THE EFFECT OF MORPHOLOGICAL COMPLEXITY ON SHORT-TERM MEMORY CAPACITY*

DEZSŐ NÉMETH[a] – ROZÁLIA ESZTER IVÁDY[b] –
ALESSANDRO GUIDA[c] – MÁRTON MIHÁLTZ[d] –
DONALD PECKHAM[e] – ATTILA KRAJCSI[a, f] – CSABA PLÉH[b]

[a]Institute of Psychology, University of Szeged, Hungary

[b]Department of Cognitive Science, Budapest University of Technology
and Economics, Hungary

[c]University Rennes 2, France

[d]GeoX Ltd., Hungary

[e]Department of English Language Teacher Education and Applied Linguistics,
University of Szeged, Hungary

[f]Department of Radiology, Brigham Women Hospital, Harvard Medical School

Corresponding author: Dezső Németh, address: Egyetem u. 2.,
H–6722, Szeged, Hungary, e-mail: nemethd@edpsy.u-szeged.hu

Abstract: The main purpose of this study was to explore the relationship between verbal short-term memory and the morphological complexity of words. Hungarian, as an agglutinative language, is of special interest for psycholinguistic inquiries in morphology. The authors presented two word-list recall experiments. The recall of the word list was measured by the classical span design. The item lists consisted of two-syllable stems (base words) and two-syllable morphologically complex words (stem + suffix). Within each list the words were of the same length, the same phonological structure (CVCVC), the same frequency and the same concreteness. The same experimental design was used with three-syllable words as well. Results indicated that morphological complexity had a significant negative effect on short-term memory span, and that memory was better for derived words (e.g., *boy + hood*) than inflected words (e.g., *boy + s*), and regular than irregular words.

Keywords: morphological complexity, short-term memory, chunks, phonological loop

Morphological complexity is a common topic of psycholinguistic studies, and how we parse, access and represent morphologically complex words

in the mental lexicon are key issues, yet this has rarely been a subject of study in memory research. The present study puts forward research that crosses sub-disciplines: namely short-term memory (STM) research and psycholinguistics. How do morphologically complex forms affect STM capacity? Do morphologically parsable and unparsable words behave differently in STM tasks? To our knowledge, only very few studies have dealt with this topic at all, and thus only a limited amount of data are available regarding the immediate recall of morphologically complex words. We present results from two experiments to help bridge the gap that exists at the confluence of these two sub-disciplines.

1. STM Limitations

Two different types of verbal STM limitations seem to appear when reviewing the literature of immediate recall (e.g., Chen–Cowan 2005): a phonological limitation on the one hand a limitation in terms of chunks on the other. We will start presenting the latter one.

1.1. Limitations in terms of chunks

Miller (1956) famously proposed that the amount of information that can be kept in mind at one time is about seven chunks or meaningful units of information. The chunking mechanism was initially described by de Groot (1978) and Miller (1956), and then by Chase–Simon (1973, also see Gobet–Simon 1996b, for a revision of the chunking theory). A current definition is given by Gobet et al. (2001, 236), who refer to a chunk as "[. . .] a collection of elements having strong associations with one another, but weak associations with elements within other chunks". To explain chunks, Miller (1956) and subsequently Cowan (2001) used the same example. When the letters "fbiibm" are presented, if one knows the acronyms "FBI" and "IBM," then it is possible to simplify the information by forming two chunks ("FBI" and "IBM") in STM. Because these familiar patterns exist in long-term memory (LTM)—FBI is the Federal Bureau of Investigation and IBM is the well-known computer company—the letters "f", "b", "i", "i", "b" and "m" can be encoded as two elements in STM instead of six.

The above example is also useful in understanding how STM capacity is determined by LTM knowledge. In fact, unlike the number of chunks,

which seems to be constant (see the constant-capacity hypothesis, Cowan et al. 2004)—even when knowledge grows (e.g., Gobet 1998; Lane et al. 2001; Gobet–Simon 2000)—and determined by attentional resources (for a review, see Cowan 2005), the size of chunks seems to be determined by one's LTM knowledge, therefore chunks can become larger with practice (Chen–Cowan 2005; Cowan et al. 2004) and expertise (e.g., Chase–Simon 1973; Gobet–Simon 1996a;b), increasing the size of STM capacity.

Although Miller's (1956) famous paper set STM limits in terms of chunks, only few studies (e.g., Johnson 1969; Slak 1970; Tulving–Patkau 1962; Zhang–Simon 1985) have investigated verbal limitations in terms of chunks, mainly because of methodological issues as put forward by Chen–Cowan (2005; see also Cowan et al. 2004). The major problem (see also for this concern, Cowan 2001; 2005) is that it is difficult to know if a participant has encoded in STM different elements as different elements or as one element by chunking the information together. As mentioned earlier, this depends on the use of LTM knowledge. But it does seem that if one controls for the impact of LTM knowledge (see Cowan's four principles, Cowan 2001) then the STM capacity is about three chunks.

1.2. Limitations in phonological terms

The alternative way of studying STM limitations is in phonological terms. This approach has been rendered popular by Baddeley's seminal model (1986; 2000), which incorporates a phonological loop. Verbal information in the phonological loop is subject to time-decay unless refreshed through rehearsal. The amount of verbal information that can be rehearsed with no decay is limited as it corresponds to people's verbal memory capacity. Baddeley et al. (1975) showed that this roughly corresponds to the amount of information that can be overtly or covertly rehearsed in 2 seconds. In the same paper, they showed that long words—which take more time to articulate—are remembered more poorly than short words, an effect they called the **word length effect**.[1] This was put forward to sup-

---

[1] It is important to mention, as put forward by Chen and Cowan (2005), that there is still a debate as to whether the word length effect results from a time-based limit due to the decay of phonological information (e.g., Cowan et al. 1997; Hulme–Tordoff 1989; Mueller et al. 2003; Schweickert–Boruff 1986) or if it is based on the amount of phonological material to recall (e.g., Caplan–Waters 1994; Cowan et al. 1997; Lovatt et al. 2002; Neath–Nairne 1995; Service 1998).

port the phonological loop model (for a more recent account of the word length effect in terms of articulation duration, see Mueller et al. 2003).

