Author(s): Brattico, Pauli

Title: On the semantic interpretation of complex causatives in Finnish: An experimental morphology approach

Year: 2009

Version: Published version

Please cite the original version:

On the semantic interpretation of complex causatives in Finnish: An experimental morphology approach

Pauli Brattico

Department of Computer Science and Information Systems, University of Jyväskylä & Department of Psychology, University of Helsinki

Abstract

It is well documented that in some languages, such as in Finnish, it is possible to stack derivational morphemes iteratively to word stems. However, such complex words are seldom used in real communication, and it is unclear whether they are interpreted compositionally in tandem with their morphological structure. Here I studied the matter by eliciting semantic interpretations from ten native speakers of Finnish for words whose complexity and morphemic content were systematically varied. The results show that although the frequency of semantic interpretation decreases linearly as a function of the number of morphemes in a word (contrary to the case of ungrammatical words where the semantic interpretation is lacking), when the participants provided semantic analyses of complex words, these analyses were compositional or nearly compositional. Even three iteratively stacked causative morphemes were analyzed as a true triple causative, i.e., 'when some person makes another person to get a third person to eliminate that person'. I conclude that while speakers possess accurate linguistic knowledge of the semantic properties of iteratively formed words, some linguistic or extra-linguistic factors make the extraction of such meanings difficult. Some possible reasons for the existence of such limitations are discussed. Furthermore, I discuss the relevance of these findings to the theory of word formation and suggest that in addition to strict rules, word formation is subject to graded, soft constraints.

Keywords: causatives; double causatives; triple causatives; Finnish; word formation; complexity; linguistic complexity; complexity effect; indirect causation; morphology

1. Introduction

Human languages are based on at least two fundamental cognitive components: a lexicon, which contains a finite list of words or idioms that belong to a particular language, and a creative rule component, which generates an infinite number of expressions by applying the syntactic rules to the lexicon (and to the resulting complex linguistic elements, recursively). Word formation, i.e., our ability to form new words, falls somewhere between these two components. While in many languages it is possible to form novel words creatively by combining morphemes (i.e., *iraqification* in English), this ability is not as productive as syntax in that very complex words are infrequent in use and perceived as odd by native speakers. Moreover, while all languages in the world contain a lexicon and a fully productive syntax, there is much variation in productivity in word formation. For instance, Finnish is quite productive in its word formation (Brattico 2005, Brattico, Leinonen & Krause 2007, Hankamer 1989, Hakulinen et al. 2004:§155, Kytömäki 1977, 1992, Lehtonen et al. 2006, Niemi, Laine & Tuominen 1994), while Chinese, Vietnamese or Burmese are less so (Sagart 2001). Thus, I will say that word formation possess 'limited generativity' instead of the full generativity of the syntax or strict finiteness of the lexicon.

Another factor that adds to this problem is that complex words are, according to some authors, not interpreted like complex phrases in several respects. First, the interpretation of complex words is not always systematic. For instance, while *drug campaign* is interpreted as a campaign against drugs, *health campaign* is interpreted as a campaign in favor of health, although both are similar in their syntactic properties. Similarly, a girlfriend is something in addition to being merely a girl and a friend. The opposite happens when some morphemes within the word are not interpreted at all. Consider

Finnish causativization, as illustrated in (1a-d). The triple causative element (1d) is taken from a real Finnish text. [1]

(1)
a. tee-n
    make-SG1
    'I make'

b. tee-tä-n
    make-CAU-SG1
    'I cause somebody else to make'

c. tee-tä-tä-n
    make-CAU-CAU-SG1
    'I cause somebody else to cause somebody else to make'

d. tee-tä-ty-tä-n
    make-CAU-CAU-CAU-SG1
    'I cause ... to make'

It is unclear whether speakers of Finnish interpret Finnish triple causatives, such as in (1d), as containing three distinct causative morphemes. In other words, it is unclear whether and how they distinguish (1a-d) semantically from each other. Some linguists thus think that word formation is subject to the property that the semantic properties of the whole item cannot be predicted (completely) from the meanings of its constituents. Another interesting question concerning the status of (1d) is whether these items belong to Finnish at all, and if not, whether they violate native speakers' intuitions as strongly as words which are ungrammatical (e.g., *punainen-taa, red-CAU, 'to cause to be/have red color').

In this article, I address one aspect of this question, namely, whether simple, double and triple causative suffixes are interpreted compositionally both when they are merged to simple root stems and to complex stems. In other words, the topic under investigation is 'limited generativity' in the semantic sense. Semantic analyses were elicited from a pool of native speakers of Finnish for morphologically complex Finnish words that were generated specifically for the purpose of the experiment, and the semantic protocols then constituted the basic corpus of the study. More specifically, sixteen (16) root stems, both verbs and nouns, were first selected. After the participants (10 altogether) completed the analysis of the given root stem, a random derivational morpheme was added to the word, and another semantic analysis was then elicited for the more complex word form. Many of these morphemes were causatives since causatives constituted the main research question of this study; the rest are viewed here as fillers to prevent the participants from detecting the underlying experimental manipulation. (However, the whole corpus of the semantic protocols elicited in the study is provided together with this report, and thus the semantic properties of other derivational morphemes, and their many combinations, can be potentially extracted from the same source.) Altogether four derivational morphemes were added to each word in this way. After four derivational morphemes had been added to the stem, the next root stem was provided and the process continued. The whole experiment took less than one hour for each subject.

Since the morphemes were added one by one to the stem, the morphological parsing of the stimulus words was made as explicit as possible. Thus, the participants were aware that a word such as *juoksu-ttaa* 'to cause to run' was composed from a root stem *juoksu-* 'to run' and a causative suffix -*ttA*- because they analyzed the meaning of *juoksu-* before analysing *juoksu-ttaa*. This method was used here to control for parsing problems with complex words. See Brattico, Leinonen & Krause (2007) for a similar experiment where all stimulus items were randomized and no explicit parsing was available.

I expected three possible outcomes of this experiment. According to one outcome, the speakers would simply refuse to analyze the meanings of words after they had become too complex, and so while simple causatives would be analyzed as causatives, complex causatives would not be analyzed at all. The threshold for such an abrupt loss of performance would then be determined by the experiment itself. If this turns out to be the case, I would conclude that word complexity in Finnish cannot exceed a given finite limit. I do not expect the threshold to be dichotomous, in that for all stimulus items of a certain complexity, all informants would be unable to analyze its meaning; rather, the hypothesis is that the performance drops in a non-linear manner. This hypothesis can be tested by appropriate statistical methods.

According to the second possibility, an increase in the morphological complexity of a word would not decrease its semantic interpretability, but the semantic analyses provided by the participants would not be compositional, i.e., the speakers would ignore the meanings of many morphemes, or provide deviant meanings that are not related to the morphological structure of the word. This would indicate that morphological complexity leads to a failure of analysis rather than to a lack of analysis. It would then be interesting to analyze the failures at an in-depth level. For instance, it may appear as plausible that the morphemes in the middle of a word would be more likely to be ignored than the morphemes close to the root and to the word boundary.

According to the third possibility, the analyses provided would faithfully represent the morphological structure of the words irrespective of their complexity. Under this scheme, I would conclude that the speakers possess accurate knowledge of the morphological structure of complex words irrespective of their morphological complexity.

The use of several informants allows us to tackle individual variation. For instance, it is possible that while some speakers are able to analyze morphologically complex words compositionally (third hypothesis), others would fail either compositionality (second hypothesis) or fail the whole analysis (first hypothesis). In that case, the three hypotheses should be relativized to the speaker, while such variations must be attributed to some other source (education, age, gender, and others).

