

# PROCESSING OF MULTIMORPHEMIC WORDS IN HUNGARIAN<sup>1</sup>

Csaba Pléh and Levente Juhász  
Department of General Psychology  
Loránd Eötvös University, Budapest  
pleh@izabell.elte.hu

*Acta Linguistica Hungarica*, 1995, 43, 211-230

## 1. Basic issues in morphological processing in Hungarian

Hungarian provides good testing grounds to study some of the basic issues in lexical access and morphological decomposition in processing and representation. With its rich agglutinative structure accompanied with more and less productive allomorphy patterns it offers ample opportunities to raise and test quite straightforwardly some of the issues that are central in the contemporary literature on morphology processing. The following list is presented here as a sort of a general research program. The experimental work presented thereafter only touches upon a subsample of these issues up to now.

**1.1 Segmentation.** The parsing system in all languages can be expected to decide upon word boundaries at least for two reasons. First, in order to allow lexical access the system 'has to know' what string to look for in its lexical memory. Even if the search is incremental as some models like the cohort model of Marslen-Wilson and Tyler (1980) suggest word boundaries would be useful in deciding about the maximum possible string to be looked for. Second, segmentation would be of use in deciding what length of a sign-string the mental parser has to combine to arrive to the meaning and syntactic function of the given word form. This second function is an especially burdensome issue in an agglutinative language. Agglutinative languages may support these efforts by their structural features. E.g. the fixed first syllable stress in Hungarian gives a cue concerning access of the given stem and a 'backward sign' for compilation.

Spoken based models of lexical access all emphasize the importance of first syllable in access, which is our first proposed factor here (see Marslen-Wilson and Tyler, 1980, and the

---

<sup>1</sup> The experiments reported here were supported by the Hungarian National Science Foundation (OTKA) under project T 018173 to the first author. The authors profited from comments on earlier drafts of this paper by Wolfgang U. Dressler, Lőszly Kőlmön and Ferenc Kiefer.

volumes edited by Fraunfelder and Komisarejevsky-Tyler, 1987 and by Marslen-Wilson, 1989). Several observations also highlight the importance of stressed syllables on lexical search (Grosjean and Gee, 1987, Gósy, 1993). We suggest that the Hungarian system uses both principles - relying in search on first and on stressed syllables. Segmentation is possibly also supported by vowel harmony as a possible cue in the sense that all suffixes belonging to the same word form 'have to harmonize'. The appearance of a non harmonizing syllable as well as a word stress may serve as instructions to start a new word in the analysis of the input. It is noteworthy in this context that Peters (1985) in her review of segmentation issues in language acquisition also highlights the importance of first syllables, stable stress, and vowel harmony in the segmentation task facing the child.

**1.2. Lexical access.** It is a central theoretical issue in Hungarian processing what weight the system should allocate to access and to decomposition respectively. The usual processing issue of holistic and analytic approaches, total access or decomposition (for reviews see Aitchinson, 1987, Clark, 1991, and the volume edited by Sandra and Taft, 1994) has rather interesting implications if applied to Hungarian (Gergely and Pl;h, 1994). It seems to be natural that an entirely holistic approach to access in Hungarian would become enormously resource sensitive due to the large number of word forms around a given stem and also due to the unsaturation of the paradigms in language use. (Even in relatively large corpora only a few dozen forms of even the most frequent stems do show up, as Kornai, 1992 has recently pointed out). That would favor an analytic model. One could suggest, however, that even in the framework of a generally analytic model in some cases holistic access might be used. Several mixed models have been proposed in the literature e.g. for Slavic languages (Lukatela, Carello and Turvey, 1987, Kostic, 1994) and the mixed model has also been proposed as a general one favoring the holistic representation of frequent forms (Sternberger and MacWhinney, 1986).

The proposed mixed models are regularly related to derivation-inflection differences as well as suggesting that derived words would be processed as holistic units (for a review see Clark, 1991) especially in the case of non-transparent derivations (Marslen-Wilson *et al.*, 1994) while inflection would be treated in an analytic manner. It is of some interest to see whether there would be similar differences between derivational and inflectional suffixes in an agglutinative language (where these morphemes are subject to unifying morphotactic rules) in the preferred mode of access. Niemi *et al.* (1994) mainly on the basis of pathological data claim for Finnish

that while nominal inflections are treated in an analytic way, derivational suffixes form a holistic entry together with their stem.

One of the basic issues in our studies up to now was to see how differentially is morphological decomposition prevalent with different types of affixes, and whether there are signs of preferential processing orders in suffixes (like the 'bathtub effect' proposed by Aitchinson, 1987). Hungarian word formation rules allow us to study affixes that differ both positionally (prefixes and ordered suffixes), and functionally (derivational suffixes, plural and possessive markers, and case markers).

**1.3 Formal combinatorics.** The issue of analytic and holistic (as well as mixed) access is complicated by a further concern in agglutinative languages characterized by long multimorphemic words. How are morphemes integrated over a time scale during understanding? For a characterization of some of the theoretical possibilities listed here, see Gergely and Pléh (1994).

