Modality Specificities in Lexical Architecture?

William D. Marslen-Wilson, Mike Ford, & Xiaolin Zhou
Centre for Speech and Language,
Psychology Department, Birkbeck College,
Malet Street, London WC1E 7HX, England
{w.marslen-wilson, m.ford, x.zhou}@psyc.bbk.ac.uk

Abstract

This paper argues for asymmetries in lexical architecture and function, based on a series of repetition priming experiments examining the representation and access of morphologically complex forms in English. These results point to modality differences in representation at the level of the lexical entry, and to marked differences in access from speech and from text. We argue that speech inputs can map directly onto abstract morphemic representations, while input from text seems to involve mediated access, via intervening orthographic representations of word form.

Introduction

In thinking about how the human language system is organised to mediate the relationship between internally represented knowledge and the input and output systems dedicated to the access and use of this knowledge, it is natural to assume that the system has a considerable degree of functional and architectural symmetry. Current models of the organisation of the mental lexicon (e.g., Miceli, 1994; Seidenberg, 1995) normally assume a modality-independent store of lexical content, and parallel sets of input lexica for the two principal modalities (speech and vision).

There are two principal claims being made here. The first is *architectural*: that the basic organisation of the language system distinguishes modality-independent representations (the common core of word meaning) from modality-specific but architecturally parallel access systems. The second is *functional*: that there are basic functional parallelisms between the two modalities, so that lexical representations are accessed by qualitatively similar procedures, irrespective of the modalities involved.

We believe both these claims are wrong, and that access from the auditory, or phonological, route is neither architecturally nor functionally equivalent to access from the visual, or orthographic, route. There are three components to our claims here:

(i) That central representations are not, strictly speaking, modality independent. Rather, we are dealing here with a level of representation we can call the *lexical entry*, where this is a triplet of abstract semantic, syntactic, and phonological information.

(ii) That phonological inputs can map directly onto these underlying, central representations, without the need for intermediate access representations. In fact, as we will argue, this involves *direct access* to decomposed morphemically organised representations.

(iii) That orthographic inputs do not have the same kind of direct access to the lexical entry, and access may well involve some form of *mediating* representation.

We will organise the argument for this view around the evidence for these last two claims, summarizing the evidence for direct auditory lexical access, and then going on to examine parallel sets of experiments in the visual domain, revealing an input system with quite different properties. In all of these experiments we rely on morphologically complex words. To make explicit the internal properties of the English mental lexicon, as an abstract, combinatorial, and morphemically organised system, it is necessary to use stimulus materials that allow us to dissociate morphemes from words, and surface form from underlying representation.

Direct Access From Speech

The first part of our argument is that phonological inputs can map directly onto morphemically organised representations at the level of the lexical entry, where these representations include an abstract specification of the phonological properties of the words involved, and where the term 'phonological input' refers to the acoustic-phonetic analysis of the speech input stream, most likely delivered to the lexical level in featural form (for a closer look at these issues, see Marslen-Wilson & Warren, 1994). The evidence for direct mapping comes, in the first instance, from experiments involving morphologically related words where the relationship between these words varies in its phonological transparency.

In our original experiments (Marslen-Wilson, Tyler, Waksler, and Older, 1994) we used a cross-modal version of the priming task, where a spoken prime word - for example, *friendly* - is immediately followed by a visually presented target word - for example, *friend* - to which the listener has to make a speeded lexical decision response (that is, decide as quickly as possible whether the target item is a word or not). We find in this task that response times are generally

faster when the target word is lexically related to the prime word, allowing us to use the task as an index of processing and representational relationships in the mental lexicon.

Most significantly, we found equally strong priming between morphologically related pairs (+Morph), irrespective of variations in the phonological transparency of the relationship between the stem and the derived form. Priming was no stronger for prime/target pairs like *friendly/friend*, where the relationship is transparent, than it was for pairs like *elusive/elude* or *decision/decide*, where the relationship is more phonologically opaque

Our preliminary interpretation of these results (Marslen-Wilson et al., 1994) was in terms of abstract underlying representations which were *underspecified* for the feature dimension that was varying. In earlier work (Lahiri & Marslen-Wilson, 1991) we had made proposals of this sort to explain phonological variation in morphologically simple words, focusing on the perceptual consequences of feature-spreading processes such as vowel nasalisation and place assimilation. Subsequently we expanded this kind of analysis to explain cases of allomorphic variation, where the phonological changes in the form of a stem are morphologically triggered.