While discussing the STM limitations in phonological terms, one should also consider the effect of LTM (e.g., Jefferies et al. 2004), exactly like above in the case of chunks. In this case, the effect associated with LTM knowledge is lexicality (word-form). This effect can be observed, for example, with tasks in which two lists of unintelligible words are to be memorized, one comprised of sequences that match the phonological pattern of the mother tongue, the other one comprised of sequences that do not match. The first list is easier to reproduce. The same effect is also observed when contrasting actual words with pseudo-words, the first are easier to recall than the second. The process whereby lost phonological material is restored on the basis of materials stored earlier (that is, knowledge) or with the help of lexical or phonological features of a language is called redintegration (Schweickert 1993). This reconstruction process may take place in the storage and recall phases of memory (Gathercole 1999; Hulme et al. 1999; Thorn et al. 2005).

Although the two types of STM limitations we have presented—chunks and phonological effects—are indeed related to the same STM, only few experiments addressed both of them in a same set of experiments (e.g., Zhang–Simon 1985) specifically to study the interplay of these limitations. Recently, Chen and Cowan (2005) have tried to do so. They investigated this question by teaching new paired associations between words to create 2-word chunks (slightly modifying the methodology of Cowan et al. 2004). Chen and Cowan (2005) showed that the importance of chunk-based versus phonological length-based limits in recall depends on a combination of scoring, nature of recall (free vs. serial recall) and list length. For example, results are best described by the chunking hypothesis in free recall for 6 chunks (6 pairs of words) and, in contrast, results are best described by the phonological length-limit hypothesis for a serial recall (with a strict serial order scoring) of 4 chunks (4 pairs of words). But intermediate results are also obtained. Chen and Cowan (*op.cit.*, 1246) seemed to suggest that STM limitations could be of a dual nature (chunk-based vs. phonological length-based) "perhaps not unlike the finding in physics that light behaves like a wave under some circumstances and like a particle under other circumstances".

In conclusion, it appears that the concepts of chunk and phonological representation are important in order to give a comprehensive explanation of STM limitations.

2. Morphological complexity

Although the effect of word frequency—for example—on STM is well known (e.g., Hulme et al. 1997), little is known about the effect of morphological complexity. The definition of morphological complexity relies on the number and type of morphemes which comprise a word. Morphemes, the smallest units of language having a meaning or signifying grammatical information, are the basic building blocks of words. Base words themselves are comprised of one or more morphemes which can, in most cases, stand alone, while affixes only appear attached to base words. Affixes can traditionally be described as derivational or inflectional in nature. This distinction between types of affixes is important, as they signify different morphological processes: when derivational morphemes are added to a word, a new word is created, typically of a different word class (e.g., $read + er$), while when inflectional morphemes are added to a word, additional grammatical information is given which does not change the class of the word itself (e.g., $read + ing$). Relevant to the present study are Hungarian examples showing derivational and inflectional morphology, which can be seen in (1) (for further details on Hungarian morphology, the reader can refer to Kiefer 2000; MacWhinney 1985). (1a) illustrates how derivational endings change the word class and the basic meaning of derived words, while (1b) shows that the inflectional endings provide grammatical meaning such as plurality and location. In our study only derivational and inflectional suffixes (rather than prefixes) were used, representing by far the most common method of affixation for both types of morphemes in Hungarian. It should also be noted, as is reflected in the examples, that Hungarian is an agglutinative language, where each morpheme is generally associated with one meaning.

(1) (a) Words created through derivation:[2]

igaz 'true' → igaz + ság 'truth'
igazság 'truth' → igaz + ság + os 'fair'
olvas 'read' → olvas + ó 'reader'
olvasó 'reader' → olvasó + i 'of the reader'

(b) The addition of inflectional morphology:

kalap 'hat' → kalap + ok 'hat + pl' ('hats')
kalapok 'hats' → kalap + ok + ban 'hat + pl + sup' ('in hats')
bor 'wine' → bor + ok 'wine + PL' ('wines')
borok 'wines' → bor + ok + at 'wine + pl + acc' ('wines')

---

[2] *-ság/-ség* and *-ó* are morphemes creating nouns, while *-os* and *-i* create adjectives. pl = plural, sup = superessive case, and acc = accusative case.

Concerning the use of rules in morphological word formation, in agglutinative languages like Hungarian with rich and generous morphology, an important issue is whether all possible forms of a word (with their derivational and inflectional affixes attached) are represented in the lexicon or whether the stem of a word is viewed as a starting point for rules to be applied in order to generate further forms. In addition, there is an issue as to whether allomorphs, i.e., alternative stem (or suffix) forms, are represented separately (see (2) below for examples in Hungarian). According to the full form model (also called the holistic model), each inflected form has a separate lexicon input (Butterworth 1983; Rubin et al. 1979). Thus, referring back to (1) above, our mental lexicon would represent each inflected form of *kalap* 'hat' separately. This would generate an extremely large lexicon, especially in a language like Hungarian which has as many as 26 separate case suffixes for nouns (Kenesei et al. 1998). Therefore, storing all of possible words of a language with a rich morphology is a problematical hypothesis to defend. Conversely, the decompositional models (the analytical view) admit separate representations only for word stems, while all other forms are generated by applying rules to the stem.[3]

Of course, a number of solutions offering a compromise have been proposed which approach the shift between decomposition (analytical) and the full form model (holistic) in terms of frequency, type of suffix, and semantic transparency. For example, recently Järvikivi and Niemi (2002) have argued, on the basis of their experiments on priming in Finnish, that noun stem variants work as units of input in this agglutinative language, and that they have a predominant intermediating role between lexemes and word forms. According to Pinker (1991; 1997; 1998) and Ullman (2001), following the dual route model, the analytical–holistic feature is also related to regularity. Grammar is a rule-based entity which corresponds to a kind of procedural memory, and it includes the manipulation of morphologically regular forms. In contrast, irregular forms and, perhaps, a part of generated words (those with vague meanings, cf. Marslen-Wilson et al. 1994) are instead stored holistically in line with associative storage, presumably by way of the associative system of the temporal lobe (see for example Lukács–Pléh 1999).

