Antonym adjective pairs and prosodic iconicity: Evidence from letter replications in an English blogger corpus

Susanne Fuchs[1], Egor Savin[1,2], Stephanie Solt[1], Cornelia Ebert[1], Manfred Krifka[1,2]

[1]Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin, Germany

[2]Humboldt Universität zu Berlin, Germany

Susanne Fuchs ORCID: 0000-0001-6751-9286

Egor Savin ORCID: 0000-0003-4209-1003

Stephanie Solt ORCID: 0000-0002-3269-8461

Cornelia Ebert ORCID: 0000-0002-4833-8655

Manfred Krifka ORCID: 0000-0002-9610-8352

Abstract: While the general assumption has long been that natural languages exhibit an arbitrary pairing of form and meaning, there is increasing empirical evidence that iconicity in language is not uncommon. One example from spoken language involves iconic prosodic modulation, i.e. the changing of prosodic features such as duration and fundamental frequency to express meanings such as size and speed. In this paper, we use data from an English social media corpus, with 140 million words written by 19320 bloggers, to investigate a counterpart to iconic prosodic modulation in written language, namely letter replications (e.g. loooong). We examine pairs of gradable adjectives such as *short/long*, *tiny/huge* and *fast/slow*, finding a higher frequency of letter replications for adjectives associated with greater size or spatial/temporal extent. We did not find an iconic effect on the number of replicated letters. Our results show evidence for iconic prosody in written language, and further demonstrate that social media databases offer an excellent opportunity to investigate naturalistic written language.

Keywords: Iconic lengthening, prosody, social media, English, scalarity

1 Introduction

1.1 Iconicity

It is frequently assumed that one "design feature" of language is that there is no relation between sound and meaning (Hockett 1960), which is what makes languages so diverse and flexible. However, there is accumulating empirical evidence that this statement is only partially true, and that some degree of iconicity, i.e. an association between form and sign, may have been necessary specifically in certain periods of mankind and human interaction. Perniss and Vigliocco argue that "iconicity is a fundamental property of language, representing an adaptation to a critical constraint on the phylogenesis, ontogenesis and use of language, namely the need to map linguistic form to human (sensory, motor and affective) experience" (Perniss and Vigliocco 2014: 2).

One well known example of iconicity involves systematic shape-sound correspondences such as the so-called 'bouba-kiki' effect, where rounded and non-rounded mouth shapes and tongue movements in production are systematically associated with rounded and pointy shapes, respectively (Köhler 1929; Ramachandran and Hubbard 2001; Maurer et al. 2006). As evidence of the widespread nature of iconicity, Blasi et al. (2016) analyzed 100 basic vocabulary items in 4298 different languages and 359 lineages, finding a large proportion to exhibit non-arbitrary sound-meaning relations which cannot be explained due to language contact. For example, words for tongue tend to contain /l/, while those for nose tend to contain /n/. Thus, iconicity in spoken language may be less exceptional than expected. It has even been studied in the context of consumer judgements in brand names (Yorkston and Menon 2004). Brand names for ice cream differing only in the two phonemes /i/ and /ɛ:/ or /ʌ/[1] revealed different consumer judgements on richness, smoothness, and creaminess. Brand names with /ɛ:/ were associated with heavier, richer, and creamier ice cream. This association is based on sound symbolism, specifically that more front vowels would be associated with "smallness, lightness,

[1] The authors describe [ä] as a phoneme, but this is not a phoneme in the IPA, hence we replaced it by a phoneme which we thought might correspond. However, at a later place the authors provide an example for a corresponding word <but> and according to this one, the [ä] might be [ʌ].

mildness, thinness, fastness, coldness, bitterness, femininity, weakness, lightness, and prettiness" (Yorkston and Menon 2004: 44) in comparison to more back vowels. The authors suggest that the preference for one or another product based on sound symbolism is below the customer's awareness.

1.2 Adjectives and scalar meaning as a semantic domain for iconicity

A semantic domain where iconicity is well documented involves scalar dimensions such as size, duration and speed. Such meanings are expressed by words of a variety of grammatical categories, including nouns, verbs and prepositions, but most typically represent the semantic content of gradable adjectives such as *large/small, long/short* and *high/low*. Gradable adjectives as a class can be identified by their occurrence in the comparative form (e.g. *larger*, *longer*). Formally, they may be analyzed as lexicalizing mappings from individuals to degrees on scales (Bierwisch 1987). Typically, such adjectives occur in antonym pairs, where one adjective is associated with the "large" direction on the scale while its antonym is associated with the "small" direction. Often, the adjective that corresponds to the "large" direction is the unmarked one. This manifests itself in several ways, one being the interpretation of questions: "How long is X?" does not presuppose that X is long, whereas "How short is X?" presupposes that X is short. However, this is not always the case. There are adjectives denoting very high degrees, like *huge*, that are marked, and there are antonym pairs where neither member is unmarked (e.g., *poor* and *rich*). Below we will see a further instance of divergence between largeness and unmarkedness that will prove relevant for the understanding of iconic effects in this domain.

A variety of factors make scalar meaning a natural place to expect iconicity in language, including humans' ability to associate magnitudes across modalities (Cuskley and Kirby 2013) as well as regular co-occurrences in the external world, such as that between the size of animals and the fundamental frequency and resonance of their vocalizations (Ohala 1984, 1994, 1997). Larger animals produce lower vocalizations, because their vocal folds tend to be longer and heavier, resulting in a lower voice in comparison to smaller ones (Ohala 1994, citing Morton 1977). Hence based on the acoustic

properties of the vocalizations, animals can roughly estimate the size of a potential aggressor and behave in the appropriate manner.

