Speaking of shape: The effects of language-specific encoding on semantic representations

PAMELA PERNISS[#], DAVID VINSON[#], FRANK SEIFART[~] AND GABRIELLA VIGLIOCCO[#*]

[#]*Deafness, Cognition, and Language Research Centre (DCAL), UCL, London*
[~]*Max Planck Institute for Evolutionary Anthropology, Leipzig*

Abstract

The question of whether different linguistic patterns differentially influence semantic and conceptual representations is of central interest in cognitive science. In this paper, we investigate whether the regular encoding of shape within a nominal classification system leads to an increased salience of shape in speakers' semantic representations by comparing English, (Amazonian) Spanish, and Bora, a shape-based classifier language spoken in the Amazonian regions of Columbia and Peru. Crucially, in displaying obligatory use, pervasiveness in grammar, high discourse frequency, and phonological variability of forms corresponding to particular shape features, the Bora classifier system differs in important ways from those in previous studies investigating effects of nominal classification, thereby allowing better control of factors that may have influenced previous findings. In addition, the inclusion of Spanish monolinguals living in the Bora village allowed control for the possibility that differences found between English and Bora speakers may be attributed to their very different living environments. We found that shape is more salient in the semantic representation of objects for speakers of Bora, which systematically encodes shape, than for speakers of English and Spanish, which do not. Our results are consistent with assumptions that semantic representations are shaped and modulated by our specific linguistic experiences.

* Correspondence address: Pamela Perniss, Deafness, Cognition and Language Research Centre (DCAL), 49 Gordon Square, London WC1H 0PD, UK. E-mail: p.perniss@ucl.ac.uk. This work was supported by a Marie Curie Fellowship (MEIF-CT-2006-041886) awarded to the first author and an Economic and Social Research Council (ESRC) of Great Britain grant (Grant RES-620-28-6001) awarded to the Deafness, Cognition and Language Research Centre (DCAL). We would like to thank Doris Fagua for the collection of most of the Spanish data, and Bernard Comrie, Olivier Le Guen, Lorena Pool Balam, and two reviewers for helpful comments on an earlier version of the paper.

Language and Cognition 4–3 (2012), 223–242
DOI 10.1515/langcog-2012-0012
1866–9808/12/0004–0223
© Walter de Gruyter

https://doi.org/10.1515/langcog-2012-0012 Published online by Cambridge University Press

Keywords
semantic representation, nominal classification, shape-based semantics, semantic similarity, linguistic relativity

1. Introduction

The question of whether different linguistic patterns differentially influence patterns of semantic and conceptual representation is of central interest in cognitive and linguistic sciences. Languages exhibit vast diversity in morphological, semantic, and syntactic structure (cf. Evans and Levinson 2009), and it thus stands to reason that the regular encoding of a certain feature (e.g. shape) in a language would make it particularly salient for speakers of that language compared to speakers of another language in which this feature is not encoded. As a result, speakers may exhibit differences in the extent to which they use this feature in performing tasks requiring judgments about semantic similarity, categorization, or memory recall. An area of particular interest is the linguistic categorization imposed by systems of nominal classification, or classifier systems. Classifiers are words or affixes that occur in collocation with a noun and categorize nouns into distinctive semantic classes according to certain semantic properties. Across classifier languages, which are prominent in East and Southeast Asia, Meso- and South America, and Austronesia (Aikhenvald 2000; Allan 1977; Dixon 1986; Senft 2000), shape is a primary semantic feature by which entities are classified.[1] For example, in Mandarin Chinese, the classifier *tiao* typically occurs with nouns denoting entities that are long and thin in shape (e.g. ropes, snakes, belts) (Gao and Malt 2009). Given the prominence of shape as a semantic feature of classification, a straightforward prediction is that shape would be particularly salient for speakers of a classifier language in semantic representation, such that they would rely on shape to a greater extent than speakers of a non-classifier language in tasks involving similarity judgments or categorization.

This question has been investigated in a number of previous studies, yielding equivocal results, with some studies finding a greater reliance on shape by speakers of a classifier language compared to speakers of a non-classifier language (Gao and Malt 2009; Imai et al. 2010; Saalbach and Imai 2007; Schmitt and Zhang 1998; Srinivasan 2010; Zhang and Schmitt 1998), others finding less reliance (Imai and Gentner 1997; Imai and Mazuka 2007; Lucy 1992), and still others finding no difference between speakers of classifier vs. non-classifier languages (Mazuka and Friedman 2000). Differences in the cultural

1. Other features include size, animacy, function, rigidity, flexibility, and social or cultural role.

https://doi.org/10.1515/langcog-2012-0012 Published online by Cambridge University Press

or socio-linguistic background of speaker groups may play a role in whether cross-linguistic differences are observed or not. However, more crucially, differences in the classifier systems that have been investigated may have been decisive in determining the type of results obtained. These differences, described in some detail below, include variation in the obligatoriness and frequency of use of classifiers, the semantic coherence of classifier categories, and the role of classifiers in specifying the reference of their associated nouns.

Many classifier languages fall under the *numeral classifier* type, which means that classifiers are used in counting contexts to specify the entities that are counted. For example, to say 'one rope', speakers of Mandarin Chinese must say *yi* ***tiao*** *shengzi*, where the classifier *tiao* (for long, thin entities) occurs between the numeral *yi* 'one' and the object noun *shengzi* 'rope' (Gao and Malt 2009: 1125). Numeral classifier languages differ, however, in the referential behavior of nouns. In some numeral classifier languages, e.g. Yucatec Maya and Japanese, nouns behave like mass nouns, referring to substances. For example, in Yucatec Maya, the noun *kib* is best translated into English as 'wax'. Individuating the *kib* substance into countable entities like 'candles' (i.e. long, thin objects made out of wax) requires the use of a classifier meaning 'long, thin' together with a numeral, e.g. *hun-**ts'íit** kib* (one-**long thin** wax) 'one candle' and *ká'a-**ts'íit** kib* (two-**long thin** wax) 'two candles'. In contrast, nouns in Mandarin Chinese refer not to substances, but to individuated objects. Nouns occur with the classifier corresponding to their noun class, but refer unambiguously to specific entity types. For example, the Mandarin Chinese noun *xiangjiao* 'banana' takes the one-dimensional classifier, while the complex noun *xiangjiao ye* 'banana leaf' takes the two-dimensional classifier (Kuo and Sera 2009: 18).

