Organising Principles in Lexical Representation: Evidence from Polish

https://escholarship.org/uc/item/04c982zn

Proceedings of the Annual Meeting of the Cognitive Science Society, 22(22)

1069-7977

Reid, Agnieszka
Marslen-Wilson, William

2000

Peer reviewed
Organising Principles in Lexical Representation: Evidence from Polish

Agnieszka Reid (agnieszka.reid@mrc-cbu.cam.ac.uk)

William Marslen-Wilson (william.marslen-wilson@mrc-cbu.cam.ac.uk)

MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, UK.

Abstract

Cross-linguistic research into the structure of the mental lexicon potentially allows us to deconfound factors which are language specific from factors which are cross-linguistically universal. In a series of three experiments we provide preliminary evidence for the structure of the Polish lexicon, which belongs to the Slavonic language family. As in English, semantic compositionality plays a crucial role, so that semantically compositional, morphologically complex words are stored in a combinatorial fashion, and semantically opaque words seem to be represented as full forms. At the same time, clear evidence is found for priming between derivational and inflectional affixes, and for interference effects between suffixed words competing for the same underlying stem. Overall the data support a combinatorial and decompositional approach to lexical representation.

Introduction

To begin to discover the organising principles underlying the representation and processing of lexical knowledge, it is necessary to conduct comparable research programmes across a variety of different languages. In the studies reported here, we take as a starting point a body of research on English (Marslen-Wilson, Tyler, Waksler & Older, 1994), and ask whether the general properties that seem to emerge for English can be found to operate for Polish, a language with a much more complex and developed morphological system.

Two kinds of claim are made for English. The first is that underlying representations of morphologically complex forms, both derivational and inflectional, are fundamentally decompositional and combinatorial in nature. Evidence for this came from three main sources. Marslen-Wilson et al. (1994) report extensive priming, in an immediate cross-modal repetition priming task, between suffixed and prefixed words sharing the same stem. At the same time, they also report the phenomenon of suffix-suffix interference, where semantically transparent pairs such as government - governor do not prime, despite sharing the same stem. Marslen-Wilson et al. (1994) interpreted this as evidence for competition between different affixes for attachment to the same underlying stem. Thirdly, and perhaps most compellingly, Marslen-Wilson, Ford, Older & Zhou (1996) demonstrate strong priming between derivational affixes, as in pairs like toughness/darkness and rearrange/rethink. Affixes like 'ness' or 're-' appear to be isolable and independent structures in the mental lexicon, participating in a dynamic and combinatorial manner in the representation of many different words.

The second important claim is that, cutting across this evidence for decompositional morphemically based representation, the further factor of semantic transparency plays a crucial role in determining the representation of morphologically complex words. Marslen-Wilson et al. (1994) found that semantically transparent morphologically complex words, such as darkness-dark prime each other, but that morphologically related, semantically opaque pairs, such as department - depart, do not, indicating that words such as department are stored as full forms. Marslen-Wilson et al. (1994) made the argument that this reflects choices made during the language acquisition process, where the language learner rejects a decompositional analysis of department (as depart + ment) on the grounds that this delivers the incorrect semantics.

Cross-linguistic research

The broader status of these claims about the structure of lexical representation – as fundamentally decompositional but conditioned by semantic factors – remains hard to interpret unless comparable bodies of research, using parallel techniques, are conducted across a typologically contrasting sample of the world’s languages. Research of this type is only now starting to emerge, and is already suggesting illuminating contrasts with the patterns proposed for English.

A salient example is the contrasting importance of semantic factors in Semitic languages, such as Hebrew and Arabic, as opposed to English. Hebrew and Arabic are characterised by non-linear morphological processes which operate on roots and word patterns. The most striking feature of this morphological system is that morphemes are not combined linearly, but a root, which usually consists of three consonants, is interleaved in a discontinuous manner with a word pattern, to create the phonetic surface form. Deutsch & Frost (1998) demonstrated that in Hebrew, words which are morphologically but not semantically related, prime each other strongly, in contrast to the findings on English. More recently Boudelaa & Marslen-Wilson (2000) demonstrate comparable findings for Arabic, using both cross-modal and masked priming tasks, and finding equally strong priming between prime target pairs sharing the same roots, irrespective of semantic transparency.