Recently, Knoeferle et al. 2017 tested the role of the acoustic properties in humans. They visually presented participants objects of different size with audible nonsense speech material that differed with respect to vowel duration, formants, intensity and fundamental frequency. Participants had to judge which object would fit with the audible sound. Their results revealed an iconic effect of object size on vowel duration and the first formant (F1), but not in the other parameters. Larger objects were associated with longer duration and higher F1 values.

1.3 Produced and perceived iconicity in the prosody of adjectives

Prosody can also contribute to iconicity in spoken language, for example by an extra lengthening of the vowel in *long* in referring to the experience of an event taking a long time. Schlenker (2017b) has recently proposed that such lengthening is an iconic enrichment of arbitrary encoding of linguistic meaning. Iconic prosodic modulation consists of changing certain prosodic features such as duration, fundamental frequency (F0), or amplitude to express additional meaning. In this line, Nygaard et al. (2009) investigated the relation between prosody and meaning in two experiments. Three speakers had to read novel words in an infant-directed speech style to stimulate a situation of engagement. The novel words, embedded in a frame sentence, were first read with a relatively neutral prosody as a baseline. Subsequently, the words were presented together with pictures that referred to the meaning of an adjective from the pairs: *happy/sad, hot/cold, big/small, tall/short, yummy/yucky*, and *strong/weak*. Speakers had to read the words again. Differences in acoustic properties were found (mean F0, F0 variation, duration, and amplitude) depending on the adjective. The authors also investigated whether listeners could reliably infer the meaning of these novel words. For this purpose, listeners saw two pictures representing an antonym pair, heard one of the previously recorded sentences, and had to choose the picture that would correspond to the perceived novel word. Listeners were significantly better in choosing the right picture when listening to the speaker's

meaningful prosody than to the speaker's neutral prosody as well as when prosody matched than when it mismatched. The authors suggest that prosody could augment, disambiguate, or reinforce meaning.

Perlman (2010) showed video clips involving fast and slow events to experimental participants, who had to retell the different events; they did so by generally talking faster for the faster events and slower for the slower ones, without being instructed to do so. In a more recent study, Perlman et al. (2015) extended these findings from the manner of motion to the size of an entity and additionally from concrete to abstract meanings (e.g. concrete: a *fast* drive; abstract: *slow* career progress). Speakers had to read short stories involving one of these semantic dimensions to a partner. The authors predicted that stories with different manners of motion would elicit corresponding prosodic variation in duration, but not fundamental frequency, while the reverse would be the case for stories varying adjectives corresponding to the dimension of size. Their findings show that stories in the small condition were read with higher F0 than stories in the large condition, for both abstract and concrete meanings. Moreover, stories in the fast condition were read within a shorter duration than stories in the slow condition, and no differences in F0 were found. Thus, different acoustic parameters may be used to mark different semantic dimensions.

In three experiments, Shintel et al. (2006) recorded speakers who described either a dot moving in upward and downward directions or dots moving with different speed. They could show that motions in the vertical dimension go hand in hand with changes in F0 in a similar direction, whereas changes in speed coincided with changes in speech rate. In follow-up studies, Shintel and Nusbaum (2008) could also provide evidence that the speed of recorded instructions influenced the time of a listener's response. In Shintel et al. (2014) the findings were extended to novel word learning, showing that congruent prosody has a positive effect on memory consolidation.

The iconic representation of scalar meaning is also not limited to the spoken modality. In signed languages including American Sign Language (ASL) and Italian Sign Language (LIS), adjectival and verbal

scales are in some cases visually represented in the signing space, a pattern that has been characterized as iconic (Wilbur 2012; Aristodeomo and Geraci 2017; Kuhn 2017). In LIS, for example, comparative *taller* can be signed via hand movement in the upward direction. Furthermore, the production of signs may be modulated with semantic effect, as when the slower-than-normal signing of a verb conveys that the corresponding event was a slow one (Wilbur 2008). Similar effects are discussed in Schlenker et al. (2013). In ASL, the sign GROW can be realized with different speed and different maximal distance of the hands. In the ASL translation of the sentence *My group has been growing*, the named parameters – speed and amplitude – are decisive for the interpretation of the sentence. Depending on these parameters and the iconic mapping, it can express that the group grew quickly or slowly and a lot or only a bit.

Here, the question naturally arises as to what exactly should be counted as an ASL sign and what is actually gesture (see Goldin-Meadow and Brentari 2017 for an overview and for discussion). Are all these individual manipulations gestural modifications of the actual sign or is this mode of realization more on the grammatical side, belonging to the sign as such? In spoken language, speech-accompanying gestures can take over this part of modifying what is said in iconic ways (see McNeill 1992 and Kendon 1980 for discussion of speech-accompanying gestures and their iconic character). For example, when talking about a painting, the utterance can be accompanied by an oval-gesture or a rectangular-gesture, thus indicating whether the painting is actually oval or rectangular.

Schlenker (2017a) distinguishes between external and internal (i.e. syntactically (in)eliminable) enrichments. Speech accompanying gestures in spoken language would be of the former kind, while the prosodic modulations we discussed in the beginning would be of the latter kind, as would the above-discussed modulations of speed and amplitude in sign languages.

1.4 New prospects on prosody with recent technological developments

Along with progress in technology (e.g. computer, smartphones, tablets, fiber optic cables, satellites), digital communication has come to have an enormous impact on our daily life and our communication

tools and styles. Social media platforms (e.g. twitter, instagram, facebook, chats, whatsapp, blogs) have been developing in parallel.