Furthermore, concerning regularity in word formation, Hungarian is a language that also merits consideration as some of its irregular

---

[3] A comprehensive summary of the decomposition and the full form model is offered in Feldman (1995).

and unproductive cases differ from their English counterparts in several respects. In English, irregular forms, in particular strong verbs, are usually those used most frequently (e.g., *go → went; do → did; eat → ate, is/are → was/were*). In Hungarian, this is not always the case, for example, the word *szú* 'wormwood' → *szuv + ak* 'wormwood + pl' ('wormwoods') is used quite rarely.

(2)   Examples of regularity and irregularity in Hungarian inflectional
       suffix attachment

       (a)   Regular, productive suffixation

              No change in stem: ablak 'window' → ablak + ok 'window + pl' ('windows')
              Low vowel lengthening: róka 'fox' → róká + k 'fox + pl' ('foxes')

       (b)   Irregular, unproductive suffixation

              Root final syllable shortening: madár 'bird' → madar+ak 'bird + pl' ('birds')
              /v/ insertion: kő 'stone' → köv + ek 'stone + pl' ('stones')
              Stem epenthesis: bokor 'bush' → bokr + ok 'bush + pl' ('bushes')

Lukács and Pléh (1999) demonstrated that variation in frequency, irregular paradigm, and the size of the set acted independently of one another. Their priming results suggests that Hungarian affixation uses processing based not on just a single rule, but rather on the formation of frequent but irregular nouns (like the vowel-shortening and the vowel-dropping nouns), probably assisted by application of a number of sub-rules.

3. STM capacity and morphological complexity

With two experiments, the present study aimed at assessing the effects of various morphological factors on STM capacity, as well as the effects of word length and word frequency (both stem and surface frequency). The morphological factors considered were the complexity of words, types of affixes (derivational vs. inflectional affixes), and additionally, in the case of inflectional suffixes, the regular or irregular character of affixation on word span. In addition, we tried to compensate for the effects of other presumed factors. These included word frequency, word concreteness and related phonological structure (affecting ease of pronunciation).

   From a psycholinguistic point of view, the basic issue in experiments 1 and 2 was whether morphological processing, i.e., the decomposition of words into morphemes, took place in cases where this was not required owing to the fact that morphological structure is irrelevant in STM tasks

as no grammatical decisions are to be made. If this is the case, one should expect a shorter STM span in the case of decomposable words (suffixed words) when compared with STM span for less morphologically complex words (stems), which should put a lower load on STM.

This hypothesis is interesting because it allows the separation of Baddeley's (1986) tripartite model from the chunking theory (Chase–Simon 1973; Gobet et al. 2001; also see a revision of the chunking theory, Gobet–Simon 1996b) in terms of expectancy. In fact, Baddeley's phonological loop does not take morphological complexity into account, and therefore an effect of morphology cannot be accounted for by Baddeley's phonological loop, whereas it is compatible with the chunking theory given the fact that the material used (stems vs. suffixed words) had comparable phonology (words being balanced in terms of word length, ease of pronunciation, frequency, and concreteness) but differed in terms of the number of morphemes (i.e., units of language having a meaning or signifying grammatical information), which can also be analyzed as a difference in terms of chunks (i.e., meaningful units of information).

In our first experiment, we also manipulated the type of frequency with which the lists were matched, that is, in terms of stem frequency and in terms of surface form frequency. This was done in order to find the best matching method in order to use only one matching technique in the second experiment. Similarly, we also presented the words in an auditory fashion and in a visual fashion in order to find the most suitable presentation for the purpose of the study.

In experiment 2, the study of decomposable words (suffixed words) has been taken further in order to analyze the impact of derivational morphology vs. inflectional morphology on STM performance and the impact of regular morphological forms vs. irregular morphological forms.

Concerning the latter contrast, our hypothesis stems from Pinker's dual model (Pinker 1997; Ullman 2001). In line with this model, irregular forms, as representing separately stored and non-synthesized units, should apply a lighter load on STM in English, while regular forms as the latter presumably require the operation of a procedural memory process as well—in other words, we can memorize fewer regular forms owing to greater computational demand. This hypothesis could also be compatible with the chunking theory (Chase–Simon 1973; Gobet et al. 2001), since irregular forms are composed of fewer chunks than regular ones, while it runs against Baddeley's (1986; 2000) phonological loop, since the words were balanced in terms of word length, ease of pronunciation,

frequency, and concreteness. However, in Hungarian the processing of irregular forms is more difficult than that of regular ones. Hungarian irregular nouns exhibit idiosyncratic morpho-phonological properties, that is, the change(s) that the stem undergoes when combining with a suffix are item-specific/non-productive. Thus, as compared to inflected regular nouns, inflected irregular nouns involve not only combination with a suffix but, additionally, an idiosyncratic change in the stem. For example, to create the plural form *lovak* 'horses', the uninflected stem *ló* 'horse' must undergo modification to *lov* prior to the attachment of the suffix *-k*. So in Hungarian the irregular forms may load the STM more than in English.

Concerning the derivational vs. inflectional contrast, we expected that our derived words would cause better performance than the inflected words on the STM task, since derived words count as one chunk (one element in the mental lexicon), while in inflected ones there are as many chunks as morphemes, or in other words, as seen above, inflected words are supposed to put an extra load on the procedural system.[4]

4. Experiment 1

4.1. Method

4.1.1. Participants

164 undergraduate students at the University of Szeged participated in the study as partial fulfilment of course requirements. 113 took part in the auditory condition and 53 in the visual condition. Their ages ranged between 19–23 years, all of them had good hearing, and none had language disorders. They were all native speakers of Hungarian.

4.1.2. Word stimuli

When designing the word span task, various aspects were taken into consideration: frequency was controlled for, as well as the semantic attributes and phonetic attributes of the words. The frequency of the words

---

[4] We thank an anonymous reviewer for proposing this hypothesis and his valuable comment that this hypothesis is true only if the words in the task (which is the case in our experiment) are not derived from the same stem like *asztalos*, *asztali*, *asztalság* and if in the inflectional condition, they are not inflectional variants of the same stem, like *asztalok*, *asztalon*, *asztalnál*.

was similar, based on a spoken and written corpus, they had the same syllabic structure (e.g., all were CVCVC, or CVCVCVC in the case of three syllable words), they were all nouns with a concrete (not abstract) meaning, and lists were designed so as not to contain words of similar sound structure (i.e., excluding rhyme and alliteration or other similarities). Thus the word material for the experiments had to satisfy three constraints, which were solved by using computational methods. First, words in the lists had to be matched in frequency of occurrence in colloquial Hungarian. This was established by using frequency counts from the 150-million-word Hungarian National Corpus (Váradi 2002). The words used had a frequency of 950–4500 in the corpus. The second restriction called for semantic symmetry among items: words should have relatively equal pictorial value. We solved this by filtering the nouns in the list by only allowing items present in certain concrete semantic categories in the Hungarian WordNet electronic thesaurus (Miháltz et al. 2008). Finally, the third constraint—phonological balance—required items in lists to have similar phonological structure. We achieved this by the further filtering of the lists by simple pattern matching that grouped words together with similar consonant–vowel syllable structure. At the end of the process, the lists were checked manually in order to single out potential problems originating from the automatic filtering methods. (See the Appendix for a list of example stimuli.)

