This is an Author's Original Manuscript of an article whose final and definitive form has been published in Angelika Zirker, Matthias Bauer, Olga Fischer, Christina Ljungberg (eds.): Dimensions of Iconicity (2017). pages 4–14. ©John Benjamins, available online at:
https://benjamins.com/#catalog/books/ill.15/main
doi 10.1075/ill.15.01dia

The Effect of Iconicity Flash Blindness—An Empirical Study

Vojtěch Diatka and Jiří Milička[1]

(Charles University, Prague)

Abstract

In our experiment, the Saussurean postulate of arbitrariness has been empirically tested in order to see whether this postulate can be applied to all words to the same extent. Three hundred participants were asked to match Czech words with their Hindi translations. One set of words was randomly chosen from a Hindi corpus (set A); the second set consisted of both randomly chosen words and words categorized as ideophones (set B). The participants were successful in matching both sets (the lower level of the confidence interval is about 7% above random guessing), and their performance showed unexpected patterns: For one, not only iconic properties (the sound qualities) but also iconicity itself is an important distinctive feature and recipients are able to exploit this. Moreover, even words considered to be non-iconic (set A) apparently contain a degree of iconicity, which participants are able to draw upon. However, participants appear to lose this ability when non-iconic words are presented in the context of words with evident and abundant iconicity (set B). The effect resembles the accommodation process which is known for other senses; therefore, we call the effect "Iconicity flash blindness".

1. Background

It is stated in the *Course in General Linguistics* that the relationship between the signifier and the signified is arbitrary (Saussure 1959: 74). This means that there does not have to be any traceable relationship between the form (the string of phonemes) and the meaning of a word. Saussure considered onomatopoeic words as a counterexample to this statement[2] as their form

[1] Both the authors are the corresponding authors (<vojta.diatka@gmail.cz>; <jiri@milicka.cz>). The contribution by each author is as follows: V. D. prepared the sets of words and is responsible for the Background, Data, Participants and Hypotheses sections. J. M. developed the BlackSquare software and is responsible for the Experiment, Results, and Conclusions sections. Both authors participated in the experiment design and conducted the experiments and collected data. In addition, the overall outlines of the paper are the result of joint discussions. The study results from the subproject *Empirical Study of Arbitrariness* #2013FF004426 (Faculty of Arts, Charles University, Prague), supported by Specific University Research 2015 funding. We would like to thank our colleague Hana Kalábová who helped us with the experiments.

[2] For a more detailed review of his views and the historical background see (Ahlner and Zlatev 2010: 302)

resembles their meaning to some extent. However, according to Saussure, onomatopoeic words are a peripheral part of language, which is true at least in the sense of low frequency in some languages. However, there have been many linguists who have challenged this view, and the arbitrariness of language has been empirically tested many times. The first to do so was Sapir with the pseudo words *mil* and *mal* (1929).[3] The best-known experiment of this kind is the one conducted by Köhler (1929: 225), who presented his informants with two pictures, one with a pointy shape and another with a round one. The participants were asked to match those shapes with two quasi-words—"maluma" and "takete". The participants tended to match the "maluma" word with the round shape and "takete" with the pointy shape. His experiment was not valid according to current standards,[4] but it inspired many future experiments, especially during the 1950s.[5] What these experiments clearly show is that language contains a mechanism that allows speakers to successfully match words with some meanings even though they have just encountered these words for the first time. This indicates that there has to be at least sometimes a relationship between the signifier and the signified. We assume that this relationship is based on iconicity. When speaking of an 'iconic relationship' between the signifier and the signified, we assume that there is some kind of weaker or stronger motivation for a signifier to have a particular phonetic shape and that it is not arbitrary.

This statement raises a question: Are speakers better at matching iconic words than at matching ordinary words? Our experiment shows that the answer is much less straightforward than one would expect on the basis of a long tradition in the study of iconic words.

2. Data

Ideophones in Hindi were chosen as examples of words that bear some iconicity compared to ordinary words. Ideophones are "marked words that depict sensory imagery" (Dingemanse 2011: 25). What interests us most is the fact that they are marked in terms of form. In Hindi,

[3] The experiment as described by Sapir (1929: 227): "For example, the meaningless words *mal* and *mil* were pronounced in that order and given the arbitrary meaning 'table.' The subject decided whether *mal* seemed to symbolize a large or a small table as contrasted with the word *mil*." The participants tended to match the large table with *mal* and the small one with *mil*.