Social media data have some common features with spoken language, since writers do not follow all the formal rules of traditional written norms, and they are further enriched with icons of emotional expressions (emojis). These platforms provide a great opportunity to investigate the dynamics and creativity in the use of written language beyond prescriptive rules (Kaye et al. 2017; Huang et al. Grieve 2016). Using social network databases has the advantage of getting a vast amount of data in completely natural settings, with participants who would not feel constrained as they might in a laboratory experiment. It has, however, the disadvantage that the data might be confounded by many unknown factors. So far, most studies have tested for iconicity in prosody using contrastive settings instead. Perlman and colleagues write that "Remarkably little is known about how speakers use iconic prosody in the wild" (Perlman et al. 2015: 1349). Our work contributes in this respect by using a social media corpus.

Among others, the idea of using social media for a better understanding of prosody goes back to Brody and Diakopoulos (2011), who considered word lengthening by letter replications as a substitute for prosodic emphasis. These letter replications are a way of signaling the writer's sentiment and emotion in written text where some properties of spoken language, like intonation, are partially absent. Letter replications also represent a possible candidate for a feature of written language that may have an iconic effect, in that the lengthened pronunciation of a word such as *long* in spoken speech can be reflected in writing via replications, as in *looonng*. To date, however, this has not been systematically researched for antonym pairs.

With this work, we argue that lengthening a word (as in *looonng*) has an iconic effect. In other words, a speaker or the author of a written text uses word lengthening to express some iconic feature of the object, property, or event under discussion (here: that something was extremely long). We do not, however, intend to exclude the possibility that there might be other reasons for word lengthening in

spoken and written language as well; in particular this might be to express amplification or emphasis (see Kawahara and Braver 2014; Fuchs et al. 2018).

1.5 Research questions and expectations

In this paper, we investigate the presence of iconic prosody in written language. Based on the literature on prosody and iconicity, we focus on scalar meaning as expressed by gradable adjectives, and on letter replications as a prosodic feature with a potentially iconic effect. The following specific research questions were asked:

1. In which adjectives does lengthening (letter replications) occur?

Here we distinguish two hypotheses. H1) If, as we assume, prosody and letter replications have some degree of iconicity, we predict that letter replications will occur more frequently for adjectives that express the larger size, e.g., in *long* rather than in *short*. H2) If, on the other hand, letter replications are used exclusively as an expression of emphasis or prosodic amplification, we predict a comparable frequency of letter replications in both members of a pair of antonyms.

2. If replications occur, of how many letters do they consist?

Consistent with the view that letter replications have an iconic component, we expect longer words (i.e., a greater number of replicated letters) for the larger size dimension.

2 Methodology

2.1 The blogger corpus

An English social media corpus was used. The corpus is freely available for non-commercial use. It consists of approximately 140 million words written by 19320 bloggers in August 2004 (Schler et al. 2006). The age of the bloggers ranges from 13-47 (in three age groups) with an equal number of males and females. The profession of the blogger is also provided. In the present paper we do not investigate potential age and/or gender effects in the frequency of letter replications with gradable adjectives;

we refer the reader to Fuchs et al. (2018) for evidence that in other domains it is younger bloggers who make the most use of letter replications.

2.2 Selected adjectival antonym pairs

Our starting point for selecting antonym pairs for investigation was the literature on iconic prosody in spoken language discussed in the introduction. In particular, we took from Shintel et al. (2014) all those adjective pairs with a dimensional meaning (i.e., a meaning relating to size or spatial/temporal extent). In order to ensure that our data set was sufficiently comprehensive, we augmented this set of pairs by extracting all words tagged as adjectives from the list of the top 5000 words/lemmas in the Corpus of Contemporary American English (Davies 2008)[2], and further restricting this list to those adjectives judged to have a dimensional meaning. The resulting complete set of adjectives was organized into antonym pairs. In some cases, this procedure resulted in two or more adjectives being associated with a single antonym; for example, *long* and *tall* share the antonym *short*, while *huge*, *giant* and *enormous* share the antonym *tiny*. In order to arrive at distinct pairs for analysis and avoid double counting of certain adjectives, we selected in each case one adjective for the "large" direction and one adjective for the "small" direction. In most cases where a choice had to be made, the adjective selected was the one occurring most frequently in the blogger corpus used in the present study (e.g., *long* was selected for pairing with *short* because it was more frequent than *tall*), with the following exceptions: i) *little* was excluded as a possible antonym for *big* because it has a quantificational and degree modifier use in addition to the adjectival one; ii) *close* was excluded as an antonym for *far* because it also has a use as a verb; iii) *young* rather than *new* was selected as the antonym for *old* for consistency with Shintel et al. (2014). The final list of antonym pairs is shown in Table 1.

[2] The full word frequency list is available at https://www.wordfrequency.info.

Table 1: Adjectival antonym pairs

Antonym pairs	
Smaller degree	*Larger degree*
short	long
small	big
tiny	huge
near	far
fast	slow
thin	fat
narrow	wide
low	high
young	old
shallow	deep

In each case, one member of the pair expresses a greater degree of size, extent or duration, while the other expresses a lesser degree. Here we note that the ordering of the pairs in Table 1 is based on "largeness", not markedness. As discussed in the introduction, these properties typically coincide. However, the pair *slow* and *fast* represents an exception: *fast* is arguably the unmarked term (cf. "How fast did he walk?" vs. "How slow did he walk?"), but it is instead *slow* that is associated with larger temporal extents, because a slow event requires more time than a fast one. We thus expect *slow* to be targeted for lengthening more often than *fast*.

Note also that the adjective pairs *short/long*, *near/far*, *fast/slow*, *narrow/wide*, *thin/fat* and *young/old* correspond to the horizontal axis, whereas *low/high* and *shallow/deep* correspond to the vertical axis and *small/big* and *tiny/huge* do not make specific reference to the axis but are instead general size properties.

2.3 Data extraction and preprocessing

The NLTK toolkit was used as a natural language processing environment to tokenize the corpus (http://www.nltk.org/api/nltk.tokenize.html). All lower- and upper-case tokens were considered together. To further process the data, we used Python 2.7. and R Core Team (2017).

