ORGANISING PRINCIPLES IN LEXICAL ACCESS AND REPRESENTATION?
A VIEW ACROSS LANGUAGES
William D. Marslen-Wilson
MRC Cognition and Brain Sciences Unit, Cambridge, UK.

ABSTRACT
The notion "word" is investigated cross-linguistically in a series of studies systematically comparing lexical representation and processing in English, Polish, Arabic, and Mandarin Chinese, using a variety of priming techniques. The studies so far reveal considerable diversity, with languages differing widely in types of lexical organization. Mandarin appears to rely primarily on non-combinatorial representations, while English and Polish employ in addition a decompositional, morphemically based system. The non-concatenative morphology of Arabic is also highly combinatorial, and plays an obligatory morpho-phonological structural role. We find little evidence for specific cross-linguistic constraints on lexical structure and content.

1. INTRODUCTION

An adequate account of "spoken word access processes" is going to require, if nothing else, a convincing definition of the notion "word". Without knowing what is being accessed, it is going to be hard either to construct a theory of how access is conducted, or to determine what are the appropriate tests of such a theory and its competitors. The nature of an account of the access system will differ, in truly fundamental ways, depending on whether what is being accessed is a stored representation of a word's acoustic-phonetic form, a decomposed representation of its constituent morphemes, or even a direct mapping to lexical semantic representations, without mediation by any conventional form of phonological representation.

An answer to the question of what is being accessed will require a better understanding of the structure of the mental lexicon – what are its basic units of representation and analysis, and how are these organised in relationship to each other. This perspective on the study of the mental lexicon is immediately confronted with the immense variety of lexical arrangements across languages. The research I summarise here represents some preliminary attempts to explore and describe this cross-linguistic variation. Do we find evidence for common underlying principles, that might place useful constraints on the notion "word" in the context of lexical access from speech? The results we have so far suggest that lexical systems are as notable for their differences as for their similarities.

2. STARTING POINTS

The starting point for this enterprise is a view of lexical organisation that developed initially in the context of English derivational morphology [5]. This is not because English has any special status but because any comparative enterprise has to have a starting point, and English was the one we had. Derivational morphology, in a language like English, is the concatenation of a base form (a root or a stem) with one or more derivational affixes. These may be suffixes or prefixes, as in forms like happiness, analysable as {happy} + {-ness}, or rethink, analysable as {re-} + {think}. These derivational processes change the meaning, and often the form-class of the stem, and are generally thought of as generating new lexical items.

Our primary methodology for studying the representation of English derived forms is a repetition priming technique, and our strategy is to use parallel tasks and contrasts, as far as possible, across the other languages we are looking at. The key assumption, in using priming to probe the structure of the mental lexicon, is that priming is obtained between morphologically related items if they both share the same lexical entry – if darkness primes dark, we assume this is because the prime darkness activates the underlying morpheme dark, and this facilitates the lexical decision response when dark itself is subsequently presented. This is in distinction to semantic priming, between pairs like cello and violin, where priming is assumed to involve interactions between two otherwise separate lexical representations.

Using a variety of priming techniques, we have developed a view of the English mental lexicon along the following lines. The primary claim is that the mental lexicon is not like a print dictionary in many crucial respects. It is not a static storehouse of listed words, their pronunciations, and their meanings. Instead, it is a much more dynamic cognitive entity, distinguished by three core properties:

1. Morphemic and decompositional: The unit of representation is the morpheme, and complex words are represented in terms of their constituent morphemes.
2. Combinatorial: The same morpheme combines with other morphemes across a morphological family; the morpheme {dark} in darkness is the same lexical and cognitive entity as the {dark} in darkly. Similarly, the {--ness} in darkness is the same as the {--ness} in toughness.
3. Dependent on semantic transparency: Complex forms are only represented in decomposed, morphemic format if they are synchronically semantically transparent. The form punishment is semantically transparent and represented as {punish} + {ment}. The form department is semantically opaque and cannot be represented as {depart} + {ment} because this would give the wrong semantics; therefore it must be represented as the non-decomposed whole form {department}.

1 An important methodological point is that the separation of morphological and semantic effects is not always straightforward in languages like English, where morphologically related words are typically also closely related in meaning and in form. One way round this is to use versions of the priming task which are less sensitive to semantic effects, such as masked priming and delayed repetition priming.
The evidence for these claims comes from a set of priming effects, the most salient of which are the following:

- **Stem priming**: This is the priming effect, found in masked priming and in immediate and delayed repetition priming [5,6], between a semantically transparent complex form and its stem, as in prime/target pairs like **darkness**/dark. Critically, semantically opaque pairs like department/depart do not prime.