One option is a stepwise incremental left-to-right view: multimorphemic words 'are glued together' right away as the different morphemes come in. There is a constant search in a morpheme store (may be even in two stores, one corresponding to lexical words the other bound grammatical morphemes) and the accessed morphemes are integrated together right away. Another option is integration initiated by the word boundary marker. In this case compilation would wait until the end of the word. As one extreme one could even imagine to start lexical search for the stem also only after a word-end signal was received. This would of course be most inefficient. Its analysis, however, points out that in the temporal organization of access interesting asymmetries might be present in Hungarian. With regard to prefixes (a most prevalent affix type in Hungarian, see below) a stripping-access-combination cycle might be imagined while regarding suffixes the initiation of search would not require an 'active stripping' merely a lexical search for the 'remainder' of the word body.

The possible sequential differences also relate to the issue of possible typologically specific processing mechanisms as proposed by Gergely and Pléh (1994, Gergely, 1991, Pléh, 1989, 1990) suggesting a more analytic mechanism in Hungarian compared to e.g. English, or by Berwick (1991) claiming that non-configurational languages still have an order based strategy but on the level of words rather than phrases.

**1.4. Semantic integration.** Along with integration of forms one also needs to integrate the meaning of multimorphemic words. (As a matter of fact it is an empirical issue whether this goes on simultaneously with the integration of form or subsequent to it.) Some of the theoretically interesting empirical issues relate here as well to the temporal course of events. In listening to or reading a multimorphemic word like (1) are the corresponding conceptual elements such as units corresponding to 'plurality', 'Ego' and so on already active at the given earliest decision points, or are they activated only later on. Does their activation depend on transparency or on change in part of speech category (compare 1 and 2).

(1) *barát -sága -i -m -ban*  
friend-ship -Pl-1stSg-Inessive

(2) *ház- as- sága -i -m -ban*  
house-Adjder.-Nderiv - Pl-1stSg-Inessive  
married  
marriage

Does 'house' get activated at all in a non-transparent case like (2)? Is there an on line semantic integration corresponding to the proposed incremental formal compiler irrespective of semantic transparency ?

It is quite natural to expect differences related to morpheme types. The studies by Marslen-Wilson et al. (1994) had shown in English that in a cross-modal priming paradigm there were no signs of integration between the stem and the derived form in the case of non transparent derivations. They have also raised the issue what is the relationship between access/integration on the one hand and lexical representation on the other. It is possible in principle to have a direct access to *barátság* 'friendship' in (1) with no need to build it up starting from the two morphemes while on the level of lexical representation this form might still have a pointer to the word *barát* 'friend' that would fill it up with semantic content during speech understanding. However, does *házas* 'married' in (2) have a pointer towards the word *ház* 'house' even in a secondary way i.e. on the level of representation? From the point of view of language structure this suggests that in dealing with multimorphemic word forms psycholinguists have to differentiate between access and representation issues allowing most of all holistic access combined with analytic representation, and some kind of control over this by apparent (or naive) motivation of

derivational semantic relations (Dressler, 1989). Principles like the one proposed by Kiefer (1992, p. 197.) to the effect that opaque derivations are of the same order of semantic complexity as transparent ones may hold for representation but not for access. In access, opaque derivations might be very well less complex than transparent ones.

The general idea of a differentiation between access and representation issues may be rather directly connected to our experiments up to now. We were running lexical decision tasks that are most likely targeting slow (secondary) representation processes rather than access itself. Thus our results may not tell too much about fast access routes.

**1.5 Allomorphy.** Hungarian is infamous for its rich allomorphy relations (for linguistic characterizations see Papp, 1982, Kálmán, 1985, Nádasdy, 1985, for a psycholinguistic one MacWhinney, 1978). Quite a few studies on acquisition patterns - overgeneralizations, rules versus exceptions or items etc. - have applied different existing approaches to rule learning to Hungarian (MacWhinney, 1976, 1978, 1994, Réger, 1979). Some studies (MacWhinney, Pléh, and Bates, 1985, Pléh, 1989, 1990, Pléh, Jarovinskij, and Balajan, 1987) also showed that sentence processing was related to allomorphy: in interpreting simple transitive sentences both preschoolers and adults reverted to order based strategies - and thereby sometimes to mistakes - when accusatives of allomorphs that result in processing difficulties (consonant clusters) were in a non- prototypical linear position (as sentence initial objects).

The allomorphy issue, however, should also be taken up with regard to access and representation as a function of allomorphy. Some trivial issues to be studied are listed here. They should come as no surprise knowing the child language data, the history of stem classes and their possible relationship to the rule versus exception issue so central not only in contemporary linguistics at large but in psycholinguistic studies on morphology in particular (Pinker, 1991, Pinker and Prince, 1994, MacWhinney, 1994, Rebrus, 1992).

Are **allomorphs always mapped onto the same citation form** during access or are there different access files depending on frequency and phonetic motivation? E.g. just to take the accusative (-t) is the mapping the same in the case of *mókust* MÓKUS 'squirrel', *kutyát* KUTYA 'dog', a very productive lengthening type, *kezet* KÉZ 'hand', a no more productive shortening type, and *havat* - HÓ 'snow', a closed class of '-v insertion' that has less than a dozen members.

Is there a primary access of allomorph relevant to the given ending in entries with multiple allomorphs or are word forms directly mapped onto a dominant or abstract allomorph?

