Inflectional morphology and phonological regularity in the English mental lexicon

William Marslen-Wilson, Mary Hare, & Lianne Older
Birkbeck College, University of London
Center for Speech and Language
Malet St., London WC1E 7HX
Email: ubjta38@cu.bbk.ac.uk

Abstract

We used a cross-modal repetition priming task to investigate the mental representation of regular and irregular past tense forms in English. Subjects heard a spoken prime (such as *walked*) immediately followed by lexical decision to a visual probe (such as *walk*). We contrasted three types of English verbs, varying in the phonological and morphological regularity of their past tense inflection. These were (i) Regular verbs (*jump/jumped*), with the regular /-d/ inflection and no stem change, (ii) Semi-Weak verbs (*burn/burnt*, *feel/felt*), with irregular alveolar inflection and some phonologically regular stem vowel change, and (iii) Vowel Change verbs (*sing/sang*, *give/gave*), which mark past tense through phonologically irregular changes in the stem vowel. The stem forms of these verbs were presented in three prime conditions -- preceded either by the Identity Prime, a Past Tense Prime, or a Control Prime. The Identity Prime significantly facilitated lexical decision responses for all three verb classes, but the Past Tense Prime, while significantly facilitating responses in the Regular verb class, produced no overall effect for the Semi-Weak verbs, and significant interference for the Vowel Change verbs. We conclude that phonological irregularity in the relation between a stem and an inflected form can lead to very different lexical structures than we find for more regular phonological relationships.

Introduction

English inflectional morphology, and in particular, the study of English verbs and how they form their past tenses, has become the battleground for a series of major controversies in experimental psycholinguistics and in connectionist cognitive modelling.

Due to the accidents of its historical development, English past tenses offer a contrast between regular and irregular inflections. The majority of English verbs form their past tense by adding the *regular* affix *-d* to the stem, which is otherwise unchanged. These are verbs like *jump/jumped*, *agree/agreed*, *seem/seemed*, etc. But about 160 verbs -- many of them among the most common verbs in the language - take some other form of past tense inflection, often of an idiosyncratic nature. These may involve no stem change at all, as in verbs like *hit/hit*, but the majority change the stem in some way, either the vowel (as in *sing/sang* or *drive/drove*), or the final consonant (as in *make/made*), or indeed both (as in *keep/kept*, *feel/felt*).

The current controversies are concerned with how these contrasting properties of the verb inflection system are represented in the adult mental lexicon, and how they are learnt in the process of language acquisition.

The traditional linguistic and psycholinguistic view has been that regular inflected forms are derived from general rules, while unpredictable, irregular forms are stored in the lexicon as a rote list. The focus from the language acquisition side has been on the process whereby the appropriate rule-based generalisations are formed, with the phenomenon of *over-regularisation* playing a special role. This is the widely reported phase in children's acquisition of past-tense inflections where they tend to regularise irregular past tense forms (Brown, 1973; Ervin, 1964; Kuczaj, 1977). The child may, for example, produce the past tense of *hit* as *hitted* (instead of *hit*), of *steal* as *stealed* (instead of *stole*), of *go* as *goed* (instead of *went*), and so on.

Historically, this phenomenon has played a special role in post-Chomskian psycho-linguistics, and, indeed, in the cognitivist revolution of the 1960's and 70's, since it seemed to provide compelling evidence that language acquisition is a process of acquiring rules, demonstrating that people really do have rules in their heads and not just associative chains. The child never hears forms like *stealed* or *hitted*. The only plausible explanation for their appearance seemed to be that the child had learnt the regular past-tense rule, and was now misapplying it to the stems of irregular forms.

This special significance of past-tense acquisition helps to explain the vigorous counter-attack that was mounted against Rumelhart & McClelland's (1986) apparent demonstration that a simple single-layer connectionist network could not only learn the present tense to past tense mapping, but also exhibit over-regularisation behaviours, and all without the benefit of explicit rules; just superimposition of patterns of weights in the network during the learning process. These claims were immediately and extensively disputed by, among others, Pinker and Prince (1988) and Fodor and Pylyshyn (1988).