2. Word formation: syntax or not?

The study of word formation can be categorized into two types of research agendas that I want to keep distinct, but which are related, if somewhat indirectly, to each other. One of these research agendas is concerned with the rules which define a set of words in some language currently in frequent (and possibly infrequent) use. We can think of such research as aiming at a complete description of some kind of external or autonomous language entity, at a particular moment of time, with respect to its lexical and morphological inventory. Thus, through this route we would arrive at the word formation rules of a geographically, temporally, and stylistically restricted notion of Finnish (see Hakulinen et al. 2004:18-21). Of course, such a set exists in an enumerative form whenever we have a comprehensive dictionary at our disposal; thus, the research project can aim only at compressing that set by means of regularities. When seen from this perspective, the fact that word formation possesses limited generativity concerns the compressed description of that set of expressions: words as they occur in this domain, unlike sentences, have a strict upper bound on their length. (Note that this set may be, and often is, more structured than a set of expressions; it can contain pairs of expressions and meanings/uses, for instance.)

The present research agenda is based on a different goal, the description of the cognitive and neurocognitive mechanism underlying the use of complex and simple words. Sometimes this perspective is said to be interested in the 'knowledge of language' or an internal language (I-language) instead of language as an external entity or as a set of expressions. Seen from this perspective, the fact that word formation is based on limited generativity should be understood as an assertion concerning the cognitive mechanisms underlying word formation; their generativity falls somewhere between fully productive operations, such as syntax, and fully non-productive operations, such as table-lookup memory. Under this approach the set of words in use in present-day Finnish is not the

target of description or data compression. First of all, we are attempting to list all the words that could be used and understood by native speakers irrespective of whether they have been introduced to the language at some particular point in time or place.

When seen from the latter perspective, linguists have not reached agreement on how to best describe and explain such limited generativity. While some authors conceive word formation as an extension of syntax proper (Baker 1988, Brattico 2005, Julien 2002, Lieber 1992, Marantz 1997, Selkirk 1982, Ullman 2001), others disagree (Anderson 1982, Chomsky 1970, Karlsson 1983). Those who disagree have still plenty of options, for instance, processes based on analogical reasoning, autonomous morphological modules, and so forth. Much of the controversy revolves around the fact that word formation is both like and unlike syntax: "The last decades have witnessed the development of two competing lines of thinking about morphology and its relation to syntax", one which says that word formation is part of syntax and another which denies this assertion, while the "schism stems from the trivial observation that morphology and syntax are seemingly separate domains of grammar which still to some extent must interact" (Julien 2002:8). Note that this disappointing state of affairs is not in any sense peculiar to morphology: a similar situation obtains currently in most if not in all cognitive domains, which is expected given the complexity of the human brain and restrictions on the methodologies that could be applied to this subject matter; the nervous system of the fruit fly remains a profound mystery even though it represents a problem of much smaller scale and posits fewer methodological barriers. [2]

Dowty (1979) provides another useful discussion on the matter. He points out that "it is universally agreed that principles of word formation are real enough principles that must be described in any account of a native speaker's knowledge of his language, yet these principles are everywhere subject to exceptions [...] both in the matter of 'productivity' [...] and in semantics" (p. 295). He also points out what I regard as an important cognitive observation that derived words are subject to another peculiarity that people are 'cautious' about using a derived word they have never heard before, although the word would still be intuited as possible. This is in clear contrast with phrasal syntax, where we use new phrases all the time without being cautious about the process at all. Hence, in the case of words, memory traces of previous use are an important factor, while in the case of syntactic phrases, this is not so. This is another factor that has to be explained.

Taking formal and semantic productivity as a criterion for making the distinction between syntactic and lexical rules, Dowty then points out that this distinction nevertheless crosscuts the traditional domains of morphology and syntax. Although there are clear instances of syntax which use productive syntactic rules (e.g., permutations of phonological words within an expression) and morphology which uses less productive lexical rules (e.g., zero derivation, certain derivational word formation), there are also instances of syntax which uses less productive rules (e.g., idiom formation, the formation of lexical units of more than one word) and instances of word formation that uses productive rules. Of the last category, he mentions inflectional morphology and derivational morphology when this is particularly productive, as in the case of polysynthetic languages. I further think that this whole classification should be partitioned into two separate groups, one where these properties concern the formal aspects of the process and another where they concern the semantic aspect of the process. In the case of compounding, for instance, the process is formally productive while semantically less so. In any case, given the existence of such languages as Eskimo where the speakers can form extremely complex words with transparently productive rules, we are led to believe that productivity is not a peculiarity of words, but of certain linguistic operations which are perhaps only typically applied to the morphological domain (with significant variation across languages). Finnish causativization is interesting in this respect since it falls somewhere between these two extremes: it is certainly based on a recursive causativization rule, but it seems on the other hand that this potential cannot be used to its full extent if compared to the potential in compounding and in the syntax proper. Thus, we have a case of limited generativity that falls somewhere between productive and non-productive rules.

3. Causatives and causativization

Causation is a concept, or a phenomenon, that has a tendency to find its way into the grammar of natural languages. First, many transitive verbs involve causation in the sense that when, for example, John boils the water, he also causes the water to boil. The same is true of a number of lexical entailments, such as *kill → cause to die*, *open → cause to open*, *sink → cause to sink*, *feed → cause to eat* and so on (e.g., Comrie 1976, 1985, Kemmer & Verhagen 1994, Shibatani 1976, Song 1995). Let us call these lexical causatives. Although the picture is more complex than what these simple examples suggest, most linguists agree that this feature cannot be accidental and that there must be a reason why many words in the world's languages involve causation and not, say, wanting or needing. [3]

Another way in which causation enters the grammar is via productive word formation. In the simplest case, causation is expressed via a causative morpheme (e.g., *large - enlarge*). This is also true of the Finnish causatives investigated in the present study. In Finnish, this type of morphological causativization is also productive in that after the word has been causativized, it can be causativized again (Karlsson 1983, Kytömäki 1992; examples 1a-d above). Let us call these morphological causatives. Finally, languages contain many lexical items which denote causation in various forms, such as English *cause* (*John caused the door to open*), *make* (*John made the door open*) and *have* (*John had the door open*). I call these analytical or periphrastic causatives. The distinction between lexical, morphological and analytical causatives is an idealization and ignores both other construction types, such as the Romance-type serial verb causatives (see, e.g., Achard 2002, Aissen 1974, Kayne 1975, Zubizarreta 1985) and finer distinctions among the major categories (see Dixon 2000 and Song 1995). [4]

Turning to Finnish morphological causatives, specifically, the present study concentrates on the recursive (T)TA-causatives which are formed by suffixing a nominal, verbal or adjectival stem with the causative morpheme *-(t)tA-*. Some examples are provided in (2a-c).

(2)
a. *juoksu-ttaa*, run-CAU, 'to cause to run'
b. *paalu-ttaa*, pole-CAU, 'to cause to be/have poles'
c. *syö-ttää*, eat-CAU, 'to feed'

However, the class of TTA-causatives, as in the classification of 'causatives' more generally, is far from unproblematic. The main reason for this obstacle is that there is no isomorphism between meaning and form in the case of natural languages. Thus, if we take the concrete forms as the basis of classification, we get heterogeneous and often irrelevant semantics into the target class; on the other hand, it is very hard to classify objects on the basics of their semantics due to the fact that 'semantic interpretation' is not any kind of concrete entity. On the contrary, the notion of semantic interpretation, even in the case of causatives alone, is often a much more mysterious phenomenon than the original expression itself. [5] I satisfy myself with a heuristic classification.