The cross-modal results on their own, however, are not sufficient to exclude an alternative account in terms of a mediated access system. In particular, since priming in this task falls at the level of the lexical entry, we cannot exclude the possibility that inputs in different modalities can access this level via modality-specific access representations, in both visual and auditory input domains. In subsequent research, therefore, we tested the same stimuli in an intra-modal, auditory/auditory version of the task, where both prime and target are auditorily presented (Marslen-Wilson & Zhou, 1993; 1996b). This is again immediate repetition priming, where the target, now spoken, follows the prime word with a minimal interval of 150 msec.

If anything, the pattern of priming across conditions varying in phonological transparency was stronger and more stable than for the cross-modal study. Priming was just as strong for phonologically opaque conditions (*decision/decide*) as it was for the transparent conditions (*friendly/friend*), whereas no priming was found between prime/target pairs, such as *bulletin/bullet* or *forty/fort*, where prime and target were phonologically but not lexically related.

This is strong evidence both for an account of repetition priming in terms of lexical relations between primes and targets, and for a direct access account of lexical access in the auditory domain, where forms like *sanity* prime their stems because the underlying morpheme, at the level of the lexical entry, abstracts away from regular allomorphic variation in the surface form of the stem. It is hard to see how alternative views, where access is mediated *via* whole-word access representations of allomorphic derived forms, could explain the equal strength of priming across the (+Morph) conditions.

Suffix-Suffix Interference and Direct Lexical Access

A second kind of evidence for direct access in the auditory domain comes from the phenomenon of *suffix-suffix interference*. This is the finding, originally established in cross-modal priming (Marslen-Wilson et al., 1994), that two suffixed forms transparently derived from the same stem, as in pairs like *excitable/excitement*, do not prime each other, even though they are strongly semantically related, and despite the facilitatory effects of sharing the same underlying stem. We explain this in terms of competitive interference between suffixes attached to the same stem. When a particular combination of stem and suffix is heard - for example, *excitement* - this seems to inhibit temporarily the combination of this stem with any other derivational suffix. Thus, when a related suffixed form (such as *excitable*) immediately follows, processing and recognition of this form will be slowed.

This interference effect can be exploited as an index of the architectural properties of underlying representations, allowing a strong test of our claims about direct lexical access and the processing and treatment of allomorphic forms. We argue that the effect is specifically due to competition between suffixes locally attached to the same underlying stem morpheme. On the direct access model, this is true not only for non-allomorphic stems (like excitement/excitable), but also, crucially, for allomorphic pairs like *sanely/sanity* or *decide/decision*. The derivational suffixes {-ly} and {-ity} attach to the underlying morpheme {sane}, despite the allomorphy in the phonetic expression of these forms. This predicts that we should find reduced priming for pairs like *sanely/sanity*, just as we did for transparent pairs like *excitement/excitable*. As either *sanely* or *sanity* is heard, this will activate the underlying morpheme {sane} and the derivational suffixes attached to it. The subsequent choice of one of these suffixes will inhibit later processing of other derived suffixed forms sharing the same stem, irrespective of their surface allomorphy.

A further experiment, therefore, directly tested this prediction in an auditory-auditory priming task (Marslen-Wilson & Zhou, 1993; 1996b). The crucial comparisons here were between non-allomorphic (e.g., *abruptly/abruptness*) and allomorphic (e.g., *sincerely/sincerity*) derived/derived pairs. We compared these with other conditions where we expected priming to occur - derived/stem and stem/derived pairs such as *attractive/attract* and *calm/calmness* - and with matched phonological control conditions, containing pairs like *treaty/treatment* and *apartment/apart*, where we expected no priming. The results were very clear. There was strong priming in the two stem conditions, averaging 47 msec, no priming in the phonological control conditions (averaging 8 msec), and no priming in either of the derived/derived conditions. Priming averaged 2 msec in the non-allophonic condition, and 14 msec in the allophonic condition. This is despite the fact that the semantic relatedness of the primes and targets is just as great for the

derived/derived pairs as it for the derived/stem and stem/derived pairs.