[4] It is rather anecdotal evidence than a real experiment. The author describes the result as "most people answer without hesitation" (Köhler 1929: 225) without any information about the experiment design, participants or results.

[5] A good summary of these experiments can be found in Ciccotosto (1991: 141-188), especially in the section on natural lexicons in sound symbolism experiments (165-174).

ideophones are often reduplicated, which makes them stand out from other words. They also depict sensory imagery because they often occur in vocabulary used to talk about perceptions (visual, olfactory, audible, etc.).

In our experiment, we used two datasets to conduct experiments similar to Köhler's and Sapir's mentioned above. One dataset comprises only non-iconic words,[6] and one is mixed containing both iconic and non-iconic words. The mixed dataset is of our primary concern; the non-iconic dataset serves as the control group, as will be explained in the "Hypotheses" section (see below). The iconic words, i.e. ideophones, were taken from Diatka (2014). His collection comprises 577 ideophones belonging to seven categories – auditory, interoceptive,7 kinesthetic, olfactory, tactile, gustatory, and visual. This set of ideophones shows a difference in their distribution in categories – there are 311 auditory ideophones and only three olfactory ones. We have taken this to be an illustration of the distribution of ideophones in the language as a whole, since this set is our only source. We tried to maintain the proportions from this set in our datasets, meaning there were approximately 100 times more auditory than olfactory ideophones. Non-iconic words were randomly taken from all frequency levels. We assume that their Czech translations can be more iconic than ordinary Czech words, although we did not test this quality in any way.

The non-iconic words were chosen randomly from a Hindi corpus so that both frequent and infrequent words were included in the sample. The random set of words includes three Czech-Hindi cognates (words that share Indo-European etymology which is quite transparent for an educated speaker). Those words were used within the experiment but were excluded from the final analysis of the results. The set of non-iconic words amounted to 68 words and the mixed set amounted to 139 iconic and 67 non-iconic words.

[6] By the term "non-iconic word" we mean an ordinary word that is traditionally not considered to be iconic although the conclusions of this paper suggest that even "non-iconic words" bear some degree of iconicity. We have chosen the words randomly from a Hindi corpus described in Bojar et al. (2014).

7 Interoceptive ideophones capture sensations of inner states and feelings of human beings, such as anger, fear, happiness, etc. The term is based on the source of a sensation: "A common Western folk model of sensory perception has it that perception is about taking in information from the outside world through sensory modalities, of which (in this folk model) there are five: vision, hearing, touch, taste and smell. Twentieth century scientific taxonomies are more inclusive, including not just extero-receptors (the traditional five) but also intero-receptors…" (Dingemanse 2011: 28)

This is an Author's Original Manuscript of an article whose final and definitive form has been published in
Angelika Zirker, Matthias Bauer, Olga Fischer, Christina Ljungberg (eds.): Dimensions of Iconicity (2017). pages 4–14. ©John Benjamins, available online at:
https://benjamins.com/#catalog/books/ill.15/main
doi 10.1075/ill.15.01dia

3. Participants

To obtain representative results, we needed a high number of participants. We gathered 300 participants, all of whom were native Czech speakers and did not know any Hindi, and included various participants – elementary school pupils, secondary school students, university students (not students of languages), middle-aged working people, and seniors. Both genders were represented proportionally (55% of the participants were women). The age group distribution does not correspond to the distribution in Czech society as young people (up to 30 years) were overrepresented.

The participants were motivated to give the best answers possible because the experiment was presented as a contest, and a reward was promised for the winners. Moreover, the participants were rigorously checked so they could not cheat (by looking up words on a smartphone etc.).

4. Experiment

To conduct the experiment, a user-friendly application providing random word pairs from a given set was needed. For this purpose, we developed BlackSquare,[8] a software for linguistic experiments that displays a set of four words on a device screen. The first word pair is given in one language and the second word pair in the other. The latter pair is the translation of the former pair. The order is randomized so that participants do not know which words semantically belong together. The actual user interface can be seen in Figure 1. The task which the participants had to accomplish is very simple—they were asked to match Czech words with their corresponding words in Hindi. Our participants were asked to rely on their language intuition, but they were not advised to use any specific decision-making method, and the term *iconicity* was not mentioned during the experiment.