Concerning the matter from the perspective of linguistic form, in some cases the causative suffix TTA surfaces in a slightly different form, such as *-stA-* (*hampaalli-staa*, teeth-CAU, 'to cause to have teeth'). I will assume here that the choice between *-ttA-* and *-stA-* is predictable (thus, forms such as **hampaallisttaa*, **hampaallittaa*, **hampaallinentaa* are impossible for various reasons). In some cases what looks like the *-ttA-* suffix has been lexicalized to the stem (*roko-ttaa*, 'to vaccinate, lit. to cause to have pox'). I will assume that as a polymorphemic word lexicalizes, it obtains non-compositional, emergent and idiomatic semantic features. In the present experiment, there was one item that could be classified in this way, *häivy-ttää*, 'vanish-CAU'. As pointed out by Dowty (1979), the correct theory of word formation should incorporate both aspects: a fully predicted but approximate meaning of the word and the meaning that the word obtains through concrete usage.

From the meaning side, as described by many linguists, the TTA-forms are often intuited to refer to indirect causation in that 'somebody else' is asked/made/caused to do something (Hakulinen et al.

2004:§311-314, Kytömäki 1977, 1992). In fact, these forms seem to denote more indirect causation than the analytical causatives which often function as their paraphrases. I present three reasons to support this assertion since it may look unintuitive that morphological causatives are more indirect than analytical ones; the present experiment will later confirm these observations. First, morphological causatives, unlike analytical ones, are incompatible with a reflexive reading where the agent performed the action described in the complement sentence by him- or herself:

(3)
a. *Pekka*   *valmistu-tt-i*   *työ-t*
    Pekka   ready-CAU-PAST   work-ACC
    'Pekka caused somebody else to do the work' (indirect causation)
    *'Pekka did the work by himself' (direct causation)

b. *Pekka*   *aiheutt-i*   *töide-n*   *valmistumi-sen*
    Pekka   cause-PAST   work-GEN   completion-ACC
    'Pekka caused the work to be completed by asking somebody else to do it' (indirect causation)
    'Pekka did the work by himself' (more direct causation)

This hypothesis is supported by the fact that morphological TTA-causatives in Finnish allow the causee to appear as an oblique argument (4a). This is not equally possible in the case of analytical (4b) causatives, and totally impossible in the case of lexical causatives (4c). In all the examples, the main verb is in the past tense.

(4)
a. *johtaja*   *hoida-tti*   *likaise-t*   *työ-t*   *alaisi-lla*
    manager-NOM made-CAU dirty-ACC work-ACC employees-ADE
    'The manager made the employees do all the dirty work.'

b. ??*johtaja*   *aiheu-tti*   *töi-den*   *tekemisen alaisi-lla*
    manager-NOM caused   work-GEN do-ACC   employees-ADE
    'The manager made the employees do the work.'

c. *johtaja*   *tek-i*   *työ-t*   *alaisi-lla*
    manager-NOM did-PAST   work-ACC employees-ADE
    *'The manager made the employees do the work.'

Thus, although both analytical and morphological causatives in Finnish are biclausal, they differ in that the causative morpheme in Finnish is associated with an implicit participant, whereas the presence of such an implicit agent is optional in the case of the analytical causative. Further evidence which supports this contention comes from the scope properties. For reference, consider the following analytical causative:

(5) *Pekka*   *käski*   *koira-n*   *juosta nopeasti*
     Pekka-NOM ordered dog-GEN run   fast
     'Pekka ordered the dog to run fast.'

The standard interpretation of this sentence is that the adverb *nopeasti* 'fast' is associated with the running of the dog. If the adverb occurs immediately after the matrix verb, it is associated with the matrix verb, so that now the asking was fast, not the running. If the adverb is topicalized and occurs at

the left periphery of the clause, the expression becomes ambiguous, and the adverb may be associated either with the matrix verb or the embedded verb:

(6)   *nopeasti*   *Pekka*   *käski*   *koira-n*   *juosta*
       fast   Pekka-NOM   ordered   dog-GEN   run
       fast: Pekka ordered the dog to run (WIDE SCOPE INTERPRETATION)
       Pekka ordered *fast*: the dog to run (NARROW SCOPE INTERPRETATION)

Let us apply this test to morphological and lexical causatives. As for the morphological causatives, the relevant sentence is (7). This sentence has again two possible interpretations, comparable to the analytical causative:

(7)   *nopeasti*   *Pekka*   *juoksu-tt-i*   *koira-a*
       fast   Pekka-NOM   run-CAU-PAST   dog-PRT
       fast: Pekka caused the dog to run (WIDE SCOPE INTERPRETATION)
       Pekka caused *fast*: the dog to run (NARROW SCOPE INTERPRETATION)

This means that Finnish morphological causatives contain two dissociated events to which the adverb may be associated: either the 'inner' event where the dog runs, or the 'outer' event where Pekka caused something. Lexical causatives, however, are never ambiguous in this way:

(8)   *nopeasti*   *Pekka*   *tappo-i*   *kärpä-sen*
       fast   Pekka   killed-PAST   fly-ACC
       fast: Pekka killed the fly (WIDE SCOPE INTERPRETATION)
       *Pekka caused *fast*: the fly to die (NARROW SCOPE INTERPRETATION)

Morphological causatives in Finnish are comparable to analytical causatives in that adverbs have two events to which they may be associated, the causer event (wide scope) and the cause event (narrow scope).

To summarize these observations, lexical causatives are the most direct in the underlying causative bond between the cause and the caused event. Analytical causatives are less direct than the lexical causatives, but they appear to be more direct than the morphological causatives, which are the most indirect in my sample. These observations are true of a number of causatives, but they are also violated by many morphological causatives that have been lexicalized. It is conceivable that such lexicalization shifts the interpretation from the morphological causative towards the lexical one, and correspondingly, the causative bond tightens. A complete list of stimulus words can be found from the corpus containing the protocols obtained from the participants in the experiment. Finally, the experiment itself allows us to assess to what extent the participants interpret the morphological causatives as implying very indirect causation.

4. Methods and stimulus materials

4.1 General comments

I used several native speakers to elicit semantic analyses of words, and analyzed these protocols both quantitatively and qualitatively in order to assess whether the interpretations were compositional. I think that this methodology is superior to alternatives when it comes to the linguistic study of novel complex words. First, the intuition of one linguist - a frequently used method for data gathering in linguistics - can hardly provide reliable grammatical data concerning very complex words. Consider, for instance, whether the item in (1d) should be said to be a grammatical or semantical (/interpretable) word in Finnish. One may at first rule it out as 'impossible', but on the other hand the grammaticality or

semanticality between items in (1a-d) seems graded rather than binary. At which point should we say that any of these words are completely impossible? The present experiment also reveals that subjects are often conscious of this conflict in their intuitions, as shown by the following reply to an analysis of the Finnish word *villa-ttaa* 'to cause to be/have wool':

(9)   Mä varmaan veikkaisin, että tää ois joku semmonen termi, että tällä niinku tehdään sitä villaa. Mutta tota, jos joku käyttäis tätä termiä, niin mä oikeesti ajattelisin että tää on jotain outoo murretta, niin ehkä mä en sit tajuis tätä.

         'I guess that this could be a term which refers to something which is used to make wool [causative]. On the other hand, if somebody used this term I would think that it represents some strange dialect that I do not understand.'

In contrast, in the case of ungrammatical words (e.g., *punainen-taa* 'redden'), native speakers do not hesitate to analyze the word as completely unsemantical and ungrammatical (Brattico, Leinonen & Krause 2007), the term 'unsemantical' referring to the fact that they could not analyze the meaning of the word.

Second, corpus analysis or a search from a dictionary is impossible since the token and type frequencies of novel words are virtually zero. In my opinion, then, one reasonable way to approach this problem is to gather enough native speakers in a behavioral experiment and somehow elicit semantic analyses or other adequate behavioral responses to these items. Dowty (1979:311) mentions other potential methodologies, such as overgeneralization by children, utterances of certain types of aphasics, and the creative use of word formation by poets. All the stimulus words and the corresponding semantic protocols obtained from the ten subjects are provided together with this research report. You can view the data as a web page or download the data as a Word document. In that database, the protocols obtained for each subject and each stimulus word, 772 items altogether, are presented. However, as of now this corpus has not been annotated with English glosses, and it is thus usable only for a native speaker of Finnish.

