

Do Language-Specific Categories Shape Conceptual Processing? Mandarin Classifier Distinctions Influence Eye Gaze Behavior, but only During Linguistic Processing

Falk Huettig^{a,*}, Jidong Chen^b, Melissa Bowerman^a and Asifa Majid^a

^a Max Planck Institute for Psycholinguistics, P.O. Box 310,
6500 AH Nijmegen, The Netherlands

^b California State University, Fresno, CA, USA

* Corresponding author, e-mail: falk.huettig@mpi.nl

Abstract

In two eye-tracking studies we investigated the influence of Mandarin numeral classifiers – a grammatical category in the language – on online overt attention. Mandarin speakers were presented with simple sentences through headphones while their eye-movements to objects presented on a computer screen were monitored. The crucial question is what participants look at while listening to a pre-specified target noun. If classifier categories influence Mandarin speakers' general conceptual processing, then on hearing the target noun they should look at objects that are members of the same classifier category – even when the classifier is not explicitly present (*cf.*, Huettig and Altmann, 2005). The data show that when participants heard a classifier (e.g., *ba3*, Experiment 1) they shifted overt attention significantly more to classifier-match objects (e.g., chair) than to distractor objects, but when the classifier was not explicitly presented in speech, overt attention to classifier-match objects and distractor objects did not differ (Experiment 2). This suggests that although classifier distinctions do influence eye-gaze behavior, they do so only during linguistic processing of that distinction and not in moment-to-moment general conceptual processing.

Keywords

Attention, classifiers, eye movements, linguistic relativity

A guy goes to the pub and has four beers. The next day he wants to tell his friend about his night out. If he is a speaker of Mandarin, he has to use a specific sort of expression. He cannot just say “I had four beers”. Instead, he has to include a special category of word before the noun that serves to “individuate” the referent, for example, “I had four *bei1* beer” or “four *ping2* beer”, where *bei1* specifies glasses and *ping2* bottles. This might not seem so unusual

or exotic. But Mandarin speakers must use this construction every time they want to specify the number of entities.¹ So if the same Mandarin speaker had walked past two rivers on the way home from the pub and wished to describe how many he had seen, he would have had to say “two *tiao2* river”, where *tiao2* designates long, rope-like objects such as rivers, snakes and legs.

Bei1, *ping2* and *tiao2* are numeral classifiers, a grammatical category in Mandarin. There are several hundred such classifiers in Mandarin (e.g., Huang and Ahrens, 2003; Zhang, 2007). The semantics of classifiers have usually been characterized in terms of abstract properties such as animacy, shape, function, and size². But it is not unusual for members of a classifier category to be apparently unrelated to other members (Allan, 1977). For example, the classifier *tiao2*, which is used with rivers, snakes and legs, can also be used for dogs, underwear and a piece of news. The question we address in this paper is whether these language-specific classifier categories influence Mandarin speakers’ moment-by-moment conceptual processing or only engage processing when they are overtly deployed.

Mandarin speakers appear to be exquisitely sensitive to classifier categories in a variety of situations, including inductive reasoning, making preference judgments, remembering objects and judging object similarity (e.g., Schmitt and Zhang, 1998; Zhang and Schmitt, 1998; Saalbach and Imai, 2007). For example, Zhang and Schmitt asked whether objects that take the classifier *ba3* for “graspable objects” would become more associated with “graspability”, such that speakers direct attention to this property, even outside the linguistic realm. They constructed mock advertisements for objects and asked participants questions such as how good the product was, how much they liked it and how easy it was to use. Crucially, the advertisements consisted of photographs of objects that took either the *ba3* (graspable) classifier or a different (control) classifier (*tiao2*, for long thin things), and that appeared with or without a hand in the picture. Zhang and Schmitt found that Mandarin speakers showed more preference for an umbrella (which takes *ba3*) when a hand was in a picture but this manipulation did not influence preference for a rope (which takes *tiao2*). For non-Mandarin speakers the “with or without a hand” manipulation made no difference for any product. These findings suggest that classifiers play an important role in conceptual processing for Mandarin speakers.

¹ Classifiers occur with numerals, demonstratives (e.g., *zhei4* ‘this’, *nei4* ‘that’) and certain quantifiers (Li and Thompson, 1981).

² In Mandarin most nouns can also occur with the general classifier *ge4*. Analysts consider this to be a semantically vacuous morpheme (Meyers, 2000, although see Ahrens, 1994).

Whereas research using offline measures has found a strong influence of Mandarin classifiers on conceptualization, research using more online (or implicit) measures (e.g., Saalbach and Imai, 2007; Gao and Malt, 2009; Experiment 4) has revealed either no classifier effects or only modest ones. Saalbach and Imai (2007), for example, conducted a priming experiment with printed word primes and picture targets. Participants were asked to judge whether the printed word and the visual object matched. These authors found that classifier relations affected the response latencies of neither Mandarin speakers nor a control group of German speakers. Interestingly, an offline task using the same materials did reveal a language-specific effect. Mandarin speakers rated same-classifier pairs as significantly more similar than control pairs. Although a control group of German speakers also judged same-classifier pairs as more similar than control pairs, the Mandarin speakers' ratings for the same-classifier pairs were significantly higher than those of the German control group. Previous data on the influence of Mandarin classifiers on conceptual processing have, thus, at times been contradictory.

In the present study our question is whether classifier categories are typically recruited during online attentional processing, which we measure by tracking eye movements. If the categories of language do channel attention to particular properties of the world, we should be able to measure this by tracking eye movements over time. We also control for phonological, semantic/conceptual and visual similarity to exclude the possibility that our results are influenced by similarities that are not language-specific.