In a first step, all replications of letters were removed from the selected adjectives (cf. Table 1). This also included replications which are the orthographic norm (e.g., double *l* in *small*) and resulted in an order of strings. Hereafter, we searched for these string orders including repetitions of the same strings. From the resulting corpus the following words were eliminated: all words that do not involve all the letters of the original word (e.g. *narow* instead of *narrow*), all words that differed in just one letter and could potentially be typos (e.g. *thinn* instead of *thin*), and finally all words that might have a different meaning (e.g. *tinny* instead of *tiny*).

Furthermore, we calculated the overall number of cases for each adjective with and without letter replication. The number of cases without letter replications served as a baseline to calculate how often bloggers wrote the specific word in the orthographic norm. The number of cases including additionally letter replications was set to 100 percent for each adjective to calculate the frequency at which bloggers wrote the selected adjectives with letter replications.

In addition, we calculated the length of the words as the number of all letters as well as the number of replicated letters.

3 Results

3.1 Percentage of adjectives with letter replications

Table 2: Dimensional adjective antonym pairs

Antonym pairs				**Fisher's exact test**	
Smaller degree	*Larger degree*	*Percentage (small)*	*Percentage (large)*		
short (n=10848)	*long* (n=44738)	0.0184%	1.3476%	**p<**	**0.00001**
small (n=12487)	*big* (n=29007)	0.024%	0.155%	**p<**	**0.0001**
tiny (n=2317)	*huge* (n=8556)	0.1295%	0.7246%	**p<**	**0.0003**
near (n=6557)	*far* (n=19424)	0.0152%	0.0721%	p=	0.1364
fast	*slow*	0.1022%	0.958%	**p<**	**0.00001**

(n=7827)	(n=4696)				
thin	*fat*	0.0701%	0.0679%	p=	1
(n=1427)	(n=5886)				
narrow	*wide*	0%	0.2088%	p=	1
(n=458)	(n=1915)				
low	*high*	0.0597%	0.0371%	p=	0.4508
(n=5022)	(n=16159)				
shallow	*deep*	0.1490%	0.2359%	p=	1
(n=671)	(n=5933)				
young	*old*	0%	0.0414%	p=	0.0867
(n=7538)	(n=31354)				

Almost all antonym pairs show a numerically higher frequency of letter replications in the adjective corresponding to the "larger" scalar direction (cf. Table 2). Letter replications are significantly more often used in the adjectives *long*, *slow, big* and *huge* than in their respective antonyms. For example, out of all occurrences of *long* in the corpus (n=44819), in 1.35 % of the cases the word was spelled with letter replications, i.e. different from the orthographic norm. In the case of *far* and *old*, the difference falls short of significant. But importantly, in no case was the frequency of replications higher for the adjective corresponding to the "smaller" scalar dimension. Thus, overall we find support for our first hypothesis H1 rather than the second hypothesis H2. We note however that the effect was not found consistently across all pairs tested, but instead was strongest in pairs relating to the dimensions of overall size (*small/big, tiny/huge*) as well as temporal extent and spatial extent in the horizontal direction (*long/short, fast/slow*), and not evident for extent in the vertical direction (*low/high*). We return to this point in the conclusion section.

The percentage of occurrences of letter replications can be seen in highly frequent words in the corpus (e.g., *long* with n=44819), but also in less frequent words (e.g., in *slow* with n=4696). The group of adjectives with larger degrees is, however, realized much more frequently than the group of adjectives with smaller degrees (cf. Table 2).

