ISCA Archive
http://www.isca-speech.org/archive

ITRW on Disfluency in
Spontaneous Speech (DiSS'01)
Edinburgh, Scotland, UK
August 29-31, 2001

Filled pauses and their status in the mental lexicon

Richard Shillcock‡, Simon Kirby†, Scott McDonald‡ & Chris Brew***

*Department of Psychology, ‡Division of Informatics, †Department of Theoretical an Applied Linguistics,
University of Edinburgh, Scotland
**Department of Linguistics,
Ohio State University
rcs@cogsci.ed.ac.uk, simon@ling.ed.ac.uk, scottm@cogsci.ed.ac.uk, cbrew@ling.ohio-state.edu

Abstract

We report a study of the relationship between form and meaning in the most frequent monosyllabic words in the lexicon of English. There is a small but significant correlation between the phonological distance and the semantic distance between each pair of words. To this extent, words that have similar meanings tend to sound similar. Words differ as to the size of this meaning-form correlation in their relationship with all of the other words. When the words are ranked according to the size of this correlation we find that the words which appear towards the top of the ranking are the communicatively important words. When we look at the position in the ranking of the speech editing terms, such as *er*, *oh* and *um*, we find that they are at the very top of the ranking. We argue that this position reflects the communicative importance of these items, and that it therefore makes sense to treat them as a proper part of the mental lexicon.

1. Introduction

Intuitively, speech editing terms like *um* and *er* seem to be barely worthy of inclusion in the lexicon at all. Even though they have a clear, important communicative function [1,2], it is easy to think of them as default noises of the speech production system. In this paper we explore the implications of considering such speech editing terms as lexical entries and we investigate their relationship with the rest of the lexicon.

We adopt a new perspective on the structure of the mental lexicon, which we motivate in terms of the representational strategy predominantly found in the mammalian brain. We will characterize this strategy as "structure-preserving". We see this strategy clearly in the processing of the visual world, which generates numerous topographic representations within the relevant areas of the brain. One of the characteristics of a topographic representation is that the representation of any one thing may be "triangulated" in terms of the representations of two similar things. To the extent that *C* is quite like *A* and a little like *B* in the world, the representation of *C* will be quite like *A* and a little like *B* in the brain. Language differs from, for instance, the visual world in that its processing requirements may be expected to reflect much more the representational predispositions of the human brain. The structure of the vocal apparatus and the physical medium of speech transmission are givens, and so is the scale of the problem of reference – the sheer number of different things to which we need to refer – but beyond these constraints human language is a process by which two brains communicate and we may expect to find the processing propensities reflected in the structure found in human language.

Human language is replete with structure. There are numerous possible levels of description, from the phonological to the pragmatic. Linguists have attempted to describe the structure found at each of these levels, but one crucial domain has remained outside this enterprise: the relationship between form and meaning. It has been widely accepted that the relationship between form and meaning is arbitrary. This assumption is codified in Saussure's "arbitrariness of the sign" [3]: in principle, English could carry on just as well if the referring expressions *chair* and *swan* were interchanged, *ceteris paribus.* There is nothing essentially chair-like that makes *chair* the best expression to use to refer to chairs. It is widely accepted that there are only marginal exceptions, such as onomatopoeia, to the arbitrary relationship between form and meaning. We will argue below that this belief reflects only the absence of any means of studying this relationship.

We can begin by asking what relationship between form and meaning the brain might be expected to prefer. We claim that the default relationship should be a transparent, structure-preserving one in which words sound similar to the extent that they mean similar things. If the language were built on this principle it would be maximally easy to learn and the mental lexicon would be maximally easy to organize. To the extent that a lizard resembles a snake and also a crocodile, the phonological form of the word for lizard would resemble those for snake and crocodile. The drawback to this principle lies in the fact that adult speakers need to refer to tens of thousands of different things. If it were possible to impose a structure-preserving principle comprehensively on the mental lexicon, it would mean that the resulting words would be too long to be useful. We need, therefore, to make a weaker claim, along the lines that the structure-preserving principle may be only weakly expressed in the relationship between form and meaning, or that it may only be apparent for certain words that may be taken to constitute, in some way, the "core" of the lexicon.