Thus, would there be more facilitation in cross-modal priming between the prime and the congruent allomorphs (*a* types in 3-4) compared to the non-congruent allomorphs (*b* types).

PRIME	TARGET	
(3) <i>ló-nál</i>	(a) <i>ló-hoz</i>	(b) <i>lov-ak</i> 'horse' with "at" "to", "PluralNom"
(4) <i>lo-vat</i>	(a) <i>lov-ak</i>	(b) <i>ló-nál</i> 'horse' with "Accus", "Plural", "at"

Is frequency relevant here? According to data from a Hungarian frequency dictionary (Füredi and Kelemen, 1989), in some forms the non-citation allomorph is more frequent. To take shortening stems as an example, with KENYÉR 'bread' the shortening allomorph *kenyere-* is twice as frequent as the citation allomorph *kenyér*, while with SÁR 'mud' the citation allomorph (*sár*) is much more frequent than the shortening one (*sara-*). The issue of primary access could be tested by using delays in priming (of a few hundred ms magnitude). If priming depends in the case of identical and different allomorphs on delays that would imply a primary access of allomorphs and a secondary access of the theoretical underlying form if any. Of course the whole issue is becoming even more intriguing in Hungarian if we consider that due to the morphemic-analytic nature of orthography in some cases we find a non-transparency (though a systemic one) between the acoustic and the orthographic forms. For English, Taft (1985) showed some effects of orthography. In Hungarian there are interesting regularities in orthography that should be exploited from a psycholinguistic point of view. Usually, orthography at morpheme boundaries is analytic in Hungarian as well as in other languages, i.e. it follows morphemic structure even if there is assimilation through regular morphonology. If the access is influenced by acoustics would expect to find a large priming difference between (5) and (6), the first one deviating from morpheme-grapheme-phoneme correspondence while being transparent in writing. Of course one would expect even more difference in cross-modal priming.

(5) *szabadság* [sabatsa:g] *szabad* [sabad] 'freedom - free'

(6) *butaság* [butasa:g] *buta* [buta] 'stupidity - stupid'

Do predictability relations related to the paradigms (Papp, 1982) have an effect on processing load and speed? E.g. in a shortening stem like KÉZ 'hand' the *keze-* and *kéz-* allomorphs have some differentiating value. Singular cases except the accusative go with *kéz-*

while *keze-* goes beside the singular accusative with all the plural forms plus with most of the possessive paradigm. In the productive lengthening paradigm like KUTYA 'dog', however, the *kutyá-* lengthened allomorph appears in all word forms but the singular nominative where the short allomorph is used. Thus, in this case, the use of one allomorph has little predictive value while in other paradigms it has. This might be relevant in the light of the approach taken by Kostic (1994) that tries to relate all morphology processing issues to information load considerations.

**1.6. The primary aim of our experiments.** Out of all these relevant issues of morphology processing in Hungarian our studies up to now mainly deal with the problems outlined under 1.2. and 1.3. Using rather slow processing measures and only visual presentation we were looking for **representation** differences between different morphemes both as a function of linear position and morpheme type. Since we were using a lexical decision type of task our results are also relevant to the problem of how 'formal compilation' goes on in processing isolated words.

We used the traditional *lexical decision paradigm* introduced to the study of morphological processing by Taft (1979). Words are presented one by one on a computer screen and the task of the subject is to tell (by pressing one of two buttons) whether they are indeed words in the language. This procedure has its well-known limitations: it is of course isolated word processing and only written processing. As the use of the cross-modal priming paradigm by Marslen-Wilson et al. (1994) indicates, however, quite similar effects show up in acoustic processing as well.

## **2. Experiment one: Differential treatment of different affixes**

In his pioneering studies on the decision task Taft (1979) has outlined the rather sophisticated negative type of argumentation characterizing the entire field since that time. He has shown that deciding about real, but still 'abstract stems' like *scribe* that they were not words took longer than deciding about non-stems like *lish*. The reason for this in the original model of Taft was straightforward: as (7a) shows *scribe* corresponds to a series of prefixed forms with the virtual stem while as (7b) indicates *lish* only appears in one form. Thus, for the (a) forms there is

a prefix-stripping strategy when we process the prefixed items. The subjects would in fact strip the prefix, take the remaining stem, decide about its legitimacy and then decide about the correctness of the combination of prefix and stem. This would show up in slower rejection times with forms like (7c) where only the combination is illegal but both morphemes are legal.

#### STEM Combinations

(7a) *scribe de-scribe, sub-scribe, pre-scribe*

(7b) *lish relish*

(7c) *scribe re-scribe*

It is noteworthy that compared to the separable and entirely productive prefixes in Hungarian and German most of the studies following Taft in English, French (Beuvillain, 1994), or Italian (Laudanna, Burani, and Cermele, 1994) are using abstract stems that are normally not free forms and at the same time the 'prefixes' (*de-*, *re-*, *sub-* etc.) are not free forms either. That might be quite relevant in interpreting some of our results.