4.1.3. Word span task

Word lists were constructed in lengths from 3 to 7 words, two lists for each length, with separate lists for each of the following five word types: (1) two-syllable base words; (2) two-syllable inflected words; (3) three-syllable base words; (4) three-syllable inflected words comprised of a two-syllable stem + 1 suffix; and (5) three-syllable words comprised of a one-syllable stem + 2 suffixes.

There were two conditions in the word span task, auditory and visual. In the auditory condition the lists of words were read out loud, while in the visual condition stimuli were presented one by one on the screen of a PC in 44-point, black, Times New Roman letters on a white background with the help of the Microsoft Power Point computer program.

Scoring followed the classical span method. The score indicates the correctly recalled list length. For example, 4 points was given when the subject correctly recalled the 4-word lists every time, but he or she failed to recall the lists containing 5 words. 4.3 points was given when a subject

correctly recalled all the lists containing 4 words, and one of the 5-word lists but, failed on the other two 5-word lists. This kind of fractional span score is useful for obtaining greater resolution.

4.1.4. Experimental design

A word-span measuring procedure was used with a $2 * 2 * 2$ factor design involving the following independent variables: word length (2 vs. 3 syllables), matching frequency type (lists based on matching by surface form frequency vs. stem frequency) and morphological complexity (stems vs. suffixed words). There was also a between-subjects factor: modality, whether words were either presented auditorily (read aloud) or visually (on the computer screen). The dependent variable was the number of words subjects could recall correctly (see above for the scoring).

4.1.5. Procedure

The serial order of the tasks was counterbalanced. Before the word-span test we gave the participants the following instructions: "You will hear/read a list of words one by one. Listen/watch carefully, because I want you to repeat this list as accurately as you can and in the same order". The experimenter read aloud the lists of words, or alternatively the subjects read them on a computer monitor. In the case when the experimenter read the word lists, there was a one second pause between the words, while in the case when they had to read it from the computer screen, there was a one second pause between words, which were visible for 2 seconds on the screen. The last word in the list was followed by a picture that indicated to the participants that they had to repeat the list immediately in the same order and using the exact morphological form. Results were recorded by the experimenter by writing the **number of words** and the **number of stems** that were recalled correctly. The latter was needed because in pilot studies it became apparent that subjects can frequently recall stems, but mix up or mis-attach suffixes. If there was an error in the order of the words it was considered to be a completely wrong answer.

4.2. Results

We tested whether the factors of type of frequency, word length, morphological complexity and modality had an effect on the number of words subjects could recall correctly.

The ANOVA analysis showed that the effect of morphological complexity was significant, $F(1, 116) = 227.48$, $p < 0.001$, $\eta_p^2 = 0.58$. The average STM performance for stems was higher than the performance for suffixed words (number of recalled words: 4.54 vs. 3.85). This confirmed our expectation that suffixed words were more difficult to recall. Results concerning the word length factor showed a better recall for 2 syllable words than for 3 syllable words (4.07 vs. 3.79), $F(1, 116) = 43.42$, $p < 0.001$, $\eta_p^2 = 0.26$. There was also an effect of modality, $F(1, 113) = 9.26$, $p < 0.001$, $\eta_p^2 = 0.23$; subjects performed better in the auditory condition than in the visual condition (4.32 vs. 3.77). The results are summarized in Figure 1.
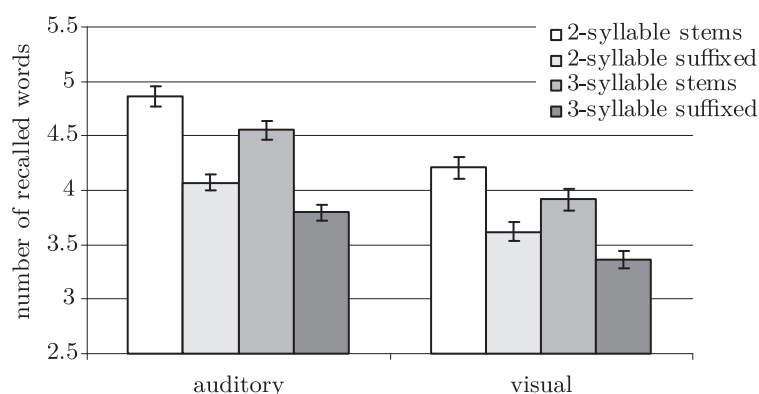

Fig. 1

Differences in word span—modality, word length and suffixes (these results are matched on surface form frequency). Post-hoc analyses revealed that all variables differed from each other in the auditory and visual modalities separately (all p's < 0.02). Comparing the two modalities, there were differences in all cases (all p's < 0.001). Error bars indicate standard error of mean (SEM).

Our results also indicated that the type of frequency was also important: matching members of the list on surface form frequency resulted in better memory performance than matching on stem frequency (3.80 vs. 3.47), $F(1, 162) = 60.01$, $p < 0.001$, $\eta_p^2 = 0.27$. It is important to point out —with regards to the second experiment—that there was a significant interaction between frequency type and the number of suffixes as well, $F(2, 65) = 29.18, p < 0.001$, $\eta_p^2 = 0.27$.