These results support our analysis of auditory lexical access in two ways. First, the replication, for non-allomorphic pairs like *abruptly/abruptness*, of the suffix interference effect previously found in cross-modal priming (Marslen-Wilson et al, 1994), is renewed evidence for the convergence of the two complex forms onto the same underlying unit - unless the {abrupt} in *abruptly* is the same as the {abrupt} in *abruptness*, there is no reason to expect any interference. Secondly, the presence of an interference effect for allomorphic pairs as well is evidence for the more radical hypothesis that the speech input projects directly onto a level of phonological representation which abstracts away from surface variation in phonetic form. Again, unless the same underlying lexical unit is involved in processing of both prime and target, there is no reason to expect interference.

In a final sequence of experiments in the auditory modality, we used a delayed repetition priming task to determine the time-course - and general robustness - of the suffix interference effect (Marslen-Wilson & Zhou, 1996a). Delayed (as opposed to immediate) repetition is a task where the stimulus words are not presented in prime/target pairs, but individually in a series, with one or more other word or nonword intervening between the presentation of prime and target. The subject is required to make a lexical decision to every item heard. We ran two versions of the task, one with a short delay (one intervening item) of about 6 seconds between prime and target, and one with a longer delay (eight intervening items) of about 25 seconds. If the interference effect was the result of some sort of inhibitory influence between competing lexical elements, then we should expect it to diminish over time, along with other transient effects in lexical access.

The outcome was straightforward. Although there was an overall drop in stem priming, to 30 msec in the eight lag condition, the interference effect remained remarkably constant over time. At zero lag the difference between the stem and derived condition averaged 50 msec and at a lag of eight intervening items the difference was unchanged, at 49 msec. These patterns are reflected in the analyses of variance on these data, which show no interactions of lag with stimulus type.

Mediated Access From Text?

We turn now to an examination of the properties of the visual, orthographically based access system. We do so armed with two distinctive patterns of responses in the auditory domain. The first is the insensitivity of morphological priming to variation in the form similarity of prime and target, and the second is the suffix-suffix interference effect, which we take as an index of structural and processing relations in a decomposed abstract lexicon. We consider first how the system responds to visually presented prime/target pairs that vary along parallel lines to the materials used in earlier experiments to evaluate phonological overlap (Marslen-Wilson & Zhou, 1996a).

These new materials, as illustrated in Table 1, hold phonological overlap constant and vary orthographic deviation over three levels, for both morphologically related [+Morph, +Sem] and unrelated [-Morph, -Sem] stimulus sets. In the Zero [+Orth] condition, the target is orthographically completely contained in the target; in the Minimum [-Orth] condition it deviates by one letter only (usually the final e, as in pairs like *criminal/crime*), and in the Maximum [-Orth] condition it deviates by an average of two letters.[1] These materials were run in a visual-visual immediate repetition task, where the prime word, in lower case letters, was exposed for 200 msec, followed by the target (in upper case) for 500 msec. The subject's task was to make a lexical decision response to the target word.

Condition		**Prime**	**Target**
1. [+M +O -P]	Zero	pressure	PRESS
2. [+M -O -P]	Min	sincerity	SINCERE
3. [+M -O -P]	Max	collision	COLLIDE
4. [-M +O -P]	Zero	mission	MISS
5. [-M -O -P]	Min	paradise	PARADE
6 [-M -O -P]	Max	agitate	AGILE

Table 1: Co-varying orthographic transparency (±O) and morphological relatedness (±M): Sample stimuli.

The results show a marked sensitivity to variation in surface form, with priming dropping off linearly across the three divergence conditions, but only for the [+Morph] stimuli. Priming is a non-significant 12 msec for the [+Morph] Max condition, increasing to a marginally significant 27 msec for the Min condition, and to a robust 47 msec in the Zero divergence condition. The [-Morph] conditions, in contrast, show no priming in any condition, even when the target is fully contained in the target, as in *mission/miss* or *flagrant/flag*.

