Pauli Brattico, Alina Leinonen and Christina M. Krause

On the Limits of Productive Word Formation: Experimental Data from Finnish[1]

Abstract

In some languages, such as Turkish, Hungarian or Finnish, word formation can be said to be characteristically creative. In these languages, it is quite normal to create novel words by merging several suffixes after the stem. This process may be iterative, allowing recursive stacking of morphemes to the stem. Although it is well-known that in some languages word formation is in this way productive relative to other languages, there is little behavioral data concerning the limits of such productivity. We report an experimental investigation of the limits of word formation in Finnish and argue on the basis of our results that the word formation in Finnish is fully productive, even recursive, but it is counterbalanced by a strong complexity effect which restricts speakers' ability to understand complex words.

1. Introduction

Word formation refers to our ability to form new words from previously learned or assembled parts. Speakers of English can readily understand polymorphemic words such as *iraqification* or *alienability* even if they have never seen these words before. Some languages, such as Turkish, Finnish or Hungarian, appear to make much use of such word formation potential (Brattico 2005, Karlsson 1983, Koskenniemi 1985, Niemi, Laine & Tuominen 1994, Sulkala & Karjalainen 1992). In Finnish, for instance, it is possible to causativisize nouns, adjectives and verbs recursively, so that one can find words with three causative morphemes stacked one after another (i.e. *tee-tä-ty-ttä-ä*, make-CAU-CAU-CAU-V). Apart from marginal cases, this type of process is absent in such languages as English (Plag 2003: 134) or Chinese (Packard 2000, Sagart 2001). Although it is well-

[1] This research project was financially supported by the Academy of Finland (project number 1060741) for the first author and the University of Helsinki for the second author.

known that some languages are morphologically productive, less is known about the limits of this otherwise excessive productivity. How many causative or other suffixes one can merge to one word? Is there a limit in the process and, if so, what kind of limit is it?

Linguists do not yet have a definite answer to this question; rather, the problem constitutes “a battleground, on which many important linguistic wars have been fought” (Borer 1998: 151). This problem boils down to the fact that words exhibit what could be termed as “limited generativity”: words sometimes behave as if they were generated by generative word formation rules, as in Finnish causativization, yet it is clear that this potential is also strictly limited. Thus, while some authors believe that words have a generative syntax of their own, or that they are generated by the same computational process as the syntax (Baker 1988, Lieber 1992, Hale & Keyser 1993, Marantz 1997, Selkirk 1982, Ullman 2001), others favor autonomous and often characteristically non-syntactic models of word formation (Anderson 1992, Aronoff 1976, Chomsky 1970, Karlsson 1983). Karlsson (1983), for instance, argues that despite such productivity Finnish morphology is based on a finite number of morphotactic positions, i.e., fixed morphological slots for suffixes.

Psycholinguistic and neurolinguistic research on word formation appears to be in no better position to solve this issue. This is because most of the psycholinguistic research on the processing of polymorphemic words has focused on the processing of relatively simple words, such as those containing only one or two morphemes (Emmorey & Fromkin 1988, Frost, Grainker & Rastle 2005, Hankamer 1989, McQueen & Cutler 1998). Moreover, this research thread has been concentrated mostly on words which exist in the lexicons of the speakers of the language, unlike most of the novel complex words made possible by recursive word formation rules. Indeed, this research thread has converged into the idea that there are two parallel routes through which we access words in language comprehension and language production: via lexicon and via morphological parser (Caramazza et al. 1988, Schreuder & Baayen 1995). Yet this proposal works only if there is lexical access to a polymorphemic word. While tracing lexical access is a worthwhile project in and of itself, in order to study the limits of creative word formation potential we need to expose subjects to novel words which do not yet exist in their lexicons, and hence to words which are accessible only through their morphological parser (Hankamer 1989) and whose interpretation is not facilitated by prior knowledge.

Another much used strategy for gauging productivity in word formation has been based on the examination of corpora. The basic assumption underlying this approach is that the degree of productivity of a given morphological suffix can be identified as the degree of its ability to merge with various lexeme types. A less productive suffix is one which attaches only to few lexemes, whereas a fully productive suffix is one which can attach to virtually any word, with a range of intermediate cases falling between these two extreme cases. Various statistical methods exist for the estimation of such productivity from the corpora (Baayen 1994, 2005). The problem with this method in the case of morphologically productive languages is that, given that the number of potential words made possible by iterative suffixation approaches astronomical numbers (Niemi, Laine & Tuominen 1994), their type and even token frequencies in any corpus approaches zero. Indeed, when we generated Finnish words randomly by suffixing derivational suffixes to the stems, layering two such suffixes generated words that could not be found from the largest corpus available to us. Furthermore, most morphemes in Finnish are close to being fully productive in that there are only few exceptional words that these morphemes cannot be merged to. The question of whether and how native speakers of Finnish could interpret such words if they were to confront them has to be assessed by different means. To try to study the limits of productive word formation, we rely here on psycholinguistic experimentation.

This methodological stance is related to another matter that sets the current research agenda apart from many other studies on word formation. It is not our primary interest to describe the rules which capture the set of Finnish words currently in frequent use within some geographically, temporally and socio-culturally defined area (compare Hakulinen et al. 2004). Instead, we assume that the word tokens found from the corpora or from everyday use derive ultimately from some type of linguistic/cognitive mechanism(s) implemented in the human brain, the latter constituting the subject matter of this study. What we thus mean by saying that word formation is based on “limited generativity” is that it is these cognitive word generation processes which fall somewhere between fully productive processes and simple lists, both types of mechanisms which are well studied in the cognitive science literature. This point bears some emphasis since although much is known about the frequencies of word types and tokens in actual use, there is very little agreement on the nature of the cognitive processes underlying word generation and morphological parsing, not to speak about the limits of such parsing processes. We are therefore not attempting to characterize the set of Finnish words, but the mechanisms underlying their use. Secondly, the cognitive research agenda cannot be limited to the study of existing words, as the cognitive processes under scrutiny here are those which are responsible for the parsing and use of all possible words, currently existing or not.

The background of the experiments reported here is as follows. Consider the model of Finnish word formation, as depicted in Karlsson (1983). As noted above, Finnish allows one to stack at least three causative suffixes one after another. The word formation rules seem to allow even more (i.e., *tee-tä-ty-tä-ttää*, *tee-tä-ty-tä-ty-ttää*), but triple causatives are nevertheless already extremely rare in actual use.[2] Karlsson suggests that these three causative suffixes each fill up their own morphotactic position, thus the three so-called "verbal suffix positions" which come right after the root stem. The system is therefore inherently finite since it is based on three slots for verbal suffixes. As a description of Finnish words currently in use, this stance is relatively unproblematic. However, the question still arises whether the cognitive mechanisms underlying the use of such words are recursive or not. Brattico (2005) describes the same causativization as a fully recursive process where the system allows one to add a causative suffix to an already causativized word. The rule therefore allows in principle an unlimited stacking of causative suffixes. If this latter model is correct, then *something else* than word formation rules has to explain why the process is limited to a few morphemes.

These two hypotheses make different predictions concerning the behavior of complex Finnish words. A theory which assumes that cognitive word formation is based on a finite and fixed number of morphotactic positions predicts that the grammaticality and semanticality of complex but otherwise correct words, as intuited by native speakers, should fall rather suddenly after more suffixes are merged to the word than what is allowed by the putative fixed positions. In other words, the fixed boundaries of the morphotactic positions in word formation should be behaviorally detectable. We call this the overflow effect: when the available fixed positions are filled, the rest of the material overflows from the system,

---

[2] One example of a triple causative used in Finnish can be found by using the Google search engine (29.10.2007), namely, the abovementioned teetätyttää, 'MAKE-CAU-CAU-CAU-V'. However, this word does not certainly feel deviant in any way; only the four causative morphemes create a sense of deviance.

creating a clear peak in the loss of performance. The magnitude of such non-linearity is an open issue and of some interest in itself, but this hypothesis implies that we should nevertheless see some kind of effect. In this article we report an experiment aimed at testing this prediction, i.e., finding out whether morphological productivity in Finnish stops abruptly and, if so, how soon and how abruptly this limiting barrier takes effect. Moreover, we wanted to see how many suffixes one can add to a word in Finnish to produce items which are still felt to be more grammatical than words generated by violating strict word formation rules.