A second ANOVA was computed on 3-syllable words only. It showed that the number of suffixes put a cumulative load on STM: the more suf-

fixes there were on a stem in a three-syllable word, the more difficult it was to recall it, $F(2, 65) = 129.22$, $p < 0.001$, $\eta_p^2 = 0.80$. In other words, the subjects' performance on word lists which were created with 3-syllable stem words was better than 3-syllable word containing one suffix—that is, a 2-syllable stem + 1 suffix. And this kind of list was better recalled than a list containing 3-syllable words with 2 suffixes (number of recalled words: 4.54 vs. 3.73 vs. 3.07, respectively)—that is, a 1-syllable stem + 1-syllable suffix + 1-syllable suffix (all p's < 0.02). The ANOVA also showed that matching on surface form frequency resulted in better memory performance than stem frequency (3.55 vs. 3.22), $F(1, 66) = 18.92$, $p < 0.001$, $\eta_p^2 = 0.22$. As is shown in Figure 2, the better memory performance when lists were matched on surface form comes mainly from the two-suffix words. There was also an interaction between frequency type and the number of suffixes, $F(2, 65) = 12.88$, $p < 0.001$, $\eta_p^2 = 0.28$. As suggested in Figure 2, on lists matched on stem frequency it was increasingly difficult to recall words as the words became more and more complex (stem < stem + 1 suffix < stem + 2 suffixes) in both frequency types.
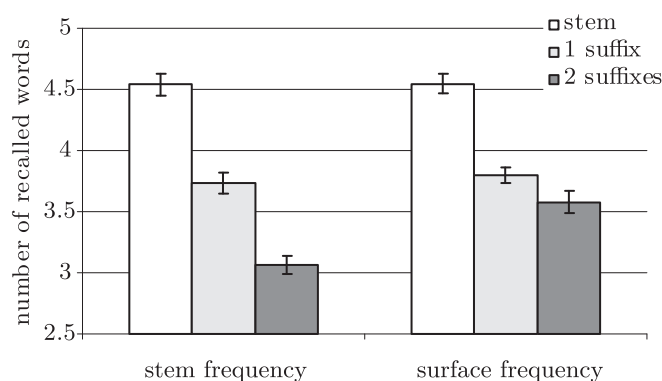

Fig. 2

The cumulative effect of suffixes on three syllable words on lists matched on stem frequency and on surface form frequency. Post-hoc analyses revealed that all variables differed from each other in the stem frequency and surface frequency conditions separately (all p's < 0.02). Comparing the two frequency types, there was a trend in the 1-suffix form ($p = 0.059$) and a significant difference in the 2-suffix form ($p < 0.001$). Error bars indicate SEM.

5. Experiment 2

In the second experiment we tried to clarify results from experiment 1. We formed different lists of words based on morphological features. We no longer created separate lists for surface form frequency and stem frequency matched words because we found differences between the frequency types in Experiment 1. These differences showed mainly in the two-suffix word list caused by the increase of variability that comes along with the increase of the number of suffixes (see the discussion below). Because we cannot control this variability effect, we matched the word list on surface frequency only.

In Experiment 2 we go beyond the previous experiment by investigating the effect of derivational/inflectional morphology and of regularity on STM performance.

5.1. Method

5.1.1. Participants

39 undergraduate students at the University of Szeged participated in the study as partial fulfilment of their course requirements. Their ages ranged between 18–24 years, all of them had good hearing and none had language disorders. They are all native speakers of Hungarian.

5.1.2. Word span task

See experiment 1.

5.1.3. Experimental design

Our independent variables were morphological complexity (zero, one or two suffixes), type of morphological suffixes (inflectional or derivational), and regularity (regular vs. irregular types). The dependent variable was the number of words subjects could recall correctly.

5.1.4. Procedure

A PC was used to present the stimuli using the same program and configuration as were used in the previous experiment. The procedure was the same as in the visual condition of the first experiment.

5.2. Results

As in Experiment 1, an ANOVA was performed on the number of correctly recalled items. Like in the previous experiment, the effect of word length was significant $F(1, 38) = 31.32$, $p < 0.001$, $\eta_p^2 = 0.45$. Three syllable words were more difficult to remember than two syllable words (number of recalled words: 3.96 vs. 3.51). There also was an effect of morphological complexity, $F(1, 38) = 49.664$, $p < 0.001$, $\eta_p^2 = 0.68$. Two-suffix words were more difficult to remember than one-suffix words, and these were more difficult to remember than stems (2.69 vs. 2.995 vs.3.71, respectively, all p's < 0.01). Thus, the replication of the results of former experiment was successfully carried out despite the smaller number of participants.

Regularity had a significant main effect: regular words were easier to recall than irregular words (3.41 vs. 3.14), $F(1, 38) = 11, 43, p = 0.002$, $\eta_p^2 = 0.23$.

Derivational vs. inflectional morphology also showed a significant difference between them, $F(2, 37) = 49.560$, $p < 0.001$, $\eta_p^2 = 0.23$: subjects were able to recall more words with derivational suffixes than with inflectional ones (4.12 vs. 3.41). As this was true for 2-syllable words and 3-syllable words as well (see Figure 3), we might think of this as a general pattern.
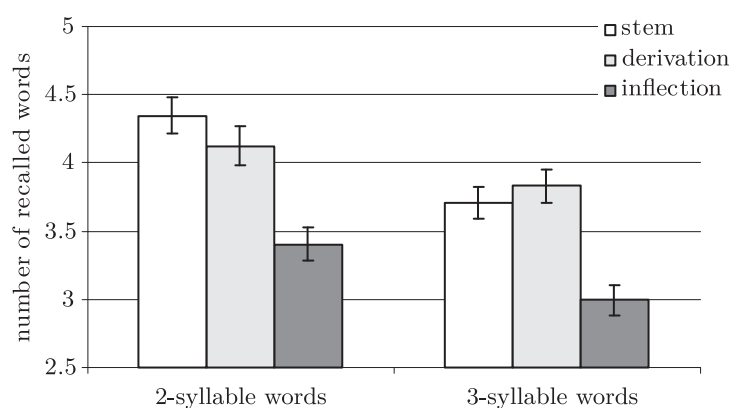

Fig. 3

Differences between derivational and inflectional morphology. Post-hoc analyses revealed that all variables differed from each other in the 2-syllable and 3-syllable conditions (all p's < 0.03), except the 3-syllable stem and derivational forms ($p = 0.22$). Comparing the two word length conditions, there was a trend in the 1-suffix form (p=0.059) and a significant difference in the 2-suffix form ($p < 0.001$). Error bars indicate SEM.

