COMPLEXITY AND ALTERNATION IN THE POLISH MENTAL LEXICON

Agnieszka Reid and William Marslen-Wilson
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, UK.

ABSTRACT
Morphological complexity, semantic transparency and allomorphy are a challenge for models of lexical representation and access. Experiment 1 contrasted Polish words varying in semantic transparency, semantic compositionality and morphological complexity. Semantic compositionality was found to be the best predictor of priming in Polish words. This suggests that semantic compositionality plays an important role in determining lexical representation in Polish, with compositional words represented in a combinatorial fashion, and non-compositional words as full forms. Experiment 2 investigated underlying representations of Polish verbs and nouns with regular and irregular alternations. The results indicate that both regular and irregular alternants map onto the same underlying morpheme, with strong priming regardless of degree and type of alternation. The theoretical implications of this are discussed.

1. INTRODUCTION
Three of the most challenging areas for the study of lexical representation, and its role in lexical access from speech, are morphological complexity, semantic transparency and allomorphy. The picture which has been emerging from the research on English indicates that the English mental lexicon is combinatorial. It was reported [1] that semantic transparency, operationalised in terms of a semantic relatedness judgement task, plays a crucial role in determining the nature of the lexical representation of morphologically complex words in English. Only words, which are semantically transparent, e.g. excitement - excite and darkness - toughness prime each other, indicating that stems and affixes are stored separately. Words which share a stem or an affix map onto the same abstract underlying lexical entry. On the other hand words which are not semantically transparent, e.g. university - universe do not exhibit any priming, indicating that they are stored in separate lexical entries. Preliminary research on Polish [2], a language characterised by a much richer and complex morphological system than English, showed a similar pattern of results. Semantically transparent, morphologically complex words, with qualitatively different affixes from English affixes, such as derivational-aspectual prefixes e.g. nagroda ‘to heat up, Perfective’ - nakrota ‘she cut, Perfective’ and words which share the same diminutive suffix, e.g. kopek ‘a little cat’ - ogroddek ‘a little garden’ exhibit priming when treated as a group, indicating that the affixes are stored in a combinatorial way. Similarly, semantically transparent words which share the same stem, e.g. szyce ‘sewing’ - szyż ‘to sew’ prime each other indicating that stems are stored in a combinatorial fashion. Finally, Polish semantically opaque words, such as jalowicę ‘juniper’ - jalousy ‘futile’ do not prime. This evidence is in contrast to data collected on Hebrew and Arabic [3], where priming was found for semantically transparent as well as for semantically opaque items.
In this paper (experiment 1) we follow our previous study and ask more detailed questions about whether factors such as semantic transparency, semantic compositionality and morphological complexity determine the representation of a wider range of Polish words and in what way.

The other issue which is challenging to the combinatorial mental lexicon are stems which exhibit various alternations. Marslen-Wilson & Zhou [4] investigated the mental representation of pairs of words where the stem exhibited either consonantal, e.g. elusive and elude or vocalic, e.g. serenity and serene, regular, phonological alternation, using cross-modal as well as immediate and delayed auditory-auditory priming. A paradigm check condition consisted of pairs with no alternation, e.g. friendly and friend. Priming was found in both the auditory-auditory and cross-modal experiments and there were no differences in the magnitude of facilitation between all three conditions. Marslen-Wilson and colleagues took these results as support for a model with alternating stems stored as a phonologically abstract underspecified representation in the same lexical entry. Subsequent research found that irregular pairs, such as taught and teach prime each other in the auditory-auditory immediate and delayed priming experiment, which indicates that they too are stored in the same lexical entry. The authors accounted for this finding by postulating the listing of all alternants within the same lexical entry. In this paper (experiment 2) we ask how nominal and verbal stems with regular and irregular alternations are represented in the mental lexicon of Polish, a language where two thirds of the words exhibit alternations.

2. EXPERIMENT ONE
In previous research [2] following the investigation of English, we looked at two extremes of the semantic transparency continuum - semantically transparent and semantically opaque words. Reliable priming was found for the first group, but no priming was found for the second group. In the current experiment, in addition to semantic transparency, we systematically looked at two other factors which are particularly relevant in the research on Polish: semantic compositionality and morphological complexity. The details on Semantic Transparency can be found in Table 1 in the ST column. Se-
semantic compositionality was defined as the quality of a word which allows one to determine its meaning from the constituent morphemes. This variable can play an important role in determining the lexical representation of words in Polish, where there is a very rich continuum of semantically related words, but where a derivative and its base can have a relatively high semantic transparency, but they might not be semantically compositional. The semantic compositionality measure was based on the judgement of the first author and three judges, who were all native speakers of Polish. As the majority of Polish words consists of one or two derivational affixes, the morphological complexity factor was also included. Morphologically complex words consist of a stem and a derivational suffix and morphologically very complex words of a stem and two derivational affixes.