The finding that semantic transparency is a crucial factor in the structure of the English mental lexicon, but seems to play no role in the morphological decomposition of Semitic words, is hard to interpret on its own, because of the many
ways in which languages like Hebrew and Arabic contrast with a language like English. One of the goals of the research reported below is to add another typologically distinct data point to these contrasts, asking for Polish not only whether there is comparable evidence here for decompositional representation, but also whether semantic factors play a critical role in determining whether or not complex forms are represented in decompositional format.

**Research on Polish**

A striking characteristic of Polish, a member of the Slavonic language family, is the richness of its morphological systems. Almost every word in Polish exists within a very rich paradigm, declensional for nouns, adjectives, numerals and pronouns or conjugational for verbs. The derivational morphology is comparable to English, being based on concatenative processes of prefixation and suffixation, but includes a number of qualitatively very different affixes, for instance verbal aspectual prefixes, aspectual-derivational prefixes and diminutival suffixes. Also, as far as derivational suffixes are concerned, they are considerably more numerous. Polish permits the formation of morphologically very complex words, such as secondary imperfectives described below, which allows a challenging test of claims about combinatorial representation and access.

**Experiment 1**

The main goals of the first experiment we report here were to investigate morphological phenomena that are absent in English, as well as to investigate parallel phenomena in the two languages. To do this we used the cross-modal immediate lexical decision task (Marslen-Wilson et al. 1994). In this task subjects hear an auditory prime, at the offset of which, they immediately see a visual target (for 500 ms) and have to decide, by pressing an appropriate button, whether a target word is a real word or a non-word.

Taking advantage of the range of qualitatively different affixes in Polish, we probed their representation in pairs of semantically transparent words, which share the same affixes. The stimuli included (a) 24 pairs of verbs which share the same aspectual prefix, e.g. skorzyć na 'to benefit, Perfective' - stracyńy 'they lost, Perfective'; (b) 22 verbs which share the same aspectual-derivational prefix, for instance: nagrać 'to heat up, Perfective' - natręta 'she cut, Perfective'; (c) 18 nouns which share the same diminutive suffix, e.g. kotek 'a little cat' - ogródka 'a little garden' and (d) 24 nouns which have the same derivational suffix, e.g. kucharz 'a cook' - piłkarz 'a footballer'. Also, having in mind the difference in findings on English and Hebrew/Arabic regarding words which are morphologically related, but semantically opaque, we included (e) 22 pairs, such as jarzyniec 'juniper' - jarzynowy 'futile', as a test of whether semantically opaque words prime each other. It seemed plausible that these pairs would prime in a language such as Polish, where the dynamics of morphological processing are much stronger than in English. We will refer to them as [Morph, - Sem]. We also included a condition (f) 20 pairs which share the same stem, e.g., szycie 'sewing' - szyć 'to sew'. Because many studies document a robust effect of stem priming, this condition served as a test of the procedure in our experiment. Finally, we included (g) 20 semantically related pairs, e.g., kokos 'coconut' - banan 'banana', which also served as a test of the experimental procedure. Many experiments on English found semantic priming in the cross-modal priming. We will refer to them as [-Morph, +Sem]. In addition, to investigate whether any observed priming in affix conditions was due to pure phonological overlap, we included two control conditions where the stimuli were phonologically but not morphologically or semantically related: (h) 18 words with phonological overlap at the onset, e.g. numer 'a number' - nuda 'boredom' and (i) 18 with overlap at the word offset, e.g. halas 'noise' - szalas 'a shelter'.

**Results**

6 subjects from version 1 and 4 subjects from version 2 were discarded from the analysis, because of high error percentage on real words (equal to or above 15%) or/and slow mean reaction times to real words (equal to or above 1000 ms). A total of 20 subjects per version was entered into the analysis. All subjects were in their twenties, and were native Polish speakers living and studying in Poland. 7 items were removed from the analysis: 3 because of high error percentage (equal to or above 30% on both versions or equal to or above 40% on one version) and 4 because of homophony. Every reaction time was inversely transformed in order to reduce the influence of outliers. The inversely transformed data were analysed in a Repeated Measures ANOVA separately for items (F2) and for subjects (F1). See Figure 1 for details of the descriptive statistics.