2. Methods

To investigate the limits of morphological competence, we need to expose subjects to complex enough words to be able to see the shape and the location of the non-linear overflow effect if it exists. Thus, we need stimulus words with varying complexity, starting from few morphemes up to possibly quite many. One problem is to find a way to construct such hypercomplex stimulus items. A corpus search is out of question, since words commonly found from corpora are in active use in the language, and hence they are most likely to be lexicalized to varying degrees. Thus, speakers have prior knowledge of such items. The second possibility is that the stimulus words are constructed by native speakers of the language. Such items are nevertheless likely to be contaminated by uncontrolled biases, thus unknown variables which could either make the items systematically easier or more difficult to process. Instead, in this study we selected the stimulus words by selecting the suffixes randomly (without violating the word formation rules of the language). Random generation ensues that other variables besides complexity do not enter systematically into the experiment and that the stimuli is not biased to favor either one of the hypotheses. The details of this process are described in the next section.

There is an additional problem with any selection method which aims to tap the complexity limits in morphological processing. Suppose we want to craft very complex words containing, say, six derivational morphemes. It is possible that the word formation rules dictate that the morphemes that can be inserted close to the root come from a different set than those that can be inserted closer to the word boundary. Furthermore, the actual word formation rules could differ as well. This means that we would be probing the effects of different sets of morphemes and word formation rules at different complexity levels, so that the relevant experimental conditions would not differ only in terms of their complexity, i.e., number of morphemes. Although still possible, the results from such an experiment would not allow us to draw strong conclusions about complexity as such. To overcome this problem, it is reasonable to try to select morphemes which can be drawn from the same set of morphemes by the same word formation rules at every morpheme position used in the stimulus generation.

Following the above guidelines, we generated two categories of novel and complex Finnish words for the experiment. In the category of grammatical words, the stimulus words followed the word formation rules of Finnish. In the category of ungrammatical words, the words contained one morphologically illegal step. In both categories, we produced words with varying complexity, where complexity was measured in terms of the number of morphemes. Both categories had words containing two to six randomly selected morphemes, including the root stem. Thus, we manipulated two independent variables, morphological grammaticality and morphological complexity.

An additional methodological difficulty concerns the way in which reliable grammaticality and semanticality judgments can be obtained. A linguistic introspection originating from a single linguist, often used alone, carries little weight in the present case, if only because of the labile and graded nature of well-formedness intuitions in the case of complex words. Moreover, to detect non-linearities in grammaticality and semanticality judgments a more fine-grained method is called for. These methodological concerns have been discussed recently by Bard et al. (1996), Cowart (1997) and Schütze (1996), and on the basis of these studies (Schütze 1996 in particular) we concluded that (1) the grammaticality and semanticality judgments must be elicited from a statistically sufficiently large pool of linguistically naïve subjects, (2) the order of the presentation of the stimulus materials needs to be randomized individually for each subject, (3) context effects, such as time pressure, need to be minimized, (4) task instructions must remain invariant across subjects, (5) between-subject and within-subject variation should be reported and handled statistically, (6) grammaticality and semanticality should be measured along a five-point scale, rather than on a forced binary choice (grammatical vs. ungrammatical), due to the possible graded nature of the grammaticality judgment, and (7) the experiment should be recorded on a video tape for additional analyses. Finally, since grammaticality and semanticality judgments are sensitive to a number of contextual factors, such as task instructions, the experiment was designed so that instead of interpreting absolute grammaticality values, the results are interpreted by comparing the experimental groups with each other. Thus, the group of ungrammatical words provides a baseline against which we interpret the effects of complexity. The results could then be replicated and compared with other task instructions, for example. Following these guidelines, we first asked subjects to provide a semantic analysis and a grammaticality judgment on a numerical scale from one to five for each of the target items.

Since it is uncommon to use statistical inference in the case of grammaticality and semanticality judgments, or in the linguistic argumentation more generally, we comment on this methodology briefly here and on how the present results should be interpreted. There are several reasons why linguists do not use statistical methods. In some cases, the grammaticality or semanticality of a target expression is assessed by a single linguist alone, who perceives the expression to be grammatical or ungrammatical and reports this as data. Implicit in this methodology is the assumption that if several subjects would have been used in the same task, the judgments would have converged into the same result. Thus, the benefits for running several experiments at once by a single linguist clearly outweights the extra reliability that we would obtain if we test each expression with several native speakers in a controlled environment. While generally speaking there is nothing wrong with this method, in some cases it does not produce convincing results. This is especially true when grammaticality/semanticality judgments fluctuate or when they are graded rather than clear-cut. In that case, it is clearly better to take the trouble to use several native speakers as informants and gather many grammaticality/semanticality judgments for each target expression to see whether they all converge into some representative value, such as the mean or the median, and whether and how much the judgments differ from each other. In a normal situation, we would get a mean grammaticality judgment for each expression with some amount of variance associated with it, and these means are very often different from each other due to many random factors which influence individual judgments. The purpose of statistical inference is to find out whether the differences in the observed means are due to pure chance or whether the difference is due to some systematic factor (the nature of these factors depending on the experimental design). The results are reported in terms of p-values, which represent the probability that the difference between two means is obtained by pure chance. The smaller the p-value, the higher the chances are that the difference is due to a real result.

3. Experiment

3.1 Stimuli

In this section we describe the model behind stimulus generation. This model is not essential for the interpretation of the results. First, the experiment itself could be replicated by using any model of word formation, keeping the above guidelines in mind. Second, all the stimulus words and their grammaticality and semanticality estimations obtained from the experiment are provided in the appendix, making it possible to interpret the results under any other theoretical framework or under any other model of Finnish word formation.

The underlying model here is based on the idea that Finnish word formation dissolves into two layers or strata, but this idea itself is controversial and should not be taken for granted. Rather, the experimental design is such that the model itself can be verified or falsified by the experiment.[3]

The stimulus words were generated according to a model of Finnish derivational morphology (Brattico 2005). Brattico follows Marantz (1997, 2000) in the contention that word formation in Finnish is a product of the fully generative engine in the brain. The model was selected here because it allows us to iterate relatively freely certain word formation rules of Finnish. The model borrows from Giegerich (1999) and much earlier literature (Allen 1978, Kiparsky 1982, for a recent review, see McMahon 2000: 1–53) in assuming that from the perspective of linguistic competence, word formation is constituted by two layers of word formation. The first layer, corresponding roughly to derivational

[3] This is because one of the independent variables was Grammaticality. If the model predicts the distinction between grammatical or ungrammatical words wrongly, this should become evident in the experiment itself in that the subjects should rate grammatical words as deviant, hence ungrammatical, and ungrammatical words as grammatical. In that case, the results of the experiment could not be interpreted at all. If, on the other hand, subjects' judgments agree with Grammaticality, we can conclude that the word formation model behind the experiment is in reasonable agreement with reality. This applies to any putative word formation model used in the experiment.

morphology, produces lexemes by applying the generative engine so that the process is constrained only by semantic conditions. Recursivity guarantees that there is no upper bound of word complexity with regard to competence, i.e., the knowledge of language as opposed to its use. The second layer, corresponding roughly to inflectional morphology, applies word formation rules on the basis of the syntactic context of the lexeme in the sentence as a whole. This layer consists of morphemes carrying information about lexical category (e.g., noun, verb, adjective), agreement features (e.g., first person plural), and case (e.g., nominative, accusative) features, among others. It is essential to this model that the derivation proceeds from layer one to layer two, and never in reverse order. Layer two morphemes are closing suffixes from the point of word formation.

After the word has been derived by means of morphological processes, it is subject to phonological and semantic interpretation. This results in a number of morphological and phonological readjustment rules, which try to produce a well-formed word by applying allomorphy selection, morphophonological rules and, finally, phonological rules to the output of syntax. This assumption is characteristic of Distributed Morphology (Halle & Marantz 1993). These rules are described later. All in all, the model we purport to use here can be depicted in Figure 1.