As the dust from these initial skirmishes begins to settle, two kinds of account have started to emerge of the acquisition and representation of English past-tense inflections, each rejecting the traditional account of rules for regular inflections and rote learning of irregular forms as a simple list.

One is a *uniform mechanism* account, arguing that an analogical connectionist network can handle both regular and irregular past-tense formation, much as Rumelhart and McClelland orginally claimed, but using more powerful modelling techniques (typically three-layer networks with a hidden unit layer) and making more acceptable representational assumptions (e.g., Plunkett & Marchman, 1991).

The second is a *dual mechanism* account, chiefly associated with Pinker and his associates (e.g., Pinker & Prince, 1991), and arguing that a rule-based symbolic processor is responsible for the generation of regular forms, while an analogical network based system is required for the irregulars. In Pinker's particular approach, the rule-based component is part of the genetically specified human language faculty.

We will argue here that both these accounts, as well as the traditional linguistic account, are insufficient, and based on incomplete psycholinguistic models of the mental representation of morphologically complex words in English. We believe that an important additional factor determining the mental representation of English inflected forms is the *phonological regularity* of the relationship between the present and the past tense forms. This has major consequences for how the language learner chooses to represent, in the mental lexicon, the relationship between inflectional variants of the same verbal stem.

The basis for these claims is, first of all, our previous research on English derivational morphology, leading to a view of the mental lexicon as being organised around abstract morphemes, and, second, the new experiment we will report here, using an on-line priming technique to probe directly the mental representation of regular and irregular English inflectional forms.

Previous research

We will begin here with a brief reminder of the difference between inflectional and derivational morphology in English.

Broadly speaking, inflectional morphology has a primarily grammatical function, as in the tense morphemes already mentioned. Inflectional morphemes in English also mark number, on both verbs and nouns -- *dog/dogs, I jump/he jumps* -- and indicate comparatives (*big/bigger/biggest*). These morphemes do not change meaning, and, by and large, do not change the form of the stems to which they are attached - with the crucial exception of certain irregular past-tense forms.

Derivational morphemes, in contrast, do alter the meaning and often the syntactic form-class of the base forms to which they are attached, as in *manage/management, nation/national*, and so on. Unlike inflectional morphemes, derivational morphemes create new lexical items. Over historical time, these derived forms may become semantically opaque, so that their meaning is no longer compositionally derivable from the combination of the stem and the affix -- as in forms like *department* which cannot be interpreted as *depart* + *ment* (as opposed to transparent forms like *govern* + *ment*). Derivational morphemes, finally, unlike inflectional morphemes, frequently change the phonological form of the stem to which they attach, as in alternations like *sane/sanity, decide/decision*, etc.

Our earlier research focussed on the mental representation of derivationally complex words in English. This work provides the methodological and the theoretical background for the new research.

The first point is that this research is makes a basic distinction between the *lexical entry* for a given

word, and the *access route* for that word. The lexical entry for a word we define as the modality-independent core representation of a word's abstract syntactic, semantic, and phonological properties. The access route we define as the modality-specific perceptual route whereby information in the sensory input is linked to a given lexical entry. Our interest in this earlier research was in the properties of the *lexical entry*, and how this was structured for morphologically complex words.

The second point concerns the experimental task we use to investigate the properties of the lexical entry. This is the *cross-modal immediate repetition priming* task, where the subject hears a spoken prime, and immediately at the offset of this word sees a visual probe which is related in some way to the prime. The subject makes a lexical decision response to this probe, and response latency relative to a control condition, where listeners respond to the same probe following an unrelated prime, is used to measure any priming effect. Because the task is cross-modal, any priming effects are attributable to events at the level of the lexical entry, rather than to effects of lower-level overlap at the level of modality-specific access representations. Our earlier research confirms that priming in this task (for morphologically related words) is based on shared morphemes between prime and target, and not on any form overlap between them.

Using this task to study the mental representation of English derived forms, we came to two main conclusions which form the basis for our approach to the inflectional morphology. The first is that the lexical entry for *semantically transparent* complex forms is morphologically structured, in a {stem + affix} arrangement. The morpheme {govern}, for example, is common not just to the word *govern* itself, but also to related forms such as *government*, *governor*, and *misgovern*. This is reflected in the priming relations between these forms, where *government*, for example, primes *govern*, while *govern* in turn primes *government*, and so on.