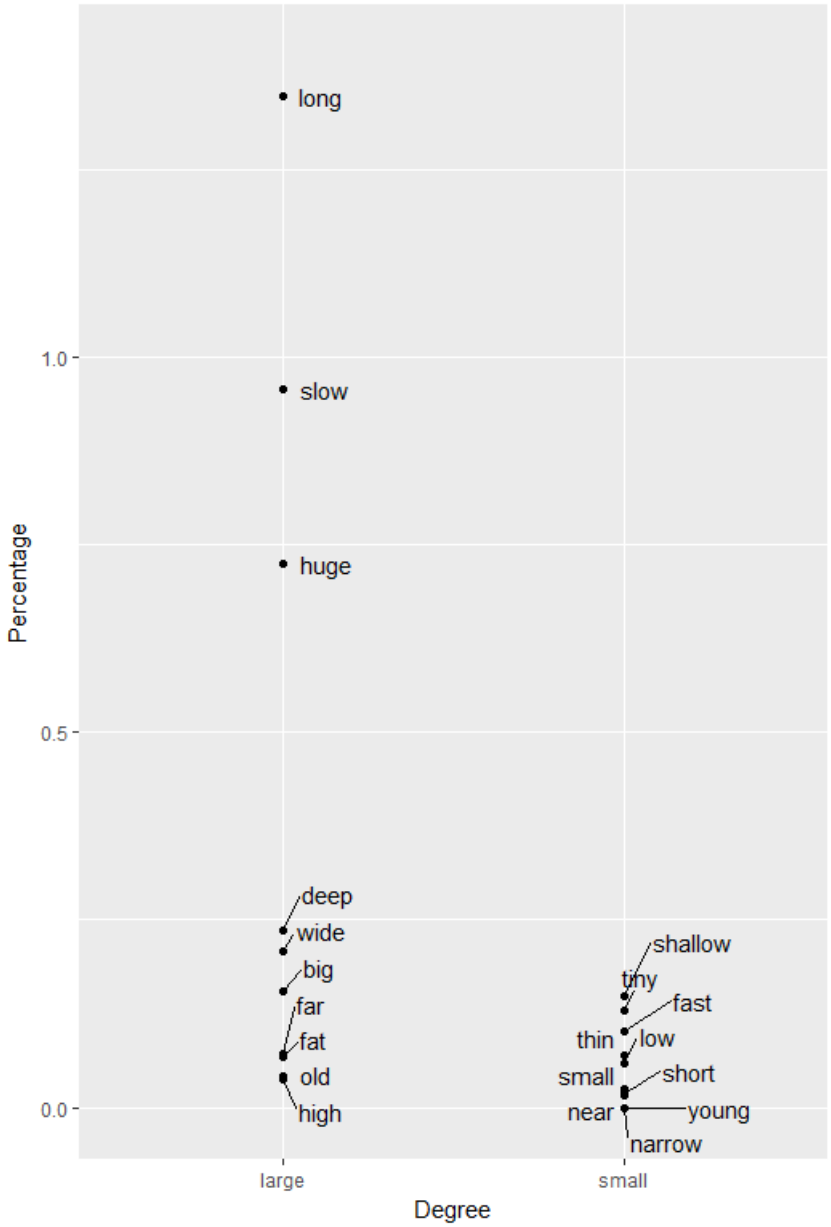

Figure 1: Plots with percentage of occurrence with letter replications in the respective adjectives with respect to the baseline (y-axis), data are split by adjective group (adjectives with larger degrees on the left and with smaller degrees on the right).

Figure 1 summarizes these findings by comparing the two adjective groups. The group of adjectives with larger degrees are not only more often found, they are also more variable than the group of adjectives with smaller degrees.

3.2 Number of repeated letters in words with letter replications

In a next step we investigated whether the number of repeated letters (in words with letter replications) would also differ among the adjectival antonym pairs (cf. Appendix I). For this purpose, we subtracted the number of letters in words with replications from the number of letters in the orthographic norm. Furthermore, we categorized all words with letter replications in categories from 2 to 10+. For example, category 2 would correspond to words with two additional letters, category 5 to words with five additional letters and category 10+ to words with ten or more additional letters. The stacked bars in Figure 2 give an overview of which categories are more often realized than others. Adjectives where we found no letter replications were not included in the plot. Results are depicted in percent. In contrast to our expectations, no consistent differences between adjectives with larger and smaller degrees are visible. In most adjectives, words are replicated with three or four additional letters.

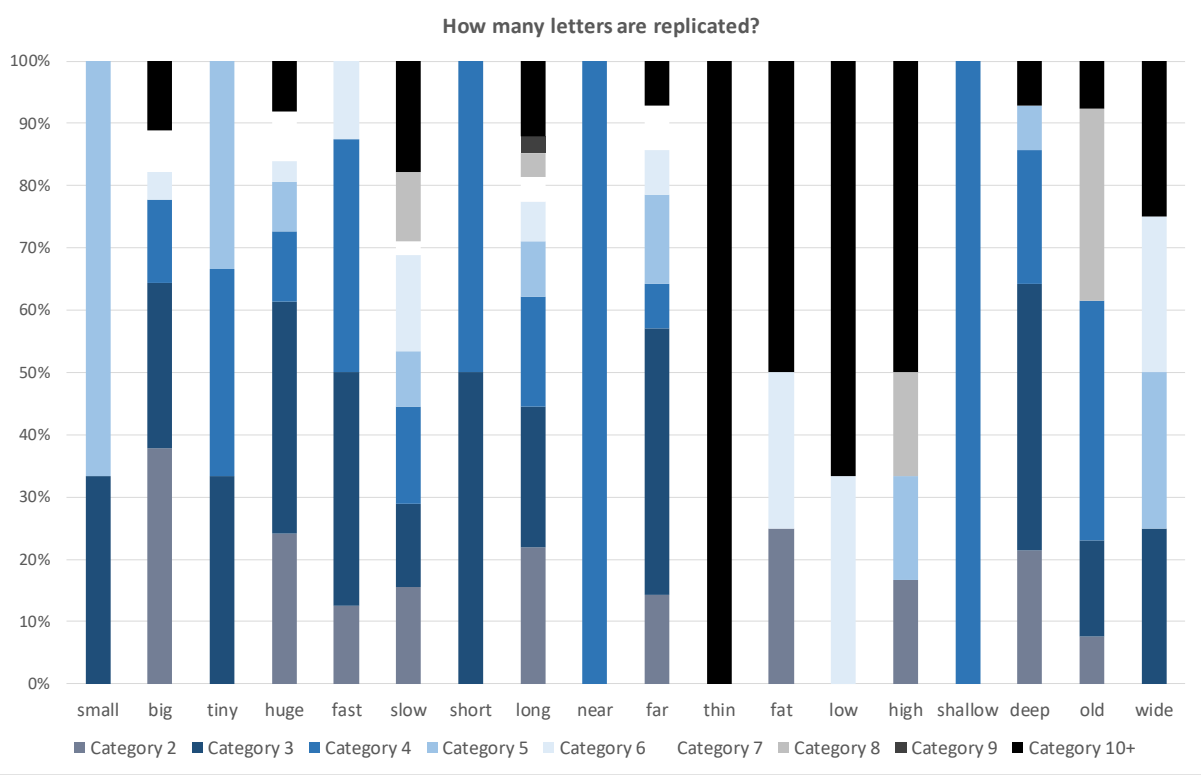

Figure 2: Stacked bar plots with replication length (number of letters for words with replications only). The number of letters for the default word was subtracted from the overall number of letters to

account for differences in word length among adjective pairs). Adjective pairs are written on the x-axis. All data are given in percent. Percentages in different colors are associated with different categories. Category 2, for instance, is a word with two more letters, category 5 with five more letters and so on.

Note also that the stacked bars corresponding to adjectives with smaller degrees include fewer data points than adjectives with larger degrees. On the basis of these findings we can exclude the possibility that writers on average lengthen all words with larger degrees to a greater extent than words with smaller degrees. Antonym pairs were however, not obtained in a contrastive context like: *fasssttt* versus *slooooooooooow*. It might well be possible that differences occur when used as antonym pairs.