This difference in the reference of nouns in numeral classifier languages may well be one main reason why different effects have been observed. Because nouns in Yucatec Maya and Japanese refer to "shapeless" substances (and thus need a classifier to give something a specific shape for counting purposes), it has been argued that speakers of these languages pay *less* attention to shape, relying on substance-based rather than shape-based categorization strategies, and possibly biasing cognitive tasks (Gaskins and Lucy 2003; Imai and Gentner 1997; Imai and Mazuka 2007; Lucy 1992; Lucy and Gaskins 2001). However, studies looking at effects of classifiers in Mandarin have argued that speakers pay *more* attention to shape because object shape is the prominent semantic basis for noun classification (Kuo and Sera 2009; Saalbach and Imai 2007; Srinivasan 2010; Zhang and Schmitt 1998). In these studies, Mandarin Chinese speakers classified objects according to shape—and thus classifier—similarity to a greater degree than English speakers.

In addition, the frequency and systematicity of classifier use within the grammar of different classifier languages have been argued to influence effects

https://doi.org/10.1515/langcog-2012-0012 Published online by Cambridge University Press

of classifiers on categorization. Imai and Saalbach (2010) found a classifier similarity effect for Mandarin speakers, but not for Japanese speakers. The authors attribute this difference to the higher frequency and greater obligatory occurrence of classifiers in Mandarin compared to Japanese. Whereas classifiers accompany nouns obligatorily in numeral as well as in demonstrative phrases in Mandarin Chinese, they occur only in numeral phrases in Japanese and can moreover be dropped when quantity can be pragmatically or contextually retrieved.

Furthermore, the semantic coherence of individual classifier categories may differ substantially within a language, such that the nature of mental representations between classifier categories may be very different. Gao and Malt (2009) defined three different types of classifier categories in Mandarin Chinese—well-defined categories, prototypical categories, and arbitrary categories—and found that speakers differed in noun recall and clustering tasks based on the type of category to which items belonged. While the classification of objects imposed by the classifier system had little *overall* impact on speakers' representations, there was a clear effect of category membership for the well-defined categories, for which a specific semantic feature (e.g. shape) fully defines category membership, compared to the other two categories, where knowledge of a specific feature defines only a category prototype or where category membership is semantically incoherent and largely arbitrary.

Differences in cultural and linguistic experience between speaker groups may also play a role in the results obtained. For example, differences in the amount of language exposure between different groups of speakers of the same classifier language have been shown to modulate effects. Kuo and Sera (2009) found differences in the degree to which Mandarin Chinese speakers pay attention to shape as a function of how regularly they used the language. Mandarin Chinese speakers who had been living in the United States for many years, and who thus had less exposure to habitual shape-encoding in classifiers, were less likely to categorize on the basis of shape than Mandarin speakers living in Taiwan. Educational background and cultural experience have also been argued to affect outcomes. In an attempt to replicate the findings of Lucy (1992), Mazuka and Friedman (2000) compared the performance of English and Japanese speakers in a semantic similarity judgment task. Contrary to what Lucy's results would have predicted, Mazuka and Friedman did not find Japanese speakers to prefer substance-based classification, but rather found both groups to prefer shape-based classification. The authors argue that Lucy's original finding may be better explained by the different backgrounds of the English and Yucatec Maya speaker groups. However, as this discussion highlights, it is also possible that the differences we have described between the classifier systems of Japanese and Yucatec Maya could explain why they did not replicate Lucy's (1992) findings.

https://doi.org/10.1515/langcog-2012-0012 Published online by Cambridge University Press

Thus, any effect of linguistic encoding of properties like shape on semantic representation, or other cognitive process, may be differentially influenced by specific characteristics of the classifier system itself, as well as by individual differences in language use. In addition, within a single language, differences between classifier categories with respect to meaningfulness and coherence may lead to differential effects on speakers' semantic representations. The variability of these factors in previous studies makes it difficult to understand the implications of the linguistic use of classifiers in our representation and categorization of shape.

Finally, in previous studies, subjects may have used shared (form-equivalent) classifiers as a strategic cue to meaning similarity (Saalbach and Imai 2007; Schmitt and Zhang 1998). (See Srinivasan (2010) for the use of a speeded task to avoid strategic use of classifier category membership.) Because classifier membership can be used strategically in this way in many languages, most previous results do not permit the conclusion that classifier categorization affects speakers' semantic representations.

In the study we report here, we investigate the question of whether regular encoding of a semantic feature (i.e. shape) leads to an increased salience of this feature in speakers' semantic representations by comparing English with Bora, a classifier language that differs in crucial ways from the classifier languages previously investigated in this domain. In Bora, an indigenous language spoken in the Amazonian regions of Colombia and Peru, classifiers are pervasive and obligatory in grammar, frequent in discourse, and predominantly and neatly characterized by shape semantics. However, shape is not the only semantic component that characterizes the inventory of Bora classifiers, and there is not always a transparent correspondence between the shape of an object and the classifier used in the noun referring to the object. For these reasons, Bora provides the opportunity for better control of the linguistic factors that have differentially influenced results in previous studies. In addition, by also comparing the Bora speakers to monolingual Spanish speakers living in the same community (as the Bora speakers), we were able to limit differences in living environments.

In Bora, classifiers appear obligatorily on the nouns themselves, as part of the derivational morphology of nouns.[2] This means that the noun form in Bora does not exist independently of the classifier (in contrast to the numeral classifiers in Chinese, Japanese, or Yucatec Maya), but is rather an integral part of the nominal lexeme. For example, the Bora word *acúúve**-wa*** 'bench' consists of the nominalized verb *acúúve* 'sitting' and the classifier ***-wa*** 'plank-like'.