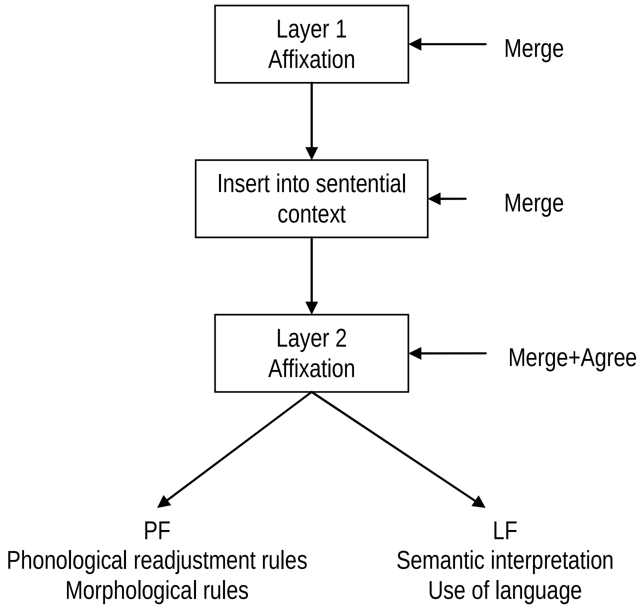

Figure 1. The basic architecture of word formation, as depicted by Brattico (2005). In layer one, morphemes are concatenated to each other to result in complex lexemes. This process is implemented by Merge. The derivation is closed by suffixing the word with a layer two morpheme. Only layer two morphemes are compatible with Agree, which combines the word with morphosyntactic features. The outputs of this process cannot be fed back to layer one. Rather, at this point the whole element is subjected to phonological and semantic processing.

The stimulus words were classified into two categories, those which did not violate word formation rules and those which violated them (ungrammatical words). Both categories contained five complexity levels, which were defined by the number of suffixes. This variable ranged thus from one to five. We selected five suffixes as the upper bound, since there are only few marginal words in Finnish containing five derivational morphemes (Karlsson 1983). Complexity level one in the category of grammatical words consisted of control words consisting of a root stem with one legible random suffix. In the category of grammatical words, complexity levels 2, 3, 4 and 5 were constructed by layering level one morphemes and concluding the word with one layer two morpheme. In the ungrammatical condition, the same rules were used, but each word began with one ungrammatical derivation from layer two into layer one. This 2>1 transition violated the rules of grammar, as predicted by Brattico (2005).

The stimuli are summarized in Table 1, where > represents an ungrammatical combination.

	CONTROL	STIMULI
GRAMMATICAL	root + 2	root+1+2, root+1+1+2, root+1+1+1+2, root+1+1+1+1+2
UNGRAMMATICAL		root+2>2, root+2>1+2, root+2>1+1+2, root+2>1+1+1+2

Table 1. Summary of the stimuli used in the experiment. The suffixes were selected randomly by following the rules of Brattico (2005). Symbol “1” refers to layer one morpheme, “2” refers to layer two morpheme.

Note that the ungrammatical 2>1 derivation does not violate any phonological rules of Finnish. We controlled for the frequencies and semantic properties of the base roots such that half of the base roots were randomly selected from high frequency words (lemma frequency range 3267–126 (frequencies per million)) and low frequency range (lemma frequency range 10–7 (frequencies per million)). Half of each group was then divided such that in the first group, the base root was a verbal root (*love*-) and in the second group it was a nominal root (*house*-). To test the generative capacity, each word in complexity levels 2–5 was tested against a large corpus of Finnish texts to ensure that it was not in use and with all likelihood was not confronted before the experiment.[4] Bimorphemic words in complexity level 1 constitute an exception to this rule, since some of these items look like regular Finnish words, others less so. Secondly, this complexity does not exist in the category of ungrammatical words, since in order to produce an illegal suffix combination, a minimum of two suffixes are needed. Because of these facts, grammatical words from complexity level 1 are not taken into account when the results are discussed and analyzed, but they nevertheless constitute a control group which should be fully grammatical.

Brattico (2005) does not provide an explicit list of layer one and layer one morphemes, but examines only a few examples. Rather, he proposes

[4] A Finnish corpus composed by the Research Institute for the Languages of Finland, the Finnish IT Centre for Science and Department of General Linguistics, University of Helsinki. The corpus was used through WWW-Lemmie 2.0 at Finnish IT centre for science, obtainable from www.csc.fi/kielipankki.

that if the causative morpheme can be suffixed to some affix (and hence to some stem), then that suffix belongs to the same layer as the causative morpheme (namely, layer 1). If it does not, then either the affix is a layer two affix, or some independent (i.e., morphophonological) rule prevents the output. Based on this test, we selected the following morphemes for our grammar of Finnish words (see Table 2).

LAYER 1	S-SELECTION	MEANING	ALLOMORPHS
CAU[eve]	[ref][eve]	‘to cause to –’	*(t)ta, sta, ta*
FRE[eve]	[ref][eve]	‘to do habitually –’	*ele, ile, eile, skele*
EVE[ref]	[eve]	‘an event of –’	*o, u, y*
REF[ref]	[eve]	‘to become –’	*u*
US[ref]	[ref]	‘the property of –’	*(u)us*
POSS[ref]	[ref]	‘something which has –’	*ll*
COL[ref]	[ref]	‘a collection of –’	*(i)sto*
LAYER 2	**S-SELECTION**	**MEANING**	**ALLOMORPHS**
MINEN[N]	[eve]	‘the property of –ing’	*minen*
MA[A]	[eve]	‘the result of –ing’	*ma*
VA[A]	[eve]	‘something which does –’	*va*
VAINEN[A]	[eve]	‘something which does –’	*vainen*
A[V]	[eve]	non-finite verb	*Ca*
MASSA[V]	[eve]	non-finite verb	*massa*
IMP[V]	[eve]	imperative verb	*∅*
∅[N]	[ref]	zero derived noun	*∅*
INEN[A]	[ref]	‘somebody who has the property of being –’	*(i)nen*
MAINEN[A]	[ref]	‘somebody who resembles –’	*mainen*
SUUS[N]	[ref]	abstract noun	*(i)suus*
KE[N]	[ref]	noun affix with unclear meaning	*ke*
TAR[N]	[ref]	‘a female who is –’	*tar*
IN[N]	[eve]	‘an instrument for –’	*In*

Table 2. Morphemes selected for this study. In the left column, we list the symbol for the morpheme together with its semantic classification according to Brattico (2005), [referential] or [eventive]. The next column lists semantic selection restrictions given for the morpheme. Thus, morphemes which select for [eventive] affixes cannot be merged with referential affixes. The third column from the left gives the most typical meaning for the morpheme. This characterization is not exhaustive because many

morphemes can be interpreted in several ways. The right column lists allomorphs which were used in this study. The selection of these allomorphs is a matter of morphological readjustment rules, which we describe later.

See the appendix for the list of words generated by this method. Stimulus words were generated by selecting morphemes from the above list randomly so that only semantic selection restrictions were followed.

Note that in the category of ungrammatical words, the stimulus words were merged with two layer two suffixes to produce a morphologically impossible word. This has the consequence that the words in the group of ungrammatical words at complexity levels 2–5 are approximately one phoneme longer than the words in the category of grammatical words (mean length for grammatical words is 17.9 phonemes and for ungrammatical words 19.3 phonemes, analysis of variance for Grammaticality $F(1, 126) = 9.859$, $p = 0.002$, no interaction with Complexity) since layer two morphemes are longer and less fusional. This could be offset by reducing the length of the roots, but the bias itself would remain in the root length. We will analyze the effect of word length in a separate analysis.

After stimulus words had been generated randomly, we needed to generate a concrete phonological form for them. This requires allomorph selection and the application of phonological and in some cases morphophonological readjustment rules (PF in Figure 1). The rules used in this study were as follows:

CONSONANT GRADATION (CG). Weaken the consonant(s) in the last syllable of the lexeme if the suffixation changes the syllabification of the stem. This rule is applied also in the case of an imperative suffix, even if the suffix does not have an overt morphological exponent (1c) (Karlsson 1983: 322–324).