The issue addressed in this experiment was whether these three factors play an important role in determining the lexical representation of Polish words which vary across these factors. We used an auditory-auditory priming lexical decision paradigm with items presented every 3 seconds, with 12 items intervening between prime and target. This dissociates morphological and semantic priming. The stimuli included (1) 32 semantically compositional, morphologically complex words, which had the highest semantic transparency (Please see Table 1 for details), e.g. chodz-enie/chodz-i-ć ‘walking/to walk’ (2) 32 semantically compositional, morphologically complex words with high semantic transparency, e.g. bad-acc/bad-a-ć ‘investigator/to investigate’, (3) 32 semantically non-compositional, morphologically complex words, with medium semantic transparency, e.g. wiaź-anka/wiaź-a-ć ‘bunch/to tie’ (4) 32 semantically compositional, morphologically very complex words, with high semantic transparency, e.g. roz-lad-ow-anie/ład-ow-a-ć ‘unloading. Perfective - to load’, (5) 30 semantically non-compositional, morphologically very complex words with medium semantic transparency, e.g. do-pis-ek/pis-a-ć ‘added note/to write’ and (6) 24 semantically related but morphologically unrelated words, dom/gara ‘a house/a garage’, to dissociate the morphological and semantic priming. The experiment was run in Kraków using DMDX software.

Table 1. Priming Effects for Experiment 1.

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>ST</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comp. Morph. Complex HtST</td>
<td>32</td>
<td>7.6 (0.5)</td>
<td>21</td>
</tr>
<tr>
<td>Comp. Morph. Complex HST</td>
<td>32</td>
<td>7.1 (0.9)</td>
<td>4</td>
</tr>
<tr>
<td>Non-Comp. Morph Complex MST</td>
<td>32</td>
<td>6.5 (0.9)</td>
<td>4</td>
</tr>
<tr>
<td>Comp. Morph V. Complex HST</td>
<td>32</td>
<td>6.5 (0.6)</td>
<td>33</td>
</tr>
<tr>
<td>Non-Comp.Morph V.Complex MST</td>
<td>30</td>
<td>5.3 (1.0)</td>
<td>15</td>
</tr>
</tbody>
</table>

ST- Semantic Transparency denotes a mean score (across 16 participants) on a Semantic Transparency pre-test, where native speakers of Polish judged on a 9-point scale (where 9 is very related), to what degree a given pair of words is semantically related. Semantic Transparency: HtST=Highest, HST=High & MST=Medium.

SDs are shown in parenthesis. P denotes priming in ms.

analysed in a Repeated Measures ANOVA, separately for items (F2) and for subjects (F1). See Table 1 for descriptive statistics. The overall analysis with Prime (2 levels), and Condition (6 levels) and Version (2 levels) revealed a main effect of Prime F2(1,166)=10.405, p<0.01, F1(1,45)=10.951, p<0.01. The main effect of Condition was not significant in the item analysis F2(5,166)=0.972, p>0.05, but it was significant in the subject analysis, F1(5,225)=19.925, p<0.001. More importantly, the two way interaction of Prime and Condition was significant F2(5,166)=2.36, p<0.05 F1(5,225)=3.543, p<0.01. The analysis of simple effects of Prime on every level of Condition revealed that there was a significant effect of Prime in conditions 1 and 4, but not in the other conditions.

No priming in Condition 2 was unpredicted outcome because the words in this condition were judged as semantically compositional. As pointed out by Smoczyńska (personal communication) the stimuli in this condition included foreign borrowings, which were imported as full forms and could be represented as such in the Polish mental lexicon. For the purpose of investigating the factors in this experiment further, condition 2 was excluded from further analysis and investigated in a subsequent experiment (see discussion). The General Factorial ANCOVA with Semantic Compositionality and Morphological Complexity, as factors and Semantic Relatedness as a Covariate on conditions: 1, 3, 4 &5 revealed no significant main effect or interactions. Individual predictions of Priming for each of the variables were investigated to test whether the effect of some variables was cancelled by the effect of other variables. Semantic Compositionality was a significant predictor of priming in a linear regression t=-2.205, p<0.05. Separate linear regressions of Morphological Complexity and Semantic Transparency on priming revealed no significant effects. However, if Morphological Complexity was entered together with Semantic Transparency to the regression analysis, they were both significant predictors of priming, t=-2.284, p<0.05 and t=-2.257, p<0.05, respectively. So, the null effect in ANCOVA was due to Morphological Complexity and Semantic Transparency together cancelling effect of Semantic Compositionality.