- **Affix priming**: This is the priming effect between semantically unrelated prime-target pairs which share the same affix, as in **darkness**/toughness and rebuild/rethink [4]. This is strongest for productive affixes, and is interpreted as the combinatorial re-use of the same bound morpheme in both prime and target.

- **Suffix-suffix interference**: This is the interference effect observed between semantically transparent pairs sharing the same stem but different suffixes, as in **darkness**/darkly [5,6]. The absence of priming between these highly semantically and morphologically related pairs is interpreted as inhibition between two affixes competing for linkage to the same stem.

The co-occurrence of these three effects we take to be diagnostic of a decompositional and combinatorial system. How far do we find similar effects in other languages?

### 3. POLISH

The first language we report on is Polish, a Slavic language with a an exceptionally rich morphological system, and which also, like English, employs a concatenative derivational morphology, combining stems with sequences of prefixes and affixes. Unlike English, essentially all surface forms are morphologically complex, combining a bound stem with one or more suffixes. Thus the word *dzieci**cy**na* 'a girl' consists of a stem *dzieci*- and an inflectional ending *-a* which indicates the nominative singular feminine; the word *prz**bi**eg**l**am*** I run up consists of the stem *bieg-*; the derivational-aseptual prefix *prz-* and the inflectional ending *-am* which denotes the 1st person singular feminine past tense. In a series of studies, Chiefly using delayed repetition priming tasks to reduce semantic effects, we have found a profile of results that is very similar overall to English [7].

- **Stem priming**: We find strong stem priming in both simple and complex forms, ranging from pairs like *chodzi**c**ie/chodzi-i-ć* (walking/to walk) to highly complex forms like *błąk-o-pis-ar-siwo/pis-a-ć* (fable-writing/to write). In the latter case, the shared morpheme (the stem *pis-*) is embedded in a multi-component complex form, and yet is able to elicit priming at a delay of 12 intervening items, under conditions where no priming is obtained between purely semantically related items, such as *dom/garaz* (house/garage) over the same delay.

- **Affix priming**: We also find priming for pairs sharing the same affix, especially for more complex forms. In immediate repetition priming we find effects, on grouped analyses, for sets of pairs sharing derivational affixes, as in *kotek/logidek* (a little cat/a little garden), sharing the diminutive suffix *-ek* or *kuch-

### arz/pil**k**-arz* (a cook/footballer) sharing the agentive suffix *-arz*. Affix priming effects come through as well in delayed repetition priming, using pairs like *roz**pakow**y**w**a-l-em* (to unwrap, l**p** person sing., masculine, past tense) and *roz**walkow**y**w**a-c* (to flatten something using a rolling-pin). These words share a derivational aspectual prefix *roz-`* and the secondary imperfective suffix *-ya-*, and show strong priming even with 12 items intervening between prime and target.

- **Suffix-suffix interference**: Polish shows strong interference effects of this type. Pairs like *pis-anie/pis-**arz*** (writing/writer) and *balon-owny/balon-ik* (balloon-like/a little balloon) show no priming at all in delayed repetition, despite the close morphological and semantic relationship between prime and target, and despite the fact that inflectionally related pairs, like *mysł-**my**ś**l**c*** (I think/to think) show strong priming effects in the same experiment.

### Semantic transparency

Finally, Polish shows strong effects of semantic transparency. As in English, there is no priming for semantically opaque pairs that historically shared the same stem, as in pairs like *ja**łowic**ja**ł**owy* (juniper/futile). Unlike English, however, Polish also shows no priming between pairs where the meaning of the prime, although semantically transparent, is not fully semantically compositional. In pairs like *więz-an**k**a/więz-a-ć* (bunch/to tie), the word *więz**k**a* is rated as being highly semantically related to the verb *więzac*. Nonetheless, the meaning of the derived form is not compositional, in the sense that it is not fully predictable from the meaning of the stem combined with the nominalising affix *-anka*. Strict semantic compositionality may play a stronger role in determining representation than in English, although research is needed to clarify this.

The profile of results for the different kinds of priming relationships suggest that Polish and English have a great deal of common. Although they are very different languages in many important respects, they both fit an overall template that we interpret in terms of a morphemically organised, decompositional, and combinatorial mental lexicon.