(1) a.	*lo**tt**a*	>	*lo**t**a -n,*	*lo**t**a -lla,*	*lo**tt**a-mainen*
	lotta.NOM	>	Lotta-GEN	lotta -ELA	lotta-like
	‘proper name	>	Lotta’s,	in Lotta’s possession,	like a Lotta’

b.	*paalu-**tt**a -a*	>	*paalu -**t**a -tta -a,*	*paalu-**tt**a -minen*	
	pole -cau -v	>	pole -cau -cau -v,	pole -cau -n	
	‘to pole	>	to cause to pole,	causing to pole’	

c.	*tö**kk**i -ä*	>	*tö**k**i -ttä -ä,*	*tö**kk**i -minen,*	*tö**k**i!*
	push -v	>	push -CAU -V	push -ING	push-IMP
	‘to push,	>	to cause to push,	pushing,	push!’

VOWEL FUSION (VF). If the morpheme begins with a vowel, it replaces the vowels (if any) at the end of the previous morpheme.

(2) a. *juoksu -tt**a** -**i**n > juoksut**i**n, juoks**e-u** > juoks**u***
run -CAU -INST > run-U
'an instrument to cause to run, a run'

b. *monist**a-e** > monist**e***
copy-E > copy-E
'the result of copying'

c. *last**aa-e** > last**e***
load-E
'the result of loading'

VOWEL INSERTION (VI). If the merging of two morphemes produces an impossible consonant cluster, such as /sll/, insert vowel /i/ or /e/ between.

(3) a. *hevo**s**-mies, hevo**s**-**ll**inen > hevo**s** -**e** -**ll**inen*
horse -man, horse -LLINEN > horse -e -LLINEN
'a horse man, a horse owner'

b. *talous -ennuste, talou**s** -**ll**inen > talou**d**-**e**-llinen*
economy -forecast, economy -LLINEN
'economy forecast, economical'

VOWEL HARMONY (VH). Vowels in two adjacent morphemes undergo vowel harmony (Karlsson 1983: 98–104).

(4) a. *tal**o** -ss**a**, p**ää** -ss**ä***
house-INE head -INE
'in the house, in the head'

b. *paal**u**-tt**a** -**a**, p**ää** -tt**ä** -**ä***
pole -CAU -V, head -CAU -V
'to cause to pole, to cause to have / be a head / to decide'

IDIOMATIC RULES (IR). (a) -us+ele- = -uskele-. (b) one-syllable root+ele = root-skele.

A total of 9 x 16 = 144 words were first generated. Eight pseudowords which were suffixed with layer two morphemes were used as filler items. A total of 152 words were used, but three were removed from the analysis since later it was found that there was an error in the generation. A list of all stimulus words can be found in the appendix.

3.2 Procedure

The experiment consisted of two separate tasks: one grammaticality and one semantic judgment task. Half of the subjects performed first the semantic judgment task for all words and thereafter the grammaticality judgment task, likewise for all words; the second half performed the semantic judgment task and the grammaticality task in the opposite order. Both word lists were randomized for each subject. The same stimuli were used in both tasks. Prior to each test, the instructions were given on paper and on a computer screen. In both tasks the visual stimuli were presented one at a time on a PC computer screen, commanded by a script written in Presentation 9.90 (Neurobehavioral Systems, Albany, USA). In both tasks, each word was centrally displayed on the monitor, formatted with black, 72-point Times New Roman font on a gray screen.

During the semantic judgment task, the subjects' task was to assess the meaning of each word by describing verbally one or more situations in which that word could be used. After the word was presented on the screen, the subjects described the meaning of that word and then pressed the green key on the keyboard. They were instructed to press the red key if they could not give any meaning for the word. In each trial, the response wait time would time out after one minute. Following the answer or time-out, the blank grey screen was displayed for 1500 ms before the next word was presented. The verbal meaning descriptions were recorded using an ElectroVoice MC100 microphone (Telex Communications Inc., USA), which was connected to a Sony Digital Handycam DCR-VX1000E video camera. The semantic judgment test comprised 152 trials and lasted approximately 45 minutes.

During the grammaticality judgment task, the subjects were instructed to press one of five keys on a keyboard after the stimulus was presented. Key 5 was to be pressed when a stimulus was assessed as a grammatical Finnish word and key 1 to be pressed when a stimulus was judged as an ungrammatical Finnish word. The trials were constructed such that a response wait time for each stimulus would time out after one minute (as in the semantic judgment task), and the next stimulus was presented 1500 ms after the response or time-out. The task consisted of 152 trials, and this task lasted approximately ten minutes.

3.3 Subjects

Twenty six adults, 13 men and 13 women, volunteered as participants. 22 participants were university students, while four additional participants had a university degree. The mean age of the participants was 26.3 years (SD = 9.6, range 19–58 years). All participants had normal or corrected-to-normal vision and none of them reported any linguistic dysfunctions. All subjects were native speakers of Finnish. The participants received two movie tickets for participation.

3.4 Statistical methods

The effects of two independent variables, Complexity (4 levels, 1, 2, 3 and 4) and Grammaticality (2 levels, grammatical and ungrammatical), were assessed using three dependent variables, grammaticality judgments, semanticality judgments and reaction times, both with an analysis of variance (ANOVA) in an item-based analysis and with a repeated measures ANOVA in a participant-based analysis (see below). Within-subject contrasts in the repeated measures ANOVA were used to assess whether the relationship between complexity and grammaticality was linear or non-linear. Post-hoc comparisons were computed with ANOVA for both grammatical and ungrammatical words, separately. Reaction times were correlated with grammaticality judgments within each complexity level in the category of grammatical words. Finally, we construed a linear regression model to assess how much complexity and reaction time explain the variance in grammaticality judgments.

The data were analysed both on a item-based analysis, where each stimulus item was attributed one grammaticality and semanticality value, as averaged from all subjects, and on a participant-based analysis, where each subject was attributed one grammaticality and semanticality value in each experimental condition (Grammaticality x Complexity). An item-based analysis is suitable when the interest is on the linguistic properties of the words themselves, and this is the most relevant for the present purposes since we aim at comparing two linguistic models of word formation. In this analysis, the source of variation comes from the stimulus items whose properties the theoretical models are trying to predict. The values obtained are reported in the appendix along with the stimulus words along with figures are drawn based on these results. In the participant-based analysis, the source of variation comes from the between-subject differences and thus reflects the amount that the subjects differ from each other in the experimental task.

The semanticality value (0–1) represents the frequency that the given word was assigned a semantic interpretation, whereas the grammaticality value represents the mean grammaticality value provided by the subjects (1–5). The statistical analyses assume that the obtained numerical scales, 0–1 for semanticality and 1–5 for grammaticality, are interpreted as interval scales. This assumption holds for the frequency of semantic interpretation, but not necessarily for the given grammaticality intuitions as it is not clear how the various values (1, 2, 3, 4 and 5) relate to each other. For the grammaticality assessments, the minimal assumption is that the given numbers come from an ordinal scale. Nonparametric tests should be used if the assumption of the interval scale does not hold, and therefore the data were additionally analyzed with the Kruskall-Wallis non-parametric test.

In the semantic task, three subjects pressed the same button for each stimulus item presumably because they forgot the task instructions. These data were removed before analysing the frequency of semantic interpretation. However, their grammaticality judgments and reaction times were in line with the other subjects, and thus the grammaticality judgment data was not removed.

4. Results and discussion

Item-based analysis. The main results of this experiment are shown in Figure 2 and described below in detail.