Eye Movements and Conceptual Structure

To investigate this issue we measure participants' eye movements to visual stimuli while they listen to speech. This method provides closely time-locked and fine-grained measures of ongoing cognitive processing over time, in the form of fixations to different locations in the visual display. For example, Cooper (1974) showed that when participants were presented with a visual display containing a number of common objects, they tended to spontaneously fixate on the visual referents of words while they heard their names. On hearing 'lion', for example, participants were more likely to fixate on the picture of a lion than on pictures of unrelated objects.

Language-mediated eye movements are also highly sensitive to conceptual structure (Huettig and Altmann, 2005). For example, Huettig and Altmann (2005) found that participants directed overt attention towards a depicted object (such as a trumpet) when a conceptually, but not associatively, related

target word (e.g., ‘piano’) unfolded acoustically. Increased attention to conceptually-related items is, moreover, proportional to the degree of conceptual overlap. Huettig et al. (2006) provided further evidence that proximities among auditory and visual stimuli in a multidimensional conceptual space determine shifts in overt attention. They showed that corpus-based contextual similarity measures predict eye movement behavior well. Participants’ overt attention is also guided to objects that are related in visual form – e.g., on hearing ‘snake’ they look at a cable (snake and cable have a similar global shape; Huettig and Altmann, 2004, 2007).

In short, these studies show that eye movements are a highly sensitive indicator of conceptual processing and conceptual similarity of various kinds. Although Saalbach and Imai (2007: 498) concluded that “the classifier organization system does not organize speakers’ concepts of objects”, they proposed that “the classifier effect found among Chinese speakers is perhaps best characterized as a magnified sensitivity to semantic features underlying classifier categories developed through the habitual use of classifiers in association with the names of objects”. If Mandarin speakers have “a magnified sensitivity to semantic features underlying classifier categories”, we ought to be able to find evidence for this in Mandarin speakers’ language-mediated eye-movements.

Current Study

We presented Mandarin speakers with simple Mandarin sentences through headphones while monitoring their eye-movements to objects presented on a computer screen. The crucial question is what they look at while listening to a pre-specified target noun that is embedded in a sentence.³ If classifier categories influence general conceptual processing, then on hearing a target noun participants should shift overt attention to objects that share the same classifier even when the classifier is not explicitly present. For example, on hearing the Mandarin word for scissors, they should look more at a picture of a chair than at a picture of an unrelated object because the nouns for scissors and chair share the classifier *ba3*. This would be consistent with the strong Whorfian position that language-specific categories are a major determinant of conceptual thought (Sapir, 1921; Whorf, 1956; Lucy, 1992; Gumperz and Levinson, 1996).

Of course, it is not obvious that classifiers have such a strong effect on conceptual processing, especially considering some of the findings from pre-

³ The target word was not indicated to the participants.

vious studies. We may ask whether classifier information can ever be used to direct overt attention. It is at least logically possible that classifiers are purely formal markers (see Greenberg, 1972; Lehman, 1979) that do not influence attentional processing even when they are explicitly present. Experiment 1 was therefore designed as a test to determine whether our method is sensitive enough for the present purpose. If classifiers can influence overt attention, then Mandarin speakers should look at potential matches on hearing the classifier in spoken sentences. A shift of eye-gaze to a chair on hearing the classifier *ba3* would mean that Mandarin classifiers can influence overt attention at least when they are present (see Discussion of Experiment 1).

To summarize, the present research investigates whether language-specific classifier categories influence Mandarin speakers' moment-by-moment conceptual processing. According to a strong Whorfian perspective classifier categories shape conceptual processing, so eye-movements should be drawn to exemplars of a classifier category even when the classifier is not presented (*cf.* Huettig and Altmann, 2005).

Experiment 1: Classifier Present

We first conducted Experiment 1 to validate our choice of method. In Experiment 1 the classifier was explicitly mentioned in the spoken sentence. We predicted that on hearing the classifier (e.g., *ba3*) participants would shift overt attention to the target object (e.g., chair) because according to our pre-tests the noun chair frequently occurs with the classifier *ba3*.

Method

Similarity-rating Pre-tests

In order to select appropriate items for the eye-tracking experiments we conducted two norming studies. Their purpose was to exclude item pairs that shared not only a classifier but also visual similarity (e.g., snake and river) or semantic/conceptual similarity (e.g., snake and dog) independently of the classifier. We chose Dutch subjects for this rating task because if we had asked Mandarin speakers, their similarity ratings might have been influenced by precisely the phenomena we wanted to study – whether membership in the same classifier category induces perceived similarity. Of course, excluding items that Dutch speakers judge to be similar may also reduce the likelihood of conceptually-mediated eye movements in Mandarin speakers. But if classifier categories are indeed a major determinant of conceptual processing in Mandarin speakers – and if they are automatically recruited during

online processing – then we should still observe conceptually-mediated shifts in eye gaze with this control in place.

Participants were presented with a booklet containing the experimental materials. A critical word was printed at the bottom center of each page, with four pictures presented directly above it (displayed just as in the eye-tracking experiments to be conducted). Pictures were black-and-white line drawings selected from the Max Planck Institute picture database.

The noun naming one of the pictures shared a classifier with the critical word, while the names for other three did not. (This was separately established through a naming study with native Mandarin speakers using the internet. None of these participants took part in the main experiments.) For example, if the written (critical) word on the bottom of the page was “scissors”, the four pictures were of a chair (the noun for which shares the classifier *ba3* with “scissors”), a framed painting, a nose and a candle (these nouns do not share a classifier with “scissors”).

The objects in the visual displays to be rated were pre-selected by an experimenter so that the names of the pictures within a display all started with different phonemes (i.e., no phonological competitors were present). In Pre-test 1 twelve participants were asked to rate how similar the object indicated by the critical word was in physical shape to each of the depicted objects using an 11-point scale (0=“absolutely no similarity in physical shape”; 10=“identical physical shape”). They were told to concentrate on shape and ignore similarity in meaning. In Pre-test 2, a different set of twelve participants used the same 11-point scale to judge meaning similarity while ignoring shape similarity. For the eye-tracking studies we selected the 38 (out of 48) visual displays for which there was no significant difference in semantic/conceptual and visual similarity ratings between the classifier match and the unrelated distractor referents.