[8] The software is more universal than described here. There are many settings and open possibilities to tweak it. It also records much more than presented—nearly every interaction of the participant with the device is recorded along with the precise time of the interaction. The particular settings and the set of words that were used can be found on <http://milicka.cz/en/blacksquare/experiment2015.zip> so that our experiment can be repeated in exactly the same conditions and thus verified. The software can be downloaded from <milicka.cz/en/blacksquare>, where the documentation can also be found (alternative repository: <https://sourceforge.net/projects/blacksquare>).

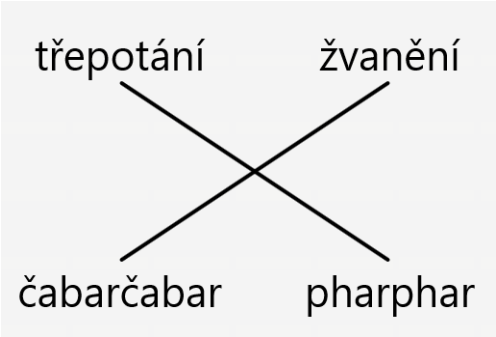

Figure 1. A typical assignment as displayed on the device screen. The participants were asked to connect the words using lines. The lines in the picture show the correct solution of the task. The upper pair of words is in Czech (the first one means "flutter", the second one means "blabber) and their Hindi translations are at the bottom.

We chose the written form to represent Hindi words as the phonemes contained in the words have a very similar pronunciation to that of Czech phonemes. As most of our participants were not experts, we needed some simple transcription. Instead of a phonetic transcription, we decided to transcribe words using Czech letters which fit the Hindi phonemes well, which means that we did not need to explain the phonetic alphabet. Those words containing sounds that are exotic to Czech speakers were omitted. In order to exclude the influence of prosody on the experiment, we decided not to use sound stimuli.

The use of tablets in the experiment enabled us to conduct a genuine field study and to investigate participants in their habitat. Each participant was given 40 assignments (i.e. distinct foursomes of words) randomly chosen from the mixed set of words or 34 random assignments from the set of the non-iconic words; each participant tried only one set of words.

5. Hypotheses

Before conducting the experiment, we formulated two hypotheses:

Hypothesis 1: The number of correct assignments of translations to unknown iconic words is significantly higher than chance expectation. We expect not mere statistical significance but also a strong effect of iconicity. This hypothesis is the traditional one that has been demonstrated many times before (Ahlner and Zlatev 2010: 309-311). However, an experiment that measures differences between a set of iconic and a set of non-iconic words has never been conducted. Furthermore, all the relevant preceding studies were based on classical

statistical significance testing[9] – but the mere fact that iconicity plays a role in language should not be surprising, at least for a typical reader of this volume; thus, we are interested rather in the effect of iconicity, and accordingly use confidence intervals[10] to describe the hypothetical population.

Hypothesis 2: Words do not have to be contrasted with only one physical dimension. This hypothesis requires some further explanation. Ahlner and Zlatev (2010:311), who summarized the empirical experiments dealing with iconicity very well, formulated two conditions which hold for all of them:

"First, participants are to match a pair of *familiar* words, with a pair of unfamiliar ones.

Second, the familiar words should contrast along a given dimension [...] and thus form antonym pairs, for example, small–big, round–flat, bright–dark."

Both conditions were deduced from the Brackbill-Little (1957) experiment, which was very carefully designed but whose implementation was seriously flawed.[11] We decided to test the second condition, according to which the pairs should be antonyms and differ in a specific physical property. We think that there does not need to be a particular physical dimension along which the two words differ and that they do not need to be antonymous.

6. Results

Results 1: Our Hypotheses

We conducted the experiment on both sets of words: the non-iconic ones and the mixed ones. The data in the chart in Figure 2 shows that there is at least a 59.6% probability that a Czech speaker would correctly assign the corresponding pairs of the non-iconic Hindi word (at 95% confidence level), which is quite a strong effect compared to the 50% baseline that would be

[9] I.e. the results were traditionally compared to the random model by t-tests, X^2 tests etc. However, the inference statistical testing is not really suitable for these kinds of research questions since a high p-value (in these circumstances) rather means that the sample is small and not that the effect of iconicity does not exist. Thus, we use the confidence intervals to reveal the size of the effect. For further explanation see e.g. Cohen (1994).