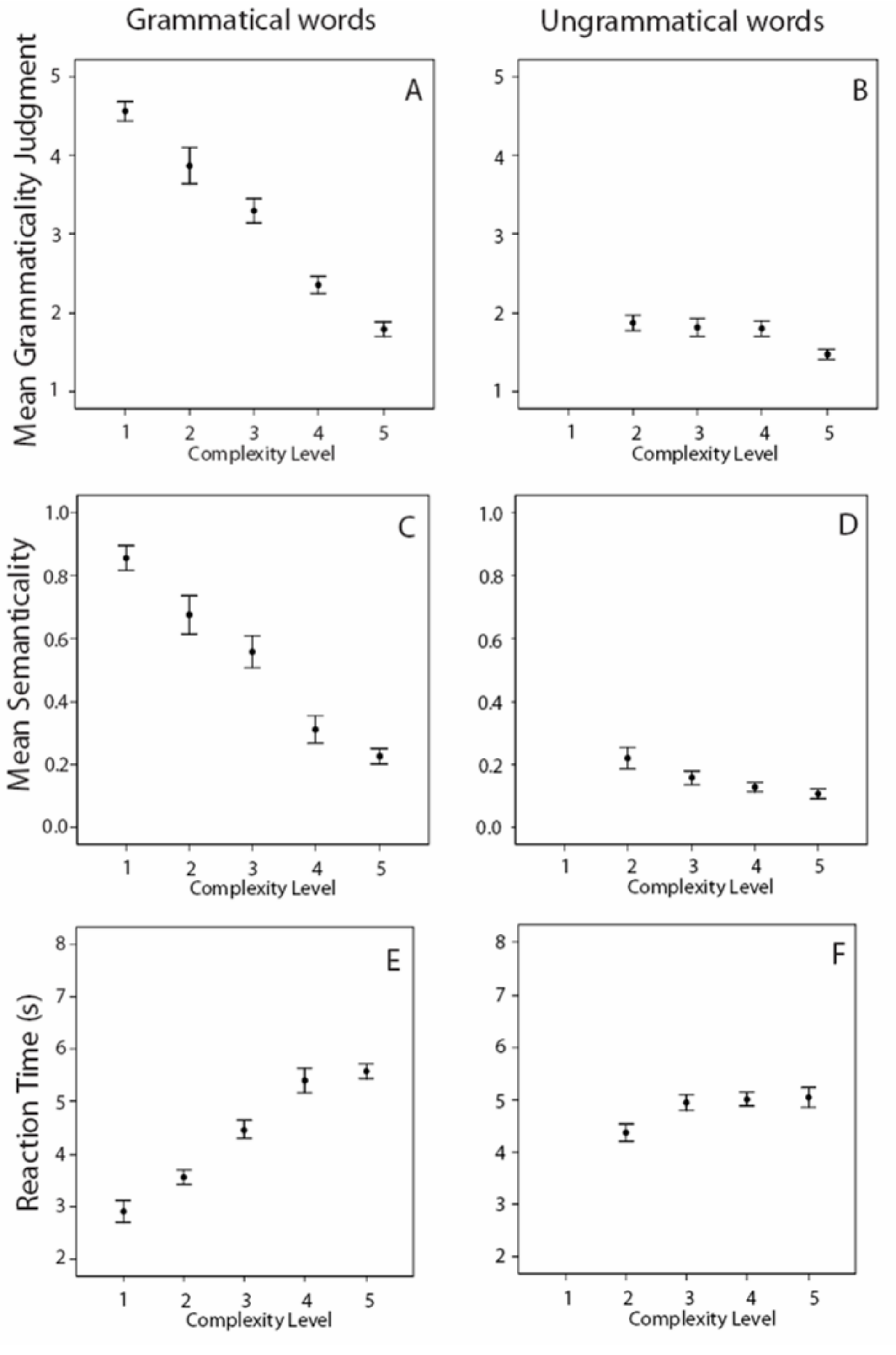

Figure 2. The main results of the present experiment. Top row. The mean grammaticality judgments (with the standard error of the mean) for both grammatical (A) and ungrammatical (B) words as a function of Complexity. Middle row. The mean semanticality (with s.e.m.) for both grammatical (C) and ungrammatical (D) words as a

function of Complexity. Semanticality refers to the frequency that the words in a given category obtained a semantic analysis. Bottom row. Mean reaction times in seconds for grammatical (E) and ungrammatical (F) words in a grammaticality task. The means and the standard error of the mean are taken from the participant-based analysis, reported later.

With respect of grammaticality judgments, ANOVA revealed a statistically significant main effect for Complexity [$F(3, 118) = 36.388$, $p < 0.001$] due to the fact that grammaticality values decreased as a function of complexity. The main effect for Grammaticality was also significant [$F(1, 118) = 142.226$, $p < 0.001$] due to the fact that morphologically ungrammatical words were judged as highly deviant. The interaction between Complexity and Grammaticality was significant [$F(3, 118) = 18.706$, $p < 0.001$] due to the fact that, contrary to the category of grammatical words where the effect was clear and linear, complexity had only sporadic and marginal effect on grammaticality judgments in the category of ungrammatical words. (Post-hoc comparisons with ANOVA showed that there was a significant difference in the grammaticality judgments between complexity levels two and five in the category of ungrammatical words [$p = 0.028$], but no other significant differences were found. In contrast, in the category of grammatical words the same test revealed that the difference between complexity levels two and three [$p = 0.094$] and four and five [$p = 0.08$] was not significant, but the rest of the comparisons were significant. These results were Bonferroni corrected for multiple comparisons.) Within-subjects contrasts of the repeated measurements analysis of variance (rmANOVA) were used to assess whether grammaticality falls in a non-linear manner as a function of complexity. This analysis confirms that the relationship between complexity and grammaticality in the category of grammatical words was linear [$F(1, 25) = 603.274$, $p < 0.001$] and neither quadratic [$F(1, 25) = 1.566$, $p = 0.222$] nor cubic [$F(1, 25) = 2.431$, $p = 0.132$].

The mean grammaticality estimation in the group of ungrammatical words was 1.74, whereas the mean grammaticality value in the category of grammatical words in complexity level 5 was 1.79. These groups did not differ from each other in terms of their grammaticality [Grammaticality $F(1, 78) = 0.123$, $p = 0.727$]. The mean grammaticality in the category of grammatical words at complexity level 4 was 2.36, while this group differed from the group of ungrammatical words [Grammaticality $F(1, 78) = 24.876$, $p < 0.001$]. Thus, the limit of morphological productivity in

Finnish settles down at six morphemes (five semantically random suffixes), which agrees well with other assessments (Karlsson 1983).

However, the assumption of equal variance did not hold for our data ($p < 0.001$) and it is neither obvious nor reasonable that the participants understood the 1–5 scale as an interval scale. Non-parametric tests were used for this reason. These tests showed a significant effect of Complexity to the grammaticality intuitions in the category of grammatical words [$\chi^2(4) = 60.026$, $p < 0.001$]. Also, Grammaticality had a significant effect on grammaticality intuitions [$\chi^2(1) = 57.268$, $p < 0.001$] in that ungrammatical words were rated as less grammatical than the Grammatical words.

With respect of semanticality judgments, i.e. the frequency that the given stimulus word was provided a semantic analysis, there was a statistical significant effect for Complexity [$F(3, 118) = 25.214$, $p < 0.001$] because the more complex words were interpreted less frequently, Grammaticality [$F(1, 118) = 125.069$, $p < 0.001$] since ungrammatical words were virtually never provided a semantic interpretation, and for the interaction between Complexity and Grammaticality [$F(3, 118) = 9.773$, $p < 0.001$] due to the fact that Complexity had only a slight effect in the category of ungrammatical words. (Post-hoc comparisons with ANOVA for ungrammatical words revealed that there was a significant difference only between complexity levels two and four [$p = 0.037$] and two and five [$p = 0.005$]. In contrast, the same test for grammatical words revealed that there were three cases where no significant difference was observed, complexity levels one versus two [$p = 0.054$], two versus three [$p = 0.764$] and four versus six [$p = 1$]. These results were Bonferroni corrected for multiple comparisons.)