I used my previous model of Finnish word formation (Brattico 2005; Brattico, Leinonen & Krause 2007) to generate the stimulus words mainly because I was familiar with the model and I had a formal stimulus-generating algorithm readily available. The details are not essential for understanding the experiment and the results. The morphemes used in this study and their selection restrictions are provided in Table 1. Note once more that these suffixes were selected randomly when generating the words; thus no linguistic intuition was used to bias the selection.

Table 1. The 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. The right column lists the allomorphs which were used in this study. The selection of these allomorphs is a matter of morphological readjustment rules which are not described in this article.

Layer 1	S-selection	Meaning	Allomorphs
CAU[eve]	[ref][eve]	'to cause to -'	(t)ta, sta
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

The causative morpheme has also a habituative reading in which it means approximately 'to feel like doing something'. Examples of each of these morphemes are given in (10).

(10)

a.   *paalu > paalu-ttaa* (CAU)
     'pole, to cause to have a pole'

b.   *syö > syö-ttää* (CAU)
     'eat, to feed'

c.   *paalu > paalu-ile* (FRE)
     'pole, to be a pole habitually'

d.   *juokse > juoks-u* (EVE)
     'run, an event of running'

e.   *sisältä- > sisälty-* (REF)
     'to include, to become included'

f.   *hyvä > hyvyys* (US)
     'good, goodness'

g.   *auto > auto-ll-* (POSS)
     'car, somebody who possesses a car'

h.   *talo > talo-sto* (COL)
     'house, a collection of houses'

Although these details are somewhat inessential to the present study, the main idea of Brattico (2005) is that there are category neutral stems in Finnish word formation. These are morphemes which do not belong to any of the full lexical categories, such as verbs, nouns, or adjectives. There were several reasons I thought this would be useful in understanding Finnish word formation. First, derivational and inflectional suffixes are often productively merged to a 'weak' noun stem that does not constitute a phonological word alone. For instance, the word *vede-ssä* 'in the water' is crafted by suffixing the inessive suffix *-ssA* to the stem *vede-*, where the latter cannot be used as a phonological word. Furthermore, this is the productive root stem to which most suffixes are merged (e.g., *vede-llinen* 'to have water', *vede-stä* 'from the water', *vede-n* genitive/accusative). The same is true of many verbal stems (*vaati-minen > *vaati-*, *autta-minen > *autta-*; note the effects of constant gradation, **autta- > auta-n*, *lautta > lauta-n*). I argued that the stem *vede-* is a category neutral morpheme, while the exceptional nominative singular *vesi* (categorized as the 'strong stem' by traditional grammarians), which can be used as a phonological word, is provided by suppletion. For instance, note that when the above suffixes are merged to *vesi*, a plural interpretation is triggered. Thus, *vesi-stä* means 'from the waters' while *vede-stä* means 'from the water'. This means that *vesi-* is a plural form of the category neutral root *vede-*.

Second, the same appears to be true of many morphologically complex stems. Consider the causative morpheme TTA 'to cause to'. When merged to another stem, this suffix generates a complex stem such as *juokse-tta-* 'to cause to run' that cannot be used as a phonological word (where consonant gradation changes, e.g., *juokse-tta-n* into *juokse-ta-n*, compare *lautta > lauta-n*). Thus, the causative morpheme does not make the stem a full verb, noun, or an adjective; further morphological operations are needed to achieve this result (e.g., noun *juokse-tta-minen*, adjective *juokse-tta-va*, adjective *juokse-tta-ma*, non-finite verb *juokse-tta-a*, participle verb *juokse-tta-van*, imperative *juokse-ta!*). Thus, I argued that the causative morpheme is a category neutral morpheme and that further grammatical operations were needed to categorize the stem either as a full verb, noun or an adjective. The same is true of other derivational 'verbal' suffixes in Finnish as well. It is perhaps important to note that implicit in this argument was the assumption that the underlying stem in a composition 'stem+suffix' is determined, or at least could or should be determined, on the basis of productivity. Thus, in the case

of 'verbal' stems, these are determined, for example, by extracting the productive nominalizer *-minen* (or other productive suffix) from the stem. We obtain non-word stems such as those marked with brackets in (11a-d).

(11)

a.   [*juokse-tta*]-*minen*(/ *-a*, *-va*, *-ja*, ...)
        run-CAU-N
        'the event of causing to run'

b.   [[*hypä*]-*htä*]-*minen*(/ *-ä*, *-va*, *-jä*, ...)
        jump-MOM-N
        'the event of jumping momentarily'

c.   [*laula-ele*]-*minen*(/ *-la*, *-va*, *-ija*, ...)
        sing-FRE-N
        'to sing frequently'

d.   [[*sisältä*]-*y*]-*minen*(/ *-ä*, *-vä*, *-jä*, ...)
        contain-REF-N
        'the event of being contained in something'

This approach still allows us to apply consonant gradation to the category neutral stems irrespective of their categorical status (verbs *juokse-tta-a* 'to cause to run' > *juokse-ta-n* 'I cause to run', nouns *lautta* 'board' > *lau-ta-n* 'board's', adjectives *törttö mies* 'badly behaving man' > *tör-tö-n miehen* 'badly behaving man's').

Third, some suffixes productively associated with certain lexical categories behave in such a manner that when they are extracted from their host stems, what is left does not constitute a phonological word. This concerns, among others, the productive nominal/adjectival suffix *-inen*. Thus, we can suffix *-inen* to a noun to result in an adjective (e.g., *talo-inen*, 'something like a house'), but when extracted from many words, the resulting stem does not constitute a phonological word (e.g., *hevo-*, *puna-*, *lähtemi-*, *talo-lli-*). I presented arguments that we should analyze *-inen* as a separate morpheme which is suffixed to category neutral stems; I will not repeat these points here. This is not to deny, however, that in some cases the morpheme may be frozen to some stem or to some suffix, as we have to separate lexicalization from productive morphology.

The main motivation for the assumption that full lexical categories are not bestowed on many morphemes does not derive from the particulars of Finnish word formation, however, but from the following underlying problem: assuming that virtually all phonological words are labelled with some categorical feature (+N, +V, +A, or whatever these may ultimately be), where does this feature composition take place, and what are the lexical items to which the categorical features are attached? Plainly, the answer to the second question must be that the lexical items are categorically neutral roots. The answer to the first question is controversial, but since the process itself is real, it makes sense to ask whether it could be completely invisible to all linguistic processes, perhaps part of some sort of extra-linguistic conceptual system that can be ignored in all linguistic studies, without any loss of explanatory power. [6] I tried to argue that the categorization process where a root is tagged with its categorical feature should not be ignored either in Finnish word formation or in Finnish syntax.

4.2 Procedure and participants

Ten participants, two men and eight women, were recruited for this experiment (mean age 27.6, S.D. = 3.2, range 23-33). Eight of the participants were university students, and all were linguistically naïve native speakers of Finnish. The instructions for the experiment were presented in written form. After reading the instructions and completing a questionnaire gathering personal details about the participants, the experiment began. Each stimulus word was presented, one at the time, on a computer screen. The presentation order of the root stems was randomized for each subject, but the

derivational morphemes were added to each root stem always in the same order, i.e., in the order they appear in the word after the stem. 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. After the word was presented on the screen, the subjects described the meaning of the stimulus 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. Following the answer, 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 experimenter was present during the experiment, but did not interfere with the process until the experiment was over. The whole experiment took approximately 30 minutes. Due to a problem with the recording instrument, we could not analyse some of the verbal protocols obtained from two participants (two stimulus words were left unanalyzed for one subject, three for another; these gaps are visible in the corpus).