Several studies have extended the lexical decision task and the prefix stripping model to the processing of suffixes. In general, most research has come up with interesting asymmetries suggesting e.g. that stripping was not characteristic of (at least inflectional) suffixes (Taft, 1985), that in slow secondary measures (comparing the identity of two visually presented words) the stem was a more important initiator for search than prefixes (Beuvillain, 1994). Some general models were also suggested claiming on the basis of mostly lexical decision studies that there is a general search economy difference between prefixes and suffixes. It is profitable to develop a strategy to strip prefixes to assist access because stem based search is more economical. Laudanna *et al.* (1994) would add that this is especially true if the given prefix is very frequent and the given word initial string is statistically very rare as a non prefix is. However, with suffixes that would not work: the stem would already be there directing search before we would be able to start stripping. General stripping with suffixes would delay search.

This leaves open, however, two basic issues that motivated our studies. First, what happens if you compare lexical decisions over comparable items in single study in a language

where the same forms can be combined with a rich variety of suffixes and prefixes. Second, even if suffixes are not stripped for search, some kind of decomposition might be important in the formal combination of the given word form into a 'legal form'.

## Methods

**Subjects and procedure:** 20 undergraduate psychology students (12 females, 8 males) between 18 and 23 participated in the experiment. They were instructed to read 'words' on a computer screen and decide if the string was indeed a word. They had to make decisions by pressing the I (*Igen*, yes) or N (*Nem*, no) letter on the keyboard of the computer.<sup>2</sup>

Stimuli were presented in a random order. Each stimulus stayed on the screen until the button pressing response. It was followed by the next stimulus with a delay of 1 sec. The program measured RTs in units of a hundredth of second. Subjects had to read 100 strings. 50 words were correct, 50 somewhere spoiled. Only data for the rejection times are used here.

### General structure of the words

All stimulus items (correct and incorrect alike) were formed on the structure of deverbal nouns, with prefixes, derivational suffixes, possessive markers and case, as (8) shows.

---

<sup>2</sup> An IBM PC-286 type computer was used with colored EGA screen. The experimental program that recorded reaction times was prepared by Zoltón Reményi and Tamás Szabó under the direction of Lószly Bernóth supported by the foundation 'For Hungarian Higher Education and Research'. We thank their courtesy here.

(8) *ki - véd- és - é - re* 'on its defense'

Pref V deverbil Poss Case

Regarding the linear structure of the words the ordering of the different morphemes was strict. This ordering is of course always fixed in Hungarian. Our subjects could build up specific strong expectations. The prefix always preceded the verbal stem, and after the stem the deverbil suffix - possessive marker - case marker order was strictly obligatory. It is important to note that all our words - as well as the non words - had basically the same structure and also the same length. This could have led to strong expectation effects and also to long reaction times with all words being six syllable long. That is much longer than the words used in most of present studies on morphology processing. All the 'spoiled' words contained only one mistake and the mistake was always changing one letter compared to the correct target.

**Stimulus list.** Table 1 shows examples and the logical structure of the stimulus list together with the number of tokens the given type was represented by and also the correct 'target word'. The table shows in bold the entire morpheme that was spoiled but it can be seen that in fact only one letter was changed.

Types of mistakes	spoiled	target	n	RT
1. Prefix comb.	<b>meg</b> -intézkedésében	??	5	2.66
2. Prefix nonexist.	<b>mag</b> -bízóimnak	<b>meg</b> -bízóimnak	5	1.92
		'to my trusters'		
3. Stem bad	be- <b>súd</b> -óinkat	be- <b>súg</b> -óinkat	5	2.33
	'our informantsACC'			
4. Derivation bad	elolva- <b>sán</b> -ával	elolva- <b>sás</b> -ával	5	2.36
	'by reading it'			
5. Poss incorre.	kifaragó- <b>e</b> -ként	kifaragó- <b>ja</b> -ként	5	<b>2.00</b>
	'as its carver'			
6. Case non-harm.	lemondásá- <b>ról</b>	lemondásá- <b>ról</b>	5	<b>2.37</b>

		'of his resignation'			
<b>7. Case misspelled</b>	elválásod- <b>gan</b>		elválásod- <b>ban</b>	5	2.18
		'in your separation'			
<b>8. Miscallenuous</b>	elválákodban	elválásodban	15	2.13	
<b>9. Correct</b>	elindulásakor			50	2.20

---

**Table 1:** The different types of stimuli used in Experiment I.

In (1) an existing prefix was combined with a form that is semantically banned. In (2) the prefix was misspelled resulting in a nonexisting morph. In (3) the same was done for the stem, and in (4) for the derivational suffix, in (5) for the possessive marker. We had two types of mistakes for case markers. (6) used an existing case marker that was, however, breaking vowel harmony, while (7) was a 'typo' leading to a non-existing suffix. Category (8) was in a way a filler. We tried to have exactly as many mistakes as we had correct forms. Therefore 15 further mistakes were introduced in order to supplement the 35 that were introduced systematically. It is slightly misleading to call them 'miscellaneous' because in an abstract sense they could all be classified under one of the above types.

## Results and discussion

Reaction time data were analyzed using multiple one-sample *t*-tests with the 3D program of the BMDP program package (Dixon and Horton, 1979). This rather questionable process was used because the comparisons were made only on the basis of correct responses. Therefore actual *N*s were different in different comparisons. Only rejection times will be considered. It is still worth to notice from the mean values presented in Table 1 two general facts. RTs were extremely slow the overall mean being well over 2 sec. If we compare this with the average times in similar experiments of the magnitude of 6-800 msec with three syllable words in Laudanna *et al.* (1994), or 1100-1330 msec with four syllable non-words in Taft (1994) it seems to be evident that our task was rather difficult. For the systematically varied rejections subjects needed over 450 msec for each syllable, while in the studies quoted a syllable needed 2-300 msec. However, in our material, each new syllable was a new morpheme. The long reaction times probably were spent both in reading and in morphological analysis. But across the board, these slow procedures certainly suggest that here we have data mainly on representation rather than access.

