Final Draft

Morphological representation as a correlation between form and meaning.

Marcus Taft

University of New South Wales, Sydney, Australia

Mailing address: Marcus Taft
School of Psychology
University of NSW
Sydney NSW 2052
AUSTRALIA
Email: M.Taft@unsw.edu.au

Acknowledgements

The research reported in this paper was supported by a grant awarded to the author by the Australian Research Council. The technical assistance of Zhu Xiaoping is gratefully acknowledged.

Lexical memory exists for the purpose of mapping meaning onto word form. It interfaces with the world via a level of representation that corresponds to the presented form of the word (either orthographic or auditory), and access to this representation makes available the semantic information associated with it. Such a form-to-meaning association has been built up over repeated occurrences of the same form referring to the same thing. That is, the lexical system captures the correlation that can be found between a form and the context in which it occurs. For example, the word form *cat* is consistently found in a context that refers to a furry, whiskered pet that says "miaow", even though this context might vary markedly in other ways (e.g., when referring to a cat stuck up a tree, a cat owned in one's childhood, or a cat chasing a mouse).

The ability to capture a correlation between two levels of stimulus characteristics (e.g., form and meaning) is a feature of connectionist models where weightings on connections between the two levels are strengthened in response to the correlations between those levels. Whether each of the units found at the different levels of analysis represent a complete identity (i.e., are "localist representations") or contribute to a pattern of activation across a number of units (i.e., are "distributed representations"), the correlation between levels can still be captured (see e.g., Page, 1999). The framework to be used here is a version of the connectionist approach that is localist (e.g., Grainger & Jacobs, 1996; Norris, 1994; Schreuder & Baayen, 1995; Taft, 1991) rather than distributed (e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989). The reason for this is that the localist version provides a more concrete metaphor for lexical processing in that the units of activation can be explicitly described. Even if it turns out that the brain works in a distributed fashion, the localist framework presented here might be thought of as a model of this distributed system that allows one to understand the nature of the lexical representation in more concrete terms.

Such a framework is presented in Fig 1, which depicts how the system might represent the correlation between form and meaning, at least for visually presented words that are the focus of this paper. In this figure, each form representation constitutes a unit that is activated by the presence of relevant orthographic features in the stimulus. Each concept unit represents the coalition of a set of semantic features.

Fig 1 about here

Note that it is possible for a single form unit to be associated with more than one concept unit, namely, when it is an ambiguous word. For example, the word *dog* can refer to a verb that means "to follow closely", which has few, if any, semantic features in common with its "animal" meaning. Therefore, there must be some way of keeping these two sets of features separate from each other and this is achieved by the existence of two separate concept units, one for each meaning of the word *dog*. If semantic features were linked directly to the form units, there would be no way of distinguishing features that are relevant only to the "animal" meaning from those relevant only to the "follow" meaning. The ability to distinguish between the two meanings of an ambiguous word provides one important reason for proposing the existence of a concept level separate from the level of semantic features. In other words, it is the concept unit that brings together a set of semantic features and links it to the word form.

The alternative to this would be the postulation of two separate but identical form units for each meaning of a word (e.g., two form representations for *dog*), but this would seem to be unnecessarily redundant. Of course, redundancy is not always a bad thing in and of itself, because it might provide a back-up mechanism should there be disruption to the system. However, it is hard to see the advantage of having two identical form representations if they are relevant to different meanings. That is, the redundant information does not in this case provide an alternative route to the same information.

Concept units (or **lemmas**, cf. Baayen, Dijkstra, & Schreuder, 1997; Schreuder, Burani, & Baayen, *this volume*) can actually be seen as the units that capture the correlation between form and meaning. As such, they take on the role of the "hidden units" that are proposed in distributed models to mediate between input and output (e.g., Seidenberg & McClelland, 1989), and are equivalent to what Plaut et al describe as "attractor patterns". What this means is that lemmas can vary in their precision depending on the degree of correlation between form and meaning. The greater the variation of meaning that a word has in different contexts, the less distinctive its lemma representation will be (i.e., the more widely tuned the attractor pattern will be). Such a proposal is consistent with the finding that concrete words, whose meaning is very

precise, are recognized more readily than abstract words (e.g., Bleasdale, 1987; Schwanenflugel & Shoben, 1983).

Representation of polymorphemic words

The concern of this chapter, however, is the role of morphology in lexical processing. To examine this issue, attention will first be turned to the nature of the form representation in visual word recognition. Does each form-level unit really represent a word, or does it represent a morpheme or even a syllable? Most connectionist models of lexical processing are based solely on monosyllabic words, like *dog* or *cat* (e.g., Coltheart, Curtis, Atkins, & Haller, 1993; Grainger & Jacobs, 1996; Plaut et al., 1996; Seidenberg & McClelland, 1989), where word, morpheme and syllable are the same thing. When modeling the processing of more complex words, however, the nature of the form representation needs to be established because the three possibilities are quite different, as depicted in Figs 2a, 2b, and 2c.

Figs 2a, 2b, & 2c about here

If the form representations are those of whole words (Fig 2a), then polymorphemic words (like *fireman*) and monomorphemic polysyllabic words (like *picnic*) will be represented in full. Alternatively, if the form representations are morphemic in nature (Fig 2b), then only monomorphemic words will be represented even when they are polysyllabic (like *picnic*). The concept unit for a polymorphemic word will be activated by the combined activation of the form units for each morpheme, with the combination of activation following a particular order (or else *gunshot* could not be differentiated from *shotgun*). This is arbitrarily depicted in the figure as numbered links. Polysyllabic monomorphemic words (like *picnic*) will be broken down further only if the form units represent syllables (Fig 2c).

An additional possibility is a hierarchy of units such that, for example, a syllabic level sends activation up to a morpheme level which in turn sends activation up to a whole word level (see e.g., Taft, 1994). This, however, will lead to considerable redundancy in representation with, for example, *man* existing at all three levels. Some of this redundancy could be eliminated by specifying that the morpheme level is only for morphemes that are not free standing words (i.e., bound morphemes like the *hench* of

henchman). In this way, *man* will have a representation at the word level, but not at the morpheme level. Nevertheless, if a syllable level also exists, there would still be redundancy between this and the other two levels.

In fact, the way in which bound stem morphemes are represented is an important issue because it addresses the relationship between the representation of morphemes and syllables, as well as that between morphemes and words. That is, a bound stem morpheme can be distinguished from a purely form-based syllable because some sort of meaning can be ascribed to it, and it can be distinguished from a word because it cannot stand alone. For this reason, bound morphemes are the central focus of the present inquiry into lexical representation.

Representation of bound morphemes

The existence of a form-level representation for bound morphemes has been proposed for a long time (e.g., Emmorey, 1989; Taft, 1994; Taft & Forster, 1975), and is supported by the finding that bound morphemes are relatively hard to classify as nonwords. For example, difficulties encountered in making a lexical decision to *vive* compared to *lish* (Taft & Forster, 1975) can be interpreted as evidence that *vive* is represented in lexical memory for the purposes of recognizing *revive*, while the *lish* of *relish* is not. Such an interpretation dismisses any model that fails to incorporate any morphological analysis such as that depicted in Fig 2a (which will therefore not be considered further), but is consistent with the "morpheme" model depicted in Fig 2b, which readily incorporates bound morphemes by simply allowing the form representations to be either bound or free. In such a way, bound morphemes (e.g., *vive*, *hench*) have a status that syllables (like *pic* and *nic*) do not have.

In contrast, according to the "syllable" model depicted in Fig 2c, bound morphemes and syllables would have the same status within the form-based representations. This does not mean, however, that bound morphemes and syllables are necessarily represented in the same way in the lexical processing system. In particular, one could argue that bound morphemes are associated with a meaning and, therefore, will have a lemma representation, whereas syllables will not. Thus, bound morphemes and syllables can potentially be differentiated at the lemma level of representation in Fig 2c.