First the overall repeated measures ANOVA with Prime (related, unrelated) and Condition (1-9) was run. There was a main effect of Prime, indicating that RTs were faster for targets when preceded by a related prime than an unrelated prime, F2(1,163)=22.62, p<0.001; F1(1, 38)=37.32, p<0.001. The main effect of Condition was significant, F2 (8, 163)=17.82, p<0.001; F1(8, 304) = 162.89, p< 0.001. However, there was also a significant two-way interaction of Condition x Prime F2(8,163) = 4.49, p<0.001; F1(8, 304) =7.45, p<0.001.

The finding that there was 18 ms of priming on average in all the affix conditions treated as a group was explored further in an ANOVA. The results showed that there was a main effect of Prime, F2(1, 75)=12.06, p< 0.001, F1(1, 38)=11.77, p< 0.001. The main effect of Condition was also significant, F2(3, 75)=19.54, p<0.001, F1(3, 114)=209.88, p< 0.001, with no interaction between Condition x Prime, F2(3,75)=0.14, p>0.05, F3(3,114)=0.37, p>0.05. This result indicates that there was a facilitatory effect of Prime in all affix conditions treated as a group.

We then conducted an analysis of simple effects of Prime on each level of Condition in the remaining Conditions. The results show a facilitatory priming for [+Morph, -Sem] pairs, F2(1,20)=0.56, p>0.05; F1(1,38)=0.72, p>0.05. There was no priming for either of the Phonological Overlap conditions: Phonological Overlap at the Word Onset, F2(1,16)=0.05, p>0.05, F1(1,38)=0.62, p>0.05 and Phonological Overlap at the Word Offset F2(1,16)=2.15, p>0.05, F1(1,38)=3.98, p>0.05. On the other hand, there
was clear priming in the Stem Condition F2(1,18)=25.0, p<0.001, F1(1,38)=53.43, p<0.001 and in Semantically, but not Morphologically Related Pairs F2(1,18)=13.36, p<0.01, F1(1,38)=22.10, p<0.001.

Discussion
The results of Experiment 1 show clear priming in all the Affix conditions treated as a group as well as in the Stem condition. The absence of priming in the two phonological overlap conditions indicates that the priming obtained in the affix conditions cannot be attributed to simple phonological overlap. The results show that affixes and stems are isolable and independent structures in the Polish mental lexicon. Polish affixes, although qualitatively different from English affixes, seem to be stored in a combinatorial manner. On the other hand, the evidence shows that morphologically related, semantically opaque words do not prime each other, indicating that they are stored as full forms. This indicates that the factor of semantic compositionality determines the representation of morphologically complex words in Polish. This is in line with the findings on English, but is in contrast with Hebrew and Arabic, where semantic compositionality does not determine the representation of morphologically complex words.

The combinatorial storage of affixes in the Polish mental lexicon is also supported by evidence on a Polish Wernicke’s aphasic patient (Ulatowska and Sadowska, 1988). In tests of production, the patient occasionally made mistakes involving derivational morphology. When asked to produce a word denoting a little plate, she produced taclezk, using an existing, but incorrect diminutival suffix, instead of saying taclezyk. When verbal aspectual morphology is considered, the patient produced an incorrect form, such as z-swiad, instead of a-swiad ‘he got grey’, substituting a correct aspectual prefix ‘a-’ with an incorrect one ‘z-’ for this verb. Although the origin of these errors may be partially conditioned by phonological deficits in the patient’s language output system, it seems to be hard to account for these errors only in these terms. The errors include an incorrect combination of existing morphemes, rather than a combination of non-existing units. Hence, we take this as a further evidence in support for the combinatorial storage of words in the Polish mental lexicon.