Below, we summarise how we have tested this hypothesis that there is a non-arbitrary, structure-preserving relationship between meaning and form. We will show that there is indeed a small but significant relationship between form and meaning in the substantial fraction of the mental lexicon which we test. We go on to explore the role of the speech-editing terms within this relationship.

2. The meaning-form relationship

We required characterize both the phonological distance and the semantic distance between any two words. Only when we had a precise number for each distance for every possible pair of words could we test for any overall correlation between the two distances. We were not concerned with relationships between words that are explicitly morphologically related, such as *walk* and *walks* or *telephone* and *telegraph*, or with words that would seem to have some historical connection, such as *circle* and *circuit*. The hypothesis we were testing involved the meaning-form relationships between superficially completely unrelated words. It is not possible, or perhaps even desirable, to exclude all traces of historical connectedness between words. Thus, *could* and *would* appear to be part of some paradigm, and *glow*, *gloom*, *glare*, … seem to have some common proto-Indo-European ancestry. To resolve this issue we tested only the words classified in CELEX as monosyllabic and monomorphemic. .

2.1. Calculating Phonological Distance

In calculating phonological distance we were calculating the number of changes necessary to turn one word into another, to turn *chair* into swan, for instance. The phonological edit-distances were calculated using the distinctive feature representations from the Festival Speech Synthesis System[1]. Festival uses eight features: one distinguishes consonants from vowels; four classify vowels by length, height, frontness and lip-rounding; three classify consonants by type (taking the values of stop, fricative, affricate, nasal, lateral and approximant), voicing and place of articulation. Using these features we created a mismatch function assigning a penalty to each pair of two phones. Vowel features, such as length, are naturally scalar. For these we assigned a penalty of 1 for a small mismatch, and a penalty of 2 for a large mismatch. Consonant features, such as voicing, were treated as nominal variables: all mismatches attracted a uniform penalty of 1 on each dimension. Pairs involving one consonant and one vowel were assigned an additional penalty of 10. More complex, and more psychologically realistic penalty schedules could be defended on phonetic grounds, but this relatively simple one was chosen because it involves fewer assumptions. This penalty schedule was created on the basis of phonetic knowledge alone, without considering specific words and without knowledge of the semantic distances. It was never varied after its initial creation. This procedure produces a set of phone-phone distances. Distances between words were generated by applying the Wagner-Fischer edit distance algorithm [4] using the penalty schedule described above, augmented with a uniform penalty of 5 for deletions and insertions. This procedure provided us with a precise measure of the phonological distance between each word and every other word.

2.2. Calculating Semantic Distance

In calculating the semantic distance between any two words, we required a definition of meaning that could be applied equally to any word in the lexicon and also one which could be expressed quantitatively. The definition we employed was based on defining the meaning of each word in terms of the context in which it occurred. This approach echoes philosophical claims that word meaning can be defined in terms of usage [5,6]. The approach we employed has been used recently to model a variety of language processing behaviours [7,8,9]. Any use we make of the words "semantic" and "meaning" refers only to this definition, unless stated otherwise.

A semantic space was constructed from the distributional information contained in the 100 million words of contemporary English comprising the British National Corpus (BNC)[2]. Inflected words were replaced by the corresponding lemma (e.g. *walks* by *walk*). Lemmatisation reduces the sparseness of the distributional data but preserves the meaning shared by the inflectional variants. To create a context vector representation for a given word, we recorded the frequency with which each member of a list of 500 "context" words occurred within a window of five words before and five words after the critical word. Every word can be represented as a 500-dimensional context vector, but the reliability of the vector representations diminishes with word frequency [7]; hence we computed context vectors for only the 8000 most frequent words in the BNC. We defined the semantic distance between any two word vectors in the resulting high-dimensional semantic space as (1 - cosine of the angle between vectors). Word pairs that have similar distributional properties (i.e. they co-occur to a similar extent with the same set of words) receive a low semantic distance score, whereas the semantic distance is larger between words that have diverging distributional properties. This procedure provided us with the semantic distance between each word and every word in our study.