Even if there are similarities between derivational and inflectional morphologies, there are also differences, in fact the interaction between morphological complexity and morphological type of suffixes (inflectional or derivational) is significant, $F(2, 37) = 6.35, p = 0.004, \eta_p^2 = 0.25$. It seems that derivation is less sensitive to suffix loss or change: subjects do not attach different derivative suffixes, while they are likely to add different inflectional suffixes. Furthermore, there is a significant difference between derivation and inflection on the lemma levels as well.

6. Discussion

Experiments 1 and 2 tested if in Hungarian, morphological complexity had an effect on STM performance, in order to learn if morphological complexity was a relevant dimension in terms of STM span.

Experiment 1 and Experiment 2 clearly showed that in Hungarian, suffixed words were more difficult to remember than stem words. The same result had already been obtained by Service–Tujulin (2002) with Finnish stimuli. As an extension of Service and Tujulin's study, our material contained one and two-suffix words, therefore we were able to observe that the number of suffixes in a word had an impact on STM performance. To wit, not only was it easier for our participants to recall stem-words compared to one or two-suffix words, but the results followed a gradient. The more suffixes a word was composed of the more difficult it was to recall.

In addition, our study showed that this same pattern of results was observed both for words presented visually and auditorily. For the two modalities, the pattern of results proved to be similar, which seems to argue in favour of modality independency.

In theoretical terms, the effect of morphological complexity on STM span does not seem compatible with Baddeley's phonological loop (1986; 2000) since this effect was observed while the word length, ease of pronunciation, frequency, and concreteness dimensions were controlled. Therefore it does not seem possible to account for the impact of morphological complexity on STM performance in phonological terms. On the other hand, it does seem possible to account for the effect of morphological complexity in terms of chunks (Chase–Simon 1973; Gobet et al. 2001), since a difference in terms of the number of morphemes (one stem, stem + one suffix, and one stem + two suffixes) also implicates a difference in

terms of the number of chunks. In line with the decompositional theory of the lexicon, it is generally supposed that when comprehending a morphologically complex word we decompose it into smaller parts (Taft 1979). A morphologically complex word is thus not only a single word, but several morphemes bound together. Each supplementary morpheme (a unit of language having a meaning or signifying grammatical information) thus imposes supplementary information, which can be coded in terms of chunks. Therefore each piece of supplementary information will occupy more space thus allowing less resources in STM, something which could easily explain our results.

It is also important to mention that if an interpretation in terms of chunks can explain our results, then it is not incompatible with the new version of Baddeley's model (Baddeley 2000; 2002) which incorporates an episodic buffer. According to Baddeley (2000), an episodic buffer holds integrated representations from different modalities (phonological, spatial, visual), and binds information into a unitary episodic representation. The episodic buffer, unlike the phonological loop (Baddeley 2001, 118), "is assumed to have a limited capacity determined by number of chunks", therefore—*mutatis mutandis*—now the explanation that used to work only for the chunking theory (Chase–Simon 1973; Gobet et al. 2001) also works for Baddeley's model (2000).

Alternatively, another way of interpreting the effect of morphological complexity could be in terms of predictability cue value. When we try to recall a word list, it is always a bit of a restoration work: we might not remember the exact word as the representation is fading, leaving only some semantic residue, or the first syllable (though we do not necessarily suppose it has to be the first). Yet, suffixed words are not so easy to recall if we only remember a part of the word. Unless we remember most of the last syllable, the suffix is lost. On the other hand, if we only remember the last syllable there is no way we can find out what the word is, since that is the first syllable. So a 2-syllable stem is a lot more predictable from its parts than a 1-syllable stem + 1-syllable suffix would be—whichever part remained.

Experiments 1 and 2 also showed that word length had a large effect on the number of words one can recall, confirming that we are able to remember shorter words more easily than longer ones, which is not surprising, because it is in line with the phonological loop theory (Baddeley–Hitch 1974; Baddeley 2001).

Our results also indicated that the type of frequency had an effect on memory performance: surface form frequency resulted in better memory performance than stem frequency. However, looking at our results more precisely, it appeared that this was true mainly for two-suffix words. We interpreted this effect in terms of increase of variability that comes along with the increase of the number of suffixes. In fact, when there is only one suffix, the difference between the "stem frequency" and the "surface frequency" is only increased by one suffix; however, in the case of two suffixes, the difference between the "stem frequency" and the "surface frequency" is increased by two suffixes, which increases the probability of observing a difference.

Concerning the difference between inflection and derivation, experiment 2 has shown that derivations seemed to be more easily stored in STM. This result was expected through the framework of the chunking theory. A derived word is one chunk—one item in mental lexicon. Our result is compatible with Service and Tujulin's (2002) results in Finnish. The authors showed that inflected words reduced the span performance to a greater extent than did derived words.

Our last comment concerns the regularity vs. irregularity result. As we put forward earlier, it is generally supposed that when comprehending a morphologically complex word we decompose it into smaller parts (Taft 1979). Naturally this does not always work with irregulars, since they cannot always be decomposed (e.g., $be \rightarrow was$). One of the most accepted models is a mixed one where irregular forms (the frequent or very exceptional ones at least), non-transparent derivations, and frequently used inflectional forms are stored as a whole, while not very frequent regular forms and transparent derivations are stored analytically (Marslen-Wilson et al. 1994). Our results clearly showed that regular words were easier to recall than irregular words. This pattern seems to imply that Pinker's (1997) dual-route model and Taft's (1979) decompositional model made incorrect predictions in this case. These models emphasize the extent to which irregulars are stored by rote. The background of the difference between our findings and the predictions of dual-route models can be found in the nature of Hungarian irregular morphology. This is likely to reflect the increased difficulty involved in processing Hungarian irregulars which require both storage and combinatorial processes, as opposed to regulars which require only the latter.

7. General discussion

Theories of STM have typically been developed using uninflected words from languages like English (or in a minor way, German, Italian, Spanish, or French, for example). Restricting these models to uninflected words made sense at the beginning of the development of these theories. But now we have enough facts on the table to begin to worry about how STM theories can deal with complex words. We agree that if the target language is English or Chinese, then language itself is nothing much more than a list of uninflected words. But when you move to languages like Hungarian or Japanese, then nearly every other word has some form of complex morphology.