Experiment 2
The findings on affix priming in Polish reported in experiment 1 left us with two further questions. Firstly, does the combinatorial representation of affixed words also hold for much more complex forms? Secondly, will we get more reliable priming, in comparison to the relatively weak priming in the four individual affix conditions in Experiment 1, if two affixes are shared by the prime and target?
Highly polymorphemic, semantically compositional words, such as secondary imperfectives, which occur in Polish, are a particular challenge for the combinatorial view of the mental lexicon. On one hand their complex structure would make them potentially more difficult to parse in comprehension and assemble in production if they are represented as a combination of morphemes, rather than as full forms. On the other hand, the intuition of native speakers of Polish is that they can process highly polymorphemic forms with the same efficiency as the less complex forms. More generally, for productive complex morphological forms, it is generally accepted that simple learning of each complex form is not a plausible language acquisition procedure (e.g., Hankamer, 1989).

We used (a) 30 pairs of secondary imperfectives, which shared the same prefix and suffix, e.g. roz-pakow-ywa-l-on (prime) ‘to unwrap, 1st person sing., masculine, past tense, secondary imperf.’ and roz-walkow-ywa-ci (target) ‘to flatten something using, a rolling-pin, secondary imperf.’. These words consisted of a derivational prefix, e.g. ‘roz-’, a secondary imperfective suffix ‘-ywa-’, past tense morpheme ‘-1’ (prime only) and a morpheme ‘-em’ (prime only), which denotes the 3rd person singular, masculine. To ensure an appropriate paradigm-check we also included (b) 24 standard stem priming pairs, myśl-ç ‘I think’ - myśl-e-ci ‘to think’; and (c) 24 semantically related, but morphologically unrelated pairs, e.g. dom ‘a house’ - garą ‘a garage’, to dissociate the morphological and semantic effects.
Because we wished to avoid possible confounds with semantic priming, we used here a different task. This was an
auditory-auditory priming experiment with 12 items inter-
vening between prime word and target. At these long lags, it
is generally found that semantic priming drops away
whereas morphological priming does not (Marslen-Wilson
& Tyler, 1998).

Results

10 subjects were discarded from the analysis according to
the same criteria as in experiment 1. Data from 23 (version
1) and 24 (version 2) participants were entered into the
analysis. One item had to be discarded from the analysis,
because of high error percentage on one version. See Figure
2 for the details on the descriptive statistics.

Figure 2. Priming effects for Experiment 2

The reaction time data were prepared for the analysis as
described in experiment 1. The overall ANOVA revealed
that the main effect of Prime was significant, F2(1,71)=
16.006, p<0.001, F1(1,45)=507.888, p<0.001. The main
effect of Condition was also significant, F2(2,71)=20.426,
p<0.001, F1(2,90)=81.308, p<0.001. The two-way interac-
tion of Prime and Condition was not significant in the items
analysis, F2(2,71)=2.004, p>0.05, but it was significant in
the subject analysis, F1(2,90)=72.625, p<0.001.

We then carried out an analysis of simple effects of Prime
on every level of Condition. There was clear priming in the
Shared Affixes Condition; F2(1,28)=4.8, p<0.05, F1(1,45)
=7.58, p<0.01. The results for the paradigm-check condi-
tions were straight-forward: as predicted there was signifi-
cant priming for the Stem Priming Condition F2(1,22)
=15.48 , p<0.01; F1 (1,45)=10.98, p<0.01 and there was no
priming in the Semantically Related, but Morphologically
Unrelated Condition: F2(1,21)=1.01, p>0.05; F1
(1,45)=0.249, p>0.05.

Discussion

Firstly, the results show reliable priming for secondary im-
perfectives, e.g. roz-pakow-ywa-l-em, which indicates that
they are in fact represented in a combinatorial fashion, de-
spite their morphological complexity. Secondly, it appears
that, when two affixes are repeated in prime and target, we
obtain a more robust priming effect, of the magnitude of
stem priming, in comparison with the relatively weak
priming in the affix conditions in experiment 1. This is
consistent with claims for combinatorial underlying proc-
essing mechanisms, comparable to those claimed for Eng-
lish, and matching the claims for somewhat different forms
of underlying combinatorial systems in the non-
concatenative morphologies of Hebrew and Arabic.