2. Much of the Bora language information reported here is taken from Seifart (2005), which describes the classifier system of the close dialectal variant Miraña. The Bora and Miraña classifier systems are identical.

https://doi.org/10.1515/langcog-2012-0012 Published online by Cambridge University Press

Similarly, the word *peeté**-i*** 'candle' consists of *peeté* 'burning' and ***-i*** 'stick-like'. In addition, classifiers also appear on determiners and modifiers (similar to gender agreement in many European languages), for instance *í**-wa** acúúve**-wa*** 'this bench' vs. *í**-i** peeté**-i*** 'this candle', or *mítya**-wa** acúúve**-wa*** 'big bench' vs. *mítya**-i** peeté**-i*** 'big candle', as well as on pronouns and verbs, for purposes of agreement marking and referent tracking in discourse. Furthermore, of the total inventory of approximately 70 classifiers, the 8 classifiers that comprise the most frequently occurring core set (including ***-wa*** and ***-i***) can be coherently described in terms of object shape (see Appendix 1). Specifically, the shape-based semantics of the core set can be described in terms of salient object dimensionality: 1-dimensional (i.e. long, thin), 2-dimensional (i.e. flat, square/round), 3-dimensional (i.e. spherical, oblong). However, as mentioned above, classifier categories vary in their semantic coherence (as is the case with many nominal classification systems), such that it is not possible to predict which classifier a noun will have based solely on visuospatial properties of the objects.[3]

Thus, in contrast to the previously investigated numeral classifier systems (which perform a specific function in specific contexts and may exhibit high variability in category membership), the classifier system in Bora is characterized by pervasive and obligatory occurrence in grammar, a strong coherence of category membership around the primary semantic feature of shape, and a high frequency of use in discourse. In addition, because object shape does not straightforwardly predict classifier category membership, we can avoid a potential confound between phonological form similarity and object similarity in judgments. These facts make Bora an ideal language for investigating effects of shape-based classifiers on patterns of semantic representation.

In this study, we use a meaning similarity judgment task to investigate the effects of differences in the linguistic encoding of shape on semantic representations by comparing speakers of Bora and English. We created triads of words for objects that differed saliently in shape (specifically, in spatial dimensionality): objects were either long/thin (i.e. one-dimensional, like a pole) or flat/wide (i.e. two-dimensional, like a table). We predicted that the more prominently shape factored into the encoded semantics of a speaker's language, the more likely word pairs with shared dimensionality would be picked as similar. Compared to English speakers, we expect Bora speakers to display a heightened sensitivity to shape as a result of habitual encoding of shape-based contrasts in the language, and thus for shape to contribute more strongly to Bora speakers' semantic representations.

3. Examples of apparent mismatches with respect to dimensionality are the Bora words for 'flag' (*badééra-ba*, 'flag' (Spanish loan)-CL:3-DIMENSIONAL) and 'spear' (*aamú-ba*, hit-CL:3-DIMENSIONAL), which both use a classifier for three-dimensional objects.

https://doi.org/10.1515/langcog-2012-0012 Published online by Cambridge University Press

Critically, Bora and English (UK) also differ substantially in terms of the environment in which the languages are used (e.g. rural rainforest vs. urban European environment). Thus, we face the skeptic's criticism that any differences found between Bora and English speakers in the salience of shape as a semantic feature of objects could be attributed not to differences in the linguistic encoding of shape between the two languages but rather to the different living environments of the groups. To address this, we also compared Bora and English to (Amazonian) Spanish, a language spoken in the same context as Bora, but which is similar to English in that it does not have classifiers and does not prominently encode shape.[4] Monolingual Spanish speakers who live in the Bora community (and have done so for at least 10 years) were tested on the same meaning similarity judgment task. The Bora and Spanish monolingual villagers share a lifestyle and livelihood. They are horticulturalists, and engage in the same daily activities, working and moving about in the village, on the rivers, and in the jungle. Our prediction was that the Spanish data would resemble the English data, based on the assumption that marked differences between languages in the encoding of spatial information will have a differential effect on speakers' semantic representations.

2. Method

Object shape and dimensionality are critical conceptual features across languages and cultures, and we may expect speakers of any language to use these features in judging object similarity. The critical question we ask is whether such a general tendency might be further boosted by linguistic marking of these features. To address this question, experimental items need not only be categorized by object dimensionality (1-dimensional vs. 2-dimensional), but, importantly, must also be normed for similarity across items sharing or not sharing dimensionality. This ensures that potentially small cross-linguistic differences can be detected if present. In order to select appropriate items, we carried out the following preliminary study.

2.1. *Preliminary study: Pair-wise meaning similarity rating*

2.1.1. *Subjects.* Twenty-four participants were recruited from the University College London (UCL) subject pool. All participants were native English speakers. Ages ranged from 19–50 (mean 26).

4. English and Spanish lexemes may contain semantic information about object shape, as exhibited in the English container words *bottle, gourd, flask,* and *vial*. Indeed, Lucy (1992) takes English nouns to encode shape as part of their lexical semantics, yet crucially, this semantic shape information is not explicitly and systematically marked by dedicated linguistic forms like classifiers.

https://doi.org/10.1515/langcog-2012-0012 Published online by Cambridge University Press

2.1.2. *Materials and procedure.* From a set of 72 object words consisting of words for long/thin objects (one-dimensional; e.g. *pen, knife, broom*) and flat/wide objects (two-dimensional; e.g. *table, rug, banknote*), we created all possible word pair combinations (N = 2556) (see Appendix 2 for the full list of object words). Four different random orders of these word pairs were then assigned and divided into six lists of 426 word pairs each. Participants were asked to judge the meaning similarity between each word pair on a 1–7 scale (1 = very dissimilar; 4 = neither dissimilar nor similar; 7 = very similar), entering their responses directly into an Excel spreadsheet.

For each word pair, we calculated the average similarity rating, and categorized the pairs according to whether the two words shared or did not share object dimensionality. This allowed us to identify word pairs that had low semantic similarity despite sharing object dimensionality, and word pairs with high semantic similarity (e.g. based on thematic relations) but different dimensionality, which we used in creating the materials for the main experiment. Relatively low similarity for objects sharing dimensionality is essential if we are to observe (additional) effects of linguistic encoding of dimensionality in Bora, as shape is an important conceptual property universally, and because thematic relations have been shown to be a strong organizer of concepts in both classifier and non-classifier languages (Imai and Saalbach 2010; Kuo and Sera 2009; Saalbach and Imai 2007).