With respect of reaction times for the grammaticality judgments, there were statistical significant effects for Complexity [$F(3, 118) = 26.246$, $p < 0.001$] in that more complex words took longer to rate, and the interaction between Complexity and Grammaticality was also significant [$F(3, 118) = 6.588$, $p < 0.001$] because this trend was absent in the category of ungrammatical words. However, Grammaticality had no significant effect on the reaction times [$F(1, 118) = 1.094$, $p = 0.298$]. We also found that within complexity levels 1 and 2 reaction times correlated negatively with grammaticality judgments within each complexity level in the category of grammatical words (complexity class 1, Pearson correlation -0.72, $p = 0.002$, complexity class 2, -0.535, $p = 0.033$), suggesting that processing difficulties are directly related to grammaticality in the category of grammatical words.

On the basis of these results, a linear regression model was constructed with grammaticality as the dependent variable, and complexity and processing time as independent variables. This analysis reveals that complexity and processing time explain in a linear model R_a^2 = 78% the total variance in grammaticality judgments for grammatical words ($F(2, 76) = 142.602$, $p < .001$) and R_a^2 = 10% in the category of ungrammatical words ($F(2, 62) = 4.747$, $p = 0.012$).

Participant-based analysis. With respect of grammaticality intuitions, a repeated measures analysis of variance (rmANOVA) revealed a significant effect for Grammaticality [$F(1, 25) = 205.782$, $p < 0.001$] and Complexity [$F(3, 75) = 179.868$, $p < 0.001$]. This was because ungrammatical words were rated as highly deviant and unsemantical, and because the increase in complexity caused a linear drop in grammaticality and semanticality intuitions. Also the interaction between Grammaticality and Complexity was significant [$F(3, 75) = 93.086$, $p < 0.001$], because complexity did not have a similar effect in the category of ungrammatical words (see Figure 2). With respect of semanticality intuitions, Grammaticality [$F(1, 25) = 54.972$, $p < 0.001$], Complexity [$F(3, 75) = 48.360$, $p < 0.001$] and the interaction between Grammaticality and Complexity [$F(3, 75) = 23.222$, $p < 0.001$] were significant. With respect of reaction times in the grammaticality task, Grammaticality [$F(1, 25) = 4.573$, $p = 0.042$], Complexity [$F(3, 75) = 25.405$, $p = 0.008$] and the interaction between Grammaticality and Complexity [$F(3, 75) = 14.998$, $p < 0.001$] were significant.

To ensure that the explaining feature was complexity and processing time alone, we performed a number of other tests. We investigated whether the presence or absence of certain specific morphemes correlates with reaction time in the grammaticality judgment task, thus whether some of the specific morphemes were more difficult to process than others. No such morphemes were found. In the group of control words (complexity level 1), the dictionary form was correlated with shorter reaction times [Pearson correlation -0.639, $p = 0.008$].[5]

[5] Due to the large amount of correlations (126), we report only correlations where $p < 0.01$. Five other morphemes were associated with lower or higher reaction times at various levels, but these findings were not robust statistically ($p > 0.01$).

One possibility for the sharp difference between ungrammatical and grammatical words is that the stimulus words in the category of ungrammatical words were ungrammatical because of some quickly detectible surface property, such as a violation of a phonological rule. We took some care to keeps this from happening in the stimulus construction. Another evidence which points to the same conclusion concerns reaction times, since the processing of grammatical and ungrammatical words took the same amount of time [independent samples t-test for grammatical and ungrammatical words $t(124) = 0.533$, $p = 0.595$]. Thus, ungrammatical words were not recognized as such by means of a simple surface property; rather, the subjects had to parse these words as well.

Because morphemic complexity correlates with absolute length of the words, one could argue that the complexity effect is based on the length of the word, not the number of morphemes it contains. This by itself would not disprove the existence of a complexity effect as compared to the ungrammatical words, but it would show that our decision to count the number of morphemes was misleading. This hypothesis predicts that there should be a correlation between word length and grammaticality also *within* each complexity level. At complexity level three, longer words were associated with lower grammaticality judgments [Pearson correlations -0.575, $p = 0.031$], but no such correlations were found from other complexity levels.

Due to the stimulus generation, ungrammatical words were phonologically slightly longer than grammatical words. Thus, one could argue that the main effect of Grammaticality to the grammaticality and semanticality intuitions was due to the difference in phonological length and not due to the presence of illegal morpheme combination. This hypothesis predicts that within each complexity level, word length predicts grammaticality intuitions better than Grammaticality. We tested this hypothesis by assessing how much of the variance in grammaticality judgments within each complexity level is predicted by grammaticality and word length. This analysis shows that whereas Grammaticality predicts grammaticality judgments within complexity levels 2–5 [level 2: R_a^2 67%, $F(1, 31) = 63.480$, $p < 0.001$, level 3: $R_a^2 = 65\%$, $F(1, 29) = 61.399$, $p < 0.001$, level 4: $R_a^2 = 30\%$, $F(1, 31) = 14.527$, $p = 0.001$, level 5: $R_a^2 = 18\%$, $F(1, 31) = 7.749$, $p = 0.009$], word length does not [level 3: $R_a^2 = 2\%$, $F(1, 31) = 1.730$, $p = 0.198$, level 4: $R_a^2 = 17\ \%$, $F(1, 29) = 7.182$, $p = 0.012$, level 5: $R_a^2 = 8\%$, $F(1, 31) = 3.731$, $p = 0.063$, level 6: $R_a^2 = 1\%$, $F(1, 31) =$

1.396, $p = 0.247$]. Hence, the hypothesis that it is the word length which explains grammaticality intuitions receives no support from our data.

One possible hypothesis is that the presence of one or several specific morphemes explains variation in grammaticality, rather than the number of morphemes whatever their form and meaning. We correlated grammaticality judgments with the presence of specific layer one morphemes within each complexity level in the category of grammatical words. At complexity level 5, we found that the presence of a reflexive suffix correlated with higher grammaticality judgments [Pearson correlation 0.698, $p = 0.003$]. Two other weak correlations were found (at level 4, the presence of a possessive suffix correlated positively with the grammaticality judgments, Pearson correlation 0.531, $p = 0.041$ and at level 5 the presence of a possessive correlated negatively with the grammaticality judgments, Pearson correlation -0.523, $p = 0.037$).

These results show that morphological productivity is limited to five random derivational suffixes in Finnish. Nevertheless, the suffixes themselves were generated by fully productive and recursive rules by applying random iteration. This result agrees quite well with Karlsson (1983), who cites only few examples of Finnish words with five derivational suffixes but no words which contain additional suffixes. It disagrees with this model, however, in that we found no evidence for fixed morphotactic positions: we found no trace of a threshold effect which would drop grammaticality abruptly at certain fixed point; rather, each suffix caused a constant reduction of the grammaticality and semanticality, independent of its morphotactic position.

Because Finnish is morphologically productive, we may regard this result as an estimation of the upper bound of morphological creativity at least for derivation. This effect stands in sharp contrast with the effect of violating a strict word formation rule. Such a violation produces gibberish independent of the number of morphemes. In effect, it seems to us then that the two factors which influence grammaticality in word formation have to be distinguished. On the other hand, there are strict word formation rules which dissolve the set of phonologically possible words into those which are morphologically possible and those which are not possible. The category of morphological and phonologically possible words is further constrained by complexity, but this effect is graded rather than binary. The explanation for both factors is thus likely to be different.

5. Conclusions

The creative periphery of word formation capacity has remained a controversial and little studied topic both in linguistic and psycholinguistic research. In this paper we contrasted two models of productive word formation: one which assumes that word formation is based on a fixed number of morpheme positions and another which assumes that the underlying mechanism is generative. Our evidence supports the latter model. More specifically, we presented evidence that word formation is subject to two factors: a possibly generative word formation capacity which is regulated by hard constraints that appear to be non-violable and not sensitive to complexity, and a complexity effect which restricts the interpretation and use of complex words made possible by word formation competence alone. The complexity effect remains constant over all morpheme positions, suggesting that the addition of each morpheme to the word contributes an equal amount to the effect.

Appendix: List of stimulus words, complexity level, grammaticality and semanticality estimations

Table 3. Stimulus words in the category of grammatical words, together with the grammaticality estimation (with the standard deviation from the mean) and the frequency of semantic interpretation (with the standard deviation from the mean).

