is also common, although more complicated to reconstruct. A typical example is the evolution of VOS ordering in Proto-Austronesian, which has been inherited by over a thousand Austronesian languages or evolved further into SVO or VSO (Adelaar and Himmelmann 2005:7). It is now accepted that verb phrases in Austronesian languages evolved from nominal verbs, with a sentence such as 'The children are looking for the house' derives from a Proto-Austronesian construction meaning 'The children are the searchers of the house', when in the active voice: *humanap ng bahay ang mga bata* 'searcher GEN house the PL child' (Starosta et al. 1982). The evidence in favour of this hypothesis as presented in Starosta et al. (1982) and Kaufman (2009) is as follows: that the voice markers on verbs derive from nominalising markers, cognates of which still exist in Tagalog and other languages, such as the locative voice marker *an* which is also used for deriving place names, *aklatan* 'library' < *aklat* 'book'; that the direct object of the verb is marked with the genitive marker *ng* or put into the genitive case if a pronoun; that both nominalisation and the use of equational sentences of the form AB 'A is B' are extremely common in conservative Austronesian languages and presumably in Proto-Austronesian, allowing this frequently used construction to become a standard form of predication (Starosta et al. 1982); and that the 'subject-only' restriction on relativisation in Tagalog is straightforwardly predicted (for details see Kaufman (2009)). Thus the verb-object ordering in Austronesian languages derives simply from the noun-genitive ordering of Proto-Austronesian which is still retained in these languages. At a stroke this word order correlation is accounted for in roughly a sixth of the world's languages.

As Sasse (2009) noted in a comment on Kaufman (2009), the situation in Austronesian is 'not as exotic as appears at first sight, especially for a Semiticist or Afroasiaticist.' He notes that the Cushitic languages also replaced their finite verb forms with participles and are used with dative marking on the agent, in effect saying 'I hear it' as 'To me was hearing' (Sasse 2009); and that the dative pronouns eventually grammaticalised further to finite verbal morphology. This change also took place in the Iranian and Indo-Aryan languages, and that this process of replacing finite verbs with nominal forms is over a large linguistic area. Sasse also notes that Ancient Egyptian independently shows evidence of this development, in that agents are marked with genitive case. This is in fact a feature of many languages; Sasse cites Mayan and Inuit languages as examples. Gildea (1997) made a similar reconstruction for the Cariban language family, of which the famous OVS language Hixkaryana is an example; it has genitive marking on the object, effectively expressing 'the enemy will destroy the city' as 'it will be the city's destruction by the enemy' (Gildea 1997:153, explaining among other things why the subject is placed last, and why it has ergative marking).

One can add to this list many languages in Asia, as described in Yap et al. (2011), such as Tibeto-Burman languages that often use nominalised forms as sentences (e.g. 'goat-killing exists' for 'he is killing a goat': DeLancey 2011 p. 349), and even Japanese in which argument markers such as *ga* were originally genitive markers (Yap et al. 2011:462). Examples of Niger-Congo languages such as Ewe were given earlier and are discussed by Claudi (1994), while Heine describes how many Nilo-Saharan and Chadic languages such as Angas express the sentence 'Musa wants to harvest corn' as *Musa rot dyip ke shwe* (Musa want harvest POSS corn) 'Musa wants the harvesting of corn' (Heine 2009 p.31).

These are examples of languages whose basic verb phrases show evidence of evolving from nominalisations with a genitive object. There are further examples of this grammaticalisation pathway in subordinate clauses: Yap et al. (2011) extensively documents this phenomenon in Asia, and Heine (2009) in various African families as well as in some North American languages such as Quechua. Deutscher (2009) documents this in Akkadian, and even argues, following earlier work by Lehmann, that nominalisation may be the origin of subordination in Germanic languages.

There is thus work on a number of language families showing that these processes of grammatical change are widespread. This is a prerequisite for the grammaticalisation explanation of word order patterns to work: a strong prediction which need not have been verified but which seems increasingly borne out by historical linguistic work. The next section discusses whether this makes alternative hypotheses invalid.

1.3 Against Explanations invoking Processing

Word order correlations are often argued to make sentences easier to parse in real time. Is it possible that these factors play a role alongside grammaticalisation, for example, if processing demands 'filter out' certain

difficult-to-process constructions, as Kirby and Hurford (1997) suggest? It would seem that there is something to explain, why grammaticalisation happens to produce orderings that are easy to parse.