Semantically opaque forms, however, are not represented in a morphologically decomposed manner; *depart* does not prime *department*, nor does *department* prime *depart* (despite the amount of phonological overlap between prime and target).

The second main conclusion of this research is that the representation of lexical form at the level of the lexical entry is *phonologically abstract*; it abstracts away from surface variation in the phonetic realisation of a given stem morpheme. The morpheme {sane} for example, will surface as [sæn] in the context of the suffix {-ity}, but as [seyn] in the context of the suffix {-ly} or when it appears in isolation. This phonological opacity, as we termed it, in the relationship between the form of the stem in isolation and in a derived form had no effect on priming. *Sane* primes *sanity*, *decide* primes *decision*, (and vice versa) just as strongly as primes where the relationship is phonologically transparent (as in *happy/happiness* or *govern/government*).

We account for this by arguing that the underlying morpheme, at the level of the lexical entry, is phonologically abstract in ways which mean that it is compatible with all *regular* surface variations and which allow direct mapping from the surface form onto the lexical entry.

It is important to stress that we are dealing here exclusively with *phonologically regular* alternations; the *sane/sanity*, *clear/clarity* type of alternation is an example of an extremely widespread phonological process, usually going under the title of closed syllable shortening (e.g., Myers, 1987). This is the process whereby an abstract underlying vowel is phonetically realised as a long vowel when it occurs in an open syllable but as a short vowel when it occurs in a closed syllable (as in the /ey/-/æ/ alternation in *sane/sanity*). These are not processes that are specific to the derivational morphology, or to individual affixes or stems; it is simply the case, on this analysis, that the morphological process of adding the suffix has the phonological consequence of closing the syllable and therefore leading to vowel shortening.

If we take these conclusions, and the accompanying model and methodologies, and apply them to English inflectional morphology, this leads to conflicting expectations.

On the one hand, since the relationship between a verbal stem and its inflectional variants is always semantically transparent, there should always be priming between past and present inflected forms, irrespective of phonological opacity. A form like *gave* should prime *give* just as well as a form like *jumped* should prime *jump*, since each pair shares a common underlying morpheme at the level of the lexical entry, and since this is the basis for priming in the cross-modal repetition task. The same abstract morpheme {give} must underly both *give* and *gave*, since the semantics and syntax of *gave* is just that of *give* with the past tense applied, in the same way that *jumped* is just the underlying morpheme *jump* with the past tense operator applied.

On the other hand, our claims about phonologically abstract underlying derivational morphemes were situated in a context of phonological regularity, where there were regular and independently attested processes that allowed

one form to be derived from another (i.e. *sanity* from underlying {sane} + {ity}. But it is not clear that we can make the same move for phonologically irregular or idiosyncratic relationships, such as those between *give* and *gave* or between *bring* and *brought*. There is no regular process that would allow us to derive one from the other, nor is there is any identifiable abstract form from which both could be derived. It is not clear, therefore, that we can make the same representational moves for at least some of the English inflectional morphology as we could for the derivational morphology.

Experiment

To explore these issues further, we used the cross-modal immediate repetition priming technique to probe the on-line processing relations between inflected forms of the same verb stem. A pilot experiment confirmed that phonological regularity was indeed an important factor in determining whether or not an irregular inflectional form successfully primed its stem. In the main experiment, therefore, we distinguished between three classes of verb inflection, according to the degree of phonological and morphological regularity they exhibited.

1. **Regular**: This is the standard class of regular or *weak* verbs, using affixation - the dominant form of inflection in English - to form the past tense. All these verbs take the regular past tense suffix /-d/, which assimilates in voice to the preceeding segment, and which causes no change in the vowel stem. These are verbs like *jump, ask, beg, seem*, etc.

2. **Vowel Change**: This is the class of traditional *strong* verbs in English, where the past tense is marked by a change in the stem vowel (often referred to linguistically as *ablaut*). This is a morphological process, where the resulting alternation is usually not of a phonologically regular sort -- that is, it cannot be regarded as an example of an independently motivated phonological process (such as cllosed syllable shortening). These are verbs like *cling, give, run, dig*, etc.