5. Results

5.1 General

In the quantative analysis, I plotted the mean frequency of the semantic interpretation, as elicited from all participants, by the morphological complexity of the word, using five complexity classes ((1) root stem, (2) +1 morpheme, (3) +2 morphemes, (4) +3 morphemes and (5) +4 morphemes). Individual variation is shown by the standard error of the mean. The resulting plot is shown in Figure 1.
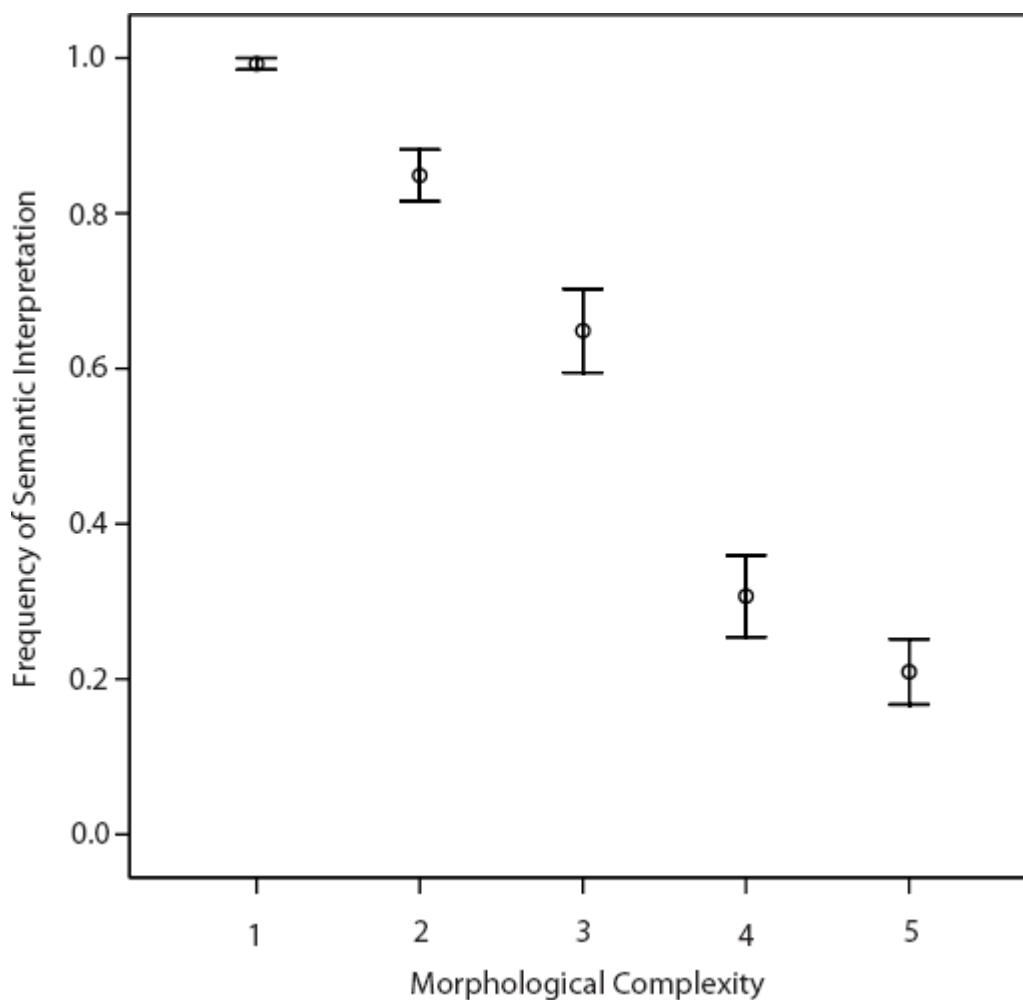
Figure 1. Mean frequency of semantic interpretation (with standard error of the mean) as a function of morphological complexity in a participant-based analysis.

As can be seen from this figure, the frequency of semantic interpretation drops linearly as a function of morphological complexity. Statistical tests were then carried out in order to validate this hypothesis. [7] In a participant-based analysis, the number of derivational morphemes had a statistically significant effect on the frequency of semantic interpretation so that the more complex words were less likely to be interpreted semantically [repeated measures ANOVA $F_{(4, 36)} = 63.458$, $p < 0.001$]. The decrease in the frequency of semantic interpretation appears to be close to a linear function of the number of derivational morphemes (=morphological complexity) [within-subjects contrasts for the linear relationship, $F_{(1, 9)} = 206.328$, $p < 0.001$]. This result was also significant in the case of the item-based analysis [ANOVA $F_{(4, 60)} = 105.162$, $p < 0.001$]. These results seem to replicate our previous findings that morphological complexity affects both the grammaticality and semanticality of the word (=frequency of semantic interpretation) in a linear fashion (Brattico, Leinonen & Krause 2007).

Consider the three hypotheses as presented in Section 1 in the light of the data in Figure 1. This data shows that hypothesis 3 must be wrong in the sense that there is some factor which makes the interpretation of a word more difficult as a function of the complexity of the word irrespective of the fact that the parsing of the word was given explicitly. The native speakers were not just registering the upcoming morpheme and analyzing it through the background of the previous analysis, but they treated each word as a new whole that may or may not have a proper meaning. Somehow, then, the causative morpheme that was attached to a simple word was analyzed as a causative, but not when applied to a complex word. To understand better the reasons why semantic interpretation is related linearly to morphological complexity, we have to look both at the individual variation and the verbal protocols.

5.2 Single causative morphemes

In this section, I will look more carefully at how the native speakers reflected on the stimulus words. Some linguists might regard the analysis of very complex words as bizarre, as such words are infrequent or outright inexistent; clearly they are strange to the native speakers. However, note once more that the point of the whole experiment is to *compare* the behavior of the native speakers within the experimental categories (complexity classes) and not to claim on any normative or empirical grounds that these items should be 'normal words in modern Finnish'. We are not attempting to study a set of words in contemporary Finnish, but to find a reason why there are such differences between the experimental groups (Figure 1).

Before looking at the complex causatives, it makes sense to make sure that the participants analyzed simple causatives as causatives. There were five stimulus words in the dataset which constituted a monomorphemic root stem suffixed with a causative suffix, *villattaa* 'to cause to have/be wool', *aamuttaa* 'to cause to have/be a morning', *lauluttaa* 'to cause to produce a song', *mitattaa* 'to cause to have/be a measure' and *häivyttää* 'to cause to disappear'. Approximately half of these items were analyzed as causatives, whereas the frequency of the habituative interpretation was ten times less, 0.05. The rest of the analyses were either completely missing, more general event descriptions, synonyms, or unclear cases. For example, the word *villattaa* was interpreted as 'to cover with a coat' (verbal report number 61 and 63-66 in the corpus) or 'to make wool' (VRN 69). There was only one habituative reading 'to feel like touching wool' (VRN 67), while two participants did not provide a semantic analysis at all. Some subjects were aware of both the causative and habituative reading (VRN 173). The causative verb *mitattaa* was analyzed as a causative by five participants (VRN 200-203, 207) and as both causative and habituative by one participant (VRN 208), while three participants did not provide an analysis at all. Only two participants perceived *aamu-ttaa* as a causative (VRN 741, 745). The rest interpreted this word as some kind of event description involving morning (VRN 737, 739-740), while four participants did not analyze this item at all.

One important observation emerging from these data is that even in the case of simpler words (e.g., *villattaa* 'to cause to be/have wool'), which are clearly possible given the word formation rules of Finnish (e.g., *paaluttaa* 'to cause to be/have poles'), some participants did not provide an analysis. The present experiment does not allow us to gauge the mental processes during such a silence to find out the reason for this problem. Instead, we should use more fine-grained psycholinguistic experimentation or brain-imaging methods which are more sensitive to processes that are invisible in the protocols.