We should be open to the possibility that this is merely a theoretical coincidence. There happen to be two possible explanations available for word order correlations, but that does not mean that they are both correct: it could be that grammaticalisation happens anyway without any guidance by processing, and the result - the illusion of economy or efficiency - would be the same. One would need independent historical evidence that processing concerns do in fact guide historical change. There are some attempts to show this in Old English, for example (e.g. Fischer 1992, Clark et al. 2008) in which the language appeared to converge on the word order correlations after a period of free word order. This would be evidence for word order correlations emerging at least in part out of processing; but there are other possibilities in this case which need to be investigated further, such as it being related to the rise of analytic verb forms and periphrastic *do*, the loss of inflections (Fischer 1992) or as a result of contact from French. The historical role of processing is unclear even in this case, and there is little conclusive cross-linguistic evidence for it. Additionally it seems to be falsified by the considerable number of mixed languages such as Chinese and the Gbe languages. In these languages grammaticalisation has gone ahead and produced a mixture of prepositions and postpositions, seemingly disregarding whether or not the resulting grammar is 'economical' or 'efficient'. Typologically mixed languages have also survived for thousands of years, without showing inclination to change, a problem for a theory such as Kirby and Hurford (1997)'s that says such inconsistencies should die out (that is their explanation for why languages are so consistent).

2. Implications for Language Evolution

Word order correlations are often invoked as evidence for universals of language acquisition or use. It is hence important to understand their historical background, which standard interpretations do not. The most pernicious abuse of Greenberg's word order correlations is perhaps in Pinker (1994) and Baker (2001), two books which popularised the Universal Grammar account of these patterns. It is still common to invoke parameters relating constructions of different kinds, in literature on language acquisition (e.g. Sugisaki 2008) and typology and syntax (e.g. Roberts 2008). The alternative processing account does not in itself have a negative influence, but is an example of a broader trend of the '*ad hoc* search for functions that match the universals to be explained', as Kirby (1993) put it (Clark et al. 2008). Attempts in the evolutionary literature to simulate processing or learning with computers, thus deriving Greenberg's word order universals (e.g. Van Everbroek 1999, Livingstone 2000), have a particularly 'just-so' flavour: all computer simulations can do is show that processing/learning preferences of individuals *can* cause these correlations to emerge over time, all other historical factors being equal, not that they are actually responsible.

The historical evidence for the grammaticalisation hypothesis by contrast is growing as different language families are investigated. The word order correlations arise during the creation of new constructions by extending old constructions, and are not in themselves functional, but a byproduct. The process these residual patterns reflect is grammaticalisation: an important process first because it seems to be instrumental in creating complex grammar, as argued in Givon and Shibatani (2009), and underlies other universals such as syntactic categories, as argued in Heine and Kuteva's (2007) book. And second, because grammaticalisation seems to be ubiquitous, and therefore is likely to reveal true universals of the language faculty, as argued in Bybee (2002).

The basic mechanisms of grammaticalisation are clear, even if it is possible that there are additional guiding factors yet to be demonstrated (processing concerns being a possibility, but unsubstantiated so far). Two processes are involved: one is semantic, whereby a new meaning is created either by metaphor (body-parts for adpositions, e.g. 'face' for 'opposite', 'back' for 'behind') or by pragmatic inference (the future reading of 'going to' arose in ambiguous sentences such as 'I am going to buy milk') (Bybee 2002). The other is syntactic, with a larger and more complex construction being created by combining smaller constructions: in this case, forming a new lexical unit *going to* + verb, out of the originally independent elements *go*, continuous affix *-ing* and *to*. This latter process has been referred to as *chunking* by Bybee (2002), describing the process by which 'sequences of units which were previously independent come to be processed as a single unit or chunk'. This may subsequently lead to reduction in the form of the unit, as in *gonna*. This process of forming larger units through repetition is argued by Bybee to be part of the production of fluent speech, and resembles chunking in other complex motor skills such as tool-making or playing the piano: and this is supported by evidence for the relationship between language and dexterity from neurology (e.g. Ramachandran 2011:182). Dexterity and the capacity for metaphor have been suggested to be two aspects of the mind crucial to the biological evolution of language in any case, Dawkins (1999:301-10) and Ramachandran (2011) being two popular accounts that stressed these: but they gain an additional importance because of their role in a fundamental part of the cultural evolution of language, grammaticalisation.

If typological universals emerge from grammaticalisation, they thus ultimately reflect the universal cognitive skills of metaphor and chunking. This is an alternative picture of the biological basis of universals, lent support by the known importance of these skills to language evolution, and supported in this paper by historical linguistic evidence.

References

- Adelaar, A. and Himmelmann, N.P. eds. (2005). *The Austronesian Languages of Asia and Madagascar*. New York: Routledge.
- Bordet, L. and Jamet, D. (2010). Are English prepositions grammatical or lexical morphemes? *Cercles Occasional Papers Series* (2010).