3. **Semi-Weak**: This is an intermediate set, labelled by linguists as *semi-weak*, but also falling squarely among the traditional irregular verbs, that marks the past tense using irregular variations of the standard /-d/ past tense affix. These affixes may be irregular (relative to the weak class) because they do not necessarily agree in voicing with the final segment (*learnt, burnt*), because they may trigger shortening of the stem vowel (*dreamt, felt, kept*), and because they may replace the stem-final consonant (*made, built*). These semi-weak forms, although less regular than the weak forms, are, nonetheless, more phonologically regular than the strong class because (1) they use affixation to mark the past tense, in common with the weak class, and because (2) any changes in the stem vowel are phonologically regular, so that the surface form can reasonably be treated as deriving from an underlying abstract form, unlike the strong verbs.

Our question was whether these three classes would behave differently in the repetition priming task. They are all equally semantically transparent, and must therefore all share common morphemes at some level of the lexical entry. Is this sufficient to guarantee priming - as it seemed to do for the derivational morphology - irrespective of the type of phonological relationship between stem and past tense form?

To test this, we constructed 21 sets of triplets for each verb type, where the uninflected form is always the target, presented visually for lexical decision, at the offset of the auditory prime, which could either be the Identity, Past Tense, or Control Prime (see Table 1).

The Control words were matched to the primes for frequency and syllable length, and the three verb classes were matched to each other for frequency. Three experimental versions were constructed, so that each target word only occurred once for each subject. 30 subjects were tested, 10 for each version.

Table 1: Example stimulus set

	Primes			*Target*
	Identity	**Past**	**Control**	
Regular	Agree	Agreed	Occur	AGREE
Semi-Weak	Sleep	Slept	Throw	SLEEP
Vowel	Draw	Drew	Catch	DRAW

The experiment included a further 190 filler prime-target pairs, so that the target was a nonword on 50% of the trials, and to keep the relatedness proportion down to 15% overall. 30 word/non-word pairs were included where there was a phonological relationship between prime and target. Half of the fillers had nouns or adjectives in prime position.

Table 2: Priming effects for different verb types (msec)

Verb Type	*Identity Prime*	*Past Tense Prime*
Regular	-36	-24
Semi-Weak	-52	-1
Vowel Change	-29	24

Results

In presenting the results, we will focus on the test-control difference scores, which show directly the extent to which responses to the stem form are speeded up or slowed down by the identity or the past tense prime, relative to the control condition, where the target is preceded by an unrelated prime.

An overall analysis of variance gave a significant interaction between Verb Type and Prime Type ($F1[4,108]=4.836$, $p<.01$; $F2[4,108]=3.330$, $p<.01$) and this is what we see reflected in Table 2. For the Identity prime, there is strong priming for all three verb classes and no significant differences in the size of the effects. There are pronounced differences between verb classes, however, for the Past Tense case. For *regular*, or weak, verbs, there is significant priming, and this does not differ from the size of the effect for the Identity case, indicating that priming is being mediated via a shared morpheme in the lexical entry.

The *vowel change* verbs show the reverse - a significant interference effect - with responses being significantly slower following the Past Tense prime than following an unrelated Prime. This is a robust effect, shown by 80% of the individual items, and significantly different from the effect for the weak verbs. It is also an effect that we have replicated in other experiments.

The *semi-weak* verbs show an intermediate result, with no overall priming relative to control, and with a big reduction relative to the Identity case. Organising the verb classes according to degree of morpho-phonological regularity in the stem/inflected form relationship seems to give a parallel ordering in the manner in which a past tense form primes its stem.

What is particularly striking here is the significant interference effect for the strong verbs. We never obtained effects like this for the derivational morphology, least of all when the semantic relationship between stem and complex form was so transparent. The fact that we do get interference here strongly suggests that there are two representations here in competition, and that each inhibits the other.