These findings clearly point to a different relationship between the surface form of prime and target than we saw in the previous experiments. Auditory-visual and auditory-auditory repetition priming show quite different patterns, with no effects of allomorphic variation on the effectiveness of morphologically related primes. This suggests that the representation of a visually presented prime has significantly different properties to those of an auditorily presented prime, and that, furthermore, priming in the visual domain is not

[1] We label this the "Maximum" condition because we could find very few cases of derivationally related prime/target pairs where the orthographic divergence was more than two letters.

dependent in the same way as priming in the auditory domain on repeated access to the same abstract lexical elements.

This is not to say that priming here does not involve lexical representations. It clearly does, since it is only in the [+Morph] conditions that we get any priming. But to explain why *pressure* primes *press* but *collision* does not prime *collide*, we have to assume that the level of representation involved is specifically orthographic in nature, and does not engage either the abstract morphological or the phonological aspects of the words involved. From both these perspectives the stimuli were equivalent across the three conditions, and should have primed equally well. The fact that they did not points, in fact, towards some form of lexically structured orthographic mediating representation, possibly corresponding to an orthographic "input lexicon" of the classical sort. If so, we can probe its properties by looking at suffix-suffix interference in the visual domain. This reflects, we believe, the consequences of two affixes trying to combine with the same stem. If the interpretation of visually presented prime words involves the same kinds of processes, then we should see the same interference effects here as well.

Condition	Prime	Target
STEM [+Morph, +Sem, +Phon]	delightful	DELIGHT
DERIVED [+Morph, +Sem, +Phon]	abruptness	ABRUPTLY
PHON [+Morph, -Sem, +Phon]	organize	ORGAN
SEMANTIC [-Morph, +Sem, -Phon]	cello	VIOLIN

Table 2: Sample stimuli for repetition priming

In a series of three experiments (Marslen-Wilson & Zhou, 1996a) we followed the same strategy as before, running the stimuli from the earlier auditory-auditory studies in immediate and delayed repetition visual-visual priming tasks. An example set of these stimuli is given in Table 2. The critical comparison is between the STEM condition, with derived/stem pairs, and the DERIVED condition, with derived/derived pairs. In addition, we include a phonological control condition (PHON), and a SEMANTIC condition, where prime/target pairs are purely semantically (and associatively related). Again we contrasted a Zero lag condition, corresponding to the immediate repetition study, with Lag 1 and Lag 8 conditions, reflecting short and long repetition delays.

The results were straightforward. There was no sign of the interference effects that dominate responses in the auditory domain. Stem priming was significant at each lag, averaging 31 msec overall. But at no point did it differ from priming in the derived-derived condition, which is significant at Lag 1 and Lag 8, and which averages 32 msec overall. Even the phonological control condition, bumping along at an average effect of 14 msec, is significant overall. The differences between performance here and performance in auditory-auditory priming come through very clearly in an overall analysis of the two sets of data, where we see a significant interaction between modality (auditory vs. visual) and target type.

Condition	AUD/AUD	VIS/VIS
STEM	40.8	31.2
DERIVED	-14.7	32.5
PHON	3.2	14.5

Table 3: Mean priming effects (collapsing across lags) for Auditory-Auditory and Visual-Visual priming.

This interaction is illustrated in Table 3, which summarises the suffix-suffix interference results for the two series of auditory-auditory and visual-visual priming experiments (we collapse here over lag, which proved not to play a statistically significant role in either modality). The two modalities are relatively similar in the derived/stem and the phonological control conditions. But we see no interference between visually presented suffixed primes and visually presented suffixed targets sharing the same stems. The presence of this interference in auditory-visual and auditory-auditory priming we take as evidence of combinatorial processes involving stem and affix morphemes. The pronounced absence of these effects in the visual domain suggests that repetition priming here is tapping into a level of representation which is not morphologically decomposed - a level of representation, in fact, which might look very much like the kind of whole-word access representation that is often argued to be the first step in moving from visual input to lexical representation.