[10] The confidence interval is obtained by the bootstrapping method and thus by a separation of the data for each participant, a calculation of its mean value; then a randomization with repetitions is performed several million times, measuring the weighted means for each randomization. The confidence intervals displayed for all results are based on the 95% confidence level. This method respects the idea that the studied population (in the statistical sense) are *speakers* of a given language, and it is more conservative than pooling the answers of all participants and conducting a binomial test for all together, as seen in the cited studies.

[11] The participants had to decide whether a given pair of words were correct or not, rather than matching correct pairs. Furthermore, the participants were not very well motivated students of psychology and their number was not high – only 40 participants; compared with 500 subjects in Sapir's experiment (1929).

This is an Author's Original Manuscript of an article whose final and definitive form has been published in
Angelika Zirker, Matthias Bauer, Olga Fischer, Christina Ljungberg (eds.): Dimensions of Iconicity (2017). pages 4–14. ©John Benjamins, available online at:
https://benjamins.com/#catalog/books/ill.15/main
doi 10.1075/ill.15.01dia
obtained for random guessing.[12] Thus, the first hypothesis is confirmed.

At the same time, these results also confirm our second hypothesis as the word pairs were chosen randomly from the set of words. Therefore, miscellaneous combinations of unrelated words naturally occurred. The Brackbill-Little (1957) experiment thus needs revision and deserves an independent replication. In case that the replication fails, all theoretical insights based on the Brackbill-Little experiment, , would need reconsideration.[13]

We did not expect any strong similarities between the size of the respective effect when presenting two words from a set of non-iconic words as opposed to presenting two words from a set of iconic words.[14] These findings deserve a closer look and more detailed analysis, which follows..

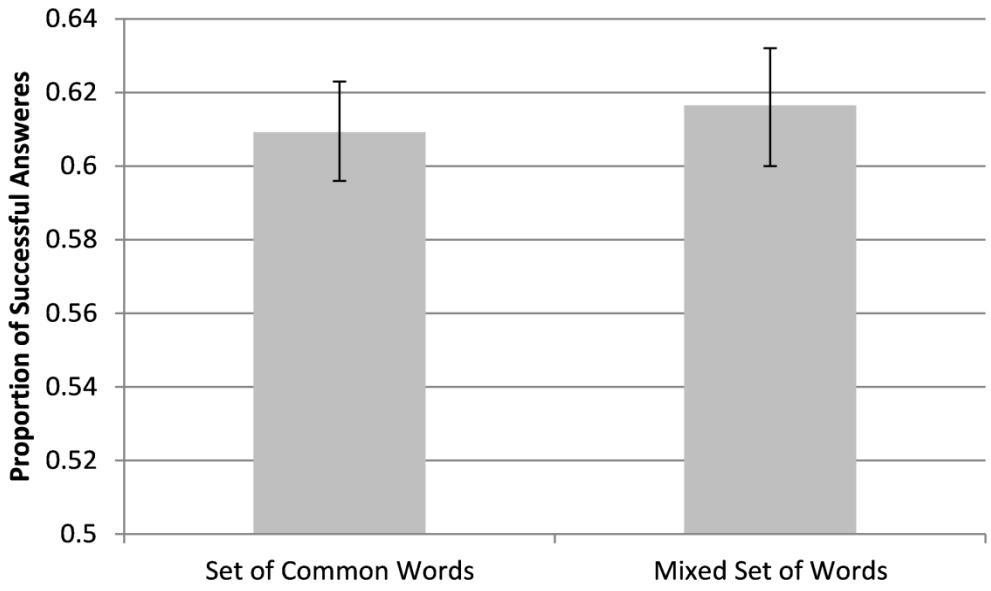

Figure 2. Comparison of the non-iconic set and the mixed set.

Results 2: Unexpected Patterns

The data presented in Figure 2 is surprising as it shows that there is only a slight difference between the results for the set of purely non-iconic words and the set comprising both iconic and non-iconic words. In both cases, the participants were quite successful in guessing the

[12] Detailed results for various sociological groups of participants and various word subsets along with some technical and methodological remarks can be found in Milička and Diatka (forthcoming).