Our next goal will be to carry out reaction time measurements as well, in order to find out the amount of cognitive resources we need when recalling suffixed lists. We also will attempt to carry out the same experiment with prefixes to see if there is a difference compared to suffixes.

References

Baddeley, Alan David 1986. Working memory. Clarendon Press, Oxford.

Baddeley, Alan David 2000. The episodic buffer: A new component of working memory? In: Trends in Cognitive Sciences 4 : 417–23.

Baddeley, Alan David 2001. The magic number and the episodic buffer (Commentary on "The magical number 4 in short-term memory: A reconsideration of mental storage capacity" by N. Cowan). In: Behavioural and Brain Sciences 24 : 117–8.

Baddeley, Alan David 2002. Is working memory still working? In: European Psychologist 7 : 85–97.

Baddeley, Alan David – Graham Hitch 1974. Working memory. In: Gordon H. Bower (ed.): The psychology of learning and motivation (Vol. 8), 47–89. Academic Press, New York.

Baddeley, Alan David. – Neil Thomson – Mary Buchanan 1975. Word length and the structure of short-term memory. In: Journal of Verbal Learning and Verbal Behavior 14 : 575–89.

Butterworth, Brian 1983. Lexical representation. In: Brian Butterworth (ed.): Language production, Volume 2, 257–94. Academic Press, London.

Caplan, David – Gloria Waters 1994. Articulatory length and phonological similarity in span tasks: A reply to Baddeley and Andrade. In: Quarterly Journal of Experimental Psychology 47 : 1055–62.

Chase, William Gary – Herbert A. Simon 1973. The mind's eye in chess. In: William Gary Chase (ed.): Visual information processing, 215–81. Academic Press, New York.

Chen, Zhihijan – Nelson Cowan 2005. Chunk limits and length limits in immediate recall: A reconciliation. In: Journal of Experimental Psychology: Learning, Memory, and Cognition 31 : 1235–49.

Cowan, Nelson 2001. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. In: Behavioral and Brain Sciences 24 : 87–185.

Cowan, Nelson (ed.) 2005. Working memory capacity. Psychology Press, New York.

Cowan, Nelson – Zhijian Chen – Jeffrey N. Rouder 2004. Constant capacity in an immediate serial-recall task: A logical sequel to Miller (1956). In: Psychological Science 15 : 634–40.

Cowan, Nelson – Noelle L. Wood – Lara D. Nugent – Michel Treisman 1997. There are two word-length effects in verbal short-term memory: Opposed effects of duration and complexity. In: Psychological Science 8 : 290–5.

Feldman, Laurie Beth (ed.) 1995. Morphological aspects of language processing. Lawrence Erlbaum, Hillsdale, New Jersey.

Gathercole, Susan E. 1999. Cognitive approaches to the development of short-term memory. In: Trends in Cognitive Sciences 3 : 410–9.

Gobet, Fernand 1998. Memory for the meaningless: How chunks help. In: Morton Ann Gernsbacher – Sharon J. Derry (eds): Proceedings of the Twentieth Meeting of the Cognitive Science Society, 398–403. Lawrence Erlbaum, Hillsdale NJ.

Gobet, Fernand – Peter C. R. Lane – Steve Croker – Peter C.-H. Cheng – Gary Jones – Iain Oliver – Julian M. Pine 2001. Chunking mechanisms in human learning. In: Trends in Cognitive Sciences 5 : 236–43.

Gobet, Fernand – Herbert A. Simon 1996a. Recall of random and distorted positions: Implications for the theory of expertise. In: Memory and Cognition 24 : 493–503.

Gobet, Fernand – Herbert A. Simon 1996b. Templates in chess memory: A mechanism for recalling several boards. In: Cognitive Psychology 31 : 1–40.

Gobet, Fernand – Herbert A. Simon 2000. Five seconds or sixty? Presentation time in expert memory. In: Cognitive Science 24 : 651–82.

Groot, Adriaan D. de 1978. Thought and choice in chess (Original publication in 1946.). Mouton de Gruyter, The Hague.

Hulme, Charles – Philip Newton – Nelson Cowan – George Stuart – Gordon Brown 1999. Think before you speak: Pauses, memory search, and trace redintegration processes in verbal memory span. In: Journal of Experimental Psychology Learning, Memory, and Cognition 25 : 447–63.

Hulme, Charles – Steven Roodenrys – Richard Schweickert – Gordon D. Brown – Sarah Martin – George Stuart 1997. Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. In: Journal of Experimental Psychology 23 : 1217–32.

Hulme, Charles – Vickie Tordoff 1989. Working memory development: The effects of speech rate, word length, and acoustic similarity on serial recall. In: Journal of Experimental Child Psychology 47 : 72–87.

Järvikivi, Juhani – Jussi Niemi 2002. Form-based representation in the mental lexicon: Priming (with) bound stem allomorphs in Finnish. In: Brain and Language 81 : 412–23.

Jefferies, Elizabeth – Ralph Matthew A. Lambon – Alan David Baddeley 2004. Automatic and control processing in sentence recall: The role of long-term and working memory. In: Journal of Language and Memory 51 : 623–43.

Johnson, Neil F. 1969. The role of chunking and organization in the process of recall. In: Gordon H. Bower – Janet T. Spence (eds): Psychology of learning and motivation (Vol. 4.), 171–247. Academic Press, Oxford.

Kenesei, István – Robert M. Vago – Anna Fenyvesi 1998. Hungarian. Routledge, London & New York.

Kiefer, Ferenc (ed.) 2000. Strukturális magyar nyelvtan 3. Morfológia [A structural grammar of Hungarian 3. Morphology]. Akadémiai Kiadó, Budapest.

Lane, Peter C. R. – Fernand Gobet – Peter C.-H. Cheng 2001. What forms the chunks in a subject's performance? Lessons from the CHREST computational model of learning (Commentary on "The magical number 4 in short-term memory: A reconsideration of mental storage capacity" by N. Cowan). In: Behavioural and Brain Sciences 24 : 128–9.

Lovatt, Peter – S. E. Avons – Jackie Masterson 2002. Output decay in immediate serial recall: Speech time revisited. In: Journal of Memory and Language 46 : 227–43.