The average similarity rating for word pairs overall (i.e. pairs sharing and not sharing dimensionality) was 1.5 ("very dissimilar"). We defined word pairs with low semantic similarity as those that received an average rating of between 1 and 1.5, while word pairs with a high similarity received an average rating of 2.5 or higher. The cut-off for high similarity was thus quite low in terms of the scale itself, but was a result of the overall low similarity ratings.

2.2. *Main experiment: Triadic meaning similarity judgment*

In the main study, speakers of Bora, English, and Spanish were presented with word triads and were asked to pick the odd one out in terms of semantic similarity. We predicted that Bora speakers would rely on shape to a greater degree than English and Spanish speakers in judging similarity, thus considering word pairs sharing dimensionality to be more similar than those not sharing dimensionality.

2.2.1. *Subjects.* (a) *English:* Forty-two participants were recruited from the UCL subject pool. All participants were native speakers of English with little or no knowledge of a foreign language. Ages ranged from 18–60 (mean 26). Participants were paid for their time.

https://doi.org/10.1515/langcog-2012-0012 Published online by Cambridge University Press

(b) *Bora:* Ten participants were recruited in the Bora village in Peru by one of the authors while doing field research within the community. All Bora participants were bilingual Bora-Spanish speakers, but considered Bora to be their first language and primary language of communication. Ages ranged from 18–63 (mean 45). Participants were paid for their time.

(c) *Spanish:* Eleven participants were recruited in the same Bora village in Peru by a linguist while doing field research within the community. All participants were monolingual speakers of Spanish, but had lived in the Bora village for at least 10 years.[5] Ages ranged from 20–60 (mean 34). Participants were paid for their time.

2.2.2. *Materials.* Word triads were created only from those words in the preliminary study that had clear English-Bora-Spanish translation equivalents and whose referents were culturally familiar to all groups.[6] For each triad, we started with a pair of words that shared object dimensionality (e.g. *bone-string*), but which did not exhibit a high degree of semantic similarity according to the ratings we obtained from English speakers in the preliminary study (in contrast to high-similarity dimension-sharing pairs like *book-paper*). Each word pair was then made into a triad by adding one additional word varying in dimensionality from the starting pair, with the requirement that the two additional (cross-dimensional) pairings had equal or greater similarity ratings (e.g. *bone-string-paper*).

A total of 710 word triads were generated by this method, from a final set of 40 object words. For English, audio recordings of all 40 object words, spoken by a native speaker of British English, were edited using Audacity (www.audacity.sourceforge.net). Matlab, version 6.5.1 (The Mathworks™), was then used to concatenate three single-word audio files together into a single audio file for each triad. The same procedure was followed to create audio files for each triad in Bora, with object words spoken by a phonetically trained linguist and Bora scholar. Similarly, all 710 triads were created in Spanish. However, due to time constraints (related to the scheduling of a field trip to the Bora village), it was not possible to create audio files for the triads in Spanish.

The triads were divided into seven lists (of 101 or 102 triads, the order of which was randomized). In addition, each triad (A-B-C) was created in the reverse order (C-B-A), resulting in two sets of seven lists. For English and

5. A few of the monolingual Spanish speakers have some vocabulary knowledge of Bora, but cannot communicate in the language, or understand an ongoing conversation, to any degree.
6. In addition, a Bora villager was asked to make a drawing of each object as an extra measure to ensure that the words in the different languages were indeed labeling the same objects.

https://doi.org/10.1515/langcog-2012-0012 Published online by Cambridge University Press

Bora, the audio files of the triads were embedded into a Microsoft PowerPoint presentation for each list using PPTmerge (www.pptools.com), with one triad per slide. For Spanish, the 14 lists of triads were saved in Microsoft Excel files.

2.2.3. *Procedure.* English and Bora participants were seated at a desktop or laptop computer, listening to triads via headphones. Spanish triads were read aloud by a native speaker of Spanish, one at a time and at a rate comparable to that in the audio files for English and Bora. Participants were instructed to listen to the words in a triad, and to then decide which two of the three words were more similar in meaning (not form), marking the odd one (i.e. the semantically least similar word) on a response sheet. Instructions stressed that participants should consider only similarity in meaning in making their decision, and not other similarities (e.g. phonological similarity).

Instructions were provided in English for the English participants, in local (Amazonian) Spanish for the Spanish participants, and in a mixture of Bora and local Spanish for the Bora participants. The instructions in Spanish used the verbs *parecerse* 'be similar' and *ser como* 'be like' to translate the notion of similarity. The instructions in Bora used the expression *X-du neéne* 'seem like X' to translate the notion of similarity.[7] All participants indicated that they clearly understood the task and that they had no further questions before starting the task.

Individual participants were allowed to complete as many lists as they wanted from one of the sets (up to 7 lists). English participants completed between 1–3 triad lists each, for a total of 50 lists. Bora participants completed 5 lists each, for a total of 50 lists. Spanish participants completed 5 lists each, for a total of 55 lists.

3. Data analysis and results

We first excluded from analysis 11 triads in which the three words did not differ in dimensionality, leaving 699 unique triads in all. We fit logistic regression models to trial-level data in order to test whether speakers of the different languages exhibited different tendencies to select word pairs that matched in

7. Any cross-linguistic task faces the problem of obtaining reliable translation equivalence. We believe that the chosen translations conveyed the notion of semantic similarity in a comparable way. All instructions were careful in formulating the difference between semantic similarity and phonological similarity (e.g. given the choice between *blue, shoe,* and *red,* the odd one out would be *shoe*, since *blue* and *red* are more similar in meaning than either *red* and *shoe* or *blue* and *shoe*, with the phonological similarity between *blue* and *shoe* being irrelevant to the task).

https://doi.org/10.1515/langcog-2012-0012 Published online by Cambridge University Press