[13] Cf. the semiotic approach by Ahlner and Zlatev (2010: 313 ff), which is partially based on the assumption that the word pair should differ along one dimension (small–big, round–flat, bright–dark) to be successfully assigned.

[14] In fact, the results are not significantly different; the 95% confidence intervals of the difference between results are -2.4 % to 1.4 %.

This is an Author's Original Manuscript of an article whose final and definitive form has been published in
Angelika Zirker, Matthias Bauer, Olga Fischer, Christina Ljungberg (eds.): Dimensions of Iconicity (2017). pages 4–14. ©John Benjamins, available online at:
https://benjamins.com/#catalog/books/ill.15/main
doi 10.1075/ill.15.01dia
correct translation for an unknown word. Does this mean that the level of iconicity is roughly the same for all words, despite our linguistic categorizing of them as iconic and non-iconic? Let us look more closely at the results obtained for the mixed set of words. The mixture of iconic and non-iconic words can yield three kinds of assignments:

The first one consists of a pair in which both words are iconic. The following table shows an example of such a task given to one participant:

Table 1. An example of two iconic words as presented to the participants in the test and the English translation of the Czech terms.

chvět se [= to tremble]	kloktat [=to gargle]
sansanáná	hakláná

The second possible assignment consists of one non-iconic and one iconic word:

Table 2. An example of one iconic and one non-iconic word as presented to the participants in the test and the English translation of the Czech terms.

drahokam [= jewel]	chvět se [= to tremble]
ratna	sansanáná

The third possible combination comprises two non-iconic words and thus it is the same for all assignments that are in a set of purely non-iconic words:

Table 3. An example of two ordinary words as presented to the participants in the test and the English translation of the Czech terms.

drahokam [= jewel]	zaměstnanec [= employee]
ratna	sevak

The results for all three combinations can be seen in Figure 3. The chart shows that the results for the purely iconic pairs (the third column) are lower than the results for the mixed pairs (the second column), which means that the participants were guided not only by the information that dwells in the iconic features of the words (in cases of two iconic words) but also by the mere iconicity itself (in cases of one iconic and non-iconic word). We have interpreted this finding by assuming that some meanings universally tend to be represented by iconic words and that this characteristic can be utilized by speakers (at least during this kind

of experiment) because of the typical formal features of iconic words such as reduplication etc. For example, when a participant sees the pair *dendár* – (*debtor*, *dlužník* in Czech) and *hakbak* – (*confused*, *zmatený* in Czech), the participant can guess that the word *confused* tends to be expressed by the iconic word rather than *debtor* and that the expression *dendár* sounds less iconic than *hakbak*.

The second unexpected pattern is even more surprising: The participants exposed to a purely non-iconic dataset could assign the right translation for two non-iconic word pairs far above the 50% baseline level (the first column in Figure 2), while the participants that were exposed to a mixed dataset performed worse when facing the same combination of two non-iconic words within the experiment (the first column in Figure 3). In fact, the results for two non-iconic words are nearly indistinguishable from random guessing. This means that the cognitive apparatus has a variable sensitivity to iconicity and that this sensitivity temporarily accommodates to the prevailing input. To put it simply: people who are exposed to a higher level of iconicity quickly become accustomed to this level and choose the strategy of focussing on this strong iconicity and not on the information provided by more subtle iconicity hidden in ordinary words.