2.1 Results

5 subjects from version 1 and 5 subjects from version 2 were rejected from the analysis because of high error percentages 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 23 participants (version 1) and 24 (version 2) were entered into the analysis. All subjects were in their twenties, they were native speakers of Polish, and were living and studying in Poland.

A experimental items had to be removed form the analysis, because of a high error percentage (equal to or above 30% on both versions or equal to or above 40% on one version. A total of 178 items were entered into the analysis. The dependent variable of interest was the reaction time (RT) to real words. Every individual RT was inversely transformed in order to reduce the influence of outliers. The inversely transformed data were
2.2 Discussion

No priming for [+Sem, -Morph] items indicates that the nature of priming for all the morphologically related items is morphological, not semantic. Priming in semantically compositional conditions 1 & 4 and no priming for semantically non-compositional conditions 3 & 5 indicates that compositionality might be an important factor which determines lexical representation in the Polish mental lexicon, similar to English and in contrast to Semitic languages. Indeed this was confirmed by the linear regression, which revealed that semantic compositionality on its own is the best predictor of priming effects and the possible underlying decomposition of Polish words. None of the other factors on its own can be used as a predictor of priming. For instance, items from condition 3 which have the semantic transparency score of 6.5 show no priming, but the items from condition 4 which have exactly the same score do show priming. Therefore semantic transparency cannot on its own be used as a predictor of lexical representation of Polish words which vary on the continuum of semantic transparency. However semantic transparency and morphological complexity when taken together can be predictors of lexical structure in Polish. As semantic compositionality seems to be the best predictor of lexical representation, the future research on the representation of Polish words needs to rely on a more sophisticated measure of semantic compositionality than the one used here. This measure would be based on a corpus statistic of Polish, taking into account, for instance, variables such as the consistency of mapping of an affix onto a meaning.

The puzzling result of no priming in condition 2, was investigated in a subsequent experiment which used the same paradigm. We used an improved set of items, excluding foreign borrowings. Significant priming of 14 ms was obtained. This result showed that lack of priming in condition 2 was most likely due to the specific items used and not to a failure of semantic compositionality to determine priming.

3. EXPERIMENT TWO

The aim of the second experiment was to investigate the representation of verbs and nouns with regular and irregular alternations. Irregularity in Polish verbs is defined on the basis of the alternation which occurs in the stem and not on the basis of which conjugational suffix is selected for a given verb. Polish allows here a tight comparison of regulars and irregulars, in contrast to English, because both regulars and irregulars share the same conjugational endings. The difference between regular and irregular verbs lies in that regular verbs exhibit an alternation which can be derived by the phonological rules of contemporary Polish and the alternations in the irregular verbs cannot.

According to the combinatorial model of the mental lexicon there should be priming in all conditions and there would be no difference in the magnitude of priming among all three verb conditions and among all three noun conditions. All alternants of the same stem, being semantically compositional would map onto the same lexical entry.

To test these predictions we included the following stimuli (1) 20 verbs with no alternation, e.g. czytasz ‘you read’/ czytać ‘to read’ were included as a paradigm check for verbs (2) 26 verbs with regular alternation, e.g. noszę ‘I carry’/ nosić ‘to carry’ were contrasted with (3) 26 verbs with irregular alternation, e.g. gnięsę ‘you bend’/ gnębić ‘to bend’. (4) 20 nouns with no alternation, e.g. placu ‘a square’, Locative/placu ‘a square, Nomina-
tive’ were included as a paradigm check for nouns, (5) 30 nouns with regular alternation, e.g. podzje ‘a leg, Da-
tive’/ podje ‘a leg, Nominate’ were contrasted with (6) 30 nouns with irregular alternation, e.g. księżyc ‘a prince, Genitive/księżyc ‘a prince, Nominate’ and (g) 20 words with phonological overlap, without semantic or morphological relationship, e.g. kotlet ‘chuck’/ kot ‘cat’, were designed to control for the phonological overlap of the target and prime.

Because no priming for English irregular past tenses was found in the cross-modal priming lexical decision paradigm, we used this technique for investigating Polish irregulars, to see whether they would correspond to English or Italian irregulars, which primed in this task. The experiment was run in Poland using DMASTER software.

3.1 Results

Three subjects from version 1 and one subject from version 2 were discarded from the analyses according to the same criteria as experiment 1. This gave 21 subjects per version. All the subjects were in their twenties and were native speakers of Polish who were studying in Poland. Out of the total of 580 items per version, 5 experimental items had to be discarded from the analyses; two because of high error percentage, two because they were erroneously classified as having a certain type of alternation and one item because of homophony. See Table 2 for the descriptive statistics.