This raises the question of whether knowledge does exist in the lexical processing system that a syllable functions as a bound morpheme or not. Do language users draw a distinction between letter-strings like *hench* and those like *pic*? The research by Taft and Forster (1975) and Taft (1994) suggests that they do, in that the non-morphemic word parts (e.g., *lish*) were often syllables. This was particularly true in the Taft (1994) study where the bound stem items (e.g., the *peccable* of *impeccable*) were matched on stress pattern with the non-stems (e.g., the *tinerary* of *itinerary*), yet were harder to classify as nonwords. It may well be, though, that the first syllable is the most important in recognizing a polysyllabic word, and we do not know whether such syllables (like *pic*) are treated any differently to bound morphemes (like *hench*). Although Taft and Forster (1976) found a delay in classifying first syllables as nonwords relative to non-syllables, they did not compare syllables to bound morphemes. Indeed, some of the syllable items used by Taft and Forster (1976) were actually bound morphemes (like *hench*).

When considering a comparison of bound morphemes and syllables, however, one must confront the difficulties that exist in defining both of these terms. In relation to the syllable, it is possible to give either a phonological or an orthographic definition. If a phonological definition is adopted, there will often be a conflict with morphemic structure (e.g., the syllable structure *ea* + *ten*, *ac* + *tor* is incompatible with the morphological structure *eat* + *en*, *act* + *or*). For this reason, Taft (1979, 1987) supported the idea that an orthographically defined syllable is important in reading, where the first syllable is includes all consonants after the first vowel without creating an illegal final consonant cluster. So, while the first syllable of *picnic* would be taken to be *pic* (because *cn* is not a permissible final cluster), the first syllable of *police*, *tiger*, and *winter* would be *pol*, *tig*, and *wint* respectively even though these violate the phonological syllable structure. This orthographically based unit was labeled the Basic Orthographic Structure (or BOSS). However, because others have failed to find support for the BOSS (e.g., Lima & Pollatsek, 1983), it is unclear what the appropriate choice of definition should be when examining the role of the syllable. There do exist clear cases that are consistent with any definition of syllabification (like the *pic* of *picnic* or the *zig* of *zigzag*), but the number of such words, at least in English, is very restricted.

Furthermore, even if syllabification were clear-cut, it is not always obvious whether

such a subunit should be considered a bound morpheme or just a meaningless syllable. Being a bound morpheme is not an all-or-none phenomenon; there are degrees of meaningfulness. A letter-string is clearly a bound morpheme when it does not stand on its own as a word, but forms a constituent in a number of different words with a shared meaning. An example might be *smith,* which is not a free-standing word in most people's vocabulary, but which recurs in several words that have overlapping meanings, like *blacksmith*, *goldsmith*, *locksmith* and *silversmith*. Thus, some sort of meaning can be extracted out for *smith* (i.e., something to do with "working with metal"), thus making it morphemic in nature.

Less clear cases are letter-strings like *hench* that only occur in a single word, but where the accompanying constituent (*man*) is used as a morpheme. That is, a henchman is a man and, therefore, *hench* might be taken to refer to "the thing that a man does in order to be a henchman". In this way, it might be possible to work out what a *henchwoman* or *henchdog* mean if those words were coined (whereas *taliswoman* or *talisdog* would be nonsensical because *man* does not function as a morpheme in the word *talisman*, being semantically unrelated to it).

There are other cases that have an even more dubious status as a bound morpheme, namely, where a syllable may be considered to have some level of meaning as a result of sound symbolism. For example, it could be argued that *zig* refers to "a sharp movement in one direction" in contrast to *zag*, which refers to "a sharp movement in the other direction", thus creating a *zigzag*. So, is it appropriate to consider that the syllables *zig* and *zag* have a morphemic status?

In the "morpheme" model of Fig 2b, such a question is relevant to the form-level representations, whereas in the "syllable" model of Fig 2c, it is relevant to the lemma level. However, if the representation of morphemic information arises from the correlation between form and meaning, neither model needs the reader to make a conscious decision about the morphemic status of any sublexical unit. In fact, a decision about whether a sublexical unit should be classified as a morpheme or not simply becomes a taxonomic exercise and not of crucial concern for the establishment of representations in the lexical processing system. The development of such representations will be based purely on the statistical relationships that exist between form and meaning.

To be more specific, the form units that develop according to the "morpheme" model, will be the smallest units that can be abstracted out from the correlation with meaning. So, *smith* would develop as a form unit because its appearance in a word is consistently associated with the idea of "working with metal". *Hench* would also develop as a form unit because it constitutes the remnant of the word *henchman* after the meaningful *man* has been abstracted out as a unit. It is unclear whether form units for *zig* and *zag* would develop on the basis of sound symbolism. However, pure syllables, like *pic* and *nic*, definitely would not. Although the existence of the letter-string *pic* is correlated with the occurrence of the meaning "an outdoor meal", this correlation will be identical with that for the unit *nic* and, therefore, having separate units would achieve nothing beyond what can be captured in a single unit for *picnic*.

While form units represent morphemes according to this model (as depicted in Fig 2b), we need to also consider the nature of the units at the lemma level. A unit at the form level that represents a free morpheme (e.g., *fire* or *picnic*) will be associated with its own lemma as well as with any lemmas for polymorphemic words in which it occurs (e.g., *fireman*). However, when a form unit represents a bound morpheme (e.g., *hench*), it may or may not develop its own lemma as a link to semantics. Fig 3 shows what the system would look like if bound morphemes were not associated with a lemma unit, but rather, linked solely to the relevant polymorphemic word lemma. By this version of the "morpheme" model, to consider a bound morpheme (like *hench* or *smith*) as being at all meaningful, there must be generalization from the meaning activated via the polymorphemic word/s in which it occurs.

Fig 3 about here

Turning now to the "syllable" model depicted in Fig 2c, the units that develop as a result of the correlation between form and meaning are at the lemma-level. That is, the form-level units are defined purely on a structural basis, independent of meaning, with the lemmas being the units that capture the correlations. So, lemmas are not restricted to free morphemes, because bound morphemes also display a correlation between form and meaning, even if not very precisely. When a letter-string occurs in several semantically related words (like *smith*), a lemma unit will develop with a precision

greater than that developed when the associated semantic features are less obvious, namely, when the letter-string occurs in only one relevant context (e.g., *hench*). In the latter case, semantic features are associated with the constituent essentially by default. That is, the lemma units for *henchman* and *man* will be linked to semantic features that overlap, and this provides the possibility for *hench* to develop its own lemma unit on the basis of the non-overlapping features associated with *henchman.* In other words, a lemma for *hench* would bring together the semantic features that are linked specifically to *henchman* and which are not relevant to the fact that it refers to a "man". Finally, because form units are based on structural grounds alone, *pic* and *nic* will be represented at the form level but not at the lemma level. This is its crucial contrast with the "morpheme" model, where pure syllables are represented at neither the form nor lemma level.

According to the pure "syllable" model, then, form units are defined entirely on form-based grounds, and these units can then participate in the correlation with meaning. However, it is easy to refute this proposal. Take, for example, the polymorphemic word *weekend*. Here, the lemmas for the constituent morphemes will be activated via the form units *week* and *end*. If form units are to be set up without any influence from higher level knowledge, there must therefore be a way of identifying these two units on a purely orthographic basis. In fact, a BOSS analysis provides the appropriate structure because the *k* would be included in the first unit. However, what then of *beekeeper*? It would have to be represented as *beek* + *eep* + *er* and this would provide no avenue to the lemmas for *bee* and *keep*. Hence it is apparent that higher level knowledge must have an impact on the form units that are stored. Indeed, it was for this reason that Taft (1979, 1987) incorporated morphology into the definition of the BOSS (i.e., "include all consonants following the vowel of the stem morpheme").

Does the fact that functional information can influence form-level representation entail that Fig 2b is correct and Fig 2c is wrong? Not entirely. It may be that a combination of the two is possible. In particular, efficient form-level representation can be seen as one that breaks complex words down into smaller, manageable units. Correlation between form and meaning provides the primary basis for this breakdown of the word, such that morphemic analysis dominates over purely form-based analysis (giving *week* + *end* and *bee* + *keeper*). Nevertheless, form-based analysis might occur when there is no morphemic basis for the analysis (as with *picnic*). This is exemplified in Fig 4, where

the existence of lemma units for bound morphemes is also depicted. The structure of this model is the same as that shown in Fig 2c. The only difference is that morphemic considerations can override syllabic ones in determining the form-level representations (as in *weekend* and *beekeeper*). For this reason, the orthographic units have been re-labelled with the general term "form" (rather than "syllables"), in order to capture the fact that the form units not only represent syllables for the recognition of monomorphemic words, but can represent morphemes for the recognition of polymorphemic words as well. Thus the model is a hybrid of a morpheme and syllable account. Morphemic considerations lead to the first form unit of *weekend* being *week* and the first form unit of *beekeeper* being *bee*, whereas syllabic considerations lead to the first form unit of *picnic* being *pic*. The basis for *hench* being a form unit is ambiguous since it is both the first syllable and first morpheme of *henchman*. However, the example of *henchman* is given in the figure in order for it to be compared directly to Fig 3.