Eye-tracking Experiment

Participants

Twenty-four Mandarin speakers, all with higher education (at least undergraduate level), participated in the experiment. Ages ranged from 20 to 38 years, with a mean age of 27. Most participants were Chinese undergraduate students studying at Dutch universities. The rest were employed at Dutch universities or multinational companies based in The Netherlands. All participants had lived in China until at least the age of eighteen. All were familiar with some other languages, such as English or Dutch, and all had normal or corrected-to-normal vision.

Materials and Design

The eye-tracking experiment consisted of 38 experimental trials (see Appendix A). On each trial, participants were presented with a visual display containing line drawings of four objects, along with a spoken sentence. To avoid anticipatory eye-movements based on the context of the sentence (cf., Altmann and Kamide, 1999), each critical word was placed in a neutral sentence context (e.g., *Do you know if there is another name for...?*) as is typical in experiments investigating lexical semantics (cf., Huettig and Altmann, 2005; Huettig and McQueen, 2007). Crucially, the sentence contained a numeral classifier (e.g., *Do you know if there is another name for one ba3 [ba3=classifier] chair*). The question was whether participants would look at the target picture of the chair more than the unrelated distractors when they heard the classifier. Throughout the experiment the participant's direction of eye-gaze was measured.

There were two experimental conditions. In the noun-match condition the spoken classifier and critical word (e.g., *ba3 chair*) matched the target picture (in this case a chair). In the classifier-match condition the spoken classifier also matched the target object but the spoken word referred to an object that was not in the display (e.g., *ba3 scissors*).

Materials were counter-balanced across conditions, with each participant receiving 19 trials in the noun-match condition and 19 trials in the classifier-match condition. Within experimental trials, the position of the pictures was counter-balanced, and randomized on every trial across four fixed positions of a (virtual) 5×5 grid (grid positions 7, 9, 16 and 18, counting left to right and top to bottom; see Fig. 1).

The spoken sentences were read aloud by a female native speaker of Mandarin in a sound-proof booth. Digital recordings of the utterances, at a sample rate of 44.1 kHz with 16-bit resolution, were stored directly on a computer.

Procedure

Participants were seated at a comfortable distance from the computer screen. One centimeter on the visual display corresponded to approximately 1° of visual arc. The eye-tracking system was mounted and calibrated. Eye-movements were monitored with an SMI Eyelink eye-tracking system, sampling at 250 Hz. Spoken sentences were presented to the participants through headphones. The parameters of each trial were as follows. First, a central fixation point appeared on the screen for 500 ms, followed by a blank screen for 600 ms. Then four pictures appeared on the screen one second before the initiation of the auditory presentation of a sentence. Participants were not asked to perform any explicit task. They were told that they should listen to the

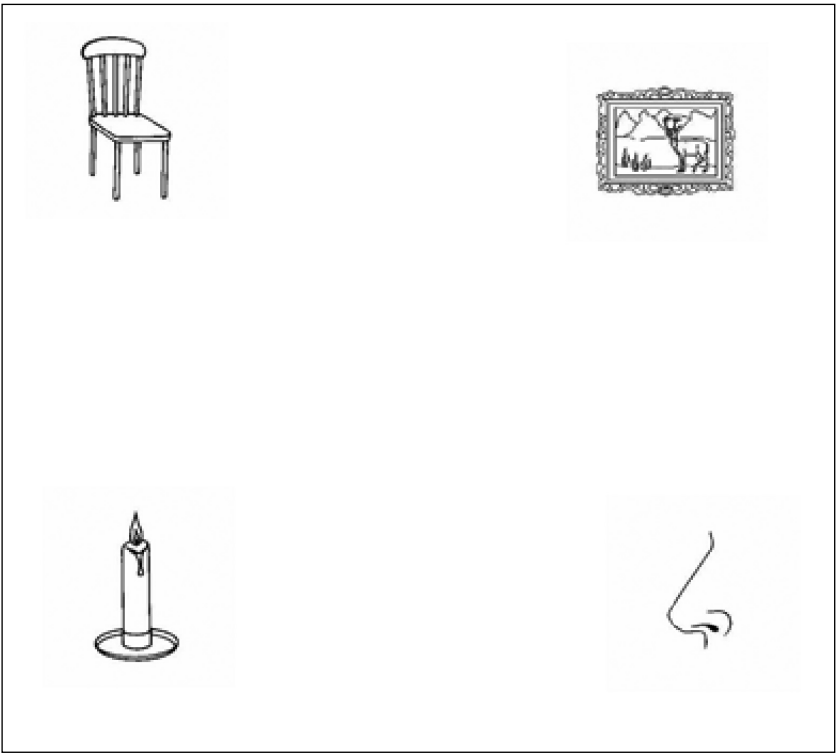


Figure 1. Example of visual stimulus (depicting target object chair and three unrelated distractors).

sentences carefully, that they could look at whatever they wanted to, but that they should not take their eyes off the screen during the experiment (see Huettig and McQueen, 2007, for further discussion of the listening-only task). Participants' fixations for the entire trial were thus completely unconstrained and participants were under no time pressure to perform any action. A central fixation point appeared on the screen after every five trials, allowing correction for some drift in the calibration.

Data coding procedure

The data from each participant's right eye were analyzed and coded in terms of fixations, saccades and blinks, using the algorithm provided in the Eyelink software. The timing of the fixations was established relative to the onset of the critical word in the spoken utterance. Graphical analysis software performed the mapping between the position of fixations and the pictures pres-

ent on each trial, and displayed them simultaneously. Each fixation was represented by a dot associated with a number that denoted the order in which the fixation had occurred. Fixations were coded as directed to the target picture, the unrelated distractor pictures, or anywhere else on the screen. Fixations that fell within the cell of the grid in which a picture was presented were coded as fixations to that picture.