Another interesting overall aspect of our data is that accepting correct words was not faster compared to all mistakes. As a matter of fact, out of the 8 paired comparisons between the correct words and the different incorrect ones only one gave a significant difference. Non-existent prefixes indeed were rejected faster than correct words were accepted ( $t(19)=2.23, p < .05$ ). This overall relative slowness of the positive judgments suggests that due to the long words and the enforced analytic strategy our subjects were forced to read all items in a careful analytic way.

**Treatment of prefixes.** Existing prefixes that violated some combinatory restriction (1) took longer to reject than nonexisting ones (2). This 700 millisecond difference was statistically very reliable ( $t(19)=4.13$ ,  $p < .001$ ). That corresponds to the general prefix stripping idea of Tafts (1979) and Clark (1991). To use the more recent formulation by Sandra (1994) regarding prefixes a language like Hungarian follows the economy according to search (fewer units) rather than that of the economy of grammatical analysis. The new model of Taft (1994) presents a 'pseudoconnectionist' analysis of the stripping effects. He claims that in English one does not have to postulate a separate prefix store. Rather, a model with separate levels of representation (morphs, word forms, conceptual units) with a strong automatic activation between levels (*vive* activates *revive*) would be sufficient. However, for languages where prefixes have their own life (they are not only productive but also separable, show up as individual stress units) this would hardly work without supposing an individual storage for forms like *meg-*. We could interpret the extreme slow times with type (1) due to the classic prefix stripping ideas: one accesses both elements and finds a rule-based mismatch. But one could be a 'mellowed connectionist' like Taft (1994) seems to be, saying that the 700 milliseconds are needed for a search starting both from *meg-* and from *-intézkedésében* in a connectionist net. But even if you are replacing here rules with a failure of pattern matching in a connectionist net you still have to suppose 'stripping' i.e. a separate existence for the prefix units.

**'Stripping of case markers'.** There was a similar trend for case markers: rejection of existing but non-harmonizing case markers took almost 200 msec longer than rejecting non-existing case markers ( $t(19)=2.03$ ,  $p < .10$ ). Though this was only a statistical trend it would support the idea that at least in representation, i.e. with slow decision processes like the ones involved here, Hungarian nouns follow the 'stripping and checking' view (Sandra, 1994). This would mean here in the case of the non-harmonizing suffix an identification of the last syllable as a case marker and then checking its 'properness' vis a vis the preceding parts of the word. On this basis of course one cannot claim any relationships between phonological (i.e. harmony), formal (e.g. using verb inflections on nouns) or syntactic and semantic coherence. In order to study that further experiments using suffixes crossing part of speech boundaries and also violating ordering rules should be used.

**Stems and derivational suffixes.** There was no difference between mistake types (3) and (4) ( $t(19)=.25$ ,  $p >.80$ ). We tend to interpret this as an indication for a holistic or separate entry approach to derived stems in the lexicon as suggested by many previous models (for reviews see Sandra, 1994, Taft, 1994). The only novelty in our finding is that this effect shows up with long words with many suffixes both preceding and following the derivational suffix. However, this result does not in fact suggest any special search for derivational suffixes. The sequences **be-súd-ó** and **elolva-sán-á...** were only trivially differentiated in the design the first one being treated as a spoiled stem and the second one as a spoiled suffix. But as a matter of fact, one could say that the processor treats the sequence *olvasán* as a non-word i.e. as a 'non-stem'.

At the same time both spoiled stems ( $t=3.81$ ,  $p <.001$ ) and derivations ( $t=2.83$ ,  $p <.01$ ) took longer to be rejected than forms with non-existing prefixes. That is indirect support for the prefix-stripping idea: the stripped early part is easier to search in a possible special limited store while searching for the 'main body' of the word takes place over a much larger store.

**Phonetically wrong possessive markers** had the second fastest rejection rate. They were significantly faster than combinatorically inappropriate prefixes ( $t=4.26$ ,  $p <.0005$ ), wrong stems ( $t=2.59$ ,  $p <.02$ ), wrong derivational suffixes ( $t=2.72$ ,  $p <.001$ ), and both types of wrong case markers ( $t=2.97$ ,  $p <.01$  and  $t=2.41$ ,  $p <.05$ ). One could suggest a post hoc explanation for this effect. In this fixed position setup there had to be a possessive marker at the given spot. This provides the subject with a predetermined very small closed set to search through. If the system accessed the stem and sees that this cannot be a case marker (because the word is still longer), it would only have a small set to check. Similar data were obtained by Gergely and Pléh (1994) in a click detection time experiment in sentential contexts. Though all suffixes increased processing load showing up in slower reaction times to target clicks, possessive markers seemed to have a relatively smaller processing load.