### 4. ARABIC

Semitic languages like Arabic and Hebrew present a richly complex morphological system that is organised on fundamentally different principles to languages like English and Polish, with their concatenative morphological processes. Semitic languages, in contrast, employ a non-concatenative morphology, where the surface phonetic form is constructed by interweaving two or more abstract morphemes. In both Hebrew and in Modern Standard Arabic, the traditional analysis is in terms of a consonantal root, carrying semantic information, and a word pattern which specifies the syntactic category and the phonological structure of the surface form. Thus, for example, the triconsonantal root {ktt} with the semantic value of <writing>, combines with the word pattern {fa*ala}², with the syntactic meaning of

² This is the traditional representation of a word pattern, with the letters "f, t, l" indicating the slots into which the consonants of the root will be inserted.
‘active verb’, to give the surface form kata=ba, meaning ‘write’. The same root, {ktb} can combine with many different word patterns, in a highly productive system, to give a range of surface words with related meanings and differing syntactic properties, as in kaati=bi (writer), maktu=ab (written), yaktu=bu (he is writing) and so forth. Conversely, the same word pattern will combine productively with many different roots, so that, for example, {fa=a}ala combines with the root {ng}I <moving> to give the active verb form naga=la ‘move’.

These are highly abstract morphemes, that never surface as phonetic forms on their own, but whose presence is inferred on the basis of compelling distributional regularities. The question we investigated was whether, as already established for Hebrew [3], there is evidence that these abstract entities function cognitively in ways comparable to stems and derivational morphemes in concatenative morphologies such as English? We cannot transpose exactly into Arabic the sets of priming contrasts that we employed for English and Polish, but there is a plausible correspondence between stem priming and root priming, and between affix-priming and word-pattern priming.

- **Root priming**: Using cross-modal and masked priming techniques, we found clear evidence for priming between pairs that shared the same consonantal root [2]. Given the role of the root in determining the syntax of the resulting word, we believe this is functionally comparable to stem priming in concatenative morphologies. Thus, for example, the prime /tuxa=xul/ (inserting) speeds responses to the target /duxu=ul/ (entering), where prime and target have in common the triconsonantal root {dx}I. Strikingly, and quite differently from English and Polish, priming is just as strong when the prime is semantically opaque, as in the form /muda=xalatul/ (interference), which also shares the root {dx}I with the target /duxu=ul/, but where the meaning of the form is not synchronically predictable. This preservation of root priming under conditions of semantic opacity shows up consistently across all our experiments, and is also found for Hebrew, in comparable priming tasks.

- **Word pattern priming**: While Arabic word patterns clearly diverge in important respects, their role in determining the syntactic characteristics of the surface form gives them some major functional characteristics in common with concatenative derivational affixes. Analogously to affix priming, we find significant effects in both cross-modal and masked priming, between pairs that share the same word-pattern but have different roots and different meanings [1]. This holds both for the verbal morphology, as in pairs like /hu=t’ama/ and /farraqa/ (demolish/scatter), sharing the word pattern (fa=I=ala), but also for the deverbal nouns, between pairs like /xudu=xul/ and /nudu=xul/ (submission/happening), sharing the word pattern {fu=xulun} (with the meaning “deverbal noun, singular”). The absence of priming between forms like /suzu=xul/ and /nudu=xul/ (prison/happening), which have word patterns that are phonologically but not morphologically identical, demonstrates the morphological nature of the effects here, and rules out an account in terms of phonological overlap between prime and target.

These results for Arabic, and the comparable results for Hebrew, suggest strong support for a decompositional, combinatorial system, with abstract morphemes combining to produce the surface form, and being separated out in the process of recognition. The complete absence, however, of a semantic transparency effect in root priming, signals an apparent fundamental difference in the principles underlying the role of morphological combination.

Semitic non-concatenative morphology arguably plays an obligatory structural role in the generation of a surface form and its interpretation. This is not the case for concatenative morphologies like English, which link together already existing phonological entities. For these languages, morphological analysis is optional; the word still exists a phonological form whether or not it is treated as a combination of morphemes or as an undifferentiated whole form. For Arabic, there is not this option. Without morphological combination, weaving together consonantal root and word pattern, there is no surface phonological form.

5. COMPOUNDING IN CHINESE AND ENGLISH

The final set of comparisons involve compounding, a quite different procedure for word-formation, and where the starting point is Mandarin Chinese, rather than English. Compounding is a highly productive means of word formation in both English and Mandarin. Unlike derivational word formation, it does not involve the combination of a stem with an affix, but the linkage of two free stems – as in the English compound houseboat, made up of the two nouns house and boat.