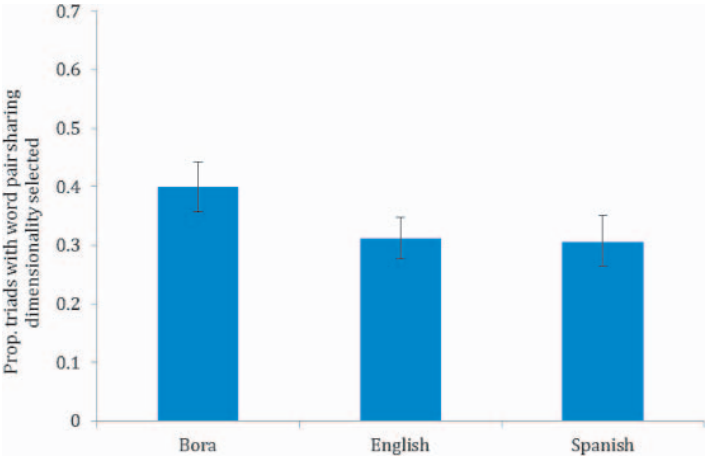


Figure 1. *Proportion of triads in which word pairs sharing object dimensionality were selected by speakers of Bora, English, and Spanish after taking into account effects of List and random intercepts for subjects and items. (Error bars reflect standard error of log-likelihood estimate, converted to proportions.)*

dimensionality. In all analyses, the dependent measure was preservation of dimension (i.e. whether a participant chose the two words in a triad that shared dimension or not). The factor of interest was Language (Bora, Spanish, English) and we also included List as a secondary factor of no theoretical interest, with subjects (n = 63) and items (n = 699: each unique triad, nested in List) as random effects. As a starting point, we compared this model to a model that did not include the factor Language but was otherwise identical. The model containing Language (log likelihood = −9560.9) performed significantly better than the model without Language (log likelihood = −9564.2); $\chi^2(2) = 6.6$, $p =$ 0369. We therefore looked specifically at differences between Bora, Spanish, and English, treating Bora as a reference condition (i.e. model intercept) and comparing it to the other two language groups (see Figure 1). Intercept log odds were −.408 (SE estimate = .179), corresponding to a probability of .399 that Bora speakers' judgments preserved dimensionality. Log odds for English speakers (intercept + log odds (English)) were −.797 (SE estimate = .167), p(preserve dimensionality) = .311; significantly less than Bora speakers ($z =$ −2.336, $p = .0195$). Log odds for Spanish speakers were −.820 (SE estimate = .202), p(preserve dimensionality) = .306; also significantly less than Bora speakers ($z = -2.038$, $p = .0415$).

These analyses show that Bora speakers' judgments preserved dimension to a greater extent than judgments by speakers of English and Spanish. However,

https://doi.org/10.1515/langcog-2012-0012 Published online by Cambridge University Press

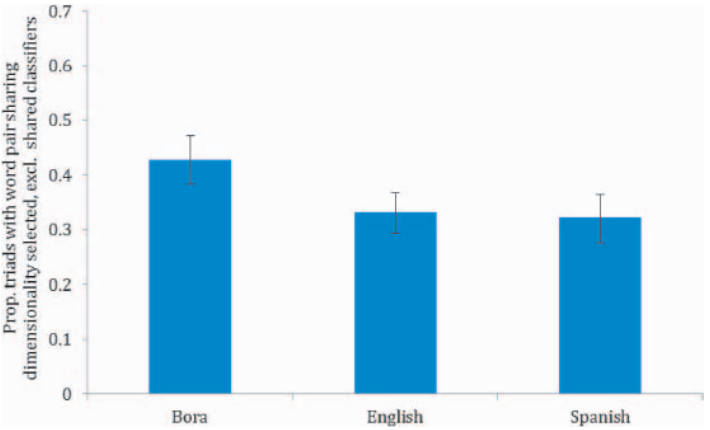


Figure 2. *Proportion of triads in which word pairs sharing object dimensionality were selected by speakers of Bora, English, and Spanish, excluding triads in which word pairs sharing object dimensionality also shared a Bora classifier, and after taking into account effects of List and random intercepts for subjects and items. (Error bars reflect standard error of log-likelihood estimate, converted to proportions.)*

we still need to rule out the possibility that this effect was driven by phonological overlap between classifiers. The 40 Bora words included 17 different classifiers (see Appendix 2). To assess the role of phonological overlap, we coded, for each triad, whether a selected word pair shared classifiers in Bora (for English and Spanish participants, this refers to the Bora translations), and excluded all such triads from analysis. We then fit exactly the same types of logistic regression models as described above to this reduced dataset. Again, a model including Language (log likelihood = −9047.2) performed significantly better than the model without Language (log likelihood = −9050.5); $\chi^2(2) = 6.45$, $p = .0397$. As before, we then looked specifically at differences between Bora, Spanish, and English, treating Bora as a reference condition (see Figure 2). Intercept log odds were −.289 (SE estimate = .179), corresponding to a probability of .428 that Bora speakers' judgments preserved dimensionality. Log odds for English speakers (intercept + log odds (English)) were −.699 (SE estimate = .167), p(preserve dimensionality) = .332; significantly less than Bora speakers (z = −2.461, p = .0139). Log odds for Spanish speakers were −.747 (SE estimate = .202), p(preserve dimensionality) = .322; also significantly less than Bora speakers (z = −2.262, p = .0237). We can therefore conclude that shared classifiers in Bora were not responsible for the greater preservation of dimension in judgments by Bora speakers, because the language difference persisted even after same-classifier pairs were removed from analysis.

https://doi.org/10.1515/langcog-2012-0012 Published online by Cambridge University Press

4. Discussion

We investigated the effect of linguistic encoding of object shape information in three spoken languages—Bora, English, and (Amazonian) Spanish. Bora differs typologically from English and (Amazonian) Spanish with respect to the use vs. non-use of classifiers (i.e. with respect to the regular linguistic encoding of shape information). The English population tested also differs from the Bora population in terms of the context in which the language is spoken (i.e. in the urban vs. rural living environments of the speakers). However, the monolingual (Amazonian) Spanish population lives in the same village as the Bora speakers, such that we were able to address a potential influence of environment on our results through a three-way comparison. The results support our prediction that shape is a more salient feature in semantic representations for speakers of a language like Bora, which systematically marks shape-based contrasts, than for speakers of languages like English and Spanish, which do not. Crucially, Spanish and English did not differ significantly from each other, but both differed significantly from Bora, suggesting that typological differences, i.e. the systematic encoding of shape information, explain the results.