Results

Figure 2 is a time-course graph that illustrates the fixation proportions at 20 ms intervals to the pictures over the course of experimental trials. In the noun-match condition the probability of a fixation to the target picture is $P(\text{target-noun-match condition})$; in the classifier-match condition it is $P(\text{target-classifier-match condition})$; and for the particular unrelated distractors it is $P(\text{distractor})$. The fixations to the three distractors were averaged. Zero represents the acoustic onset of the spoken critical word.

As can be seen from Fig. 2, $P(\text{target})$ started to diverge from $P(\text{distractor})$ right from the acoustic onset of the critical word (e.g., “chair” or “scissors”) in both experimental conditions. Since it takes about 200 ms to program and initiate a saccadic eye movement (see Altmann and Kamide, 2004), the critical information initiating this attentional shift (i.e., the classifier) was processed at least 200 ms prior to the onset of the critical word.

For the statistical analyses we computed mean fixation proportions for each type of picture from the acoustic onset of the critical word to 1000 ms later. Distractor fixations were averaged across the three distractors. We calculated the ratio between the proportion of fixations to the target picture and the sum of the individual target- and distractor-fixation proportions. We then compared the mean ratio (by participants and items) to 0.5. A ratio greater than 0.5 shows that, of all the fixations directed toward the target and the unrelated distractors, the target attracted more than half (see Huettig and McQueen, 2007).

One-sample t -tests showed that in the noun-match condition the target pictures were fixated more than the unrelated distractors (mean ratio 0.81, $t_1(23)=19.04$, $p<0.001$, $t_2(37)=25.52$, $P<0.001$). Critically, this was also true in the classifier-match condition (mean ratio 0.57, $t_1(23)=5.44$, $P<0.001$, $t_2(37)=2.12$, $P<0.05$). Thus, classifier information influenced eye-gaze.

Discussion

The purpose of Experiment 1 was to validate the eye-tracking method for research on Mandarin classifiers. Experiment 1 shows that Mandarin speakers use classifier information to direct overt attention.

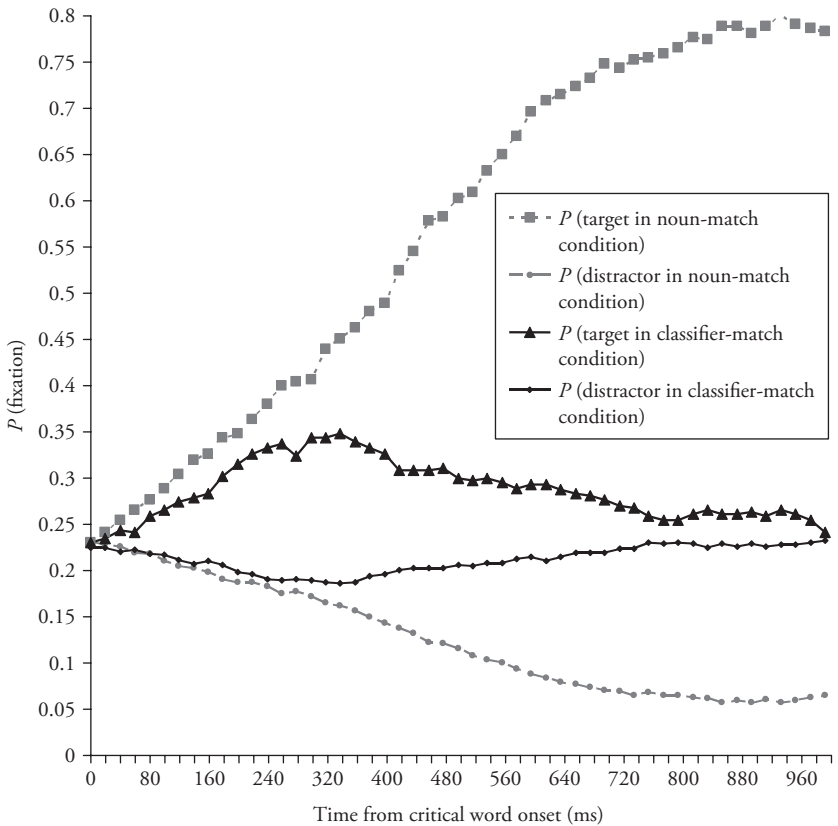


Figure 2 Mean proportion of looks to the target objects and unrelated distracters (Experiment 2, classifier present in spoken sentence).

Before we move on to the main purpose of our study, it may be worthwhile to briefly discuss the nature of the attentional effect observed in Experiment 1. These results could be compatible with Slobin’s notions of “thinking-for-speaking” and “listening-for-understanding”, according to which the structure of the language we speak exerts an influence on which aspects of the world we attend to at least when language is being produced or comprehended, even if not necessarily in cognition more generally (see Slobin, 1996, 2003).

However, when classifiers are present in the speech stream, overt attention may simply shift to objects that satisfy the classifier without necessarily being mediated by the conceptual system. At the moment of display onset, participants start to view the four objects. This causes activation of picture-derived

representations that are tied to the spatial location in the display (see Huettig and Altmann, 2007). In a series of eye-tracking experiments Huettig and McQueen (2007) showed that, with enough preview, as in the current experiments, picture-processing cascades as far as retrieval of the pictures' names even when participants are not required to name the objects. We suggest that participants' picture-derived representations also include classifier information (e.g., *ba3* chair). As participants hear the spoken classifier (e.g., *ba3*), there is overlap between the language-derived (e.g., phonological) representations and the picture-derived (e.g., phonological) representations, and this causes shifts in eye gaze to the matching object in the display (see Huettig and Altmann, 2007, for further discussion).