Experiment 3

One of the main pieces of evidence in support of the combi-
natorial approach to the English mental lexicon comes from
the finding that semantically transparent pairs which share
the same stem and have different derivational suffixes do
not prime each other. This finding has been replicated many
times in English, since the original report in Marslen-
Wilson et al. (1994). For instance Marslen-Wilson & Zhou
(1999) show that pairs which exhibit allomorphy, e.g.,
sincere-ly & sincer-ity as well as non-allomorphic pairs,
e.g., exci-t-able & exci-t-ment do not prime each other either
in a cross-modal priming task or in an auditory-auditory
lexical decision task with 0 or 8 intervening lags. The re-
results at 8 intervening lags established that the suffix inter-
ference effect is robust and can be elicited under conditions
where morphological but not semantic factors are likely to
be responsible.

Because we found evidence for the combinatorial storage
of morphologically complex, semantically compositional
words in Polish, we wanted to test whether we would find
convergent evidence from suffix interference, tested in a
language system where suffixation is one of the main deri-
valional processes.

The stimuli included (a) 32 derived - derived words which
shared the same stem, but had different derivational suffixes.
Half of the stimuli were deverbal derivatives, e.g. pis-anie
‘writing’ - pis-arz ‘a writer’. The other half were denominal
derivatives, e.g. balon-owy ‘balloon like, adj.’ - balon-ik
‘a little balloon’, SR = 8.1, SD = 0.5; (b) 32 inflected - de-
erved pairs which shared the same stem. Half of the stimuli
had an inflected verb as a prime and a deverbal derivative as
a target e.g. pho-da ‘to write, 3rd person, sing., feminine,
past tense’ - pis-arz ‘a writer’. The other half had an
inflected noun as a prime and a denominal derivative as a
target, e.g. balon-em ‘balloon, instrumental - balon-ik
‘a little balloon’, SR = 8.4, SD = 0.4; (c) 24 stem priming
pairs, as before e.g., mysł-ę ‘I think’, mysł- – to think’
were included as a paradigm check, SR = 8.2, SD = 0.2.

SR denotes a mean score (across 10 participants) on a Semantic
Relatedness pre-test, where native speakers of Polish judged on
a 9-point scale (where 9 is the highest possible score), to what
degree a given pair of words is semantically related. We use these
scores as a measure of Semantic Transparency between a prime
and target, which is highly correlated with semantic composition-
ality.
There was a strong priming in Stem Condition, \(F_2(1,20)=15.24, p<0.01, F_1(1,85)=19.874, p<0.001\) and that there was no significant priming for \([+\text{Sem}, -\text{Morph}]\) Condition, \(F_2(1,20)=3.9, p>0.05, F_1(1,85)=1.821, p>0.05\).

**Discussion**

The results are clear. There is no priming in Derived-Derived Condition nor in Inflected-Derived Condition while there is robust priming for the Stem Condition and no priming for \([-\text{Morph}, +\text{Sem}]\) pairs. The findings for the latter two conditions have been replicated in many of our experiments (e.g. experiment 2) and are in line with the predictions.

Not finding priming for semantically transparent suffixed words which share the same stem, but have different derivational suffixes indicates that there must be an inhibitory process between the suffixes. Hearing 'pis-anie’ ‘writing’, as a prime inhibits the combination of the root 'pis' with another suffix, e.g. ‘-arz’, hence the recognition of the target 'pis-arz’ ‘a writer’ is slowed, even though the root ‘pis’ is active. This finding parallels the findings reported for English (Marslen-Wilson et al. 1994).

Interestingly, the results also suggest that there is suffix interference in Inflected-Derived Condition between inflectional and derivational suffixes which are attached to the same stem in semantically compositional pairs. It is hard to see how this could be the case if both types of suffix were not represented in the lexicon.

Most influential linguistic models of word formation in generative grammar assume that inflections are not represented lexically, but they are added by syntactic rules which are outside the mental lexicon. This is supported by data from the lexical decision experiments on English, but not by the findings on Polish. One plausible source of the difference between the findings for Polish and English comes from the characteristics of the Polish inflectional system, which in contrast to English is extremely rich and carries a lot of very complex information. The findings (although from a different paradigm) on Italian (Miceli & Caramazza 1988) which is a morphologically rich language, similarly to findings on Polish support the claim that the inflectional suffixes are stored in the mental lexicon.