Turning, finally, to the associative/semantic pairs that were run both here and in auditory delayed repetition studies described earlier, we find renewed parallels with auditory access processes. Associative priming, which does not involve relations between stimuli with similar surface forms, seems to behave the same way here as it did in auditory-auditory priming. At Lag 1 we get a significant 27 msec effect for pairs like *month/year*, comparable to the 31 msec effect for the auditory studies. But at Lag 8, just as in the auditory experiment (where priming for these pairs dropped to a non-significant 1 msec), the priming between semantic/associative pairs is no longer significant, averaging 9 msec. The differences that are emerging, therefore, between the auditory and the visual access route, are not differences across the board. Under certain conditions priming effects have the same direction and the same time-course. How are we best to characterize this pattern of differences and similarities?

Symmetries and Asymmetries in Lexical Organization

To approach these issues, in the context of the research reviewed here, we need to go back to the first of the three claims we made at the outset of this paper. This was the claim that central lexical representations are not modality independent, and that the core level of representation, the lexical entry, is a tightly linked complex of abstract semantic, syntactic, and phonological properties.[2] This is a view of lexical representation that can be traced back to Saussurian concepts of the linguistic sign, and which reflects more recent treatments where the "lexical sign" is seen as a triplet of phonological, syntactic, and semantic information organised in a hierarchical feature structure (e.g., Pollard & Sag, 1987). In the current psycholinguistic framework, the move to this view of the lexical entry reflects not only the phylogenetic and ontogenetic facts of the matter - that language has evolved, and is normally learnt, as a relation between sound and meaning - but also the pattern of results we observe in auditory-auditory and auditory-visual priming tasks.

What is especially telling is that we see the same pattern, of indifference to allomorphic variation and of sensitivity to competition among suffixes, in auditory-auditory as well as in auditory-visual priming. We originally chose to use the cross-modal auditory-visual task to force the subjects to respond at a level of the system that was supra-modal, and therefore, we assumed, modality independent. We subsequently argued, on the basis of these cross-modal data, that this level was the appropriate domain for processes of morphological parsing, where the syntactic and semantic properties of individual morphemes are available to constrain and guide these processes (Marslen-Wilson et al., 1994). What the results for auditory-auditory priming now suggest is that this level is not, strictly speaking, modality independent, but includes abstract phonological specifications as well.

From this perspective, the significance of the auditory-visual experiments is not that they enable us to tap into a modality independent level of lexical representation, but that they show that visually presented inputs - the visual target words - also map onto this domain. Information extracted from an orthographic input is also interpreted at a level of representation with the abstract, combinatorial properties revealed in the auditory-auditory experiments. When the prime word is auditorily presented, the consequences for processing of subsequent target words apply equally to visual as to auditory inputs. Similarly, when the relationship between prime and target is semantically but not morphologically or linguistically mediated, as in associative/semantic priming, we find equivalent patterns with auditory and visual priming.

But does this mean that lexical access from visual inputs is therefore as direct as access from auditory inputs? There are two arguments that suggest it is not. The first follows from the assumption that the lexical entry - the representation of lexical content - is phonologically specified. This implies that access to lexical content from orthography will always involve some sort of translation process; that learning to read, at least in its initial stages, is learning to relate orthographic form to phonological form, and this transduction process remains an intrinsic part of lexical access from text. But the strength of this argument is hard to evaluate. The role of phonological factors in reading remains a controversial and unresolved issue in the visual word recognition literature (e.g., Seidenberg, 1995), compounded by further uncertainties about the computational properties of the systems involved (e.g., Van Orden, Pennington, & Stone, 1990).

Stronger arguments for mediated access can be made on the basis of the visual-visual priming effects we report. When the prime word is visually presented, then subsequent responses to a visually presented target word seem to engage a stored representation of the prime that is not morphologically decomposed and that is sensitive to orthographic but not phonological overlap between prime and target. This points to an asymmetric lexical architecture, with direct access from speech to the lexical entry, but where orthographic inputs are routed via orthographic access representations, where these are undecomposed whole-word representations, though sufficiently abstract to be indifferent to variations in case.

This kind of arrangement can account straightforwardly for the divergence between visual-visual priming and other types of repetition priming. Because the access representations are orthographic in character, any priming effects based on repeated access to this layer will be affected by orthographic similarities between morphologically related primes and targets. Hence the dominance of orthographic rather than phonological factors in responses to the derived/stem pairs varying in stem allomorphy. And because these are wholeword representations, with separate representation of forms like *excitable* and *excitement*, there are no shared morphological elements, and therefore no basis for interference when different suffixes compete for the same stem.[3]

Conclusions

Our interim conclusion, in summary, is that there are significant modality specificities in lexical architecture and function; that representations at the level of the lexical entry are not fully modality independent, and that this has consequences for the nature of the access routes from speech and from text, with speech inputs being able to project directly onto abstract, morphemically organised underlying representations of the words involved.