For the purposes of our study, nothing crucial hangs on which of these two accounts is correct. The critical point is that we can detect overt attentional shifts based on classifier information by using eye-tracking, thus indicating that our method is sensitive and appropriate.

Experiment 2: Classifier Absent

The main purpose of the present research was to investigate whether classifier categories are typically recruited during online attentional processing. According to a strong Whorfian position language-specific categories are a major determinant of conceptual processing even when they are not overtly deployed. Previous eye-tracking research strongly suggests that if language-specific classifier categories influence Mandarin speakers' general conceptual processing, then on hearing the target noun they should look at objects that are members of the same classifier category – even when the classifier is not explicitly present. Experiment 2 was identical to Experiment 1 except that the classifier was absent in the spoken sentence (i.e., there was no classifier preceding the target noun).

Method

Participants

Twenty-eight Mandarin speakers participated in the experiment. Their background was similar to that of the participants in Experiment 1. None had taken part in the previous study. All had normal or corrected-to-normal vision.

Materials and Design

The materials and design were identical to those of Experiment 1; only the spoken sentences differed. The crucial difference was that in Experiment 2

the spoken sentence did not include the classifier (see Fig. 3). For example, the sentence “Do you know if there is another name for one *ba3* [*ba3*=classifier] chair?” in Experiment 1 became “Do you know if there is another name for chair?”. In Mandarin, sentences without classifiers are frequent. The question was whether participants would look at the target picture of the chair more than the unrelated distractors when they heard the critical word “scissors”.

Procedure

The procedure was as in Experiment 1.

Results

Figure 4 is a time-course graph that illustrates the fixation proportions to the pictures at 20 ms intervals over the course of experimental trials. In the noun-match condition the probability of a fixation to the target picture is $P(\text{target-noun-match condition})$; in the classifier-match condition it is $P(\text{target-classifier-match condition})$; and for the unrelated distractors it is $P(\text{distractor})$.

)
 我在为我的诗找一个与椅子押韵的词。
 3 zai4 wei4 wo3 de shi1 zhao3 yi2 ge yu2 yi3-zi ya1-yun4 de ci2
 PROG for I GEN poem look.for one CLF with chair-noun.suffix rhyme REL word

literal translation: I am looking for my poem for one CLF with chair rhyme REL word (For my poem I am looking for a word that rhymes with chair).

)
 我在为我的诗找一个与剪刀押韵的词。
 3 zai4 wei4 wo3 de shi1 zhao3 yi2 ge yu2 jian3-dao1 ya1-yun4 de ci2
 PROG for I GEN poem look.for one CLF with scissors rhyme REL word

literal translation: I am looking for my poem for one CLF with scissors rhyme REL word (For my poem I am looking for a word that rhymes with scissors).

Figure 3. Example of spoken sentence in Experiment 2 in (a) noun-match condition and (b) classifier-match condition. CLF=classifier, PROG=progressive aspect marker, GEN=genitive, REL=relative clause marker. I PROG for I GEN poem look.for one CLF with chair-noun.suffix rhyme REL word is an example sentence in the non-classifier condition. Note that the *ge4* (general classifier) does not go with *yi3zi* ‘chair’. It goes with *ci2* ‘word’, i.e., ‘one CLF word’. The sentence fragment between *yi2 ge4* and *ci2* is a relative clause. Chinese relative clauses occur before the head noun. The syntax of a complex NP containing a relative clause is: Numeral+CLF+REL clause+de+head noun.

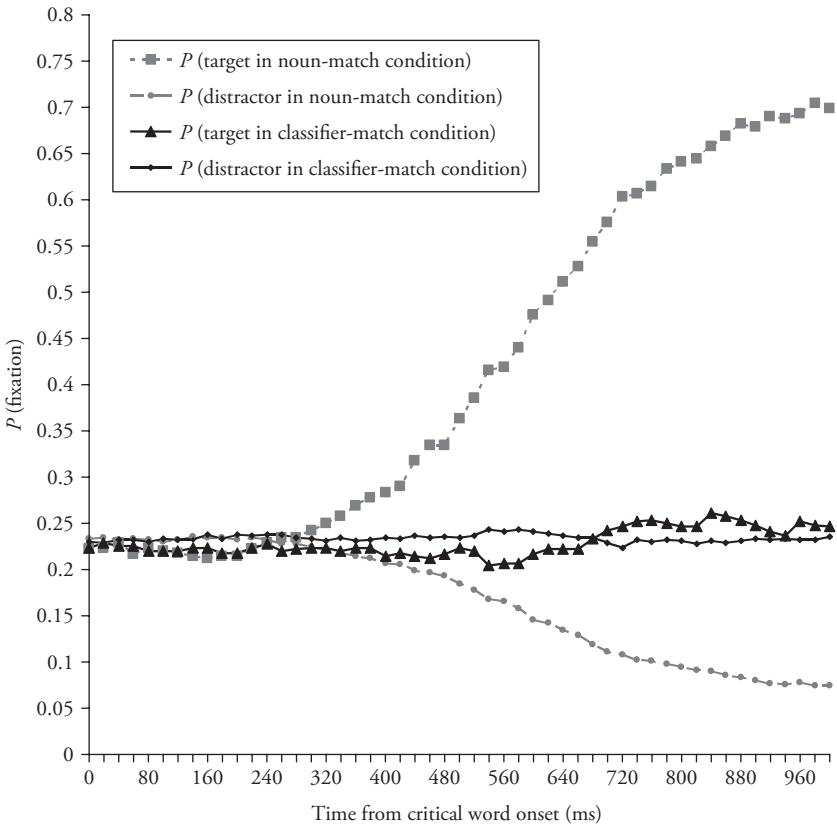


Figure 4. Mean proportion of looks to the target objects and unrelated distractors in Experiment 1 (classifier absent in spoken sentence) as a function of time (ms) after critical word onset.