Fig 4 about here

According to this hybrid model, syllables are represented at the form level in exactly the same way as are monosyllabic bound and free morphemes, but they differ by virtue of the fact that a syllable is not associated with its own individual lemma. In contrast, according the model depicted in Fig 3, syllables differ from morphemes inasmuch as the former have no form-level representations, while bound morphemes differ from free morphemes inasmuch as the former have no lemma-level representations.

What is being suggested, then, is that we can reject the "syllable" model of Fig 2c on logical grounds, but that the idea of pure syllabic analysis in lexical processing can still be incorporated if a hybrid model (Fig 4) is adopted. Whether purely syllabic representations are actually present in the word recognition system is still an open question, though research supporting the existence of the BOSS (e.g., Taft, 1979, 1987) favors such an idea.

Morpheme retrieval

In the hybrid model (as well as in the morpheme models of Fig 2b and Fig 3), higher level knowledge influences the nature of the form-level representations: In the example

of *weekend* and *beekeeper*, the form units are determined on the basis of the fact that the meaning of each unit is relevant to the meaning of the whole word. However, even if morphological considerations influence the nature of the form unit, it must be emphasized that it is still possible to activate that unit purely on a physical basis. For example, if the "beekeeper" lemma is activated via the form units for *bee*, *keep* and *er*, while the "weekend" lemma is activated via *week* and *end*, the appropriate form units will be activated simply because they are congruent with the letter groupings contained in the stimulus word.

What this means is that a number of irrelevant form units might be activated when a word is presented, but these will eventually be inhibited by virtue of the fact that they fail to activate an appropriate lemma. For example, *beetroot* contains the letter-string *bee* and, therefore, the form unit for *bee* will be activated. However, the remaining letters *troot* will fail to activate any form unit, so the alternative analysis as *beet* and *root* will eventually be accepted because it succeeds in activating the lemma for *beetroot* (see also Libben, 1994; Taft, 1979). Thus, while functional knowledge might influence the storage of form-level representations, the actual retrieval of those representations during reading can occur without the direct involvement of such higher-level information.

An experiment that distinguishes formal and functional processing

It was stated earlier that being a bound morpheme is not an all-or-none phenomenon and that this is captured within the notion of correlation between form and meaning. That is, morphemes vary in their precision or clarity of representation depending on the strength of this correlation. This variation in clarity is readily captured within lemma units because they develop as a means of characterizing the correlation between form and meaning. It was suggested earlier that lemmas representing bound morphemes like *smith* that occur in many different semantically related contexts have a reasonably precise representation, and may even be hard to distinguish from free morphemes (which also, of course, occur in a wide range of contexts). Bound morphemes like *hench*, on the other hand, which only exist within a single context, might have only an indistinct lemma representation because their meaning is harder to determine independent of the single word in which they occur.

The question that arises, however, is whether form-level units that are regulated by morphemic information (i.e., the "morpheme" model of Fig 2b and Fig 3, or the hybrid model of Fig 4), also vary in their clarity depending on their morphemic status. That is, is the form unit for *smith* "clearer" than that for *hench* which in turn is "clearer" than that for *pic*? In fact, it is not necessary to assume that this is the case as long as the variations in clarity are already captured at the lemma level. That is, form-level representations might only be influenced by their frequency of usage, whereas at the lemma level they differ in terms of the strength of correlation between the form and meaning. So, by this account, form-level representations will exist equally for *smith*, *hench*, *pic*, and even *fire*, as long as their frequency of usage is matched. Where they will differ, though, is at the level of the lemma where they vary in terms of their association with meaning. Support for this idea comes from a set of experiments that I recently conducted using Chinese characters.

Chinese words are composed of one or more physically separated characters, though a two-character structure is the most common. There were several reasons for using Chinese to examine the distinction between form-level and lemma-level processing. First, many characters exist that are bound morphemes, either occurring in many contexts, i.e., in many two-character compound words (equivalent to *smith*) or occurring in only one context (like *hench*). There are also quite a few meaningless syllables (equivalent to *pic*) that only ever occur in combination with another syllable which, in turn, only ever occurs in that same word (as is the case with *zig* and *zag* in English). Second, and most importantly, because each character in Chinese is a physically distinct unit, there are two types of judgement that can be made about such characters that cannot be differentiated if using alphabetically scripted words.

In English, one is able to say that *bench* and *pit* are words while *hench* and *pic* are nonwords. However, it is unclear whether this judgement is based on the fact that the letter-strings *bench* and *pit* exist as word forms while *hench* and *pic* do not (i.e., a form-based explanation), or on the fact that *bench* and *pit* have clearly defined meanings while *hench* and *pic* do not (i.e., a function-based explanation). Any confusion in classifying *hench* as a nonword relative to *pic* could similarly be given a form-based or a function-based explanation. That is, the morphemic status of *hench* could have its impact either at the form-level or at the lemma level. In Chinese, on the other hand, these two possibilities can be differentiated because a character can be

classified in two distinct ways; one focusing on form and the other on function.

In a "Character Decision Task" (or CDT), existing characters are discriminated from non-existing characters. Non-characters are composed of real components which are combined in non-existing ways, while real characters are either free morphemes, bound morphemes, or pure syllables. The CDT could be potentially performed purely at the form level because real characters have an orthographic representation while nonsense characters do not. Note that character frequency may well have an effect on the CDT, arising from the likelihood that activation of a form unit is affected by its frequency of usage.

In contrast to the CDT, however, it is also possible for a Chinese reader to decide whether a character exists as a free-standing word or not (a "Word Decision Task" or WDT). Here, the response is "yes" for free morphemes and "no" for the other types of characters, namely, bound morphemes and pure syllables. Thus, it can be seen that the Word Decision Task requires the processing of functional information, which is not necessary in the Character Decision Task.

The CDT and WDT tasks can therefore be contrasted to determine what is important in form-level processing and what is important at the higher functional level. First, according to the hybrid model of Fig 4 (as well as the rejected "syllable" model of Fig 2c), pure syllables, bound morphemes and free morphemes should all be equally well classified as real characters in the CDT as long as frequency is well matched. In the WDT task, on the other hand, the pure syllables should be easier to reject than the bound morphemes (because the former have no lemma representation). Moreover, the bound morphemes should be distinguishable from the free morphemes in the sense that the latter will have a more precise lemma representation (because of their more consistent meaning across contexts). The distinction between bound and free morphemes, then, should be seen in the greater likelihood of classifying the latter as a "word" rather than a "nonword" in the WDT, even though they are equally classifiable as characters in the CDT.

As mentioned, in testing these predictions, it is important that frequency be well-matched and this raises the issue of what the appropriate frequency should be. Specifically, should the frequency of a character be the number of times that that

character occurs regardless of context (i.e., token frequency) or should one take into account the number of compound words in which that character occurs (i.e., type frequency)? An answer to this would be needed if one were wanting to compare "clear" bound morphemes that occur in many compound words (like *smith*) with "unclear" ones that occur in only one compound word (like *hench*). However, because one cannot know if a token frequency based on the frequency a single compound word (like *henchman*) is equivalent to the same token frequency based on the summed frequencies of a number of different compound words (like *blacksmith*, *goldsmith*, *locksmith* etc), a direct comparison of clear and unclear bound morphemes was not carried out. On the other hand, *hench*-like bound morphemes can unquestionably be compared to *zig*-like[1] syllables because the frequencies of both sorts of item are based on the token frequency of a single word (i.e., *henchman* and *zigzag* respectively) and are therefore necessarily matched on type frequency. In addition, free morphemes (like *fire*) can be compared to *smith*-like bound morphemes because it is possible to match them on both character frequency (token frequency) and the number of compound words in which they occur (type frequency).