Our results are in line with previous studies showing that speakers of classifier languages rely on shape in categorization, similarity judgment, and memory tasks to a significantly greater degree than speakers of non-classifier languages, because object shape is prominent in the semantics of noun classification (Kuo and Sera 2009; Saalbach and Imai 2007; Srinivasan 2010; Zhang and Schmitt 1998). Moreover, the specific properties of the Bora classifier system lend additional credence to our claim that the linguistic encoding of a semantic feature (specifically, shape) amplifies its salience in speakers' semantic representations. In contrast to the classifier languages used in previous studies (i.e. Mandarin Chinese, Japanese, Yucatec Maya), Bora classifiers are not contextually restricted, but occur obligatorily on nouns and other word classes. Moreover, the semantic coherence of the highly frequent core set of Bora classifiers is particularly strong and particularly focused on shape.

The specific shape-based nature of the Bora classifier system (displaying a range of semantic and phonological variation while being primarily characterized by object dimensionality) has moreover allowed us to address issues related to the strategic use of classifier category membership. In previous studies, the effects on categorization by shape may have been due in part to a strategic reliance on formal similarity, such that objects may have been judged to be the same because they shared a classifier (Saalbach and Imai 2007; Schmitt and Zhang 1998). The substance-individuation function of classifiers (i.e. for countability) in a language like Yucatec Maya means that objects made of the same substance will share a linguistic label. As Bloom and Keil (2001) point out in a criticism of Lucy (1992), Yucatec Maya speakers may have

https://doi.org/10.1515/langcog-2012-0012 Published online by Cambridge University Press

grouped pictures of a 'candle' and 'wax' together because of the shared *kib* label, suggesting that their substance-based classification was the result of covert naming, relying on a shared linguistic label not available to English speakers. In Bora, classifiers belong to the derivational morphology of nouns, and the roots to which classifiers are affixed differ from word to word. In our stimulus materials, specific category membership did not correlate to object dimensionality nor did different classifiers ever appear with the same root, rendering impossible any reliance on formal similarity (of classifiers or roots) to perform the task. This further supports our claim that the heightened sensitivity to shape displayed by Bora speakers, compared to English and Spanish speakers, is explained by differences in the linguistic encoding of shape.

It is important to underscore this point. By ruling out the strategic use of formal similarity, our results demonstrate that the linguistic system (i.e. the Bora classifier system) has an effect over and beyond the linguistic classification itself (i.e. as determined by the classifier categories). Instead, our results indicate that it is indeed the semantics of the system (i.e. the pervasive and systematic encoding of shape information) that drives the observed pattern. Our results suggest that the salient shape semantics of the system makes Bora speakers overall more sensitive to object shape and dimensionality.

In the bigger picture, our results are consistent with assumptions that our specific linguistic experiences shape and influence our semantic representations (often framed in terms of *thinking-for-speaking* effects, Slobin 1996). Specifically, the habitual and regular exposure to a semantic feature through its encoding in the morphology of a language can influence speakers' behavior in tasks (e.g. similarity judgments) involving semantic representation and cognitive processes related to language. In our task, the greater reliance on shape as a basis for semantic classification exhibited by Bora speakers, compared to English and Spanish speakers, is predicted by the fact that shape is the primary semantic feature of Bora's classifier system and is ubiquitously present in Bora grammar. Future research must determine to what extent shape would remain a highly salient dimension of semantic and conceptual representation for Bora speakers using other tasks, including tasks using non-verbal materials. Our study demonstrates that an effect is present when similarity of *word* meaning is judged. This is a prerequisite to establishing effects of linguistic marking of dimensionality with the use of non-verbal materials. As Iwasaki et al. (2010: 194) point out, "[i]f a language-specific property [. . .] were to be found not to affect the semantic representation used in 'thinking for speaking' there would be little point in attempting to establish whether it can affect non-verbal tasks" (see also Kousta et al. 2008; Vigliocco et al. 2005).

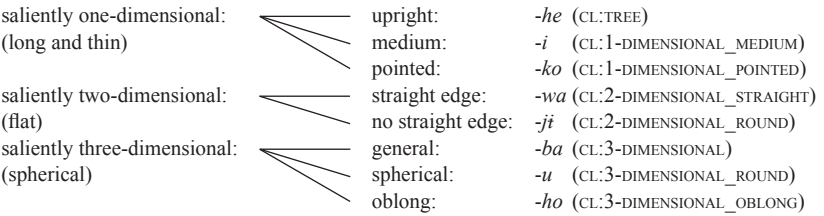
In conclusion, it is interesting to note the similarity between Bora's classifier system and classifier systems found in sign languages, i.e. in the natural visual languages that emerge in deaf communities. Sign languages share fundamental

https://doi.org/10.1515/langcog-2012-0012 Published online by Cambridge University Press

structural properties on all levels of linguistic structure (Klima and Bellugi 1979; Sandler and Lillo-Martin 2006), but their radically different modality of production uniquely privileges the representation of visuospatial (e.g. shape) information. In terms of structural and semantic linguistic encoding, however, the representation of shape information in the Bora classifier system (and in languages with comparable systems) is remarkably similar to the ubiquitous encoding of shape information in sign languages. The study of such classifier languages opens up an avenue of research that investigates not only the effects of typological differences between spoken languages, but also the effects of differences imposed by the modality of language on semantic, and possibly, conceptual representation.