### **Linearity and the bathtub effect**

Was there any clear linear effect? In a broad sense, yes. If we set aside possessives and only consider the strings at morpheme slots that are in themselves non-existent (i.e. if we only consider the search problem and not the combination) we find the following trend. **Prefix >**

**Case, Stem, Derivation.** Prefixes made for faster rejection than case markers ( $t=2.63$ ,  $p < .01$ ). At the same time it was faster to reject would be words on the basis of prefixes ( $t=3.81$ ,  $p < .001$ ,  $t=2.83$ ,  $p < .01$ ) compared to spoiled stems and derivations. Though non-existing case markers seemed to be faster their advantage over spoiled stems and word form internal derivational suffixes was not significant ( $p > .02$ ). Thus, there is some slight evidence for the 'bathtub effect' proposed by Aitchison (1987) but it mainly concerns primacy of the beginning of the word.

### 3. Experiment two: Positional effects and vowel harmony

Experiment One had several disadvantages that all have to do with the facts that search and position both were involved. The differences found between combination based (existing morphemes in wrong combination) and 'sign body' based mistakes (non-existing morphemes) raise the trivial empirical issue: what would happen if only one type of mistakes namely phonological combination would be involved in the combination mistakes.

In the second study the empirical issues represented by the design were simplified. The only spoiled forms used were strings with vowel harmony mistakes. This provided a possibility to study more neatly the linear ordering effects in morphological parsing since rejection on the basis of 'non-existence' was not possible here. Stems were always correct. Subjects had to make their decisions on the basis of non-compatibility of the morphs. The exclusive use of vowel harmony mistakes (rather than semantic or word-class combination mistakes) had a function as well: it set up a shallower processing attitude in the experiment.

## METHODS

**Subjects and procedure.** 20 undergraduate psychology students (10 males, 10 females between 19 and 23) had to decide on 84 words, half of them being correct half of them not. Details of the procedure were identical with those in Experiment One.

### General structure of the words used

(9) *nevet- és- ed ben*

V DerSuff Poss Case

The general structure of words corresponded to (9). All stimuli were multiply suffixed nouns. In comparison with the previous study, however, the material was more varied in length and complexity. Sometimes the words were prefixed (though prefixes were not experimental variables) and due to singular-plural variations in the possessive marking world length varied between 4 and 8 syllables, most of the words being 6 syllables long. Mean word length was 5.7 syllables, 5.8 for correct words and 5.7 for incorrect ones.

### **Stimulus list**

The structure of the stimulus list is shown in Table 2 together with the mean reaction times. (The 1s show the linear position of the vowel harmony mistake)

Types of mistakes	form	gloss	n	RT
000: correct	futásod-dal	'with your run'	42	1.61
001: case is bad:	futásod-del	'with your run'	6	1.73
010: poss is bad:	kiáltás-em-ra	'to my shout'	6	1.62
011: case poss :	olvasás-ük-re	'to their reading'	6	1.62
100: derivation:	ír-and-aimból	'from my writings'	6	1.60
101: der and case:	húz-and-á-ben	'in his pull'	6	1.56
110: der and poss:	fest-ás-aink-re	'to our paintings'	6	1.53
111: all suffixes:	hat-and-ed-nek	'to your impact'	6	1.42

Table 2: The structure of words in the second study together with mean rejection times

## Results

A 3 way analysis of variance was performed on the reaction times. Mean RTs for each type of target were used in each subject. The three within subject factors were the three types of suffixes: derivational suffix, possessive, and case. Each factor had two levels: right or wrong. This model has two disadvantages. It underestimates the weight of the correct items their number being 42 while that of each incorrect type being represented by 6 tokens. Second it directly compares rejection and acceptance time which is a questionable practice.

Table 3 summarizes the effects found in the analysis of variance.

Effect	df	F	p
Derivational suffix	1,19	8.03	.01
Possessive marker	1,19	7.08	.02
Case marker	1,19	< 1	n.s.
Der x Poss	1,19	< 1	n.s.
Der x Case	1,19	6.08	.02
Poss x Case	1,19	2.86	n.s.
Der x Poss x Case	1,19	< 1	n.s.

Table 3: Effects on reaction times in the vowel harmony experiment

**General observations.** RTs were quite slow in this study as well. The mean RT for rejections was 1.58 sec while for accepting the correct items the mean RT was 1.61 sec. This is however, by far not as slow as results in Experiment One. When subjects had to make rejections only on the basis of a 'shallow' phonological feature, mean times needed by syllable went down to the usual 2-300 msec range quoted above. (The actual mean by syllable was 233 ms.) Rejection of incorrect strings on the whole was of the same magnitude as acceptance of correct items. At the same time, similar to the results in Experiment One some mistakes were rejected faster than the mean correct acceptance times. Here this was mainly true for cumulative cases with several

mistakes piled up. Forms with three mistakes took 190 msec less time to reject than correct forms to be accepted.

Two of the three morpheme types had a significant main effect. If the derivational suffix was spoiled mean RTs went down to 1.53 sec compared to 1.64 if the derivational suffix was correct. If the possessive marker was spoiled again RTs were faster, 1.55 se compared to 1.62 when the possessive was correct. Only the last possible suffix had no main effect. With case markers the spoiled forms had a mean RT of 1.56 sec while the the non-spoiled forms showed a mean RT of 1.59 sec.