We should add two further riders. The first is that these issues of modality specificity and asymmetry will by no

[2] Any genuinely modality-independent aspects of meaning representation we can assign to the domain of conceptual representation. This is not the object of our enquiries here.

[3] For additional discussion of these issues, see Marslen-Wilson, Zhou, & Ford (1996).

means disappear if we approach the problems of lexical representation and processing from a more distributed, connectionist perspective. These approaches may offer quite different solutions to problems of allomorphy and phonological variation (Gaskell, Hare, & Marslen-Wilson, 1995), and may require a reinterpretation of the whole notion of lexical access and representation (Gaskell & Marslen-Wilson, 1995; Seidenberg, 1995). But they will still need to provide an account of how a system which has developed in the context of a phonology to semantics to phonology mapping learns to superimpose on this an orthography to semantics mapping, and to model appropriately the quantitative and qualitative parallels and divergences between the two domains.

Our second comment is that our arguments here should be taken as language specific. Our data and our interpretations are limited, unless proven otherwise, to the English mental lexicon. It is plausible that similar claims can be defended for languages with similar writing systems, and similar relations between phonology and orthography. But we cannot assume that the same asymmetries will hold true for languages with non-alphabetic writing systems, and, indeed, we have some evidence that for Chinese they do not (Zhou & Marslen-Wilson, 1996).

Acknowledgements

This research was supported by grants from the UK Medical Research Council and Economic and Social Research Council. We thank members of the Birkbeck Centre for Speech and Language for many stimulating discussions of morphology and the mental lexicon.

References

Gaskell, G., & Marslen-Wilson, W. (1995). Modeling the perception of spoken words. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the 17th Annual Conference of the Cognitive Science Society* (pp. 19-24). Mahwah, NJ: Erlbaum.

Gaskell, M.G., Hare, M., & Marslen-Wilson, W.D. (1995) A connectionist model of phonological representation in speech perception. *Cognitive Science*, 19, 407-439.

Lahiri, A., & Marslen-Wilson, W.D. (1991) The mental representation of lexical form: A phonological approach to the mental lexicon. *Cognition*, **38**, 245-294.

Marslen-Wilson, W.D., & Zhou, X. (1993) *Auditory morphological priming for English derived words.* Paper presented at Psychonomic Society meeting, Washington, D.C.

Marslen-Wilson, W.D., & Zhou, X. (1996a) *Modality specificity in lexical access?* Paper presented to the Experimental Psychology Society, London, January.

Marslen-Wilson, W.D. & Zhou, X. (1996b) *Abstractness and the access of lexical form.* Manuscript, Center for Speech and Language, Birkbeck College.

Marslen-Wilson, W.D., & Warren, P. Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, **101**, 653-675, 1994.

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

Marslen-Wilson, W.D., Zhou, X-L, & Ford, M. (1996) Morphology, modality, and lexical architecture. *Yearbook of Morphology.*

Miceli, G. (1994) Morphological errors and the representation of morphology in the lexical-semantic system. *Philosophical Transactions of the Royal Society of London B*, 346, 79-88.

Myers, S. (1987) Vowel shortening in English. *Natural Language & Linguistic Theory*, **5**, 487-518.

Pollard, C., & Sag, I.A. (1987) *Information-based syntax and semantics.* Stanford: CSLI.

Seidenberg, M.S. (1995) Visual word recognition: An Overview. In P.Eimas & J.L.Miller (Eds.) *Handbook of perception and cognition: Language.* New York: Academic Press.

Van Orden, G.C., Pennington, B.F., & Stone, G.O. (1990) Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review*, 97, 488-522.

Zhou, X-L., & Marslen-Wilson, W.D. (1996) Direct visual access is the only way to access the Chinese mental lexicon. *Proceedings of the 18th Annual Conference of the Cognitive Science Society.*