The effect of this, in contrast to derivation, is that compounding is not fully compositional or combinatorial in nature. The meaning of a compound is never fully predictable from the meaning of its components. For example, although a snowman is a man made of snow, a milkman is not a man made of milk but someone who delivers milk (in the UK, at least). A fireman, however, is neither made of fire, nor does he bring fire to the house; rather he puts it out. There are many examples like this, and they all make the same point – that the meaning of a compound is not reliably predictable from the combination of the meaning of its components. To know what a compound means, you need to know what it refers to. This is not true of synchronically transparent, productively generated derived forms. To know the meaning of oiliness, it is enough to know the meaning of oily and of -ness. The issue, that we addressed first in Mandarin and then in English, is whether this leads to a whole-word, rather than a decompositional and mor-

---

1 In fact, on a standard generative analysis, the word-pattern is decomposed further into two still more abstract patterns, corresponding to the vocalic component of the word pattern and to its structural component (a specific sequence of consonants and vowels determining the phonological structure of the surface word). We have some priming evidence, not reported here, to support this.

2 The word pattern {fu=xulun} again has the meaning “deverbal noun, singular” for the target word /nudu=xul/ but has the meaning “plural” for the prime /suzu=xul/.
pheme-based representation of compounds in the mental lexicon.

Compounding in Mandarin takes place in a functionally very different linguistic environment from English. Mandarin has essentially no derivational morphology, so that compounding is its only productive means of word-formation, under conditions where there is considerable pressure due to homophony at the syllabic level. Recent corpus analyses suggest that around 70% of word types in Mandarin are bisyllabic compounds.

In a series of studies [8] we addressed the issue of whether a morphemic account or a whole form account was appropriate. Mandarin seemed to be a plausible candidate for a morphemic account, because of the salience of individual morphemes in the spoken language and in the writing system. Extensive research using auditory-auditory repetition priming shows that such an account is not correct. Compounds are represented as separate lexical entries, and not as combinations of their constituent morphemes. This means that Mandarin, unlike the other languages we have studied, does not have a system of word-formation that is decompositional and combinatorial.

This raises the question of how English compounds are represented, which we investigated in a series of cross-modal experiments, looking at the priming relations between transparent (bathroom), opaque (blackmail), and pseudo (shamrock) compounds and their constituent morphemes. Two results in particular seem hard to handle for a morphemic story. The first is that we did not find priming between the first and second constituents of a compound. Thus bath, for example, does not prime tub. This is quite inconsistent with the view that compounds are represented as strengthened links between their constituent morphemes.

The second finding is that shared constituents between transparent morphemes do not lead to priming unless the compounds as a whole are semantically related. Thus headache does not prime headscarf, even though they both transparently contain the morpheme head. In contrast, teacup does prime teapot. This is because these two compounds are strongly semantically related, whereas headache and headscarf are not. This is not consistent with a morphemic, combinatorial story, where the morpheme head is a constituent of headache and headscarf in the same way that punish is a constituent of both punishment and punishable.

In summary, compounding in Mandarin and English seem to be remarkably similar, reflecting in the same way the representational consequences of the unpredictability of the meaning of compounds.

6. OVERVIEW

Despite the small sample of languages studied, we are left with a wide range of lexical arrangements. Mandarin Chinese seems to lie at one extreme, with apparently no combinatorial procedures for word-formation, and with a lexicon made up of whole forms, in which compounds and the words that make up these compounds all have separate lexical representations. English has a similar system for compounding, but also has a decompositional system of word-formation and representation, reflecting the different processing requirements of derivational procedures that operate on a combinatorial basis, and that deliver predictable and compositional meanings.

In this respect English parallels the broad characteristics of a language like Polish, which has a much richer and more complex morphological system. Both these languages, in turn, share with Arabic (and Hebrew) a combinatorial and decompositional approach to lexical representation. In these Semitic languages, however, morphological representation appears to play a more fundamental structural role, so that no surface form can be produced without some underlying process of morpho-phonological combination. This delivers both the surface form, and its basic syntactic and semantic properties.

It is possible that there may not be lexicon-specific universals.

7. ACKNOWLEDGEMENTS

This research was supported by grants from the UK MRC and ESRC. Author address: William Marslen-Wilson, MRC-CBU, 15 Chaucer Road, Cambridge CB2 2EF, UK. Email: william.marslen-wilson@mrc-cbu.cam.ac.uk

8. REFERENCES