The strong interaction of Case with Derivation, as well as the main effects found can be summarized into some meaningful patterns.

**Linearity.** If only one suffix was wrong there was a clear linearity effect observed cases taking the longest time to be rejected. The linear order was 1.60, 1.62 and 1.73 sec. This is not trivial because it does not support the slight 'bathtub effect' found in our first study. It seems to be that a bathtub effect only prevails if we consider non-existing morphemes. As we turn to incompatibility decisions based on shallow features a clear left-to-right linearity show up. The earlier a mistake is the faster it is rejected.

**Stem closeness.** If the first mistake is next to the stem (i.e. the derivational suffix does not harmonize) rejection is faster. Mean times being 1.53 and 1.66 for the relevant mistakes (for all the mistakes that do involve derivational suffixes compared to those that do not). This is the reason for the strong interaction between Derivation and Case.

## **General discussion**

The two experiments presented here are only the first steps in the more ambitious program outlined in the introductory section of this paper. The rather off-line and methodically questionable lexical decision types of studies showed a few interesting methodical and a few substantial results. Regarding methodology our studies showed that in an agglutinative language if lexical decisions involve both search and grammatical-semantic combinations reaction times become extremely slow. it is apparent that in these cases one can only study secondary representation processes rather than lexical access itself an the role of morphological parsing in it. The overall times are very sensitive to the general arrangement. If only combinatory mistakes are present and them on a rather shallow level processing becomes much faster.

With the provision that we might very well limited our scope to secondary analysis due to the nature of our task, we could still claim a few substantial results. In Hungarian both prefix stripping and case stripping were present. Also, decision times with non-existing morphemes indicate that search according to a fixed word structure is influenced by the relevant set size. That implies that people in a reanalysis of the input consult different types of stores one corresponding to lexical stems the other to grammatical morphemes.

On the basis of Experiment Two we can also claim that the important integrator of word forms in Hungarian, namely vowel harmony can become the basis of systematic decisions during reanalysis features of this reanalysis being combinatorics and linear effects.

However, in order to see whether all this combinatorics of Hungarian morphological parsing is a real element of the primary access process one would need to turn to more indirect on-line methods and to the study of continuous speech processing.

## References

- Aitchison, J. (1987): *Words in the mind*. London: Blackwell
- Beuvillain, C. (1994): Morphological structure in visual word recognition: Evidence from prefixed and suffixed words. *Language and Cognitive Processes*, 9, 317-340
- Berwick, R. C. (1991): Principle-based parsing. In: Sells, P., Shieber, S.M. and Wasow, T.: *Foundational issues in natural language processing*. Cambridge: MIT Press, 115-226
- Clark, V.: *Morphology and the mind*. London, Routledge, 1991
- Dixon, N. and Horton, L. (1979): *The BMDP Series manual*. Berkeley: University of California Press
- Dressler, W. (1989): Prototypical differences between between inflection and derivation. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung*, 42, 3-10
- Fraunfelder, U. H. and Komisarevsky Tyler, L. (1987, ed.): *Spoken word recognition*. Cambridge, Mass.: MIT Press
- Füredi, M. and Kelemen, J. (1989): *A mai magyar nyelv szépprózai gyakorisági szótára*. (Frequency dictionary of present day literary Hungarian.). Budapest: Akadémiai Kiadó

- Gergely György (1991): *Free word order and discourse interpretation*. Bp. Akadémiai
- Gergely, Gy. and Pléh, Cs. (1994): Lexical processing in an agglutinative language and the organization of the lexicon. *Folia Linguistica*, 28, 175-204
- Gósy, Mária (1993): A lexikális hozzáférés. In: Gósy, M. and Siptár, P. (eds.): *Beszédkutatás 1993*. Bp.: MTA Nyelvtudományi Intézete, 14-32
- Grosjean, F. and Gee, J. (1987): Prosodic structure and spoken word recognition. In: Fraunfelder, U. H. and Komisarevsky Tyler, L. (1987, ed.): *Spoken word recognition*. Cambridge, Mass.: MIT Press, 135-155
- Kálmán László (1985) Inflectional morphology. In: Kenesei, I. (ed.): *Approaches to Hungarian. Vol One*. Szeged: JATE, 247-263
- Kiefer, F. (1992): Hungarian derivational morphology, semantic complexity, and semantic markedness. In: Sag, I.A. and Szabolcsi, A. (eds.): *Lexical matters*. Palo Alto, Ca: Stanford University Press, 183-208
- Kornai, A. (1992): Frequency in morphology. In: Kenesei, I. and Pléh, Cs. (1992)(eds.): *Approaches to Hungarian. Vol. Four*. Szeged: JATE, 247-268
- Kostic, A. (1994): Information load constraints on processing inflected morphology. In: Feldman, L.B. (ed.): *Morphological aspects of language processing* g. Hillsdale: Erlbaum, 319-346
- Laudanna, A., Burani, C., and Cermele, A. (1994): Prefixes as processing units. *Language and Cognitive Processes*, 9, 295-316
- Lima, S.D., Corrigan, R.L. and Iverson, G.K. (1994). *The reality of linguistic rules*. Amsterdam: Benjamins
- Lukatela, G., Carello, C. and Turvey, M.T. (1987): Lexical representation of regular and irregular inflected nouns. *Language and Cognitive Processes*, 2, 59-63
- Lukatela, G., Kostic, A., Feldman, L. B. and Turvey, M. T. (1983): Grammatical priming of inflected nouns. *Memory and Cognition*, 11, 59-63.
- MacWhinney, B. (1975): Rules, rote, and analogy in morphological formations by Hungarian children. *Journal of Child Language*, 2, 65-77
- MacWhinney, B. (1978): The acquisition of morphophonology. *Monograph of the Society for Research in Child Development*. 43
- MacWhinney, B. (1994): Rules vs. cues: The baby and the bathwater. In: Lima, S.D., Corrigan, R.L. and Iverson, G.K.: *The reality of linguistic rules*. Amsterdam: Benjamins