4 Discussion and conclusion

With this work, we contribute to the growing literature that shows that natural language is to some extent iconic. Considering the four pairs that show a significant difference of the involved items in the percentage of occurrences with letter replications (*short/long, big/small, tiny/huge, fast/slow*), it is always the adjective corresponding to the "large" direction of the scale that occurs more often in a lengthened version (*long, big, huge, slow*). Other adjective pairs show a similar trend, but the difference was not significant. We did not, however, find an iconic effect on the word length for words with letter replications.

These results cannot be explained with respect to sound symbolism in the antonym pairs we selected: the *short/long* pair differs only marginally in vowel quality (/ɔ/ versus /o/), the *big/small* pair behaves in the opposite direction as one would predict according to sound symbolism (/i/ being smaller than /a/, see Shinohara and Kawahara 2010), and the *tiny/huge* pair is difficult to compare, since the first

adjective involves the diphthong /aɪ/ with two vowels differing in vowel height and backness. In the *fast/slow* pair the latter adjective involves a diphthong as well /əʊ/.

We would like to stress that it is arguably the adjective expressing larger degree that is lengthened and the relevant distinction is not between positive vs. negative or unmarked vs. marked adjectives. The pair *fast/slow* is particularly telling here. While semantically *fast* is the unmarked adjective of the two (i.e., not presupposing anything when used in the comparative), it is *slow* that occurs more often with letter replications, corresponding to the iconic effect that we expect.

We observed earlier that the adjectives pairs for which the greatest difference was found have meanings relating to overall size as well as temporal extent and extent in the horizontal direction; by contrast, pairs relating to extent in the vertical direction (in particular *high/low*) showed no clear effect. While we cannot say anything conclusive based on the present data, we hypothesize that the former sorts of dimensions lend themselves especially well to being reinforced by lengthening in written language. To explore this issue further, other methodologies such as acceptability judgement studies may prove very valuable.[3] We leave this as a topic for future research.

Finally, although letter replications were found to be very infrequent for adjectives corresponding to the "small" scalar direction, they were not entirely absent. This is consistent with the possibility that lengthening via letter replication might have other effects than the iconic mapping effect we argue for, such as e.g., emphasis in general (Kawahara and Braver 2014; Brody and Diakopoulos 2011; Fuchs et al. 2018). We are convinced though that our data show that iconicity is one important guiding factor, because otherwise the significant differences found between adjectives with smaller and larger degrees would not be explainable. The statistic effects we find are even more remarkable since our investigations are based on a naturally occurring non-elicited data set with no artificial contrasts or other contexts that could make the effects even stronger.

[3] We thank an anonymous reviewer for suggesting such studies.

Acknowledgements

This work was supported by the following grants from the German Research Council: FU 791/6-1 to S.F., C.E. and M.K. within the priority program XPRAG and SO 1157/1-2 to S.S.

References

Aristodemo, Valentina & Carlo Geraci. 2017. Visible degrees in Italian sign language. *Natural Language and Linguistic Theory*. Published online 14 November 2017 https://doi.org/10.1007/s11049-017-9389-5

Bierwisch, Manfred. 1987. Dimensionsadjektive als strukturierender Ausschnitt des Sprachverhaltens. In Manfred Bierwisch & Ewald Lang (eds.), *Grammatische und konzeptuelle Aspekte von Dimensionsadjektiven*, 1–90. Berlin: Akademie-Verlag.

Blasi, Damián E., Søren Wichmann, Harald Hammarström, Peter F. Stadler & Morten H. Christiansen. 2016. Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences*. 201605782.

Brody, Samuel & Nicholas Diakopoulos. 2011. Cooooooooooooooollllllllllll!!!!!!!!!!!!!! Using word lengthening to detect sentiment in microblogs. *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. 562–570.

Cuskley, Christine & Simon Kirby. 2013. Synesthesia, cross-modality and language evolution. In: Julia Simner & Edward Hubbard (eds.), *Oxford Handbook of Synesthesia*, 869–907. Oxford: Oxford University Press.

Davies, Mark. 2008. The Corpus of Contemporary American English (COCA): 560 million words, 1990-present. Available online at https://corpus.byu.edu/coca/.

Fuchs, Susanne, Egor Savin, Uwe D. Reichel, Cornelia Ebert & Manfred Krifka. 2018. Letter replication as prosodic amplification in social media. In Malte Belz, Susanne Fuchs, Stefanie Jannedy, Christine Mooshammer, Oxana Rasskazova & Marzena Zygis (eds.), *Proceedings of the conference: Phonetics and phonology in the German-speaking countries*, 65–68. Berlin: HU.

Goldin-Meadow, Susan & Diane Brentari. 2017. Gesture, sign, and language: The coming of age of sign language and gesture studies. *Cambridge Core in Behavioral and Brain Sciences* 40. e46.

Hockett, Charles S. 1960. The origin of speech. *Scientific American* 203. 88–106.

Huang, Yuan, Diansheng Guo, Alice Kasakoff & Jack Grieve. 2016. Understanding U.S. regional linguistic variation with Twitter data analysis. *Computers, Environment and Urban Systems* 59. 244–255.

Kawahara, Shigeto & Aaron Braver 2014. Durational properties of emphatically lengthened consonants in Japanese. *Journal of the International Phonetic Association* 44(3). 237–260.

Kaye, Linda K., Stephanie A. Malone & Helen J. Wall. 2017. Emojis: Insights, affordances, and possibilities for psychological science. *Trends in Cognitive Sciences* 21(2). 66–68.

Kendon, Adam. 1980. Gesticulation and speech: Two aspects of the process of utterance. In Mary Ritchie Key (ed.), *The Relationship of Verbal and Nonverbal Communication,* 207–227. The Hague: Mouton.