The second observation concerns the particulars of the causative analyses elicited. Although in many cases the causative analysis clearly functions in the background, the participants selected a wide variety of verbs to express the situation. For instance, the verb *villattaa* 'to cause to be/have wool' was analyzed as 'to cause to cover with wool' and 'to cause to make wool' (perhaps 'to cause to make wool appear'). This means that a possibly large variety of different causative meanings can be projected into the same verb, suggesting that we should not pick up any of these as the underlying basic form. More generally, there is no neat isomorphism between form and meaning; both function to some extent as autonomous systems.

5.3 Double causatives within simple words

Having made sure that the participants analyzed the causative morpheme as a causative, the more interesting question concerns how they analyzed double causatives. Recall that in this experiment, a double causative word was provided right after the subjects analyzed the single causatives. It thus makes sense to compare the participants' analyses of both. This data is collected into the three tables below (Tables 2, 3, and 4).

Table 2. Single causative and double causative interpretations for the word *villa-ttaa > villa-ta-ttaa*. Those participants were removed who did not provide an analysis for either of these items.

Participant	villa-tta-a 'wool-CAU'	villa-ta-tta-a 'wool-CAU-CAU'
12	*Päällystää villalla seiniä.* 'to cover a wall with wool'	
14	*Rakennusta kun eristetään, niin se voidaan villattaa ja pistää sen seiniin lasivillaa tai jotain muuta vastaavaa eristettä.* 'when a building is insulated, one could put wool or some other stuff into its wall'	*Kun tämä erityistoimenpide teetetään jollain toisella, niin se villatetaan.* 'when somebody else is asked to do that operation, then we can use this word'
15	*Niin sitten on karhunvilla. Jos joku vaikka rakentais taloa ja sit se laittais sinne vuorivilloja, niin sen vois ehkä sanoa, että se villattaa sitä.* 'then there is this particular type of wool. If somebody is building a house, and puts this kind of wool into the house, then we could use this word'	
16	*No vaikka kietoa joku villaan.* 'to wrap somebody in wool'	*Käskeä jonkun toisen kietoa joku villaan.* 'to ask somebody else to wrap somebody in wool'
17	*Peittää villalla.* 'to cover with wool'	*Laittaa joku toinen peittämään villalla.* 'to make somebody else cover with wool'
18	*Kyllä mulle tästkin tulee sellanen, sellanen voisko sanoo mielleyhtymä, että jos tekee mieli vaikka kosketella jotain villaa, niin sillon sitä*	

	ihmistä villattaa. Mut ei se nyt kyllä mitään tarkota.
	'this reminds me of an association, where it feels like one would like to touch wool. But it doesn't mean anything'
21	*Mä varmaan veikkaisin, että tää ois joku semmonen termi, että tällä niinku tehdään sitä villaa. Mutta tota, jos joku käyttäis tätä termiä, niin mä oikeesti ajattelisin että tää on jotain outoo murretta, niin ehkä mä en sit tajuis tätä.*
	'I would guess that this would be a term which refers to the making of wool. On the other hand, if somebody used this term, I would think that this is some strange dialect, so I would not probably understand it'
22	*Esimerkiksi, jos haluais villavaatteen ympärillensä kun paleltaa, niin vois villattaa.*	*Saada muut ihmiset laittamaan villatakit päälle.* 'to get other people to dress up with clothes made of wool'
	'if, for example, one wants to get wrapped in wool, then this word would apply to the situation'

As can be seen from this table, when the participants did provide an analysis (14, 16, 17, 22), it was the indirect double causative interpretation 'make somebody else do something with wool'. This provides further reason to say that Finnish TTA causatives are interpreted as indirect. The verbs the participants used for expressing causation were *teettää* 'to cause to make', *käskeä* 'order', *laittaa* 'put' and *saada* 'have', affecting the force and nature of the causation. We can conclude that this aspect of the causative interpretation is not encoded by the causative suffix; rather, the causative suffix expresses abstract causation, and the details are provided by other factors in the context.

Table 3 shows the data for *laulu-ta-tta-a*.

Table 3. Single causative and double causative interpretations for the word *laulu-ttaa > laulu-ta-ttaa*. Those participants were removed who did not provide an analysis for either of these items.

Participant	laulu-tta-a 'song-CAU'	laulu-ta-tta-a 'song-CAU-CAU'
12	*Sillon kun tekee mieli laulaa niin lauluttaa.* 'when you feel like singing, then this word applies to the situation'	*Joku pistää sut laulamaan.* 'somebody makes you sing'
14	*Kanttori lauluttaa kuorolaisia.* 'somebody causes the choir members to sing'	
16	*Tuotattaa jollain laulu, joko äänellisesti tai sitten että tuottaa sillä laulu paperilla.* 'to cause somebody to do a song, either vocally or on paper'	*Täs ois taas niinku, että tekisi mieli tuottaa laulu.* 'this refers to a situation where it feels like that one wants to produce a song'
19	*Sekään ei ehkä ole niinku virallinen sanakirjasana, mutta vois ajatella, että joskus voi... No ehkä se ois enemmän laulatuttaa siinä tapauksessa, jos haluais...* 'This is not a word that one could find from the dictionary, but one could think that one could... maybe it would be laulatuttaa in that case'	
21	*Oisko tää nyt sitten, että tässä ois taas sit tekijä*	*No en mä tätä kyl sanois suomen kielen sanaks,*

vähän se toinen kuin se laulaja, eli joku toinen lauluttaa. Tai oikeestaan ei. Lauluttaa, onks se.. Lauluttaa. Itse asiassa nyt kun mä rupeen miettimään, niin tarkottaaks se sitä, että tekee mieli laulaa. Mä sanon, mä tajusin sen itse asiassa niin, mä luin sen eka väärin vaan.	*mut kyl mä silti tajuisin, jos joku sanois mulle, että sitä laulutattaa, niin edelleen mä sanoisin, että sitä tekee oikeesti mieli laulaa. Tai laulutattaa, tekee toiselle mielen laulaa.*
'This could be so that here the agent is another person than the singer, so that somebody else is causing a person to sing. Or maybe not. In fact, now that I think about it, does this mean that I feel like singing? This is how I would interpret it'	'I would not call this a Finnish word, but I would still understand if somebody said this word to me that one feels like singing. Or somebody causes another person to feel like singing'

22	*Saada joku toinen laulamaan. Vähän käskeä toisia laulamaan.*	*Musta tää ois sama kuin edellinen. Eiku se ois sitten laulattaa. Ei tää oo musta järkevä.*
	'to make somebody to sing, or to order somebody to sing'	'this feels like the previous one [on the left]. No, this does not make sense'

The results were different compared to *villatattaa*. In the case of *laulutattaa*, no participant provided a double causative interpretation, although their analyses were compositional. Participants 12 and 21 provided a single causative interpretation because they analyzed the single causative stem as a habituative. Participant 16 interpreted the second causative suffix as a habituative and not a causative suffix. Participant 22 first identified the double causative interpretation with the single causative, but rejected this view and finished with the thought that *laulu-ta-ttaa* is not a possible word at all. The different results obtained for *villa* 'wool' and *laulu* 'a song' must trace back to the differences in the lexical items. It seems that in the latter case the double causative interpretation was overdriven by the habituative interpretation for one of the causative morphemes. This interpretation might be triggered by the fact that *laula-ttaa* 'to cause to sing / to feel like singing' is more strongly associated with the habituative reading.

Table 4. Single causative and double causative interpretations for the word *aamu-ttaa > aamu-ta-ttaa*. Those participants were removed who did not provide an analysis for either of these items.