Appendix 1: Core set of Bora classifiers

Semantic distinctions in the core set of classifiers:



Core set of Bora classifiers with examples:

Classifier	**Examples**
-ba CL:3-DIMENSIONAL (i) fruits	*nééba-ba* 'anetto fruit' *neevá-ba* 'fruit (generic)' *iihyúje-ba* 'lemon (fruit)'
-ba CL:3-DIMENSIONAL (ii) logs	*úméne-ba* 'tree trunk' *áálla-ba* '*bombona* (tree, sp.) trunk' *dsííñíja-ba* 'tree, sp. trunk'
-ba CL:3-DIMENSIONAL (iii) mushy objects	*úni-ba* 'spit' *máhni-ba* 'tar' *úji-ba* 'thick drink made from banana'
-wa CL:2-DIMENSIONAL_STRAIGHT	*nijtyú-wa* 'bar of soap' *bohdó-wa* 'paddle' *íwa-wa* 'board of *sancona* (tree, sp.)'
-jɨ CL:2-DIMENSIONAL_ROUND	*bohdó-jɨ* 'plate' *mááho-jɨ* '*cazabe* (manioc bread) loaf' *watájko-jɨ* 'hat'

https://doi.org/10.1515/langcog-2012-0012 Published online by Cambridge University Press

-i CL:1-DIMENSIONAL_MEDIUM	*koo-i* 'wooden stick' *kaanú-i* 'pestle' *kujkú-i* 'walking stick'
-ko CL:1-DIMENSIONAL_POINTED	*áámúta-ko* 'shaft of harpoon' *á-ko* 'beam' *pijjú-ko* 'fishing rod'
-u CL:3-DIMENSIONAL_ROUND	*kóómɨ-u* '*milpeso* (palm, sp.) fruit' *állu-u* 'eye' *kúnii-u* 'yams tuber'
-u CL:STRING	*áhdi-u* 'cotton string' *moohó-u* 'liana' *wáábya-u* 'string'
-he CL:TREE	*bááko-he* 'grape tree' *kóóju-he* 'avocado tree' *tehké-he* 'calabash tree'
-ho CL:3-DIMENSIONAL_OBLONG	*áá-ho* '*maraca* (tree, sp.) fruit' *náme-ho* 'penis' *túju-ho* 'nose'
-ho CL:ENCLOSURE	*mátsájke-ho* 'peanut plantation' *mɨ́jko-ho* 'room' *ííñu-ho* 'stove'

Appendix 2: Object words used in the preliminary and main tasks

(Words used in the main task have an asterisk behind them. In some cases, a different English, but synonymous Bora, word was used in the main task. This word is included after the asterisk. Dimensionality and Bora classifier indicated for words used in the main task.)

#	Object	Dimensionality	Bora classifier
1	Arrow (*Spear)	1-dimensional	*-ba* (CL:3-DIMENSIONAL)
2	Audio cassette		
3	Banknote (*)	2-dimensional	*-haamɨ* (CL:LEAF)
4	Baseball bat		
5	Bed		
6	Bench (*)	2-dimensional	*-wa* (CL:2-DIMENSIONAL_STRAIGHT)

https://doi.org/10.1515/langcog-2012-0012 Published online by Cambridge University Press

7	Blackboard (*)	2-dimensional	*-wa* (CL:2-DIMENSIONAL_STRAIGHT)
8	Blanket (*)	2-dimensional	*-ja* (CL:COVER)
9	Bone (*)	1-dimensional	*-bahku* (CL:BONE)
10	Book		
11	Bridge (*)	2-dimensional	*-aahyo* (CL:LOGS.TOGETHER)
12	Broom (*)	1-dimensional	*-cooha* (CL:BRUSH)
13	Calculator		
14	Camera (*)	2-dimensional	*-ihkyu* (CL:GADGET)
15	Candle (*)	1-dimensional	*-i* (CL:1-DIMENSIONAL_MEDIUM)
16	Cane (*)	1-dimensional	*-i* (CL:1-DIMENSIONAL_MEDIUM)
17	Cigar		
18	Cigarette		
19	Comb		
20	Door (*)	2-dimensional	*-wa* (CL:2-DIMENSIONAL_STRAIGHT)
21	Drain pipe		
22	Envelope		
23	Feather (*)	1-dimensional	*-ɨɨ* (CL:2-DIMENSIONAL_ROUND)
24	Fishing rod		
25	Flag (*)	2-dimensional	*-ba* (CL:3-DIMENSIONAL)
26	Fork (*)	1-dimensional	*-cooha* (CL:BRUSH)
27	Hammer		
28	Key		
29	Knife (*)	1-dimensional	*-wa* (CL:2-DIMENSIONAL_STRAIGHT)
30	Ladder		
31	Ladle		
32	Letter (*)	2-dimensional	*-haamɨ* (CL:LEAF)
33	Log (*)	1-dimensional	*-ba* (CL:3-DIMENSIONAL)
34	Mat		

https://doi.org/10.1515/langcog-2012-0012 Published online by Cambridge University Press

35	Matchstick (*)	1-dimensional	*-ho* (CL:3-DIMENSIONAL_OBLONG)
36	Mobile phone		
37	Nail (*)	1-dimensional	*-iihyo* (CL:POINTED)
38	Needle (*Thorn)	1-dimensional	*-jto* (CL:THORN)
39	Paintbrush (*)	1-dimensional	*-cooha* (CL:BRUSH)
40	Paper (*)	2-dimensional	*-haami* (CL:LEAF)
41	Pen (*)	1-dimensional	*-iihyo* (CL:POINTED)
42	Pencil		
43	Plank (*)	2-dimensional	*-wa* (CL:2-DIMENSIONAL_STRAIGHT)
44	Plaster		
45	Playing card		
46	Raft (*)	2-dimensional	*-ji* (CL:2-DIMENSIONAL_ROUND)
47	Rake		
48	Razor		
49	Rifle (*)	1-dimensional	*-ju* (CL:TUBE)
50	Road (*)	2-dimensional	*-ju* (CL:TUBE)
51	Roof (*)	2-dimensional	*-wa* (CL:2-DIMENSIONAL_STRAIGHT)
52	Rope (*Vine)	1-dimensional	*-u* (CL:STRING)
53	Rug (*)	2-dimensional	*-ji* (CL:2-DIMENSIONAL_ROUND)
54	Saw (*)	2-dimensional	*-ji* (CL:2-DIMENSIONAL_ROUND)
55	Scissors (*)	1-dimensional	*-wa* (CL:2-DIMENSIONAL_STRAIGHT)
56	Screw		
57	Screwdriver		
58	Shoelace		
59	Shovel		
60	Spoon (*)	1-dimensional	*-wa* (CL:2-DIMENSIONAL_STRAIGHT)
61	Straw		
62	String (*)	1-dimensional	*-u* (CL:STRING)

https://doi.org/10.1515/langcog-2012-0012 Published online by Cambridge University Press