The suffix-suffix interference leaves, at the current stage, at least one unresolved issue. All the inflectional suffixes of the primes for deverbal targets in Inflected-Derived Condition were singular person, masculine or feminine, past tense. In our previous experiments, we found priming for pairs which shared the same stem and where prime had a derivational suffix whereas the target had an infinitival ending, denoting person, tense and number, whereas the target was an infinitive and had an infinitival marker ‘č’ and (d) 24 [-Morph, +Sem] pairs, included to dissociate the morphological and semantic effects, \(SR = 8.0, SD = 1\). In order to ensure the most rigorous comparison between conditions 1 and 2, the same target was used in both conditions. Experimental items were assigned to 4 versions, so that the same target was preceded by one of the two related primes or one of the two unrelated primes, with one combination of a prime and a target per version. Because we were mainly interested in the morphological effects, we again used an auditory-auditory priming, lexical decision experiment with 12 intervening items to dissociate the morphological from the semantic effects.

**Results**

10 participants were rejected on the same criteria as in experiment 1. A total of 89 participants: 22 (version 1), 24 (version 2), 21 (version 3) and 22 (version 4) were entered into an analysis. No experimental items were removed. See Figure 3 for the details of the descriptive statistics.

The overall repeated measures ANOVA analysis with Prime (related, unrelated) and Condition (1-4) revealed that the main effect of Prime was significant \(F_2(1,96)=22.2, p<0.001, F_1(1,85)=32.48, p<0.001\). The main effect of Condition was not significant in the item analysis \(F_2(3,96)=0.537, p>0.05\), but it was significant in the subject analysis \(F_1(3,255)=6.776, p<0.001\). A two-way interaction of Prime and Condition was significant, \(F_2(1,3)=2.9, p<0.05, F_1(3,255)=2.71, p<0.05\).

The simple effects analysis was conducted to investigate effect of Prime at each level of Condition. The analysis showed that there was no significant priming in Derived-Derived Condition, \(F_2(1,28)=1.35, p>0.05, F_1(1,85)=0.548, p>0.05\), nor in Inflected-Derived Condition, \(F_2(1,28)=3.615, p>0.05, F_1(1,85)=2.702, p>0.05\). The remaining analyses on the individual conditions revealed that...
**General Discussion**

We have reported the findings on the Polish mental lexicon in a series of three experiments, in an attempt to examine the organising principles affecting the structure of the mental lexicon of a morphologically complex language from the Slavonic family. In the first experiment we concentrated on probing the representation of morphologically complex words, which included affixes which are qualitatively different from English. The findings indicated that affixes are represented in a combinatorial fashion. Secondly, the results show that semantic compositionality is an important factor in determining the lexical representation in the Polish mental lexicon. In the second experiment we confirmed that a combinatorial representation also holds for words with a much more morphologically complex structure, such as the secondary imperfectives, at the same time confirming the existence of strong priming between derivational and inflectional affixes. Finally, in the third experiment we addressed the issue of suffix interference in Polish, finding clear evidence for interference in derived-derived pairs as well as in inflected-derived pairs. This is further evidence for underlying combinatorial representations and processes.

In summary, the overall picture which has emerged as a result of our investigation of the Polish mental lexicon is that, Polish, similar to English and Hebrew is characterised by a combinatorial mental lexicon. However, different factors which condition the structure of the mental lexicon have different 'weightings' in Polish as in comparison with Hebrew and English. The factor of semantic compositionality is crucial in determining the structure of the representation of words in Polish, similarly to English, but in contrast with Hebrew (and Arabic). On the other hand, the factor of the type of inflectional morpheme is important in the structure of the Polish lexicon, in that both types of inflectional morphemes verbal and nominal seem to be represented in the Polish lexicon. This contrasts with English, where neither verbal nor nominal inflections seem to be represented as lexical processing structures.

**Acknowledgements**

The research reported here is supported in part by an ESRC research studentship to A. Reid, and in part by the UK Medical Research Council. We would like to thank K. Kowalik, M. Smoczyńska and B. Szymank for very helpful comments made during the selection of the stimuli. We also would like to thank P. Brzuski who kindly provided the facilities in Poland to run our experiments.

**References**