WORD	COMPLEXITY 1–5	GRAMMATICALITY 1–5	SEMANTICALITY 1–2
muuttaminen	1	5.00 ± 0.00	2.00 ± 0.00
tuntemassa	1	4.92 ± 0.27	1.91 ± 0.29
nostaa	1	5.00 ± 0.00	2.00 ± 0.00
pelaamassa	1	5.00 ± 0.00	2,00 ± 0,00
joutamassa	1	4.35 ± 0.98	1.74 ± 0.45
urakoida	1	4.96 ± 0.20	2.00 ± 0.00
muhia	1	4.85 ± 0.78	1.96 ± 0.21
sahaavainen	1	4.54 ± 0.71	1.83 ± 0.39
levymäinen	1	5.00 ± 0.00	2.00 ± 0.00
pisteke	1	3.38 ± 1.27	1.39 ± 0.50
valotar	1	4.69 ± 0.84	1.91 ± 0.29

kieletär	1	4.27 ± 1.12	1.78 ± 0.42
naudatar	1	4.04 ± 1.15	1.78 ± 0.42
puustomainen	1	4.73 ± 0.60	1.87 ± 0.34
häviöisyys	1	4.42 ± 0.76	1.74 ± 0.45
koppinen	1	3.85 ± 1.26	1.78 ± 0.42
antoinen	2	4.08 ± 1.06	1.74 ± 0.45
löytin (löydä-, löyt-ö, löyt-in)	2	3.42 ± 1.45	1.57 ± 0.51
jatkoisuus	2	4.58 ± 0.50	1.83 ± 0.39
esitteinen	2	4.00 ± 1.13	1.74 ± 0.45
ruotuisuus	2	4.46 ± 0.76	1.83 ± 0.39
lasteke	2	2.85 ± 1.26	1.26 ± 0.45
luistelevainen	2	4.69 ± 0.47	1.87 ± 0.34
höystemäinen	2	4.88 ± 0.33	2.00 ± 0.00
kirjastoisuus	2	4.42 ± 0.76	1.78 ± 0.42
perheistöke	2	2.46 ± 1.30	1.26 ± 0.45
koneellisuus	2	4.96 ± 0.20	1.95 ± 0.21
peliys	2	4.04 ± 1.04	1.83 ± 0.39
liemiysmäinen	2	3.04 ± 1.25	1.39 ± 0.50
rosvosto	2	4.58 ± 0.64	1.83 ± 0.39
metsoustar	2	1.88 ± 0.91	1.30 ± 0.47
aisailevainen	2	3.50 ± 1.21	1.61 ± 0.50
kestelettämässä (kestä-, kestele-)	3	2.38 ± 1.30	1.14 ± 0.35
vasteistoisuus	3	3.31 ± 1.38	1.65 ± 0.49
kasvatetin	3	2.81 ± 1.17	1.52 ± 0.51
hyppyileminen	3	3.46 ± 1.48	1.52 ± 0.51
käärittely	3	4.00 ± 1.02	1.78 ± 0.42
revestömäinen	3	2.88 ± 1.45	1.35 ± 0.49
matkuudellinen	3	2.88 ± 1.24	1.48 ± 0.51
rannastotin	3	2.77 ± 1.48	1.52 ± 0.51
metsäytys	3	3.77 ± 0.95	1.78 ± 0.42
nousutin	3	4.19 ± 0.85	1.74 ± 0.45
saapastumassa	3	4.00 ± 1.23	1.77 ± 0.43
purostoileminen	3	2.73 ± 1.04	1.39 ± 0.50

vajeuksellinen	3	3.73 ± 1.25	1.65 ± 0.49
vuotoutteleminen	3	3.15 ± 1.29	1.48 ± 0.51
näytättelyisyys	4	2.46 ± 1.24	1.50 ± 0.51
piteletteinen	4	2.04 ± 1.18	1.09 ± 0.29
koskeskelustotar	4	2.12 ± 1.03	1.23 ± 0.43
päätetin	4	3.27 ± 1.28	1.52 ± 0.51
suututatella	4	2.42 ± 1.27	1.61 ± 0.50
veistettelemä	4	2.85 ± 1.29	1.43 ± 0.51
maisteletemainen	4	1.88 ± 0.95	1.09 ± 0.29
ilahtelettamassa	4	2.46 ± 1.27	1.35 ± 0.49
tuestoustaa	4	2.00 ± 1.10	1.22 ± 0.42
miehestöiletin	4	2.15 ± 1.05	1.22 ± 0.42
ryhmyystöisyys	4	2.88 ± 1.28	1.35 ± 0.49
sotautteistovainen	4	2.15 ± 1.19	1.09 ± 0.29
riistastoileus	4	1.92 ± 1.13	1.22 ± 0.42
apinallistattamassa	4	3.00 ± 1.36	1.57 ± 0.51
vuoteuttelettaa	4	2.00 ± 1.06	1.35 ±0.49
rakoustouttaminen	4	2.15 ± 1.19	1.17 ± 0.39
mitatteuksellisuus	5	2.08 ± 1.09	1.30 ±0.47
laadiskeluiluke	5	1.65 ± 0.75	1.22 ±0.42
laulutattuke	5	1.73 ± 1.00	1.17 ±0.39
valvoiletatteleminen	5	1.73 ± 0.96	1.21 ±0.42
häivytteletätin	5	1.84 ± 1.01	1.26 ±0.45
pohjaatastoustominen	5	1.35 ± 0.63	1.08 ±0.29
alistuiletattuinen	5	1.69 ± 0.97	1.21 ±0.42
tasailullistamassa	5	2.80 ± 1.39	1.45 ±0.51
viesteilyyskelemä	5	1.80 ± 0.69	1.13 ±0.34
aamutatuttaminen	5	2.08 ± 1.32	1.30 ±0.47
ilmaileskeletattama	5	1.34 ± 1.32	1.13 ±0.34
veroiluudellistar	5	1.65 ± 0.98	1.21 ±0.42
villatatelettäväinen	5	1.19 ± 0.49	1.13 ±0.34
sotkulliskelustomainen	5	1.96 ± 0.92	1.17 ±0.39
hovillisteilevainen	5	2.08 ± 0.93	1.39 ±0.50

otsastolliskeluinen	5	1.65 ± 0.80	1.21 ±0.42

Table 4. Stimulus words in the category of ungrammatical words, together with the grammaticality estimation and the frequency of semantic interpretation.

WORD	COMPLEXITY 2–5	GRAMMATICALITY	SEMANTICALITY
odottamava	2	1.88	1.09
käyttämämässä	2	1.65	1.17
puhinsuus	2	1.96	1.13
tietäväminen	2	1.73	1.17
kutomassaa	2	2.54	1.48
kahmivainenin	2	1.58	1.22
keikkuama	2	2.69	1.39
häätäämässä	2	2.38	1.13
kellomaineninen	2	1.42	1.09
liikevainensuus	2	1.50	1.09
kenttäinenke	2	1.50	1.09
rajainenke	2	1.62	1.13
tuhkaisuustar	2	2.27	1.48
köysinenmäinen	2	1.88	1.35
noidatarke	2	1.81	1.22
kilpinenke	2	1.58	1.30
katsovaineneilemainen	3	1.31	1.04
nousemastoinen	3	2.96	1.39
jättämässää	3	1.92	1.17
sanovainenuske	3	1.23	1.09
hiertääskelemä	3	1.77	1.13
huokutareileminen	3	1.54	1.17
vahtivainenstoisuus	3	1.46	1.22
soutameilema	3	1.65	1.17
jäsentäreileke	3	1.54	1.09
vesinentäväinen	3	1.92	1.09
lakimainenillinen	3	1.65	1.17
kuvatarustar	3	1.88	1.26