The graph shows the mean percentage of fixations per distractor. Zero represents the acoustic onset of the spoken critical word.

One-sample t -tests showed that in the noun-match condition the target pictures were fixated more than the unrelated distractors (mean ratio 0.71, significant by participants, $t_1(27)=14.33$, $P<0.001$, and items $t_2(37)=14.03$, $P<0.001$). In the classifier-match condition there were no significant differences between fixation to the target pictures and the unrelated distractors (mean ratio 0.49, $t_1(27)=0.64$, $P>0.1$, $t_2(37)=0.97$, $P>0.1$).

In sum, participants shifted their eye-gaze to the target picture (e.g., chair) on hearing the critical word in the noun-match condition (e.g., “chair”), but not on hearing the critical word in the classifier-match condition (e.g., “scissors”).

General Discussion

We investigated the influence of Mandarin classifier categories on moment-by-moment attentional processing. Even though the visual display and most of the elements of the spoken sentences were identical across the two experiments, eye-movement behavior was radically different. When the classifier was explicitly presented (e.g., *ba3*, Experiment 1), participants shifted overt attention significantly more to classifier-match objects (e.g., the chair) than to distractors. But when participants heard a spoken word (e.g., “scissors”) from the same classifier category as a visually depicted object (e.g., a chair), but the classifier was not explicitly presented in the speech, overt attention to classifier-match objects (here the chair) and distractor objects did not differ (Experiment 2). Since Mandarin speakers use classifier information to direct overt attention, as shown in the (almost identical) Experiment 1, the absence of the critical attention shift in Experiment 2 is not a procedural artifact. The findings of Experiment 2 are consistent with a priming study reported by Saalbach and Imai (2007; see Gao and Malt, 2009). These researchers found that reaction times were not facilitated by providing a written prime from the same classifier category as the name of a visually-depicted target. The data, therefore, do not support the strong Whorfian hypothesis that classifiers are a major determinant of attentional (conceptual) processing even when the classifier is not explicitly present.

One might wonder whether our failure to find Whorfian effects in Experiment 2 is due to the sentence frames we used. Perhaps sentences like *I'm looking for a word that rhymes with chair* and *Do you know if there is another name for chair?* focus participants on the sound features of our target words rather than on their semantic/conceptual properties. A focus on sound features alone, but not conceptual properties, would suffice to explain why participants look at a picture of a chair on hearing the word “chair” (seeing the picture would activate the phonological properties of its name, allowing a match). However, sound-focused participants would have no reason to look longer at a chair on hearing “scissors”, since a match based on covert shared classifier-category membership presupposes semantic/conceptual processing.

This explanation for the absence of an effect of shared classifier-category membership in Experiment 2 can be rejected with confidence. Previous visual world paradigm studies with pictures have used almost identical neutral sentence frames (e.g., Huettig and McQueen, 2007) and obtained strong semantic/conceptual effects. In addition, a great many visual world studies (Huettig and Altmann, 2005; Huettig et al., 2006; Huettig and McQueen, 2007) have demonstrated that when a display contains pictures, participants focus on

semantic matches between the display and the spoken words that are concurrently being processed (e.g., they look at a piano on hearing ‘trumpet’). When a display consists of words naming objects rather than pictures of the objects, the situation is different: orthographic representations of the sound forms of words do prompt participants to focus on possible phonological matches (Huettig and McQueen, 2007). In the present experiment our displays consisted of pictures, not written words. It seems highly unlikely, then, that the sentence frames we used focused listeners merely on phonological rather than conceptual properties of the target words.

Our current study represents a strict test of the Whorfian hypothesis. We selected the materials to minimize the likelihood that non-language-specific factors would influence the results. Thus, we excluded item pairs that shared visual similarity (e.g., snake and river) or semantic similarity (e.g., snake and dog), as judged by native speakers of Dutch. These strict criteria also, of course, *reduced* the likelihood that Mandarin speakers will shift their attention to the chosen classifier-match objects. If we had used more lenient criteria to select materials (i.e., allowing visually similar items: snake and river; or semantically similar items: snake and dog), it is possible that we could have found a language-specific effect: Mandarin speakers might have shown a significantly greater shift of overt attention towards the picture of a river on hearing ‘snake’ than a control group of (for example) Dutch speakers (*cf.*, the judgment tasks of Saalbach and Imai, 2007). River and snake share global visual similarity for both Mandarin and Dutch speakers (but in addition they share a classifier category in Mandarin). Maybe classifiers require some core (non-language-specific) visual or semantic similarity in the first place for language-specific effects to emerge (see Davidson and Gelman, 1990). Though we cannot rule out this possibility, we believe our test of the Whorfian hypothesis is the most appropriate since it eliminates possible confounding variables, such as cultural or experiential factors.

Our findings add to the current body of work that goes beyond the simplistic dichotomy of whether language affects thought to try to establish how, when, and under what conditions language-specific distinctions affect human behavior. Our data suggest that previous findings demonstrating a strong role for classifiers in the everyday cognition of Mandarin speakers (Zhang and Schmitt, 1998; Schmitt and Zhang, 1998) may have been mediated by a strategic recruitment of classifier categories during the decision-making stage of reasoning tasks. Our similarity-rating pre-tests and Saalbach and Imai (2007) have demonstrated that although Mandarin speakers are sensitive to similarities between objects sharing a classifier, even Dutch and German speakers are attuned to this likeness, suggesting that classifiers are picking up

on perceptual/conceptual resemblances already present between objects. Studies such as Zhang and Schmitt's (1998) are also potentially problematic in the dependent measures they use. Participants are typically asked for explicit judgments, such as similarity judgments or word-picture matching, but similarity judgments are known to be particularly prone to strategic biases (Goldstone, 1994). For example, participants could be recruiting the classifier category strategically during the tasks, so any similarity effect could be due to having a shared label ("These two things are similar because I describe them in the same way") rather than to a difference in the underlying representation per se (Bloom and Keil, 2001; Goldstone et al., 2001).