Knoeferle, Klemens, Jixing Li, Emanuela Maggioni & Charles Spence. 2017. What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings. *Scientific Reports* 7. 5562.

Köhler, Wolfgang. 1929. *Gestalt Psychology*. New York: Liveright.

Kuhn, Jeremy. 2017. Telicity and iconic scales in ASL. https://www.semanticsarchive.net/Archive/Tc2ZmRiO/

Maurer, Daphne, Thanujeni Pathman & Catherine J. Mondloch. 2006. The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science* 9.3. 316–322.

McNeill, David. 1992. *Hand and mind: What gestures reveal about thought*. Chicago and London: The University of Chicago Press.

Morton, Eugene S. 1977. On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist* 111. 855–869.

Nygaard, Lynne C., Debora S. Herold & Laura L. Namy. 2009. The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science* 33.1. 127–146.

Ohala, John J. 1984. An ethological perspective on common cross- language utilization of F0 in voice. *Phonetica* 41. 1–16.

Ohala, John J. 1994. The frequency codes underlies the sound symbolic use of voice pitch. In Leanne Hinton, Johanna Nichols & John J. Ohala (eds.), *Sound symbolism*, 325–347. Cambridge: Cambridge University Press.

Ohala, John J. 1997. Sound symbolism. *Proceedings of the 4th Seoul International Conference on Linguistics,* 98–103.

Perlman, Marcus. 2010. Talking fast: The use of speech rate as iconic gesture. In Fey Parrill, Vera Tobin, & Mark Turner (eds.), *Meaning, form, and body*, 245–262. Stanford, CA: CSLI Publications.

Perlman, Marcus, Nathaniel Clark & Marlene Johansson Falck. 2015. Iconic prosody in story reading. *Cognitive Science* 39.6. 1348–1368.

Perniss, Pamela & Gabriella Vigliocco. 2014. The bridge of iconicity: from a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B* 369. 20130300.

R Core Team (2017). R: A language and environment for statistical computing. Version 3.4.1. R Foundation for Statistical Computing, Vienna, Austria. Computer program obtained from https://www.r-project.org/ (2017-06-30).

Ramachandran, Vilayanur S. & Edward M. Hubbard. 2001. Synaesthesia: A window into perception, thought and language. *Journal of Consciousness Studies* 8.12. 3–34.

Schler, Jonathan, Moshe Koppel, Shlomo Argamon & James Pennebaker. 2006. Effects of age and gender on blogging. *Proceedings of 2006 AAAI Spring Symposium on Computational Approaches for Analyzing Weblogs*, 1–7.

Schlenker, Philippe. 2017a. Gestural grammar. https://ling.auf.net/lingbuzz/003497

Schlenker, Philippe. 2017b. Iconic pragmatics. *Natural Language and Linguistic Theory*. 1–60.

Schlenker, Philippe, Jonathan Lamberton & Mirko Santoro. 2013. Iconic variables. *Linguistics & Philosophy* 36(2). 91–149.

Shinohara, Kazuko, & Shigeto Kawahara. 2010. A cross-linguistic study of sound symbolism: The images of size. *Annual Meeting of the Berkeley Linguistics Society.* 396–410.

Shintel, Hadas, Nathan L. Anderson & Kimberly M. Fenn. 2014. Talk this way: The effect of prosodically conveyed semantic information on memory for novel words. *Journal of Experimental Psychology: General* 143.4. 1437–1442.

Shintel, Hadas & Howard C. Nusbaum. 2008. Moving to the speed of sound: Context modulation of the effect of acoustic properties of speech. *Cognitive Science* 32.6. 1063–1074.

Shintel, Hadas, Howard C. Nusbaum & Arika Okrent. 2006. Analog acoustic expression in speech communication. *Journal of Memory and Language* 55.2. 167–177.

Wilbur, Ronnie B. 2008. Complex predicates involving events, time and aspect: is this why sign languages look so similar? In Josep Quer (ed.), *Signs of the time. Theoretical Issues in Sign Language Research*, 217–250, Hamburg: Signum Press.

Wilbur, Ronnie B., Evie Malaia & Robin A. Shay. 2012. Degree modification and intensification in American Sign Language adjectives. In Maria Aloni, Vladim Kimmelman, Floris Roelofsen, Galit W. Sassoon, Katrin Schulz & Matthijs Westera (eds.*), Logic, Language and Meaning*, 92–101. Berlin, Heidelberg: Springer.

Yorkston, Eric & Geeta Menon. 2004. A sound idea: Phonetic effects of brand names on consumer judgements. *Journal of Consumer Research* 31. 43–51.

Appendix I

Adjectives with replications	occurrence	nr. of letters
biigg	2	5
biiig	7	5
biggg	8	5
bigggg	1	6
biiiig	11	6
biiiiig	6	7
biiggggg	1	9
biiiiiig	1	9
biiiiigggg	1	10
biiiiiiig	2	10
biiiiiiiiiiiiig	1	17
biiiiiiiiiiiiiiiiig	3	22
bbbbbbbbiiiiiiiiiigggggggggggg	1	30
smmaalll	1	8
smallllllll	2	10
faaar	2	5
faaaar	2	6
farrrr	4	6
faaaaar	1	7
faaaaaar	1	8
farrrrrr	1	8
farrrrrrr	1	9
faaaaaaaar	1	10
faaaaaaaaaaaaaar	1	17
nneeaarr	1	8
faaast	1	6
faaaast	1	7
fastttt	1	7
fasssst	1	7
faaaaast	1	8
fasssssst	1	8
fasssstt	1	8
faaaaaaast	1	10
slooow	5	6
slowww	2	6
slowwww	3	7
sloooww	1	7
sloooow	2	7
slowwwww	3	8
sloooowww	1	8
slooooow	3	8
sloooooow	3	9
sloowwwww	1	9