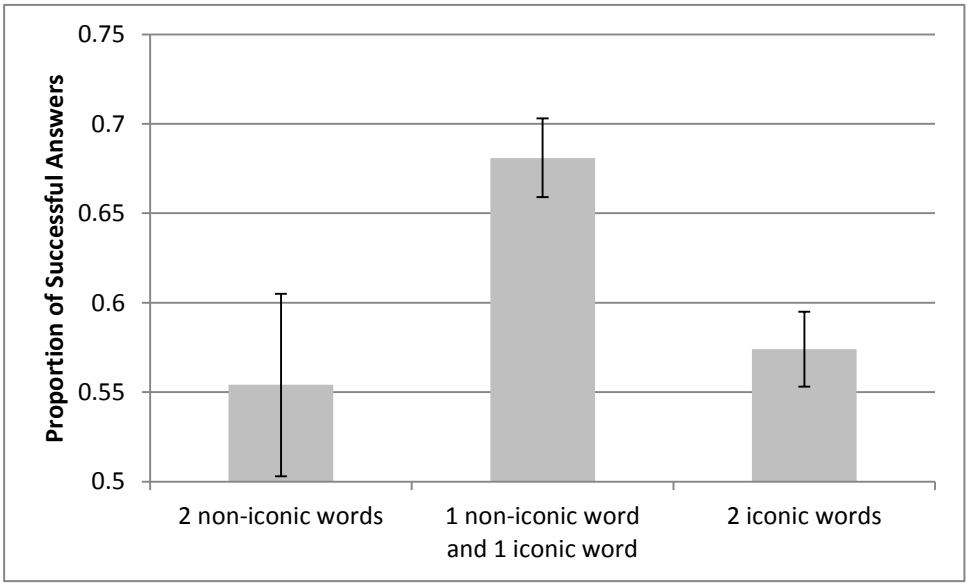

Figure 3. Three combinations that can occur in results obtained from an experiment with the mixed word set.

This effect corresponds to processing input from other human senses. Eyesight accommodates to the light level, and a quick transition from a light environment to a dark one results in poor vision capability. Calling the effect described here "iconicity flash blindness" is a rather

poetic choice of terminology but it emphasizes the similarity of the cognitive apparatus processing iconicity with that of other sensual organs. In our view, this finding deserves further experiments and better corroboration.

7. Conclusion

The participants in our test were successful at binarily matching words in an unknown language to their translations. The average success rate was at least 9% above the baseline (random guessing). The words were randomly chosen from a set and thus they did not differ along only one dimension—they were not antonyms but completely unrelated words. Ahlner and Zlatev's (2010: 311) second condition[15] is thus unnecessary for such experiments.

The mentioned 9% effect is comparable to the results measured on a set of common words and on a mixed set of common and iconic words (or: the words we have classified as iconic ones). This means that the participants were able to utilize not only the evident iconicity but also some more subtle iconic-like features of words that are not commonly considered onomatopoetic or iconic.

Not only iconic properties (the sound qualities encoded in written form) but also iconicity itself is an important distinctive feature that plays a role when a recipient is confronted with unknown words. The results of our experiment suggest that some meanings crosslinguistically tend to be represented by iconic words and that recipients are able to exploit this feature. We assume that this ability is due to the typical formal features of iconic words such as reduplication etc.

The participants were able to draw on the subtle iconicity of ordinary words to make correct decisions, although they appeared to lose this ability when these non-iconic pairs were presented in the context of words with evident iconicity. The effect resembles the accommodation process known for other senses. These two hypotheses would deserve further testing, e.g. more language pairs, more participants, various experiment designs and other kinds of iconic words.

References:

Ahlner, F., and Zlatev, J. 2010. Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign Systems Studies* 38.1: 298-348.

Bojar, O., Diatka, V., Rychlý, P., Straňák, P., Suchomel, V., Tamchyna, A., and Zeman, D.

[15] "The familiar words should contrast along a given dimension and thus form antonym pairs".

2014. HindEnCorp - Hindi-English and Hindi-only corpus for machine translation. *LREC 2014*: 3550-3555.

Brackbill, Y., and Little, K. B. 1957. Factors determining guessing of meaning of foreign words. *The Journal of Abnormal and Social Psychology* 54(3): 312-318.

Cohen, J. 1994. The earth is round (p<. 05). *American Psychologist* (49)12: 997-1003.

Ciccotosto, N. 1991. Sound symbolism in natural languages. PhD Dissertation, University of Florida.

Dingemanse, M. 2011. The meaning and use of ideophones in Siwu. PhD Dissertation, Radbound Univeristy Nijmegen. [http://thesis.ideophone.org/files/dingemanse-2011-thesis.pdf].

Diatka, V. 2014. Hindi ideophones. Master thesis, Charles University in Prague. [https://is.cuni.cz/webapps/zzp/download/120168591/?lang=en].

Köhler, W. 1929. *Gestaltpsychology*. New York: Liveright.

Milička, J., and Diatka, V. (Forthcoming). Mutual intelligibility of mutually unintelligible languages: How great is the effect of iconicity and distant kinship? Manuscript sumbitted for publication..

Saussure, F. de. 1959. *Course in General Linguistics*. New York: The Philosophical Library.

Sapir, E. 1929. A study in phonetic symbolism. *Journal of Experimental Psychology* 12: 225-239.