It could have been possible that classifiers affected judgments through automatic (rather than strategic) recruitment of labels. In the domain of color, for example, there is now converging evidence that language is recruited involuntarily during simple perceptual tasks (e.g., Gilbert et al., 2006; Tan et al., 2008; Thierry et al., 2009). Even when participants engage in a visual search task that does not require any linguistic processing (determining whether a different-colored patch is on the left or right side), participants still show an advantage when the odd-one-out is in the right visual field, which projects to the left hemisphere where language is processed (Gilbert et al., 2006).

In contrast to the apparent strong online influence of color language on color memory and perception (see also Davidoff et al., 1999; Roberson et al., 2000; Gilbert et al., 2006; Winawer et al., 2007) stands the absence of language-on-cognition effects in the domain of motion (see manner *vs.* path categorization studies of Gennari et al., 2002; Papafragou et al., 2002, 2006). Gennari et al. (2002), for example, compared English and Spanish speakers' linguistic descriptions to their performance in recognition memory and similarity judgments. These authors found no influence of language on recognition memory after either linguistic or non-linguistic encoding of events. Interestingly, they found an effect of language in the similarity judgment task after verbal encoding but not after non-verbal encoding. Gennari et al. (2002: 76) conclude that the effect in the similarity judgments "may reflect a conscious or unconscious strategy adopted by speakers, where they solved a difficult judgment by appealing to the visual dimension(s) features made salient by their description of the task". Note in this regard that Saalbach and Imai's (2007) study of Mandarin classifiers used a variety of measures in addition to similarity judgments. They observed an influence of language-specific classifiers in similarity judgment tasks but not in reaction-time priming tasks, which draw more on automatic processes. Gao and Malt (2009) also found that knowledge of Mandarin classifier categories influenced performance

when the classifier was explicitly mentioned in a memory task but not in other more implicit tasks. Gao and Malt conclude that the effect of classifiers on conceptual organization may be ‘quite modest’ under real world conditions in which people normally encounter nouns, i.e., when the nouns are embedded in spoken or written sentences that carry additional contextual meaning.

To conclude, our data demonstrate that Mandarin numeral classifiers influence online overt attention only during linguistic processing of these language-specific distinctions. This finding adds to our current understanding of where to expect language-on-cognition effects: for classifiers, effects are more likely to emerge in tasks involving some strategic deliberation than in tasks that rely on more automatic conceptual processing; for motion, even strategic effects are minimal; while in the domain of color there is growing evidence that language may affect the early stages of color perception. This highlights the need for a shift in focus for language and cognition research: from the simplistic question of whether or not there are Whorfian effects to the more nuanced elaboration of why and how language-on-cognition effects emerge in some domains rather than others.

Acknowledgements

We thank Ludy Cilissen and Nan van de Meerendonk for assistance with running the experiments and analyzing the data. We are also grateful to John Lucy and Dan Slobin for helpful discussion of this research and Henrik Saalbach for sending us his PhD dissertation and the complete list of his stimuli. Part of this research was presented at the AMLaP 2008 conference in Cambridge, UK.

References

- Ahrens, K. (1994). Classifier production in normals and aphasics. *Journal of Chinese Linguistics* 22, 202-247.
- Allan, K. (1977). Classifiers. *Language* 53, 285-311.
- Altmann, G. T. M. and Kamide, Y. (1999) Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73, 247-264.
- . (2004). Now you see it, now you don't: Mediating the mapping between language and the visual world. In Henderson, J. M. and Ferreira, F. (Eds), *The interface of language, vision and action*, pp. 347-386. Psychology Press, Hove.
- Bloom, P. and Keil, F. C. (2001). Thinking through language. *Mind and Language* 16, 351-367.
- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology* 6, 84-107.

- Davidoff, J., Davies, I. and Roberson, D. (1999). Color categories of a stone-age tribe. *Nature* 398, 203-204.
- Davidson, N. S. and Gelman, S. A. (1990). Inductions from novel categories: The role of language and conceptual structure. *Cognitive Development* 5, 151-176.
- Gao, M. Y. and Malt, B. C. (2009). Mental representation and cognitive consequences of Chinese individual classifiers. *Language and Cognitive Processes* 24, 1124–1179.
- Gennari, S. P., Sloman, S., Malt, B. and Fitch, T. (2002). Motion events in language and cognition. *Cognition* 83, 49-79.
- Gilbert, A. L., Regier, T., Kay, P. and Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences of the USA* 103, 489-494.
- Goldstone, R. L. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition* 52, 125-157.
- Goldstone, R. L., Lippa, Y. and Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition* 78, 27-43.
- Greenberg, J. H. (1972). Numeral classifiers and substantival number: Problems in the genesis of a linguistic type. *Working Papers on Language Universals* 9, 1-39.
- Gumperz, J. J. and Levinson, S. C. (1996). *Rethinking linguistic relativity*. Cambridge University Press, Cambridge.
- Huang, C.-R. and Ahrens, K. (2003). Individuals, kinds and events: Classifier coercion of nouns. *Language Sciences* 25, 353-373.
- Huettig, F. and Altmann, G. T. M. (2004). The online processing of ambiguous and unambiguous words in context: Evidence from head-mounted eye-tracking. In Carreiras, M. and Clifton, C. (Eds), *The on-line study of sentence comprehension: Eyetracking, ERP and beyond*, pp. 187-207. Psychology Press, New York, NY.
- . (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition* 96, B23-B32.
- . (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition* 15, 985-1018.
- Huettig, F. and McQueen, J. M. (2007). The tug of war between phonological, semantic, and shape information in language-mediated visual search. *Journal of Memory and Language* 54, 460-482.
- Huettig, F., Quinlan, P. T., McDonald, S. A. and Altmann, G. T. M. (2006). Models of high-dimensional semantic space predict language-mediated eye movements in the visual world. *Acta Psychologica* 121, 65-80.
- Lehman, F. K. (1979). Aspects of a formal theory of noun classifiers. *Studies in Language* 3, 153-180.
- Li, C. N. and Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar*. University of California Press, Berkeley, CA.
- Lucy, J. A. (1992). *Language diversity and thought: A reformulation of the linguistic relativity hypothesis*. Cambridge University Press, Cambridge.
- Meyers, J. (2000). Rules vs. analogy in Mandarin Chinese selection. *Chinese Language and Linguistics* 1, 187-209.
- Papafraou, A., Massey C. and Gleitman, L. (2002). Shake, rattle, ,n´ roll: The representation of motion in language and cognition. *Cognition* 84, 189-219.
- . (2006). When English proposes what Greek presupposes: The cross-linguistic encoding of motion events. *Cognition* 98, B75-B87.
- Roberson, D., Davies I. and Davidoff, J. (2000). Colour categories are not universal: Replications and new evidence from a Stone-age culture. *Journal of Experimental Psychology: General* 129, 369-398.