2.3. Calculating the Meaning-Form Correlation

We tested for the existence of a significant correlation between semantic distances and phonological distances across all the words in our sample. Of the 8000 most frequent words for which the BNC gave us reliable semantic distances, some 1733 were monosyllabic and monomorphemic. Although 1733 words only constitutes a small fraction of any adult native speaker's English lexicon, it still represents almost two-thirds of the number of word-types in the spoken part of the BNC. It is therefore a psychologically significant fragment of the lexicon of English.

It is important to realize that we were concerned with the distances between each word and every other word – with all the possible distances in the mental lexicon. For the 1733 words this means 1,500,778 pairs of phonological and semantic distances. This perspective gives us a comprehensive picture of the relationship between meaning and form in the mental lexicon. Previous psycholinguistic research has been almost exclusively concerned with the relationships between words for which there was some overt connection, such as semantically related words like *lion* and *tiger* or phonologically/orthographically related words like *hand* and *land*.

The correlation, *r*, between the semantic distances and the phonological distances was 0.061. There is thus only a very small overall correlation between meaning and form. We had predicted that only a very small overall correlation could be

[1] http://www.cstr.ed.ac.uk/projects/festival/

[2] http://info.ox.ac.uk/bnc/index.html

possible, due to the large number of necessarily short words in the lexicon. We tested, first, whether even this small correlation was statistically significant. We then tested whether this correlation was in some way concentrated in a particular subset of the lexicon.

Calculating the statistical significance of the correlation between 1,500,778 pairs of distances is problematic. First, the numbers within each set of measurements are not independent. Second, the number of pairs of measurements is very large and it is unclear how we should expect conventional tests of significance to behave with such numbers. The solution to these problems is to use a randomization test [10]. First, each word was randomly assigned a single, unique partner in the set. Second, the correlation was calculated using each word's own semantic context vector and its partner's phonological form, thereby randomising the relationship between the two types of distance. This procedure was repeated to generate the distribution of correlation coefficients expected by chance over 1000 randomisations. The veridical value of r was an outlier on this distribution ($z = 4.2$, $p < .001$ (one-tailed)). In summary, there is, therefore, a small but nonetheless significant relation between form and meaning across this substantial fraction of the lexicon of conversational English. The randomization test is tailored to the precise conditions and nature of the materials being tested and provides a maximally transparent test of the significance of the meaning-form correlation.

The overall correlation was highly significant, but it was also very small. Only a very small fraction of the overall variance within the distances was captured by the correlation between the two distances. We next tested the hypothesis that this correlation would not be evenly distributed across the words in our set, but would be concentrated in some particular subset of the lexicon.

For each word there were 1732 pairs of phonological and semantic distances relating it to all the other words in the set. We calculated this correlation for each word. Although the overall correlation was small, the correlations for the individual words ranged from $\tau = +0.135$ to -0.082, following an approximately normal curve positively displaced from zero. For r, the correlation ranged between +.189 and -.115. When these individual correlations were ranked, they revealed a striking ordering. The order was clearest in the ranking by r, but the same qualitative pattern obtained for τ. We discuss the ranking by r here.

The speech editing terms *oh*, *er* and *ah* were in the five most highly ranked words, ranked by r, followed by *eh* in 13th place. For a comprehensive report of the results and an extended discussion, see [11]. Overall, four psychologically important categories of words were clearly skewed towards the top of the ranking by r: speech editing terms, pronouns, proper names and swear words. We claim that these categories are all communicatively important.