Two experiments were therefore carried out using native Chinese speakers as participants. Fifteen such participants were given a Character Decision Task and another 15 were given a Word Decision Task. The same experimental characters were used in both tasks, though in the CDT there were also an equivalent number of non-characters used as distractors. The experimental characters included 20 bound morphemes that occur in only one word (*hench* items) matched on token frequency (about 6 per million) with 20 pure syllables that also occur in only one word (*zig* items). The distinction between these two types of item is that the character that accompanies it in the word in which it occurs either has its own meaning (as in *henchman*) or it does not (as in *zigzag*).

The other comparison was between 20 bound morphemes that occur in several

---

[1] : Terminologically speaking, analogy to *zig* is preferred over analogy to *pic*, even though the former has a dubious status as a pure syllable in English. This is because the criterion for inclusion in this condition is the fact that the item exists in only one word (true of *zig*, but not of *pic*), as does its accompanying constituent (true of *zag*, but not of *nic*).

semantically related words (*smith* items) and 20 free morphemes (*fire* items) that occur in just as many compound words. Thus the bound and free morphemes were matched on type frequency (a mean of 5 words), and they were also matched on token frequency (about 22 per million). The decision that a character could or could not be a free morpheme was based on whether or not it was found as a single character in Table 1 of the *Modern Chinese frequency dictionary* (1985) where word frequencies are listed.

In the WDT, the expected results were obtained. It was significantly harder to say that *hench* items were nonwords than to say the same of *zig* items (with 24.3% "yes" responses vs 10.7%, and a mean reaction time for the correct "no" responses of 882 ms vs 854 ms). In addition, although it proved difficult for readers to classify the putatively free morphemes (*fire* items) as free-standing words (with only 65.4% "yes" responses), this was nevertheless significantly different to the responses to *smith* items (with 40.6% "yes" responses)[2].

It is apparent from the results of the WDT that the lexical processing system represents bound morphemes differently from pure syllables (*hench* items vs *zig* items) and also from free morphemes (*smith* items vs *fire* items). Importantly, analysis of the results of the CDT indicated that this difference in representation is not at the form level. In particular, there was no significant difference in either of the comparisons when status as a character was to be determined: The error rate comparison in the CDT for *hench* and *zig* was 12.0% vs 10.7% with reaction times of 620 ms vs 618 ms; and for *fire* and *smith*, 5.0% vs 4.9% with reaction times of 568 ms vs 575 ms.

The contrast between the results from the two tasks is quite informative. First, it suggests that Fig 3 is wrong because, if a bound morpheme has no lemma, then any difference between bound morphemes and syllables must arise at the form level and, therefore, such a difference should be found in both the CDT and WDT. The fact that it is only found in the WDT, indicates that the difference between these two types of item must arise from a level higher than the form level, namely, the proposed lemma level.

---

[2] : Reaction times were not measured because the correct response was different for the free morphemes ("yes") and bound morphemes ("no").

So, it is suggested that a bound morpheme does develop a lemma representation, but one that is typically less precise than that for a free morpheme. Thus, even though free and bound morphemes are equally represented at the form level (as shown in the CDT), they can be discriminated at the lemma level, albeit with some difficulty (as shown in the WDT). Such a conclusion is compatible with the model presented in Fig 4.

The question can be raised, however, whether the conclusions drawn here might be specific to Chinese and not generalizable to alphabetically scripted languages. In particular, one of the features of Chinese is that its individual characters can be determined on the basis of physical features alone, i.e., on the basis of their physical separation from each other. For this reason, one does not need any familiarity with Chinese to isolate each form-level unit (i.e., each character) and, therefore, knowing what to store at the form level is simple. In contrast, one needs to understand English to know that *weekend* should be stored as *week* and *end* rather than as *wee* and *kend*. What this means, though, is that even in a script where form units are readily isolated (i.e., Chinese), there is still evidence that higher level units develop with a level of precision related to the correlation between the occurrence of these form units and a particular meaning.

It seems highly likely that the strength of this correlation also affects the ease with which English readers can decide whether a letter-string is a free standing English word or not (as in the standard Lexical Decision Task). Unlike the situation in Chinese, it is hard to come up with examples in English where a unit recurs in several words that are all semantically related. The example of *smith* has been used here, but it would not be ideal as a nonword item in an experiment because there is the potential for confusion with the surname *Smith*. However, one can imagine readers mistakenly classifying *venge* as a free morpheme far more than they would *ceive*, for example, because there is a clear difference in the consistency of the semantic relationship between the words in which they occur: That is, the constant meaning associated with *revenge*, *avenge*, *vengeful*, and *vengeance* is obvious, while that between *receive*, *deceive*, *perceive* and *conceive* is not and, therefore, the lemma abstracted out for the former will be more precise than that for the latter.

Furthermore, it is not hard to imagine that lexical decision responses to bound

morphemes would take longer than those to initial syllables in English just as they did in Chinese. Some indication of this can be found by looking at the items used by Taft and Forster (1976) who demonstrated a delay in classifying syllables as nonwords relative to letter-strings that were not part of a word. The few items that might be considered to be bound morphemes (i.e., *hench*, *voy*, *cran*, *yester*, *caval,* and *plat*) showed a larger RT difference relative to their controls than did the purely syllabic items (a 127 ms difference versus a 31 ms difference).

So, it seems that the results obtained in Chinese, at least for the WDT, are likely to be generalizable to alphabetic scripts as well. This therefore suggests that the notion of units that capture the correlation between form and meaning (i.e., lemmas) is a fruitful way in which to conceptualize morphological representation.

Morphological transparency

Although the hybrid model of Fig 4 appears to be a useful framework for understanding morphological processing, there are still several theoretical issues that need to be considered. As it stands, the difference between the representation and processing of a polymorphemic word (e.g., *fireman*) and that of a monomorphemic polysyllabic word (e.g., *picnic*) is whether or not there is concomitant activation of lemmas associated with the sublexical form units. That is, there are lemmas associated with *fire* and *man*, but not with *pic* or *nic*. However, it sometimes happens that the syllables of a monomorphemic word coincide with real words (e.g., the *cart* and *ridge* of *cartridge*) and, in such circumstances, lemmas for those constituents will be activated just as they would be for a polymorphemic word (see also Schreuder et al, *this volume*).

It is important, though, to explain how a difference might arise in processing a complex word whose constituent meanings are transparently related to its whole word meaning (as with *fireman*) relative to one where they are not (as with *cartridge*), because there is evidence that such transparency does have an impact on lexical decision responses (e.g., Marslen-Wilson, Tyler, Waksler, & Older, 1994, in English; Sandra, 1990, Zwitserlood, 1994, in Dutch; Taft, Liu & Zhu, 1999 in Chinese). Within the current conceptualization, the difference lies in the overlap between the semantic features associated with the lemma for the whole word and those associated with the

lemmas for its constituents. That is, the semantic features associated with both *cart* and *ridge* will be entirely different to those associated with *cartridge*, while there will be an overlap between the semantic units activated by *fire* and *man* and those activated by *fireman*. The transparency of a word's morphemic structure is indeed defined in terms of this overlap.

Any demonstration of the impact of transparency on lexical processing would therefore need to be explained in terms of the involvement of overlap at the semantic level. For example, if recognition of *man* is facilitated more by the prior presentation of *fireman* than is the recognition of *ridge* by the prior presentation of *cartridge* (Zwitserlood, 1994), this could be explained in terms of the semantic units activated via the lemma for the complex word feeding activation back to the constituent lemmas in the case of *fireman* but not in the case of *cartridge*.

Now, while it is obvious that *fireman* is semantically related to its constituent morphemes, one cannot say that its meaning is entirely predictable from the combination of the meanings of those morphemes. In particular, one needs access to information that a fireman puts out fires rather than sets fires, or that firemen wear a special uniform, and so on. This fact justifies the need for a lemma to exist for the polymorphemic word *fireman*. That is, there needs to be a unit that is associated with the semantic information specific to *fireman* because its meaning cannot be accurately generated from the meanings associated with its individual morphemes. Some of its meaning can be extracted from that of its constituents (e.g., features relevant to the fact that a fireman is a man), but other information is needed as well. If a polymorphemic word were entirely predictable from its constituents, on the other hand, there would be no need for a lemma to exist for that word. Is such a thing possible?