- MacWhinney, B. Pléh, Cs., and Bates, E. (1985): The development of sentence interpretation in Hungarian. **Cognitive Psychology**, 17, 178-209
- Marslen-Wilson, W. D. (1989, ed.): *Lexical representation and process*. Cambridge, Mass.: MIT Press
- Marslen-Wilson, W. D. and Tyler, L. K. (1980): The temporal structure of spoken language understanding. *Cognition*, 8, 1-71.
- Marslen-Wilson, W. D. Tyler, L. K. , Waksler, R., and Older, L. (1994): Morphology and meaning in the English mental lexicon. *Psychological Review*, 101, 3-33
- Nádasdy Ádám (1985): Segmental phonology and morphophonology. In: Kenesei István (ed.): *Approaches to Hungarian. Vol. One*. Szeged: JATE, 225-246
- Niemi, J., Laine, M., and Tuominen, J. (1994): Cognitive morphology in Finnish. *Language and Cognitive Processes*, 9, 423-446
- Peters, A.M. (1985): Language segmentation: Operating principles for the perception and analysis of language. In: Slobin, D. I. (ed.): *The crosslinguistic study of language acquisition. Vol. 2.*, Hillsdale, NJ: Erlbaum, 1029-1064
- Papp Ferenc (1982): *A magyar f\_név paradigmaticus rendszere*. Bp.: Akadémiai
- Pinker, S. (1991): Rules of language. *Science*, 253, 530-535
- Pinker, S. and Prince, A. (1994): Regular and irregular morphology and the psychological status of rules of grammar. In: Lima, S.D., Corrigan, R.L. and Iverson, G.K. (eds.): *The reality of linguistic rules*. Amsterdam: Benjamins, 321-350
- Pléh, Cs. (1989): The development of sentence interpretation in Hungarian. In: MacWhinney, B., and Bates, E.(Eds.): *The crosslinguistic study of sentence processing*. Cambridge: Cambridge University Press, 158-184
- Pléh, Cs. (1990): Word order and morphological factors in the development of sentence understanding in Hungarian. *Linguistics*, 28 1449-1469
- Csaba Pléh (1992): Verbal prefixes in Hungarian children. In: Kenesei István and Pléh Csaba (szerk): *Approaches to Hungarian. Vol IV*. Szeged: JATE , 269-281
- Pléh, Cs. Jarovinskij, A., and Balajan, A. (1987): Sentence comprehension in Hungarian-Russian bilingual and monolingual preschool children. **Journal of Child Langauge**, 14, 587-603
- Rebrus Péter (1994): Connectionist models and theoretical linguistics. In: Bánréti, Z. (ed.): *Papers in the theory of grammar*. Budapes: MTA Nyelvtudományi Intézete, 80-103
- Réger Zita (1979): Bilingual Gypsy children in Hungary. *International Journal for the*

*Sociology of Language*, 19, 59-82

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

Sandra, D. and Taft, M. (1994, eds.): Morphological structure, lexical representation and lexical access. *Language and Cognitive Processes*, 9. *Special Issue*

Slobin, D. (1985a, ed.): *The crosslinguistic study of language acquisition. Vol. 1-2.*

Hilssdale, NJ: Erlbaum

Slobin, D. (1985b, ed.): Why study acquisition crosslinguistically? In: Slobin, D.I. (ed.): *The crosslinguistic study of language acquisition. Vol. 1.*

Hilssdale, NJ: Erlbaum, 3-24

Slobin, D. (1985c): Crosslinguistic evidence for the language-making capacity. In: Slobin, D. I. (szerk): *The crosslinguistic study of language acquisition. Vol. 2.,*

Hilssdale, NJ: Erlbaum, 1157-1256

Sternberger, J. P., and MacWhinney, B. (1986): Frequency and lexical storage of regularly inflected forms. *Memory & Cognition*, 14, 17-26.

Stone, G.O.: Combining connectionist and symbolic properties in a single process.. In: Lima, S.D., Corrigan, R.L. and Iverson, G.K. (1994). *The reality of linguistic rules.* Amsterdam: Benjamins, 417-443

Taft, M. (1979): Recognition of affixed words and the word frequency effect. *Memory and Cognition*, 7, 263-272.

Taft, M. (1985): The decoding of words in lexical access: A review of the morphographic approach. In: Besner, D., Waller, T.G. and MacKinnon, G.E. (Eds.): *Reading research: Advances in theory and practice. Vol. V.,* New York: Academic Press

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