3. Discussion

We have shown that, contrary to the orthodox assumption concerning the arbitrariness of the sign, there is indeed a small but significant relationship between meaning and form: words that have similar meanings tend to sound similar. Although this correlation is very small overall, it is an order of magnitude larger when the relationship of individual words with the rest of the lexicon is considered. What does this individual correlation mean? What does it mean for *oh* to have a correlation of $r = .189$ between its phonological distances and its semantic distances to the other words? It means that the rest of the lexicon conspires with the phonological form of *oh* to specify its meaning, or with the meaning of *oh* to specify its form. It means that the rest of the lexicon helps to cement the meaning-form relationship for *oh*. For a word down at the other end of the ranking by this correlation, the rest of the lexicon does not play such a role. The words *friend*, *twelve* and *frank* are the last words in the list, and they receive no such support from the rest of the lexicon for their own meaning-form relationships. When we consider what factors might determine position in the ranking by meaning-form correlation, a number of possibilities suggest themselves. Indeed psycholinguists have generated a range of dimensions of lexical variation as potential determinants of processing difficulty. Such variables include word frequency, word length, concreteness, imageability, age of acquisition and so on. A casual inspection of the ranking suggests that the words that occur towards the top of the list tend to be the shorter words, and the words which have a “less propositional” meaning, although there are very prominent exceptions – *a*, *the* and *and* all occur down near the middle of the list. More detailed analysis of the ranking by meaning-form correlation shows that the ranking reflects word length most strongly, followed by a subjective measure of familiarity. Elsewhere [11] we suggest that the correlation between meaning and form is the most basic relationship that can be tested in the lexicon, and that it is functionally significant that the communicatively important words tend to be strongly skewed towards the top of the ranking. This skewing is adaptive when considered from the perspective of how any one word is stored and accessed. Words are essentially stored superpositionally in the brain. The neural substrate over which any one word's form and meaning are stored may be quite extensive and certainly participates in the storage of many other words. In this sense we can talk about the whole of the lexicon being involved in the access of any one word. A word at the top of the ranking by the meaning-form correlation is thus stored the most securely and its representation receives indirect “assistance” from the storage of the other words.

The speech editing terms finish up at the very top of the ranked list not just because they are short and not just because they have positions in semantic hyperspace characteristic of words whose context is very varied. They are at the top of the ranked list because their phonological forms are the appropriate ones for their associated meanings.

4. Conclusions

We conclude that the speech editing terms found in the BNC can be motivated as a proper part of the mental lexicon. Their individual relationships between meaning and form are such that, in the context of the rest of the frequent monosyllabic words, they take their place as some of the most communicatively important elements of the mental lexicon.

3. References

[1] Fox-Tree, J.E. & Schrock, J.C. (1999). Discourse markers in spontaneous speech: Oh what a difference an oh makes. *Journal of Memory and Language*, *40*, 280–295.

[2] Fox-Tree, J.E. (2001). Listeners’ use of um and uh in speech comprehension. *Memory and Language*, *29*, 320–326.

[3] de Saussure, F. (1916). Cours de Linguistique (eds. Bally & Sechehaye) Générale. Paris: Payot.

[4] Wagner, R.A. & Fischer, M.J. (1974). The string-to-string correction problem. J. Assoc. Computing Machinery 21, 168-173.

[5] Cruse, D.A. (1986). *Lexical semantics.* Cambridge University Press.

[6] Wittgenstein, L. (1958). *Philosophical investigations.* Oxford: Blackwell.

[7] McDonald, S. (2000). Environmental determinants of lexical processing effort. PhD thesis, University of Edinburgh.

[8] Lund, K., Burgess, C. & Atchley, R.A.. Semantic and associative priming in high-dimensional space. In (J.D. Moore & J.F. Lehman) *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*, Erlbaum: Hove, UK. pp 660–665 (1995).

[9] Lund, K. & Burgess, C. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behav. Res. Methods, Instrum., & Comput.*, **28**, 203-208 (1996).

[10] Cohen, P. R. (1995). Empirical methods for AI. MIT Press, Mass.

[11] Shillcock, R., Kirby, S., McDonald, S. & Brew, C. (*in preparation*). Exploring the structure-preserving nature of the mental lexicon.