This extreme form of transparency would seem to exist most clearly in the case of inflected words. For example, the meaning of *cats* is entirely predictable from the meaning of its stem *cat* and the plural function of its suffix *s*. This implies that the inflected word *cats* will be recognized via both the form units and the lemma units associated with its constituent morphemes *cat* and *s*[3]. The more interesting cases are

---

[3] In fact, there would exist two lemmas activated by the presence of a final *s*: One for the plural

irregularly inflected words, like *men*. Would such words require their own lemma? Not necessarily. While there would need to be independent form units for *men* and *man*, the former might simply be associated with the *man* lemma and the "plural" lemma, as depicted in Fig 5 (where the regular plural is also shown).

Fig 5 about here

However, it should be noted that if the "plural" lemma can be activated via a form unit that does not represent *s*, the idea of lemmas capturing the correlation between form and meaning is lost. That is, a "plural" lemma should develop explicitly out of the correlation that exists between plurality and the presence of a final *s*, so it should only be activated when a final *s* is present. If it were also activated by units that are entirely unrelated in form (as with *men*), one would have to say that units representing grammatical function, such as plurality, are of quite a different nature to those representing semantic function. In fact, they may not have a representation at the lemma level at all, perhaps being represented by a single unit in a separate syntactic module which is directly activated by the appropriate form-level representations (cf Taft, 1994). Such an idea is compatible with the conceptualization of lemmas as being the means by which a number of different semantic features are brought together and linked to the form level. If there is no set of semantic features to unite, as in the case of a syntactic function like "plural", then there is no need for a lemma.

Morpheme frequency

The difference between a complex word with transparent constituents, like *fireman*, and one with opaque constituents, like *cartridge*, is described in terms of semantic overlap between lemmas. However, there is a modification to the model that seems to be required and which potentially provides another way of differentiating them. In particular, there is a problem with the idea that a monomorphemic lemma unit can be directly activated by more than one syllable unit, as in the example of *picnic* being activated via *pic* and *nic*. There is at least one logical problem with this, having to do

---

inflection and one for the verbal inflection (as in *sees*), but the former will be determined to be the relevant one on the basis of the fact that the stem is a noun rather than a verb.

with the influence of morpheme frequency on polymorphemic word recognition.

Experiments have demonstrated that polymorphemic words containing high frequency stems are easier to recognize than those containing low frequency stems (e.g., Andrews, 1986; Burani, Salmaso, & Caramazza, 1984; Taft & Forster, 1976; Taft, Huang, & Zhu, 1994; Zhang & Peng, 1992). In terms of the current framework, recognition of the word *fireman* would not only be influenced by the weighting on the links that connect the form units for *fire* and *man* with the lemma for *fireman* (i.e., based on the frequency of *fireman*), but also by the ease with which the form units can be activated in the first place (i.e., based on the frequency of the individual constituents). The latter gives the morpheme frequency effect.

The problem arises, however, when a constituent of the polymorphemic word has more than one syllable. Take *policeman* for example. Activation of the lemma for this word will emanate directly from the form units for *pol*, *ice,* and *man* (assuming a BOSS analysis). Thus, it is the frequency of *pol* (and *ice* and *man*) that will affect the ease of access to *policeman*, and not the frequency of *police*. The frequency of *pol* does incorporate the frequency of *police*, but also that of *policy*, *polite*, *politics*, *polish*, etc. So, the prediction from the model is that the morpheme frequency effect will only hold for monosyllabic morphemes, and this is highly unlikely to be true.

One way to avoid this situation would be to incorporate a hierarchy of units at the form level, with representations for polysyllabic monomorphemic words in addition to their syllable representations. That is, there would exist a form unit for *police* that is activated through the form units for *pol* and *ice* and, therefore, the lemma for *policeman* would be activated via the units for *police* and *man*. In this way, the frequency of *police* can have an impact on the recognition of *policeman*.

What is being suggested by this modification, then, is that the form units that interface with the lemma units are always morphemic in nature and never purely syllabic. It also means that the form units for *cart* and *ridge* would together activate the form unit for *cartridge*, but also each independently activate their own lemmas. This is different to the situation with transparent complex words, like *fireman*, which would not have their own whole-word form unit.

By incorporating whole word form units, however, the issue of redundancy arises.

There is actually no apparent processing advantage for the system to include such units given that the lemma for the polysyllabic word can be effectively activated without it (i.e., directly from the syllable units). While it might be argued that redundancy is not necessarily a bad thing (e.g., Sandra, 1994), there is an alternative modification that avoids it. In particular, one can suggest that it is the lemma units that are hierarchically structured, and this is illustrated in Fig 6.

Fig 6 about here

It can be seen from the figure that the lemma for a polymorphemic word is activated via the lemmas for its constituent morphemes. In this way, when a polymorphemic word is presented, the lemma for a constituent morpheme is not activated merely in parallel with the lemma for the whole word, but as a crucial step in activating the lemma for the whole word. On the other hand, activation of the lemma for a polysyllabic word (e.g., *police* or *cartridge*) does not involve any lemma that might be activated by its constituent syllables (e.g., the *ice* of *police*, or the *cart* and *ridge* of *cartridge*).

While one might expect the frequency of use of a form unit to have some impact on recognition times, the morpheme frequency effect will primarily arise in this hierarchical model from the lemma level. That is, recognition of *fireman* will be influenced by the ease with which the lemma units for *fire* and *man* are activated (as determined by their frequency), and recognition of *policeman* will be influenced by the activation of the lemma units for *police* and *man*.

Previously it was suggested that lemma processing is influenced by the precision of the form-meaning correlation. How is this compatible with the idea that lemma activation is also influenced by frequency? There are various possibilities. One is that both contribute additively to the "strength" of the lemma and, therefore, only frequency effects will emerge if lemma precision is matched (as might be the case, on average, for two randomly chosen free morphemes). A second possibility is that frequency and precision are confounded for free-standing words in that the more commonly a word occurs, the greater the variety of contexts in which it is likely to appear. This, in turn, would provide more information about which semantic features are consistently

associated with the word form and which vary with context.

Another alternative is to suggest that the precision of a lemma has its effects in terms of its activation threshold, whereas frequency has its effects in terms of weightings on the links between units. For example, regardless of how precise the meaning of *fire* or *man* is, activation of their lemmas will be influenced by the weighting on the connections from their form representations to their respective lemmas. Furthermore, the frequency of *fireman* will also affect recognition responses, regardless of the precision of its lemma, as a result of the weightings on the connections to it from the *fire* and *man* lemmas. That is, the connection between *fire* and *fireman* will be stronger than that between *fire* and *firewall* because the former is more frequently used.

Note that by postulating a hierarchy of lemmas, there is the opportunity to explain the results of experiments that fail to observe morpheme frequency effects under certain conditions (e.g., Bertram, Schreuder, & Baayen, 2000). In particular, there is the possibility to introduce a second route to the lemma for a polymorphemic word, namely, one that goes directly from the form level units to that lemma (as in Fig 4). In other words, a combination of Fig 4 and Fig 6 can be suggested whereby there are two pathways to the recognition of a polymorphemic word: Via the lemmas for the constituent morphemes (giving morpheme frequency effects) and directly from the relevant form units (giving form frequency effects). Form frequency may or may not be the same as morpheme frequency depending on what other units are activated by the same form, but they will not be as marked as the effects obtained via the hierarchical route, which arise from activation at both the form and lemma levels. Whether or not strong morpheme frequency effects are observed will therefore depend on which pathway succeeds in activating the whole word lemma first, and this could depend on a number of different factors (see e.g., Bertram et al, 2000).

What is being suggested here, then, is that results that appear to reflect whole-word processing rather than morphological decomposition do not necessarily require the postulation of a whole-word form representation in addition to the morphemic representations (e.g., Bertram et al, 2000; Schreuder et al, *this volume*). Instead, it might be possible to explain such results within the suggested framework whereby form units are always smaller than the whole word when that word is polymorphemic, or even just polysyllabic. Whole word information is found at the lemma level and it is

here that "whole-word" effects are potentially explained. If there is a "race" between whole-word processing and morphological decomposition, the suggestion being made is that it takes place within the lemma system rather than between different sized form units.

The involvement of phonology

The correlation between form and meaning has been talked about here in terms of orthographic form. The reason for this was because the focus has been upon visual word recognition. However, speech is the fundamental form of language and, therefore, the correlation between spoken form and meaning may well be more important in setting up the lexical processing system than the correlation between visual form and meaning. It is possible that lemmas are developed solely on the basis of exposure to the spoken word and that orthographic forms are merely appended to these. However, it is also likely that the introduction of orthography actually modifies the system in some way.