It cannot be the case, for reasons of basic cognitive economy, that the complete syntax and semantics of a strong verb are replicated in the mental lexicon for stem and for past-tense forms. There must be a single underlying morpheme which captures the properties of a given verb - say *give* - but it looks as if the relationship between the phonological form [giv], the past tense form [geyv], and the abstract morpheme {give} cannot be handled in the same way as phonologically opaque derivational relations - or indeed relations among the weak verbs in English.

Conclusions

This seems to us to be a new finding about mental representation of English inflected forms, and it is one that poses a challenge for all existing models. Both uniform and dual mechanism accounts treat irregulars in an essentially homogeneous fashion, and provide no motivation for treating the strong verbs in a fundamentally different way. Both types of account assume a distributed network representation for the irregular verbs, and it is not immediately clear on what basis such a representation, built up using standard learning algorithms, would choose to represent vowel change verbs in such a way that priming with the past tense form would interfere with responses to the stem form. We do not need to renounce single mechanism theories of acquisition, but it needs to be shown how a uniform mechanism can nonetheless produce non-uniform representational solutions.

The problem here is likely to be that exposure to the verbal inflectional system in English will not be sufficient on its own to lead a learning model to treat vowel change verbs differently from semi-weak verbs. Of course, in so far as the training set of

English irregular verbs contains sub-regularities, then the network should pick these up -- for example, the subset of verbs with an *ing/ung* alternation (as in *swing/swung*, *fling/flung*, etc). But these are minor variations, where the underlying mode of representation is not likely to differ qualitatively from the representation of irregular verbs in smaller groupings.

Differences in phonological regularity can only be picked up when a sufficient basis is provided for drawing the correct underlying generalisations. The same transformations (such as closed syllable shortening) need to occur in different morphological environments, and not just as a correlate of a subclass of English past tense forms. The implication of this is that models of the English inflectional paradigms are going to have to pay more attention to phonological factors, and start to explore ways of embedding tasks such as the learning of inflectional morphology in the broader context of learning the phonological regularities and irregularities, across the whole language, in the relation between underlying forms and their surface realisation as phonetic strings.

Acknowledgements

This research was supported by grants to William Marslen-Wilson and Lorraine Tyler from the British ESRC, MRC, and SERC.

References

Brown, R., 1973. *A first language*. Cambridge MA: Harvard University Press.

Ervin, S., 1964. Imitation and Structural Change in Children's Language, in *New directions in the study of language* (E. Lenneberg, Ed.), Cambridge, MA: MIT Press.

Fodor, J., and Pylyshyn, Z., 1988. Connectionism and cognitive architecture: a critical analysis, *Cognition* 28:3–71.

Kuczaj, S.A., 1973. The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior* 16:589–600.

Marslen-Wilson, W.D., Tyler, L.K., Waksler, R., & Older, L. 1991, Abstractness and Transparency in the Mental Lexicon. In *The Proceedings of the 14th Annual Meeting of the Cognitive Science Society*, Princeton, N.J.: Erlbaum.

Marslen-Wilson, W.D., Tyler, L.K., Waksler, R., & Older, L. (in press), Morphology and Meaning in the English Mental Lexicon, *Psychological Review*.

Myers, R., 1987. Vowel Shortening in English, *Natural Language and Linguistic Theory* 5:485–518.

Pinker, S., and Prince, A., 1988. On Language and Connectionism: Analysis of a Parallel Distributed Processing Model of Language Acquisition, *Cognition* 28: 73–193.

Pinker, S., and Prince, A., 1991. Regular and irregular morphology and the psychological status of rules of grammar, in *Proceedings of the 17th Annual Meeting of the Berkeley Linguistics Society*, Berkeley, CA: Berkeley Linguistics Society.

Plunkett, K., and Marchman, V., 1991. U-shaped learning and frequency effects in a multilayered perceptron: Implications for child language acquisition, *Cognition* 38: 73–193

Prasada, S., and Prince, S., (1993) Generalization of Regular and Irregular Morphological Patterns, *Language and Cognitive Processes*, 8, 1–56.

Rumelhart., D.E., and McClelland, J., 1986. On Learning the Past Tense of English Verbs. In *Parallel Distributed Processing, Vol. II* (Rumelhart, D.E., and McClelland, J., Eds.), Cambridge, MA: MIT Press.