63	Table (*)	2-dimensional	*-wa* (CL:2-DIMENSIONAL_STRAIGHT)
64	Telegraph pole (*Pillar)	1-dimensional	*-hajku* (CL:PILLAR)
65	Tie		
66	Toothbrush (*)	1-dimensional	*-ta* (CL:DEVICE)
67	Umbrella (*)	2-dimensional	*-wa* (CL:2-DIMENSIONAL_STRAIGHT)
68	Walking stick		
69	Wall (*)	2-dimensional	*-ho* (CL:3-DIMENSIONAL_OBLONG)
70	Window	2-dimensional	
71	Wooden spoon		
72	Wrench		

References

Aikhenvald, A. Y. 2000. *Classifiers. A typology of noun categorization devices.* Oxford: Oxford University Press.

Allan, K. 1977. Classifiers. *Language* 53. 285–311.

Bloom, P. & F. C. Keil. 2001. Thinking through language. *Mind and Language* 16(4). 351–367.

Dixon, R. M. W. 1986. Noun classes and noun classification in typological perspective. In C. Craig (ed.), *Noun classes and categorization*, 105–112. Amsterdam: John Benjamins.

Evans, N. & S. C. Levinson. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioural and Brain Sciences* 32(5). 429–492.

Gao, M. Y. & B. C. Malt. 2009. Mental representation and cognitive consequences of Chinese individual classifiers. *Language and Cognitive Processes* 24(7). 1124–1179.

Gaskins, S. & J. A. Lucy. 2003. Interaction of language type and referent type in the development of nonverbal classification preferences. In D. Gentner & S. Goldin-Meadow (eds.), *Language in mind: Advances in the study of language and thought*, 465–492. Cambridge, MA: MIT Press.

Imai, M. & D. Gentner. 1997. A crosslinguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition* 62. 169–200.

Imai, M. & R. Mazuka. 2007. Revisiting language universals and linguistic relativity: Language-relative construal of individuation constrained by universal ontology. *Cognitive Science* 31. 385–414.

Imai, M. & H. Saalbach. 2010. Categories in mind and categories in language: Are classifier categories a reflection of the mind? In B. C. Malt & P. Wolff (eds.), *Words and the mind: How words capture human experience*, 138–164. Oxford: Oxford University Press.

Imai, M., H. Saalbach & E. Stern. 2010. Are Chinese and German children taxonomic, thematic, or shape biased? Influence of classifiers and cultural contexts. *Frontiers in Psychology* 1. doi: 10.3389/fpsyg.2010.00194.

Iwasaki, N., D. P. Vinson & G. Vigliocco. 2010. Does the grammatical count/mass distinction affect semantic representations? Evidence from experiments in English and Japanese. *Language and Cognitive Processes* 25(2). 189–223.

https://doi.org/10.1515/langcog-2012-0012 Published online by Cambridge University Press

Klima, E. S. & U. Bellugi. 1979. *The signs of language.* Cambridge, MA: Harvard University Press.

Kousta, S., D. P. Vinson & G. Vigliocco. 2008. Investigating linguistic relativity through bilingualism: The case of grammatical gender. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 34(4). 843–858.

Kuo, J. Y. & M. D. Sera. 2009. Classifier effects on human categorization: The role of shape classifiers in Mandarin Chinese. *Journal of East Asian Linguistics* 18. 1–19.

Lucy, J. A. 1992. *Grammatical categories and cognition. A case study of the linguistic relativity hypothesis.* Cambridge: Cambridge University Press.

Lucy, J. A. & S. Gaskins. 2001. Grammatical categories and the development of classification preferences: A comparative approach. In M. Bowerman & S. C. Levinson (eds.), *Language acquisition and conceptual development*, 257–283. Cambridge: Cambridge University Press.

Mazuka, R. & R. S. Friedman. 2000. Linguistic relativity in Japanese and English: Is language the primary determinant in object classification? *Journal of East Asian Linguistics* 9(4). 353–377.

Saalbach, H. & M. Imai. 2007. The scope of linguistic influence: Does a classifier system alter object concepts? *Journal of Experimental Psychology: General* 136. 485–501.

Sandler, W. & D. Lillo-Martin. 2006. *Sign language and linguistic universals.* Cambridge: Cambridge University Press.

Schmitt, B. H. & S. Zhang. 1998. Language structure and categorization: A study of classifiers in consumer cognition, judgment, and choice. *Journal of Consumer Research* 25(2). 108–122.

Seifart, F. 2005. *The structure and use of shape-based noun classes in Miraña (North West Amazon)*. MPI Series in Psycholinguistics 32. (PhD dissertation, Radboud University Nijmegen.) Wageningen: Ponsen & Looijen.

Senft, G. (ed.). 2000. *Systems of nominal classification.* Cambridge: Cambridge University Press.

Slobin, D. I. 1996. From "thought and language" to "thinking for speaking". In J. J. Gumperz & S. C. Levinson (eds.), *Rethinking linguistic relativity*, 70–96. Cambridge: Cambridge University Press.

Srinivasan, M. 2010. Do classifiers predict differences in cognitive processing? A study of nominal classification in Mandarin Chinese. *Language and Cognition* 2(2). 177–190.

Vigliocco, G., D. P. Vinson, F. Paganelli & K. Dworzynski. 2005. Grammatical gender effects on cognition: Implications for language learning and language use. *Journal of Experimental Psychology: General* 134(4). 501–520.

Zhang, S. & B. H. Schmitt. 1998. Language-dependent classification: The mental representation of classifiers in cognition, memory, and ad evaluations. *Journal of Experimental Psychology: Applied* 4. 375–385.

https://doi.org/10.1515/langcog-2012-0012 Published online by Cambridge University Press