solmutarusmainen	3	2.62	1.09
munkkinenstoinen	3	1.54	1.22
sädekeileminen	3	2.19	1.13
levätärttäväinen	3	1.88	1.09
ostatarusellike	4	1.31	1.13
johtamainentattamassa	4	1.23	1.13
menemättelemässä	4	2.46	1.17
toivominensuustaa	4	1.46	1.13
tempaintellä	4	1.46	1.04
riistäelyisyys	4	1.77	1.13
peilaameilemassa	4	2.38	1.09
laukkamellike	4	1.96	1.17
ruokakettymä	4	2.23	1.13
äänimäinentöä	4	1.50	1.17
tietomainenilleileminen	4	1.50	1.04
koirakkeilletar	4	1.77	1.13
viitakkeilettama	4	2.12	1.17
purkumainentominen	4	1.65	1.09
luolaisuususillimainen	4	1.69	1.04
purkkikellistomainen	4	2.35	1.26
laitinstoiletin	5	1.62	1.17
torjumalliustaminen	5	2.08	1.17
lentäväinenskeletättämä	5	1.27	1.04
korjaavaudellistamassa	5	1.73	1.13
huollinstousto	5	1.69	1.22
sysääväinenuskeluinen	5	1.35	1.09
hierrinillustotar	5	1.50	1.13
pestaamaustollistar	5	1.38	1.22
eläininenteusmainen	5	1.85	1.04
taitoinenskelusto	5	1.46	1.04
iskuisuudeleskelettamassa	5	1.27	1.09
vanhusmainennukselleskeleminen	5	1.50	1.04
syketärtellistar	5	1.31	1.09

lohtuinenelliustama	5	1.35	1.04
sirkussuudellittelema	5	1.19	1.13
päätyinenelliskeletin	5	1.04	1.05

References

Allen, Margaret (1978) *Morphological Investigations*. Dissertation, University of Connecticut.

Anderson, Stephen R. (1992) *A-Morphous Morphology*. Cambridge: Cambridge University Press.

Aronoff, Mark (1976) *Word formation in generative grammar*. Cambridge: MIT Press.

Baayen, Harald (1994) Productivity in language production. *Language and Cognitive Processes* 9: 447–469.

—— (2005) Morphological productivity. In Reinhard Köhler, Gabriel Altmann & Rajmund G. Piotrowski (eds.), *Quantitative linguistics: an international handbook*, pp. 243–255. Berlin: M. de Gruyter.

Baker, Mark (1988) *Incorporation: a theory of grammatical function changing.* Chicago: University of Chicago Press.

Bard Gurman, Ellen, Robertson, Dan & Sorace, Antonella (1996) Magnitude estimation of linguistic acceptability. *Language* 72: 32–68.

Borer, Hagit (1998) The morphology-syntax interface. In Andrew Spenser & Arnold M. Zwicky (eds.), *The Handbook of Morphology*, pp. 151–190. London: Basil Blackwell.

Brattico, Pauli (2005) Category-Free model of Finnish derivational morphology. *SKY Journal of Linguistics* 18: 7–45.

Caramazza, Alfonso, Laudanna, Alessandro & Romani, Christina (1988) Lexical access and inflectional morphology. *Cognition* 28: 297–332.

Chomsky, Noam (1970) Remarks on Nominalization. In Roderick Jacobs & Peter Rosenbaum (eds.), *Readings in English Transformational Grammar*, pp. 184–221. Waltham: Ginn.

Cowart, Wayne (1997) Experimental syntax: *Applying objective methods to sentence judgments*. Thousand Oaks: Sage Publications.

Emmorey, Karen D. & Fromkin, Victoria (1988) The mental lexicon. In Frederick Newmeyer (ed.) *Linguistics: The Cambridge Survey, Vol. 3*, pp. 124–149. Cambridge: Cambridge University Press.

Frost, Ram, Grainger, Jonathan & Rastle, Kathleen (2005) Current issues in morphological processing: An introduction. *Language and Cognitive Processes* 20: 1–5.

Giegerich, Heinz J. (1999) *Lexical Strata in English. Morphological Causes, Phonological Effects.* Cambridge: Cambridge University Press.

Hale, Kenneth & Keyser, Samuel J. (1993) On Argument Structure and the Lexical Expression of Syntactic Relations. In Kenneth Hale & Samuel J. Keyser (eds.), *The View from Building 20*, pp. 53–109, Cambride, MA.: MIT Press.

Halle, Morris & Marantz, Alec (1993) Distributed morphology and the pieces of inflection. In Kenneth Hale & S. Jay Keyser (eds.), *The View from Building 20*, pp. 111–176. Cambridge: MIT Press.

Hankamer, Jorge (1989) Morphological parsing and the lexicon. In William Marslen-Wilson (ed.), *Lexical Representation and Process*, pp. 392–408. Cambridge, MA: MIT Press.

Hakulinen, Auli, Vilkuna, Maria, Korhonen, Riitta, Koivisto, Vesa, Heinonen, Tarja R. & Alho, Irja (2004) *Iso suomen kielioppi*. Helsinki: SKS.

Karlsson, Fred (1983) *Suomen kielen äänne- ja muotorakenne*. Juva: WSOY.

Kiparsky, Paul (1982) Lexical Phonology and Morphology. In In-Seok Yang (ed.), *Linguistics in the Morning Calm*, pp. 3–91. Seoul: Hanshin.

Koskenniemi, Kimmo (1985) *Two-level Morphology: A general computational model of word-form recognition and production.* Publication No. 11, Department of General Linguistics. Helsinki: University of Helsinki.

Lieber, Rochelle (1992) *Deconstructing Morphology*. Chicago: University of Chicago Press.

Marantz, Alec (1997) No escape from syntax: Don't try morphological analysis in the privacy of your lexicon. In Alexis Dimitriadis, Laura Siegel, Clarissa Surek-Clark & Alexander Williams (eds.), *Proceedings of the 21st Annual Penn Linguistic Colloquium*, pp. 201–225. Pennsylvania Working Papers in Linguistics 4.2. Philadelphia: Penn Linguistics Club.

—— (1999) *Creating verbs above and below little v*. Unpublished manuscript, MIT.

—— (2000) Reconstructing the lexical domain with a single generative engine. Unpublished manuscript, MIT.

McMahon, April (2000) *Lexical phonology and the history of English*. Cambridge: Cambridge University Press.

McQueen, James & Cutler, Anne (1998) Morphology in word recognition. In Andrew Spenser & Arnold M. Zwicky (eds.), *The Handbook of Morphology*, pp. 406–427. London: Basil Blackwell.

Niemi, Jussi, Laine, Matti & Tuominen, Juhani (1994) Cognitive morphology in Finnish: Foundations of a new model. *Language and Cognitive Processes* 9: 423–446.

Packard, Jerome L. (2000) *The Morphology of Chinese: A linguistic and cognitive approach.* Cambridge: Cambridge University Press.

Plag, Ingo (2003) *Word-Formation in English*. Cambridge: Cambridge University Press.

Sagart, Laurent (2001) Vestiges of Archaic Chinese Derivational Affixes in Modern Chinese Dialects. In Chappell, H. (ed.), *Sinitic Grammar. Synchronic and Diachronic Perspectives*, pp. 123–142. Oxford: Oxford University Press.

Schreuder, Robert & Baayen, R. Harald (1995) Modeling morphological processing. In Laurie B. Feldman (ed.), *Morphological aspects of language processing*, pp. 131–154. Hillsdale: Erlbaum.

Schütze, Carson T. (1996) *The empirical base of linguistics. Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.

Selkirk, Elisabeth O. (1982) *The Syntax of Words*. Cambridge: MIT Press.

Sulkala, Helena & Karjalainen, Merja (1992) *Finnish.* London: Routledge.

Ullman, Michal T. (2001) A Neurocognitive Perspective on Language: The Declarative/Procedural Model. *Nature* 2: 717–726l.

Contact information:

Pauli Brattico
Department of Computer Science and Information Systems
00014 University of Jyväskylä
Finland
e-mail: pajubrat(at)jyu(dot)fi

Alina Leinonen and Christina M. Krause
Cognitive Science Unit
Department of Psychology
University of Helsinki
00014 Helsinki
Finland
e-mail: alina(dot)leinonen(at)helsinki(dot)fi
e-mail: christina(dot)krause(at)helsinki(dot)fi