This is particularly clear in the case of Chinese where there is an enormous amount of homophony at the syllable level. In other words, a spoken syllable might be associated with a wide range of meanings because a number of different morphemes are pronounced in the same way. However, much of this ambiguity is resolved by the introduction of orthographic information, because the different homophonic morphemes are represented by different characters. So, the involvement of orthographic information serves to greatly increase the correlation between form and meaning in Chinese.

In an alphabetic script it is possible for orthographic information to similarly increase the correlation between form and meaning. The internal form analysis of a visually presented word is not forced upon the language user by the stimulus itself, unlike the spoken word, which physically defines its syllabic structure. For this reason, a morpheme may be disguised by the syllable structure or phonetic realization of a spoken word, but orthographically available in the visually presented version of the word. For example, the morpheme *act* does not form a unit in the spoken version of *actor*, *active*, or *action,* where the first unit is the syllable /æk/. Obviously, any correlation been form and meaning will be lower if it were based on /æk/ rather than

/ækt/, since the former will include unrelated words like *accurate*, *acne*, *accent*, *accident*, and *axis*. The orthographic version of the word, on the other hand, invites the optimal correlation between form and meaning because it allows the consistent presence of the letter cluster *act* to be detected in the semantically related words.

So, the suggestion is that phonological and orthographic information both participate in the correlation with meaning, leading to a single lemma that represents the linkage between all three types of information. This means that the pathway from orthography to phonology, and vice versa, is mediated by a lemma unit, though it may also be that there are additional direct linkages between the two modalities.

Conclusions

The aim of this chapter has been to consider the implications of an attempt to incorporate polysyllabic and polymorphemic words into a lexical processing system that captures the correlations between form and function. It is suggested that morphemic information is primarily captured in lexical memory in the guise of units (i.e., lemmas) that mediate between form and meaning, with lemma distinctiveness depending on the magnitude of this correlation. The range of lemma distinctiveness goes from free-standing morphemes (e.g., *fire*) and polymorphemic words (e.g., *fireman*) that have a constant meaning within a huge number of contexts, through to bound morphemes that can only be found within a single context (e.g., *hench*). In between are cases that occur in many contexts (e.g., *smith*), but not enough to be clearly considered a free morpheme. Such a view was supported by experimental data obtained with Chinese, which were claimed to be generalizable across all languages.

Clearly, there are a number of issues that need to be explored further, like the syllabic nature of the form units, the mechanisms underlying morphemic frequency effects, and the relationship between orthography and phonology. However, what is provided here is another stage in the evolution of a framework that might prove optimal in understanding morphological representation and processing.

References

Andrews, S. (1986). Morphological influences on lexical access: Lexical or nonlexical effects? *Journal of Memory and Language, 25*, 726-740.

Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language, 37*, 94-117.

Bertram, R., Schreuder, R, & Baayen, R. H. (2000). The balance of storage and computation in morphological processing: The role of word formation type, affixal homophony, and productivity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *26*, 489-511.

Bleasdale, F. A. (1987). Concreteness-dependent associative priming: Separate lexical organization for concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 582-594.

Burani, C., Salmaso, D., & Caramazza, A. (1984). Morphological structure and lexical access. *Visible Language, 18*, 342-352.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100*, 589-608.

Emmorey, K. (1989). Auditory morphological priming in the lexicon. *Language and Cognitive Processes, 4*, 73-92.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103*, 518-565.

Libben, G. (1994). How is morphological decomposition achieved? *Language and Cognitive Processes, 9*, 369-391.

Lima, S. D., & Pollatsek, A. (1983). Lexical access via an orthographic code? The Basic Orthographic Syllabic (BOSS) reconsidered. *Journal of Verbal Learning and Verbal Behavior, 22*, 310-332.

Marslen-Wilson, W., Tyler, L.K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review*, *101*, 3-33.

Modern Chinese frequency dictionary [Xianda huayu pinlu zidian] (1985). Beijing, China: Beijing Language Institute Press.

Norris, D. (1994). A quantitative multiple-levels model of reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 1212-1232.

Page, M. (1999). Connectionist modelling in psychology: A localist manifesto. *Behavioral and Brain Sciences, 23*.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*, 56-115.

Sandra, D. (1990). On the representation and processing of compound words: Automatic access to constituent morphemes does not occur. *Quarterly Journal of Experimental Psychology*, *42A*, 529-567.

Sandra, D. (1994). The morphology of the mental lexicon: Word structure viewed from a psycholinguistic perspective. *Language and Cognitive Processes*, *9*, 227-269.

Schreuder, R., & Baayen, R. H. (1995). Modelling morphological processing. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 131-154). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schreuder, R., Burani, C., & Baayen, R. H. (*this volume*) Parsing and semantic opacity.

Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory and Cognition, 9*, 82-102.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*, 523-568.

Taft, M. (1979). Lexical access via an orthographic code: The Basic Orthographic

Syllabic Structure (BOSS). *Journal of Verbal Learning and Verbal Behavior, 18*, 21-39.

Taft, M. (1987). Morphographic processing. The BOSS re-emerges. In M. Coltheart (Ed.), *Attention and performance, XII* . London: Lawrence Erlbaum Associates Limited.

Taft, M. (1991). *Reading and the mental lexicon*. Hove, UK: Erlbaum.

Taft, M. (1994). Interactive-activation as a framework for undertanding morphological processing. *Language and Cognitive Processes, 9*, 271-294.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior, 14*, 638-647.

Taft, M., & Forster, K. I. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior, 15*, 607-620.

Taft, M., Huang, J., & Zhu, X. (1994). The influence of character frequency on word recognition responses in Chinese. In H. W. Chang, J.-T. Huang, C.-W. Hue, & O. Tzeng (Eds.), *Advances in the study of Chinese language processing* (Vol. 1, pp. 59-73). Taipei: National Taiwan University.

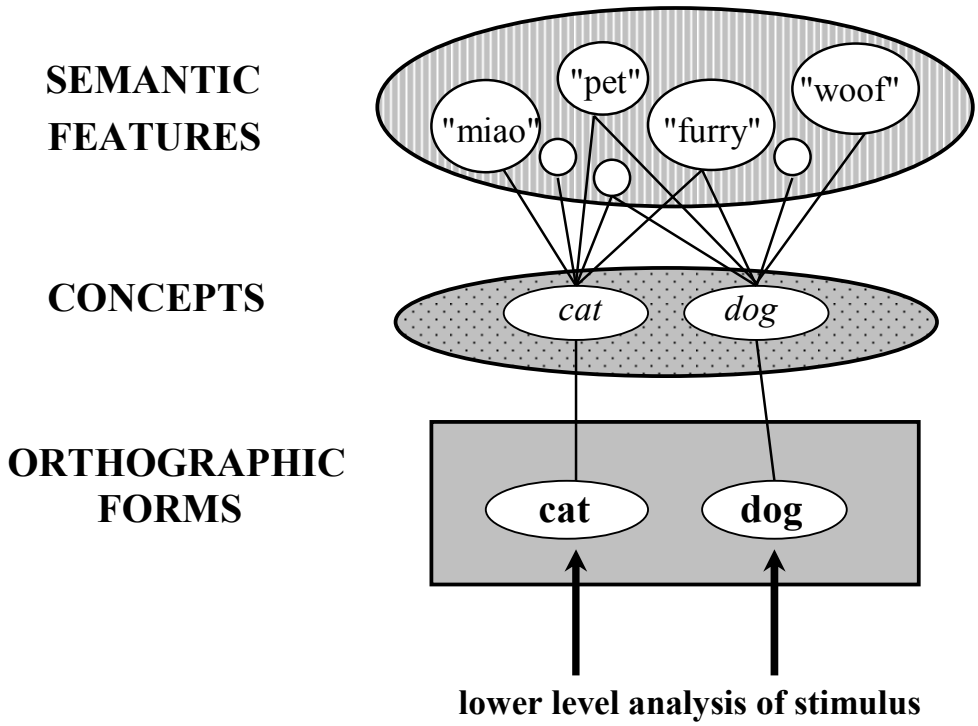
Taft, M., Liu, Y., & Zhu, X. (1999). Morphemic processing in reading Chinese. In A. Inhoff, J. Wang, & H.-C. Chen (Eds.), *Reading Chinese Script: A cognitive analysis* . New Jersey: Lawrence Erlbaum Associates.