- Bybee, J. (2002). Cognitive processes in grammaticalization. In Tomasello, M. ed. *The New Psychology of Language*, vol.2. New Jersey: Lawrence Erlbaum Associates Inc.
- Clark, B., Goldrick, M., and Konopka, K. (2008). Language change as a source of word order generalizations. In Eckardt, R., Jaeger, G. and Veenstra, T. eds. *Variation, Selection, Development: probing the evolutionary model of language change*. Berlin: Mouton de Gruyter. 75-102.
- Claudi, U. (1994). Word order change as category change. In Pagliuca, W. ed. *Perspectives on Grammaticalization*. Amsterdam: John Benjamins. 191-231.
- Croft, W. (1990). *Typology and Universals*. Cambridge: Cambridge University Press.
- Dawkins, R. (1999). *Unweaving the Rainbow: Science, Delusion and the Appetite for Wonder*. London: Penguin Books.
- DeLancey, S. (1997). Grammaticalization and the Gradience of Categories: Relator Nouns and Postpositions in Tibetan and Burmese. In Bybee, J., Haiman, J, and Thompson, S.A. eds. (1997). *Essays in Language Form and Language Type: Dedicated to T.Givon*. Amsterdam: John Benjamins. 51-70.
- Deutscher, G. (2009). Nominalization and the origin of subordination. In Givon and Shibatani (2009): 199-214.
- Dunn, M., Greenhill, S.J., Levinson, S.C., and Gray, R.D. (2011). Evolved structure of language show lineage-specific trends in word-order universals. *Nature* 473:79-82.
- Evans, N. and Levinson, S. (2009). The myth of language universals: language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, 32:5, 429-448.
- Fischer, O. (1992). Syntax. In Blake, N. ed. (1992). *The Cambridge History of the English Language*. Cambridge: Cambridge University Press.
- Frellesvig, B. (2010). *A History of the Japanese Language*. Cambridge: Cambridge University Press.
- Gildea, S. (1997). Introducing ergative word order via reanalysis: word order change in the Cariban family. In Bybee, J., Haiman, J, and Thompson, S.A. eds. (1997). *Essays in Language Form and Language Type: Dedicated to T.Givon*. Amsterdam: John Benjamins. 145-163.
- Givon, T. and Shibatani, M. eds. (2009). *Syntactic Complexity: Diachrony, acquisition, neuro-cognition, evolution*. Amsterdam: John Benjamins Publishing.
- Greenberg, J. ed. (1966). *Universals of Language*. Cambridge, MA: MIT Press.
- Haspelmath, M. (1989). From purposive to infinitive - a universal path of grammaticalization. *Folia Linguistica Historica* 10(1/2): 287-310.
- Hawkins, J. (1983). *Word Order Universals*. New York: Academic Press.
- Hawkins, J. (2004). *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
- Heine, B. and Kuteva, T. (2007). *The Genesis of Grammar*. Oxford: Oxford University Press.
- Heine, B. (2009). From nominal to clausal morphosyntax. In Givon and Shibatani (2009): 23-51.

- Kaufman, D. (2009). Austronesian nominalism and its consequences: a Tagalog case study. *Theoretical Linguistics* 35:1-49.
- Kirby, S. and Hurford, J. (1997). Learning, Culture and Evolution in the origin of linguistic constraints. In Husbands, P. and Harvey, I. eds. (1997). *ECAL97*: 493-502. Cambridge, MA: MIT Press.
- Livingstone, D. (2000). A modified-neural theory of the evolution of linguistic diversity. Paper presented at The Evolution of Language: Third International Conference. Paris, France.
- Lynch, J., Ross, M., and Crowley, T. eds. (2002). *The Oceanic Languages*. London: Routledge Publishing.
- Pinker, S. (1994). *The Language Instinct*. London: Penguin Publishing.
- Pinker, S. and Jackendoff, R. (2009). The reality of a universal language faculty. *Behavioral and Brain Sciences*, 32:5, 465-466.
- Ramachandran, V.S. (2011). *The Tell-Tale Brain*. London: William Heinemann.
- Roberts, I. (2008). *Diachronic Syntax*. Oxford: Oxford University Press.
- Sasse, H-J. (2009). Nominalism in Austronesian: A historical typological perspective. *Theoretical Linguistics* 35:167-181.
- Starosta, S., Pawley, A. and Reid, L. (1982). The evolution of focus in Austronesian. In Halim, A., Carrington, L. and Wurm, S.A. eds. *Papers from the Third International Conference in Austronesian Linguistics 1: Tracking the Travellers*, 145-170. Canberra: Pacific Linguistics.
- Sugisaki, K. (2008). Early acquisition of basic word order in Japanese. *Language Acquisition* 15:183-191.
- Van Everbroek, E. (1999). Language type, frequency and learnability: a connectionist appraisal. In *Proceedings of the Twenty-first Annual Cognitive Science Society Conference*. 755-760.
- Yap, F-H., Grunow-Harsta, K. and Wrona, J. eds. (2011). *Nominalization in Asian Languages: Diachronic and Typological Perspectives*. John Benjamins.