Lukács, Ágnes – Csaba Pléh 1999. Hungarian cross-modal priming and treatment of nonsense words supports the dual-process hypothesis (Commentary on "Lexical entries and rules of language: A multidisciplinary study of German inflection" by H. Clahsen). In: Behavioral and Brain Sciences 22 : 1030–1.

MacWhinney, Brian 1985. Hungarian language acquisition as an exemplification of a general model of grammatical development. In: Dan Isaac Slobin (ed.): The cross-cultural study of language acquisition, 1069–155. Lawrence Erlbaum, Hillsdale NJ.

Marslen-Wilson, William – W. D. Lorainne Komisarjevsky Tyler – Rachelle Waksler – Lianne Older 1994. Morphology and meaning in the English mental lexicon. In: Psychological Review 101 : 3–33.

Miháltz, Márton – Csaba Hatvani – Judit Kuti – György Szarvas – János Csirik – Gábor Prószéky – Tamás Váradi 2008. Methods and results of the Hungarian WordNet project. In: Attila Tanács – Dóra Csendes – Veronika Vincze – Christiane Fellbaum – Piek Vossen (eds): Proceedings of the Fourth Global WordNet Conference, 311–21. University of Szeged, Szeged.

Miller, George A. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. In: Psychological Review 63 : 81–97.

Mueller, Shane T. – Travis L. Seymour – David E. Kieras – David E. Meyer 2003. Theoretical implications of articulatory duration, phonological similarity, and phonological complexity in verbal working memory. In: Journal of Experimental Psychology: Learning, Memory, and Cognition 29 : 1353–80.

Neath, Ian – James S. Nairne 1995. Word-length effects in immediate memory: Overwriting trace decay theory. In: Psychonomic Bulletin and Review 2 : 429–41.

Pinker, Stephen 1991. Rules of language. In: Science 253 : 530–5.

Pinker, Stephen 1997. Words and rules in the human brain. In: Nature 387 : 547–8.

Pinker, Stephen 1998. Words and rules. In: Lingua 106 : 219–42.

Rubin, Gary S. – Curtis A. Becker – Roger H. Freeman 1979. Morphological structure and its effects on visual word recognition. In: Journal of Verbal Learning and Verbal Behaviour 18 : 757–67.

Schweickert, Richard 1993. A multinomial processing tree model for degradation and redintegration in immediate recall. In: Memory and Cognition 21 : 168–75.

Schweickert, Richard – Brian Boruff 1986. Short-term memory capacity: Magic number or magic spell? In: Journal of Experimental Psychology: Learning, Memory, and Cognition 12 : 419–25.

Service, Elisabeth 1998. The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. In: Quarterly Journal of Experimental Psychology: Section A 51 : 283–304.

Service, Elisabeth – Anna-Marie Tujulin 2002. Recall of morphologically complex forms is affected by memory task but not dyslexia. In: Brain and Language 81 : 42–54.

Slak, Stefan 1970. Phonemic recoding of digital information. In: Journal of Experimental Psychology 86 : 398–406.

Taft, Marcus 1979. Recognition on affixed words and the word-frequency effect. In: Memory and Cognition 7 : 263–72.

Thorn, Annabel. S. C. – Susan E. Gathercole – Clive R. Frankish 2005. Redintegration and the benefits of long-term knowledge in verbal short-term memory: An evaluation of Schweickert's (1993) multinomial processing tree model. In: Cognitive Psychology 50 : 133–58.

Tulving, Endel – John E. Patkau 1962. Concurrent effects of contextual constraint and word frequency on immediate recall and learning of verbal material. In: Canadian Journal of Psychology 16 : 83–95.

Ullman, Michael Thomas 2001. A neurocognitive perspective on language: The declarative/procedural model. In: Nature Review Neuroscience 2 : 717–26.

Váradi, Tamás 2002. The Hungarian National Corpus. In: Mark T. Maybury (ed.): Proceedings of the Third International Conference on Language Resources and Evaluation, 385–9. European Language Resource Association, Las Palmas.

Zhang, Guojun – Herbert A. Simon 1985. STM capacity for Chinese words and idioms: Chunking and acoustical loop hypothesis. In: Memory and Cognition 13 : 193–201.

Appendix

A sample list of two-syllable stems and suffixed words (abbreviations: acc = accusative, pl = plural, sup = superessive. Note that suffixes might have different allomorphs):

Stems	Gloss
tükör, szobor, csomag, sziget	mirror, statue, pack, island
motor, terem, kakas, szövet	motor, room, cock, cloth
talaj, folyam, lemez, patak, fodor	soil, river, disc, creek, ruffle
fotel, vödör, kavics, bokor, bagoly	armchair, bucket, pebble, bush, owl
fogoly, cukor, kalap, terep, küszöb, hotel	captive, sugar, hat, land, doorstep, hotel
nyereg, szatyor, malac, konyak, salak, szivar	saddle, bag, pig, cognac, cinder, cigar
doboz, liget, gyomor, szalon, meder, majom, gödör	box, grove, stomach, saloon, water-course, monkey, hole
leves, tücsök, rovar, mecset, robot, horog, zsilip	soup, cricket, insect, mosque, robot, hook, flood gate
szalag, bajusz, köröm, terasz, lepel, szirom, torok, kölyök	ribbon, mustache, nail, terrace, sheet, petal, throat, kid

Suffixed words	Word structure	Gloss
falat, borok, hegyen	fal-at, bor-ok, hegy-en	wall+acc, wine+pl, mountain+sup
nyakon, hajat, porok, sörök	nyak-on, haj-at, por-ok, sör-ök	neck+sup, hair+acc, powder+pl, beer+pl
botot, sasok, rizsek, tököt	bot-ot, sas-ok, rizs-ek, tök-öt	stick+acc, eagle+acc, rice+pl, marrow+acc
tejek, magon, hasat, fülön, fogak	tej-ek, mag-on, has-at, fül-ön, fog-ak	milk+pl, seed+sup, tummy+acc, ear+sup, tooth+pl
dobon, gyepen, bakot, dögöt, csapot	dob-on, gyep-en, bak-ot, dög-öt, csap-ot	drum+sup, lawn+sup, buck+acc, carcass+acc, tap+acc
halon, romon, lyukat, padon, zsebet, vajat	hal-on, rom-on, lyuk-at, pad-on, zseb-et, vaj-at	fish+sup, ruin+sup, hole+acc, bench+sup, pocket+acc, butter+acc