Participant	aamu-tta-a 'morning-CAU'	aamu-ta-tta-a 'morning-CAU-CAU'
12	*Laskea aamuja.*	
	'to count mornings'	
14	*Aamuttaminen vois olla joku aamulla tehtävä rituaali, tai ei ehkä rituaali niinkään, mutta tämmönen perinteisesti aamulla tehtävä toimenpide, joka pitää tehdä erityisesti aamulla.*	*Jos jollekin pitää tehdä tällanen aamulla tehtävä toimenpide, niin sitten tämä henkilö tai eläin tai asia pitää aamutattaa.*
	'This could be a ritual performed in the morning, or not a ritual but some act traditionally performed in the morning, something that has to be done specifically at morning'	'if this act [on the left] must be done to some other person, then this word applies to this other person'
16	*Tän vois niinku ajatella, että esmes joku tekeminen voitas niinku siirtää aamuksi, tai ajatella että tulen tekemään sen aamulla, niin voisin aamuttaa tällaisen teon.*	*Jos tekisi mieli, tai olisi mukavaa, jos olisi aamu, niin sitten mua voisi aamutattaa.*
	'here one could think of moving some activity to the morning, or that I would do it in the morning'	'if I would feel good about mornings, then this word would apply to me'
17	*Tehdä aamuksi.*	*Laittaa joku toinen tekemään aamuksi.*
	'to make something a morning'	'to cause somebody else to make something a morning'
22	*Tehdä aamuksi. Eli, no tehdä aamuksi on se,*	*Kuulostaa joltain semmosen lapsen puheelta,*

mikä tulee nyt mieleen.   *aamutattaa. Mutta ei se musta oo järkevä.*

'to make something a morning'   'sounds like child language. Doesn't make sense'

Participant 14 interpreted the double causative compositionally, but the root stem did not have a causative interpretation. The result was that the double causative was interpreted as a single causative with the root being not causative. Participant 16 interpreted the second causative suffix as a habituative, and thus the double causative interpretation was blocked. Finally, participant 17 provided the double causative interpretation with an approximate paraphrase as 'to cause somebody else to make something a morning'.

Overall, these data show that the participants are able to analyze double causatives compositionally, and when they analyze both causative morphemes as causative, they give a corresponding double causative interpretation. In several cases, however, one of the causative morphemes was analyzed as habituative. In either case, it seems to me that the analyses were compositional.

5.4 Double causatives within complex words

In the above cases, the causative suffix was merged either to a root stem or to a single causative stem. The results of those analyses can be compared to words where the causative suffix was merged to a complex word. The data presented in Figure 1 tells us that there must be a difference. As can be seen from Figure 1, very complex words were seldom analyzed at all.

One relevant word was *valvoilettaa*, which was composed from *valvoa* (root) > *valvo-illa* (frequentive) > *valvo-ile-ttaa* (causative) and which means approximately 'to cause someone to stay awake frequently'. Out of 10 participants, four participants analyzed it as composed out of these three morphemes (verbal report number 258-259, 261, 265 in the corpus). Two participants analyzed the causative suffix as habituative (VRN 256, 260). Another causative suffix was then merged to the word (*valvoilettaa* > *valvoiletattaa*). Seven of the ten participants provided no analyses, two analyzed it as a simple habituative (VRN 266, 275), while only one protocol indicated a close to a compositional analysis:

(12)   [(outer causative) Antaa jonkin tehdä tätä [(frequentive) ei niin väkinäistä [(inner causative) valveilla pitämistä]]].

'To let somebody [outer causative] keep somebody awake [inner causative] in a not so forceful manner [frequentive].'

A similar item (*alistua* > *alistuilla* > *alistuilettaa*) was analyzed so that three participants provided a habituative analysis (VRN 405, 409-410) while one participant interpreted the last suffix as a causative but was unable to provide an analysis (VRN 404). When another causative suffix was merged to the word (*alistuiletattaa*), this participant provided a causative interpretation, but this time the rest of the analysis was missing (VRN 416). This reply is interesting in that it shows that the subject was sensitive to the semantics of the last causative suffix without being able to parse the rest of the word. Other participants provided a causative analysis on the top of the habituative analysis (VRN 419-420).

The next item, *ilma* > *ilmailla* > *ilmaileskella* > *ilmaileskelettaa*, was analyzed so that three participants analyzed it as a habituative (VRN 316, 320, 324), while one analyzed it as a causative:

(13)   [(outer causative) Laittaa joku toinen [(inner event) puuhailemaan ilman kanssa [(frequentive) ei niin vakavissaan]]].

'To make somebody else [outer causative] play with air [inner event] in a not so serious manner [frequentive].'

This analysis was close to being fully compositional. A further causative morpheme was then suffixed to the word (> *ilmaileskeletattaa*), after which only two participants provided any analyses at all. One recognized this as causative, but judged it as not a word (VRN 329), while one participant attempted a causative interpretation and then analyzed the inner causative morphemes as habituative:

(14)   Täst tulee taas mieleen, että [(causative) tehdä jollekin toiselle [(habituative) sellanen mieli, [(frequentive) et se vähän haluis [(inner event) lenneskellä]]]].

'This reminds me of a process where somebody does to somebody else [causative] a state of mind [habituative] that s/he would like to fly [inner event] frequently [frequentive].'

An adjective *tasailullinen* was converted into a causative *tasailullistaa*. A causative interpretation was given by six participants (VRN 473-476, 479-480).

The word *häivyttelettää* contained two causative morphemes, one suffixed to the root stem, another as a last suffix, while there was a frequentive suffix between them (*häipyä > häivyttää > häivytellä > häivyttelettää*). There was one analysis which was almost fully compositional:

(15)   [(causative) Teettää jollain toisella se, että [(eventive) joku häipyy pois], mutta että se [(frequentive) häipyminen tapahtuu ei niin vakavissaan].

'To make somebody else [causative] disappear [inner event], but so that the disappearing is implemented in a not so serious manner [frequentive].'

What is missing in this analysis is only the inner causative *häivy-ttää*. Another causative suffix was then merged to the word (> *häivytteletättää*). Two participants understood this as a causative (VRN 522-523), while one analyzed it as a habituative (VRN 518).

(16)   Tässä siihen sotkeutuu kolmas henkilö, eli [(causative) joku henkilö saa [(causative) jonkun henkilön saamaan [(causative) jonkun kolmannen henkilön [(inner event, frequentive) häivyttelemään sen jonkun asian tai ihmisen]]]].

'Here a third person is involved, so that some person [causative] makes another person [causative] get a third person [causative] to eliminate [inner event], in a nonserious manner [frequentive], that thing or a person.'

These observations reveal the cause of the linear trend reported in Figure 1, and hence the nature of the semantic 'limited generativity' in Finnish word formation. It turns out that when the words became more complex, fewer and fewer of the participants were able to analyze the meaning of the words irrespective of the fact that they had an explicit parsing available. However, those who did analyze the words were able to represent their morphological structure in the semantic interpretation faithfully. The phenomenon is thus dichotomous at the individual level, each word being either not interpreted at all or interpreted correctly, but continuous on the aggregate level in that on average, the proportion of the former type of responses increases linearly as a function of the morphological complexity: more and more participants give up the analysis. The mean frequency of semantic analysis for each subject, as averaged over all complexity levels, ranged from 0.83 to 0.36, suggesting that the participants used different overall approaches and/or possessed different linguistic skills concerning the task.

6. Limited generativity in word formation: discussion and conclusions

The results show that native speakers of Finnish are able to analyze both single and double causatives, as part of both simple words and complex words. However, when the words become complex, the ability to extract this interpretation becomes more difficult (Figure 1). In the present

experiment, very complex causatives were analyzed only on few occasions, but these analyses were surprisingly accurate. There were very few, if any, intermediate cases where the participant would analyze the meaning of the word without reflecting on its morphological structure compositionally. Thus, these data suggest that a computational or nearly compositional interpretation exists; however, when words become more complex, it becomes less likely that these interpretations are elicited. Instead of giving an analysis, the participants then judge the word to lack semantic interpretation.