Taft, M., & Zhu, X. (1997). Using masked priming to examine lexical storage of Chinese compound words. In H.-C. Chen (Ed.), *The cognitive processing of Chinese and related Asian languages* (pp. 233-241): Chinese University Press.

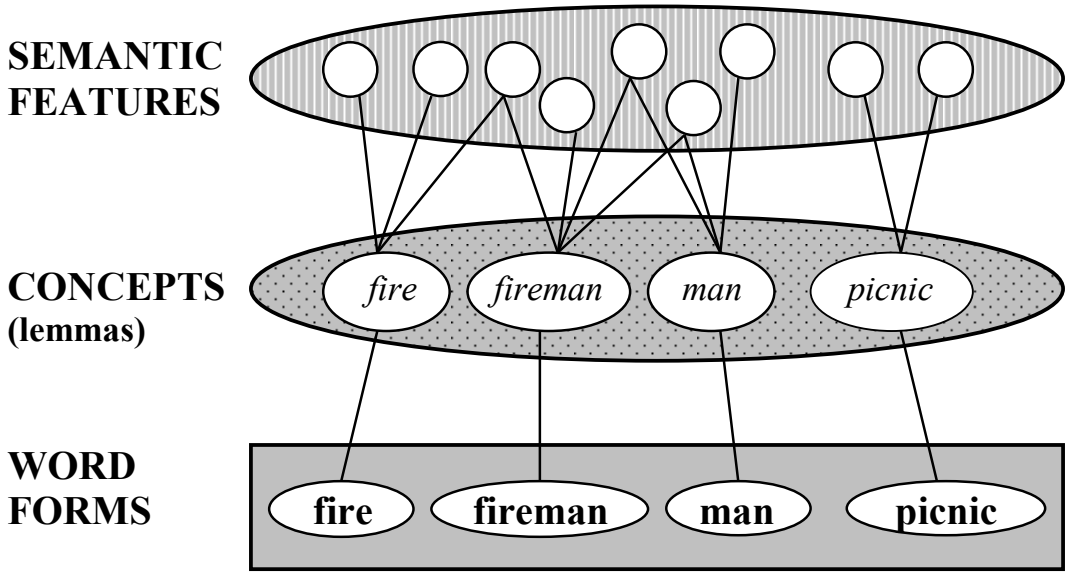
Zhang, B. Y., & Peng, D. L. (1992). Decomposed storage in the Chinese lexicon. In H.-C. Chen & O. J. L. Tzeng (Eds.), *Language processing in Chinese* (pp. 131-149). Amsterdam: North-Holland.

Zwitserlood, P. (1994). The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes, 9*, 341-368.

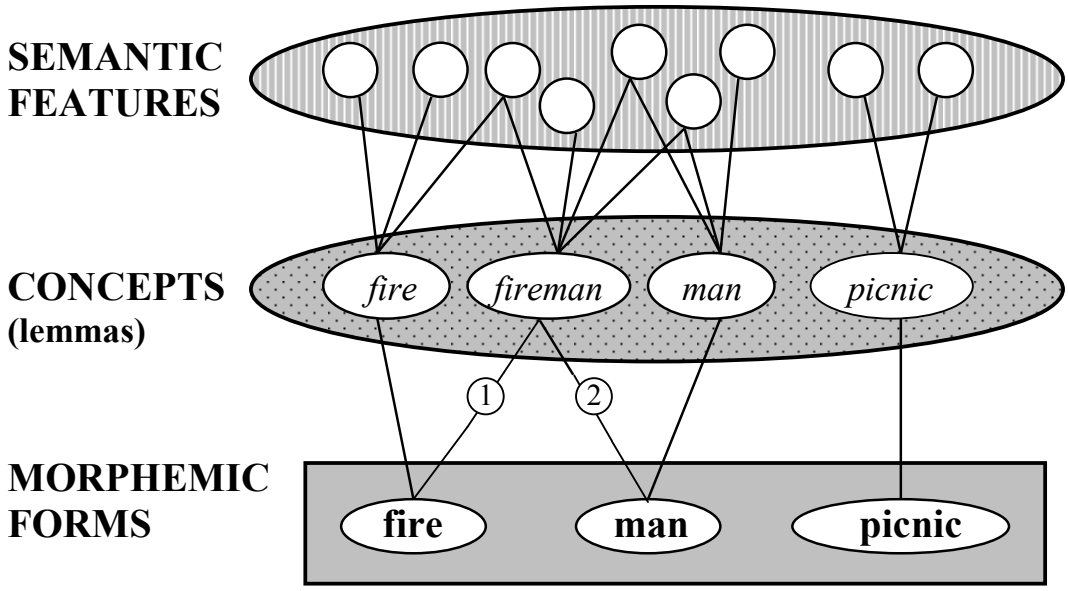
(Fig 1)


SEMANTIC FEATURES
"miao"
"pet"
"furry"
"woof"
CONCEPTS
cat
dog
ORTHOGRAPHIC FORMS
cat
dog
lower level analysis of stimulus

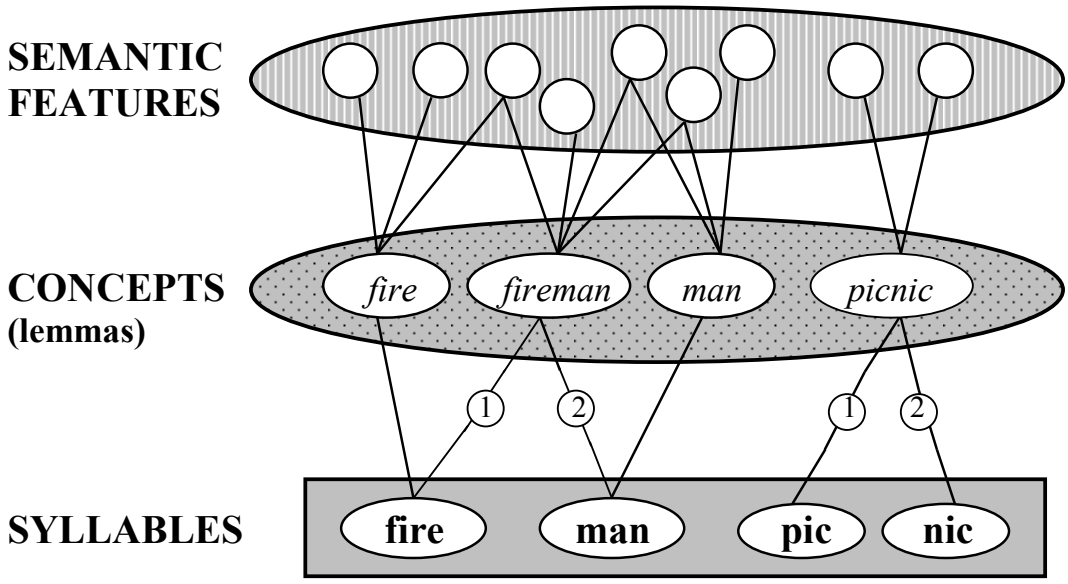
(Fig 2a)


SEMANTIC FEATURES
CONCEPTS (lemmas)
fire
fireman
man
picnic
WORD FORMS
fire
fireman
man
picnic