- Saalbach, H. and Imai, M. (2007). Scope of linguistic influence: Does a classifier system alter object concepts? *Journal of Experimental Psychology: General* 136, 485-501.
- Sapir, E. (1921). *Language: An introduction to the study of speech*. Harcourt, Brace and Company, New York, NY.
- Schmitt, B. and Zhang, S. (1998). Language structure and categorization: A study of classifiers in consumer cognition, judgment, and choice. *Journal of Consumer Research* 25, 108-122.
- Slobin, D. I. (1996). From “thought and language” to “thinking for speaking”. In Gumperz, J. J. and Levinson, S. C. (Eds), *Rethinking linguistic relativity*, pp. 70-96. Cambridge University Press, Cambridge.
- . (2003). Language and thought online: Cognitive consequences of linguistic relativity. In Gentner, D. and Goldin-Meadow, S. (Eds), *Language in mind: Advances in the investigation of language and thought*, pp. 157-191. MIT Press, Cambridge, MA.
- Tan, L. H., Chan, A. H. D., Kay, P., Khong, P.-L., Yip, L. K. C., Luke, K. K. (2008). Language affects patterns of brain activation associated with perceptual decision. *Proceedings of the National Academy of Science of the USA* 105, 4004-4009.
- Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B. and Kuipers, J.-R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Science of the USA* 106, 4567–4570.
- Winawer, J., Witthoft, N., Frank, M., Wu, L., Wade, A. and Boroditsky, L. (2007). The Russian Blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Science of the USA* 104, 7780-7785.
- Whorf, B. L. (1956). Language, thought, and reality. In: Carroll, J. B. (Ed.), *Selected writings of Benjamin Lee Whorf*. Cambridge, MA, MIT Press.
- Zhang, H. (2007). Numeral classifiers in Mandarin Chinese. *Journal of East Asian Linguistics* 16, 43-59.
- Zhang, S. and Schmitt, B. (1998). Language-dependent classification: The mental representation of classifiers in cognition, memory, and ad evaluations. *Journal of Experimental Psychology: Applied* 4, 375-385.

Appendix A: Experimental Items Used in Both Experiments

Classifier	Mandarin character	Competitor word	Target word	Target picture
<i>ba3</i>	把	scissors	chair	chair
<i>ba3</i>	把	comb	pick	pick
<i>ba3</i>	把	umbrella	knife	knife
<i>ba3</i>	把	rocking chair	shovel	shovel
<i>ba3</i>	把	key	Chinese fan	Chinese fan
<i>chuan4</i>	串	grapes	(a bunch of) keys	(a bunch of) keys
<i>chuan4</i>	串	(a bunch of) bananas	necklace	necklace
<i>dui1</i>	堆	hay	fire	fire

Appendix A: (cont.)

Classifier	Mandarin character	Competitor word	Target word	Target picture
<i>duo3</i>	朵	cloud	flower	flower
<i>fu4</i>	副	saddle	glasses	glasses
<i>fu4</i>	副	chess	handcuffs	handcuffs
<i>gen1</i>	根	cigarette	cane	cane
<i>gen1</i>	根	match	sausage	sausage
<i>jia4</i>	架	(air)plane	ladder	ladder
<i>jian4</i>	件	vest	present	present
<i>ke1</i>	棵	palm tree	lettuce	lettuce
<i>ke1</i>	棵	celery	tree	tree
<i>ke1</i>	颗	strawberry	satellite	satellite
<i>kuai4</i>	块	watermelon (slice)	field	field
<i>kuai4</i>	块	gum	(a piece of) cake	(a piece of) cake
<i>kuai4</i>	块	bread	napkin	napkin
<i>mian4</i>	面	mirror	flag	flag
<i>pan2</i>	盘	tape	spaghetti	spaghetti
<i>pian4</i>	片	leaf	snowflake	snowflake
<i>shuang1</i>	双	chopsticks	(a pair of) shoes	(a pair of) shoes
<i>tai2</i>	台	TV	microscope	microscope
<i>tiao2</i>	条	leg	scarf	scarf
<i>tiao2</i>	条	jump rope	tunnel	tunnel
<i>tiao2</i>	条	pants	snake	snake
<i>tiao2</i>	条	fish	towel	towel
<i>tiao2</i>	条	tail	bracelet	bracelet
<i>tiao2</i>	条	chain	dolphin	dolphin
<i>zhang1</i>	张	desk	paper	paper
<i>zhang1</i>	张	(one) poker	spider web	spider web
<i>zhang1</i>	张	stamp	mouth	mouth
<i>zhi1</i>	只	cat	pencil	pencil
<i>zhi1</i>	只	chicken	ear	ear
<i>zuo4</i>	座	castle	slide	slide