slowwwwwww	1	10
sloooowwww	2	10
slooooooow	1	10
sloooooowww	2	10
sllloowww	1	10
slowwwwwwww	1	11
slooooooowwww	1	12
slooooooooow	2	12
ssslllooowww	2	12
sssslllooowww	1	14
slllllooooooow	1	14
slowwwwwwwwwww	1	15
ssslloooowwwww	1	15
slooooooooooooow	1	16
sssslllllooooowww	1	17
slooooooooooooooooow	1	21
slooooooooooooooooooooooow	1	27
fattt	1	5
fatttttt	1	9
faaaaaaaaaat	1	13
faaaaaaaaaaaaaaaaaat	1	23
ttttttttttttttttttthhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhiiiinnn	1	57
highhh	1	6
highhhhhh	1	9
hiiiiiiiigh	1	12
hiiiiiiiiiigh	1	14
highhhhhhhhhhhh	1	17
higgggggggghhhhhhhhhh	1	21
looooowww	1	9
looooooowwwwwww	1	15
looooooooooooow	1	15
llongg	1	6
lonngg	2	6
looong	101	6
loonng	1	6
longgg	8	6
lonnng	2	6
lonnggg	2	7
looonng	2	7
loooong	99	7
loonngg	2	7
lonnnng	2	7
longggg	11	7
lonnnggg	2	8
loonnggg	2	8
looooong	71	8
lonnnnng	2	8

longgggg	5	8
looonnng	5	8
looonngg	1	8
lonnnngg	1	8
lonngggg	1	8
loonnngg	1	8
looooong	1	8
longggggg	1	9
looooongg	1	9
loooooong	38	9
lonnnnnng	1	9
looonnggg	1	9
looooonng	1	9
looonnngg	1	9
loonnnggg	3	9
lonnnnnnng	1	10
longgggggg	1	10
looonnnggg	4	10
looonnnngg	1	10
loooooooong	26	10
looooonnngg	1	11
looonnngggg	3	11
loooooooong	10	11
loooooooonng	1	11
lonnnnngggg	1	11
loonnnggggg	1	11
looooonnggg	3	11
llllooonnggg	1	11
looooonnnggg	3	12
llllooonnnggg	1	12
looooonnnggg	2	12
looonnnggggg	1	12
looooonnngggg	1	12
looooonnnngg	1	12
loooooooooong	11	12
looooonnnnggg	2	13
looooonnngggg	1	13
lonnnnnnggggg	1	13
looooonnnngggg	2	13
loooooooooong	6	13
longgggggggggg	1	13
looooonnnnngg	1	13
looonnnnnggggg	1	14
loooooooooooong	9	14
loonnnnggggggg	1	14
longgggggggggg	2	14
looooonnnnggggg	1	14
looooonnnggggg	1	14

loooonnnnggggggg	1	15
looooonnnnnggggg	1	15
looooooooooooong	2	15
looooonnnnnggggg	1	15
looooooooooonng	1	15
looooonnnnnggggg	1	15
llloooonnnnnggggg	1	16
longgggggggggggg	1	16
looooooooonnnnng	1	16
llloooonnnnggggg	1	16
looooooooooooong	1	16
llllloooonnnnnggg	1	17
llloooooonnnnggggg	1	17
looooooooooooooong	3	17
loooooooooooooong	5	18
looooooonnnnnggggg	1	18
looooooonnnnnggggg	1	18
loooooooooooooooong	1	19
loonnnnnnnnnnnnnggg	1	19
longgggggggggggggg	1	19
looooooooonnnnnnnng	1	19
llllloooooonnnnnggg	1	19
loooooooooooooooong	1	20
llloooonnnnnnnnggggg	2	20
llllooooonnnnnnggggg	1	21
loooooooooooooooooong	1	21
loooooooooooooooooong	2	22
loooooooooooooooooong	1	23
loooooooooooooooooooong	1	24
loooooooooooooooooooong	1	25
loooooooooooooooooooooong	1	26
loooooooooooooooooooooong	1	27
loooooooooooonnnnnnnggggggggg	1	29
lllllllllooooonnnnnnnnggggggggg	1	30
lllllloooooooooooooonnnnnnggggggg	1	33
looooooooooonnnnnnnnnnnnggggggggggg	1	34
lllllllllloooooooooonnnnnnggggggggggggg	1	39
lonnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnng	2	62
shoooort	1	8
shooooort	1	9
wiiiide	1	7
widdddeee	1	9
wiiiiiiide	1	10
wiiiidddeeeee	1	14
huggee	1	6
huuuge	10	6
hugeee	4	6
hugeeee	7	7

huuuuge	15	7
huuugee	1	7
huuuuuge	5	8
hhuuggee	1	8
huuuugee	1	8
huuuggeee	1	9
huuuuuuge	4	9
huuuuuuuge	1	10
hugeeeeeee	1	10
huggggggeeee	1	11
huuuuuuuuge	4	11
hhhhhuuuugggeee	1	15
huuuuuuuuuuuuuge	1	16
huuuuuuuuuuuuuuge	1	17
huuuuuuuuuuuuuuuge	2	19
tinnnny	1	7
tiiiiiny	1	8
tiiiiiiny	1	9
oldddd	1	6
olddddddddd	2	11
oolllld	1	6
ooooolllllld	1	11
olddd	1	5
ooooooooold	1	11
ooooooooooold	1	13
ooooold	5	7
shallooowww	1	11
deeeeeeep	1	9
deeeeeep	3	8
deeeeep	5	7
ddddddddddddeeeeeeeeeppppppppp	1	30
deeeep	3	6
deeeepp	1	7