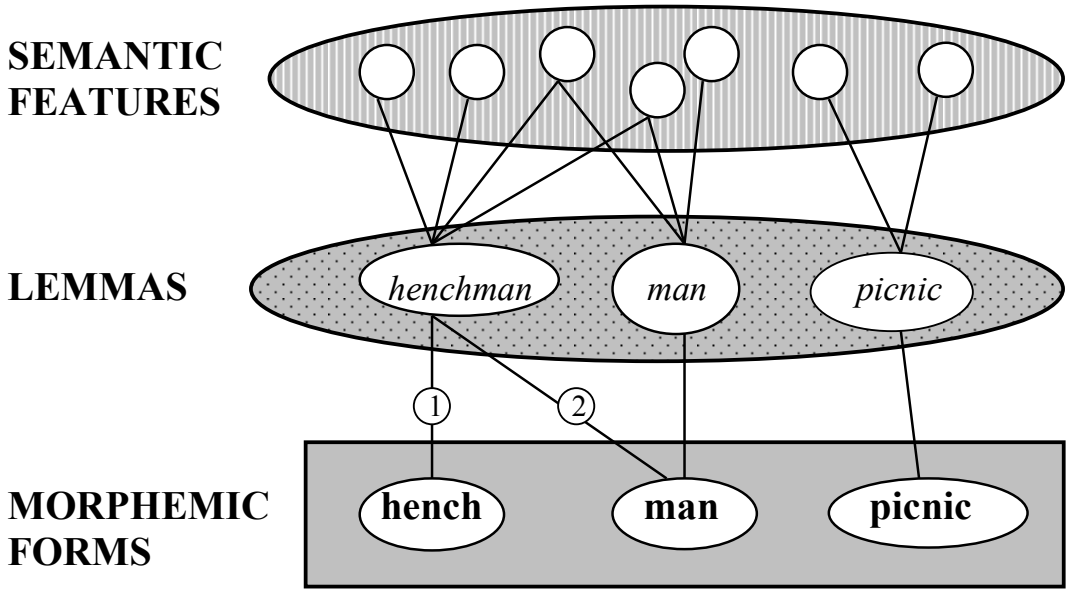
(Fig 2b)


SEMANTIC FEATURES
CONCEPTS (lemmas)
fire
fireman
man
picnic
1
2
MORPHEMIC FORMS
fire
man
picnic

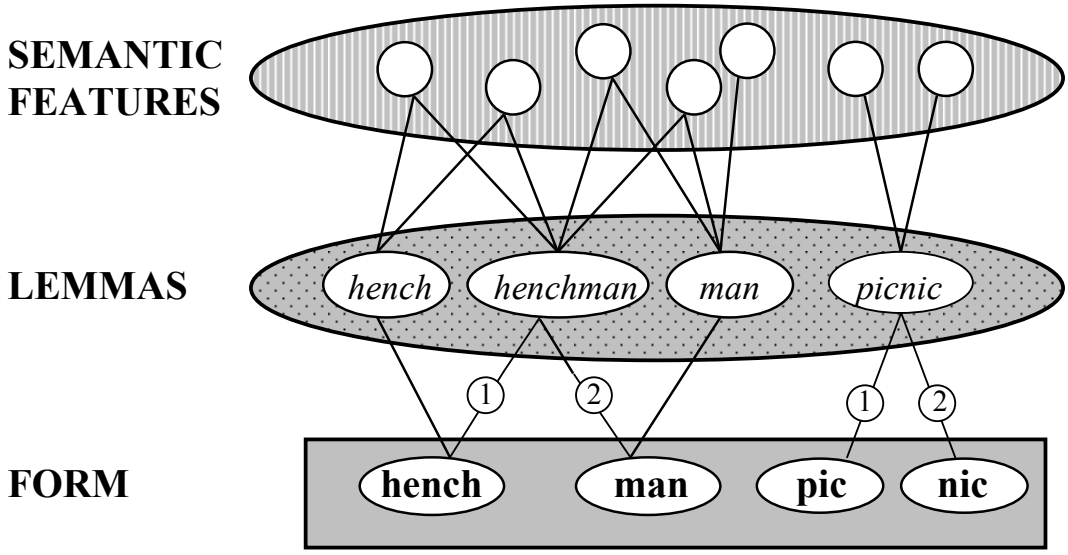
(Fig 2c)


SEMANTIC FEATURES
CONCEPTS (lemmas)
fire
fireman
man
picnic
1
2
1
2
SYLLABLES
fire
man
pic
nic

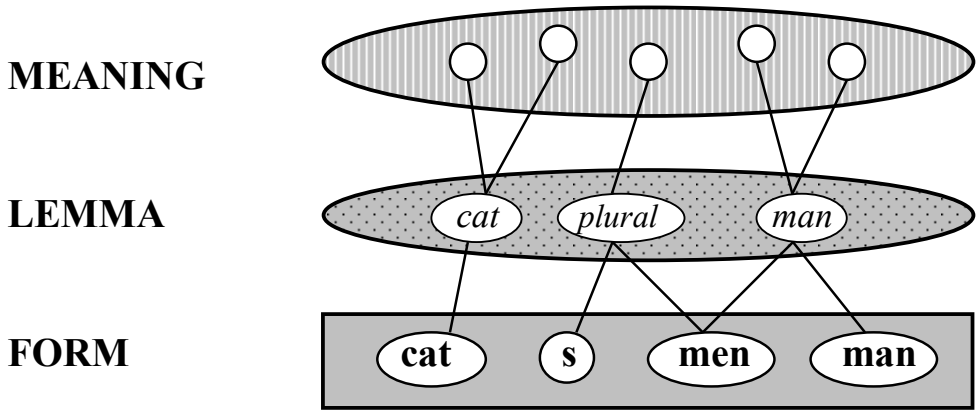
(Fig 3)


SEMANTIC FEATURES
LEMMAS
henchman
man
picnic
1
2
MORPHEMIC FORMS
hench
man
picnic

(Fig 4)


SEMANTIC FEATURES
LEMMAS
hench
henchman
man
picnic
1
2
1
2
FORM
hench
man
pic
nic

(Fig 5)



(Fig 6)

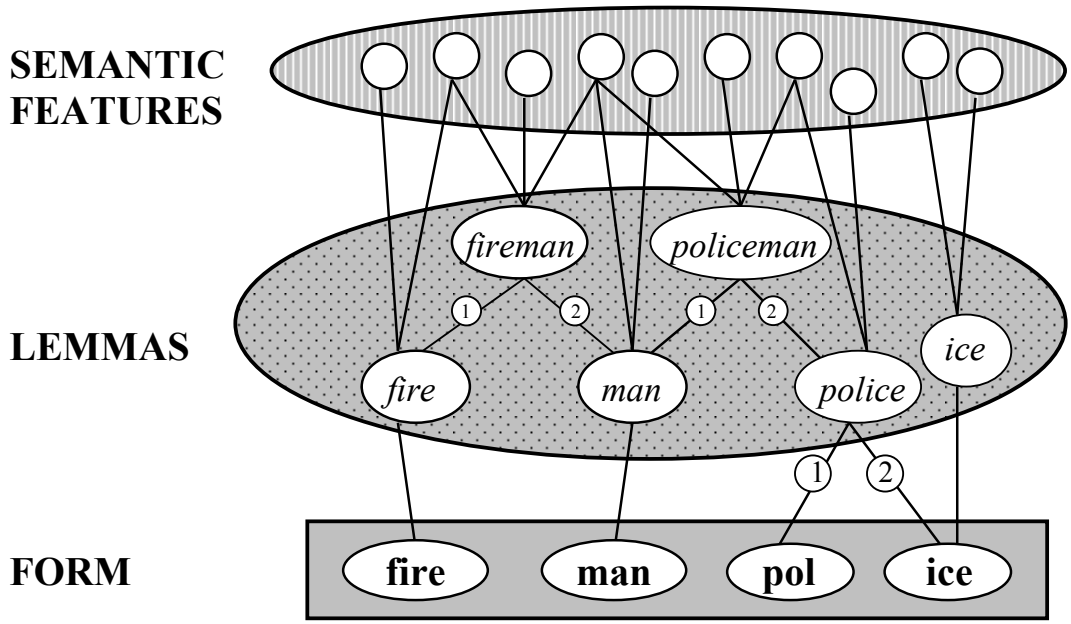

SEMANTIC FEATURES
LEMMAS
FORM
fireman
policeman
fire
man
police
ice
1
2
1
2
1
2
fire
man
pol
ice

Figure headings

Fig 1. A depiction of the connection between form and meaning in lexical memory.

Fig 2a. The representation of polymorphemic and polysyllabic words when the orthographic form level represents whole words.

Fig 2b. The representation of polymorphemic and polysyllabic words when the orthographic form level represents morphemes.

Fig 2c. The representation of polymorphemic and polysyllabic words when the orthographic form level represents syllables.

Fig 3. The representation of polymorphemic and polysyllabic words when the orthographic form level represents morphemes and bound morphemes have no lemma-level representation.

Fig 4. The representation of polymorphemic and polysyllabic words when the orthographic form level represents both syllables and morphemes, and bound morphemes have a lemma-level representation.

Fig 5. The representation of regularly and irregularly inflected words.

Fig 6. The representation of polymorphemic and polysyllabic words when lemmas are hierarchically represented.