The reaction time (RT) data were prepared for the analysis as described in experiment 1. The overall ANOVA with Prime (2 levels), Condition (7 levels) and Version (2 levels) revealed that that there was a facilitatory effect of Prime F2(1,153)=200.94, p<.001, F1(1,40)=160.22, p<.001. There was a main effect of Condition F2(6,153)=3.94, p<.01, F1(6, 240)=35.52, p<.001. More importantly, there was a two-way inter-
action of Condition and Prime F2(6, 153)=8.67, p<.001, F1(6,240)=160.22, p<.001. Subsequent analysis of the simple effects of Priming at each level of the Condition showed that there was a consistent facilitatory effect of Prime at all levels of Condition with the exception of the Phonological Overlap (p<.001 for both F1 and F2 throughout all analyses, with the exception of F1 for Irregular Alternation Verb, where p<.01). In order to explore whether there were any differences in the magnitude of priming between no alternation, regular alternation and irregular alternation the planned comparis-
sions were made separately for verbs and nouns. There was no significant difference in the magnitude of priming between all the verb and all noun conditions.

37
Table 2. Priming effects in Experiment 2.

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>PRM*</th>
<th>Priming (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 No Alternation Verb</td>
<td>20</td>
<td>0.8</td>
<td>83</td>
</tr>
<tr>
<td>2 Regular Alternation Verb</td>
<td>26</td>
<td>0.5</td>
<td>88</td>
</tr>
<tr>
<td>3 Irregular alternation Verb</td>
<td>26</td>
<td>0.3</td>
<td>73</td>
</tr>
<tr>
<td>4 No Alternation Noun</td>
<td>20</td>
<td>0.8</td>
<td>61</td>
</tr>
<tr>
<td>5 Regular Alternation Noun</td>
<td>30</td>
<td>0.6</td>
<td>62</td>
</tr>
<tr>
<td>6 Irregular alternation Verb</td>
<td>30</td>
<td>0.5</td>
<td>46</td>
</tr>
<tr>
<td>7 Phonological Overlap</td>
<td>20</td>
<td>0.7</td>
<td>-10</td>
</tr>
</tbody>
</table>

*PRM - Phonological Relatedness Measure between the primes and targets is the number of shared phonemes by prime and target divided by the mean phoneme length of the prime and target.

3.2 Discussion

Significant priming was found for regulars as well as for irregulars. More importantly, no differences were found in the magnitude of priming between items with no alternation, regular alternation and irregular alternation. No priming for the pairs with pure phonological overlap indicated that the observed priming cannot be attributed to phonological overlap. These results were in line with the predictions made by the combinatorial model of the mental lexicon. However, to make claims about the nature of the of the underlying representation one has to demonstrate that priming for all the morphologically related conditions, which in Polish, similarly to English, are also semantically related, is morphological and not for instance semantic. It is not possible to disentangle these in the current experiment which used a cross-modal paradigm.

Reid & Marslen-Wilson [2] followed up the three verb conditions in a further experiment, using an auditory-auditory delayed priming with a 12 item lag between prime and target, to dissociate the morphological and the semantic component. The experiment demonstrated that there was clear priming for all three verb conditions and no differences in the magnitude of priming between the conditions. These results allow us to make the stronger claim that the data on Polish words with alternations indicate that all alternants of the same stem map onto the same lexical entry.

It is worth asking whether these data can be accounted for by connectionist models in which morphology is not explicitly represented in the mental lexicon, but rather is an emergent property of a system where phonological representations interact with semantic representation [5]. The connectionist models apparently predict that, because significant differences were found in the phonological form between all the verb conditions and between most of the noun conditions, there would be significant differences in priming for all verb conditions, with the most priming for no alternation, less for regulars and least for irregulars. This prediction would be the same for nouns, with the exception that no difference in priming would be predicted for nouns with regular and irregular alternations. The finding of no difference in priming for forms significantly different in the phonological overlap is not directly predicted by current connectionist models.

4. GENERAL DISCUSSION

We have reported two experiments which examined the structure of the Polish mental lexicon. In the first experiment we concentrated on probing the representation of words which varied in semantic compositionality, semantic transparency and morphological complexity. The findings indicated that semantic compositionality on its own is the best predictor of priming effects and the possible underlying decomposition of Polish words similar to English and in contrast to Semitic languages. In the second experiment we investigated the representation of verbal and nominal stems with regular and irregular alternations. The results showed that there were no differences in the magnitude of priming between all the verb conditions and between all the noun conditions, despite the different types of alternation. This indicates that all alternants of a given stem map onto the same lexical entry. The results are in line with the predictions of the combinatorial mental lexicon and it is hard to see how they could be accounted for by connectionist models.

Overall, the data on the Polish mental lexicon provides support for the combinatorial mental lexicon, where morphologically complex, semantically compositional words are stored in a combinatorial fashion, with the stems and the affixes stored separately and the alternants of the same stems mapping onto the same lexical entry.

5. REFERENCES