In a previous experiment performed in our lab (Brattico et al. 2007), we found a similar linearly decreasing function between the morphological complexity and grammaticality/semanticality of the word. We also found that the reaction times in the grammaticality task increased as a function of morphological complexity and that when the stimulus words violated the strict word formation rules of Finnish, all stimulus items were judged as very ungrammatical and unsemantical irrespective of their morphological complexity. We reasoned that there must exist a 'complexity effect' which involves the use and interpretation of grammatically complex words and which stands in contrast with the violation of word formation rules. The present experiment replicated this result. However, contrary to the previous study, here the full morphological parsing of the words was explicitly given to the participants, and hence we may speculate that the morphological parsing difficulties as such cannot explain, at least completely, the reason why morphological complexity decreases linearly as a function of word complexity.

The protocols showed that the morphological causatives were analyzed as indirect causatives: they implied the presence of an implicit causee who performed the action. They also reveal that there is no common causative interpretation behind morphological causatives: the subjects used a wide variety of causative words (i.e., *made*, *let*, *have*, *force*, *ask*). If we want to maintain that the Finnish TTA encodes causation, this semantic aspect should be represented in an abstract way so as to subsume all these instances as special cases.

One explanation for the complexity effect is that long words are very infrequent and unnatural, and therefore difficult to process. While this statement is true as it stands, it cannot explain the difference between the processing of phrases and the processing of words. Certainly, most phrases that we use in everyday communication are such that their frequency of use is very low, if not virtually zero, and the fact that they are felt to be more 'natural' simply records a fact to be explained. Indeed, one way to approach the complexity effect is to ask why words with several morphemes are felt to be more 'unnatural' than phrases with several words. Further, in some polysynthetic languages word formation approaches such levels of complexity that this hypothesis collapses cross-linguistically.

Another possibility is that longer words are harder to interpret because they violate some kind of 'soft constraints' in Finnish word formation, for instance, constraints which regulate the number of syllables in a stem that takes a particular affix. While such constraints are an established fact, this hypothesis claims that these constraints are the cause rather than the consequence of the complexity effect. It is possible that, in contrast to binary 'hard violations' of strict word formation rules, violations of this type of soft constraints produce graded responses. The reason why the participants did not analyze the complex words semantically could thus be because they regarded the stimulus items as going against their intuition about what counts as a possible word in present-day Finnish. Consequently, those participants who did analyze the words were more tolerant of the soft violations. This interpretation of the data is supported by the fact that several participants reported that some of the complex items went against their judgment of what belongs to Finnish despite the fact that they were able to interpret their compositional structures:

> (17)   No en mä tätä kyl sanois suomen kielen sanaks, mut kyl mä silti tajuisin, jos joku sanois mulle, että sitä laulutattaa
>
>   'I would not call this a Finnish word, but I would still understand what it means if somebody used this word.'

On the other hand, this response was not restricted to complex words since it was also elicited in the case of simpler ones. For example, a word such as *villattaa* 'to cause to be/have wool' elicited a similar response:

(18)   Mä varmaan veikkaisin, että tää ois joku semmonen termi, että tällä niinku tehdään sitä villaa. Mutta tota, jos joku käyttäis tätä termiä, niin mä oikeesti ajattelisin että tää on jotain outoo murretta, niin ehkä mä en sit tajuis tätä.

        'I guess that this could be a term which refers to something which is used to make wool. On the other hand, if somebody used this term I would think that it represents some strange dialect that I do not understand.'

If this is the right explanation for the phenomenon, the explanation of the complexity effect would thus consist of an explanation of why such graded, soft constraints which are related to the complexity of words are part of the knowledge of morphology in Finnish. The same phenomenon of graded responses is sometimes confronted in the case of syntax as well. One potential example of such graded grammaticality is provided in (19a-d), which illustrates the results of question formation with various types of complement sentences.

(19)
a.   Kenet Merja ajatteli että Pekka löysi?
       'Who did Merja think that Pekka found?'
b.   ?Kenet Merja pohti että Pekka löysi?
       'Who did Merja wonder that Pekka found?'
c.   ?*Kenet Merja pohti miksi Pekka löysi?
       'Who did Merja wonder why Pekka found?'
d.   *Kuka Merja pohti miksi löysi Pekan?
       'Who did Merja wonder why found Pekka?'

In the generative literature, this phenomenon has sometimes been explained by assuming that the more severe violations violate more grammatical rules than the softer violations (e.g., Lasnik & Saito 1993). Perhaps in the case of word formation, the decrease in interpretability is due to the linear accumulation of soft constraints.

Another interpretation of the complexity effect in the case of word formation is based on the different semantic functions performed by morphemes within a word and words within a typical linguistic phrase. Whereas complex words refer to complex predicates, many phrases, no matter how complex, refer to simple entities. Thus, when a new morpheme is added to a stem, it produces a complex predicate. The causative morpheme, for instance, produces a complex predicate 'to cause to P' from the existing predicate P. But the causative morpheme does not produce an intelligible meaning in isolation; rather, it must always be interpreted as relative to the meaning of P, and such dependency relations become complex when the word contains several morphemes since the predicate itself is complex (e.g., 'to cause somebody to understand the abstract property of owning a house'). This is not the case with most phrases with full words. For instance, in the noun phrase *se pieni talo* 'that small house', each element serves to determine one aspect of the intended but simple referent, the one small house in the concrete world. The whole meaning that emerges from the phrase is not complex, but it may be indeed a very simple entity - the house (or a mental representation of the house). On the other hand, when a linguistic phrase is formed so that its semantic constitution mirrors that of complex words, we encounter difficulties in semantic integration. Consider, for example, the following phrase whose meaning mirrors that of a complex word (*hampaallistaminen*) in Finnish: 'the abstract property of causing somebody to own a collection of teeth'. Semantic integration of these elements is perhaps more difficult, and the reason could be that the complex phrase refers to a complex property instead of a simple one.

The last explanation I would like to mention in this connection is the hypothesis that word formation is based neither on syntax nor autonomous morphological processing, but on analogical reasoning or analogy more generally. If made precise, this hypothesis could also explain why complex words are harder to interpret, namely, if the analogical processes are somehow limited in their computational power.

One could ask whether the present data has any relevance to the description of present-day Finnish as an autonomous, mind-independent entity (or a set of expressions). Most of the stimulus words do not belong to this set; hence the most we can say is that they represent something that could potentially be introduced to the language but which is currently not part of it. The problem is that as shown by the present data, the more complex the word is, the less likely it is that the speakers are willing to accept and interpret it. We do not know on the basis of the present experiment what would happen if the subjects were to encounter any of these words in the context of real linguistic communication: that has to be tested in a separate experiment. What we do know, however, is that they treat too-complex words differently from strictly ungrammatical words (Brattico, Leinonen & Krause 2007), and this distinction could also be relevant to the description of present-day Finnish. If so, then instead of modeling the set of Finnish words as a dichotomous category, we could model it as a more structured set where expressions are associated with their acceptability rating and where many word formation rules modify these ratings instead of a binary class inclusion.

Acknowledgements

Notes

[1] One example from http://terrible.kupoli.net/index.php?topic=237.15;wap2 (23.10.2007) - link no longer available, see http://www.kupoli.net/ .

[2] This approach should not be confused with the claim of 'psychological reality' in synchronic linguistics. When a linguist invokes the notion of psychological reality in connection with some assertion about the structure of a language, what is meant is that the analysis should somehow reflect the knowledge of language of the native speakers. While the present approach assumes one version of this dictum as well, this criterion is also compatible with the description of a language as a set of expressions. In the simplest imaginable case, one could use native informants simply in deciding whether a given expression 'belongs to the language' with the intention of collecting a set of expressions, rather than that of understanding anything about the cognitive mechanisms underlying communication. Moreover, the notion of psychological reality has multiple meanings as well. On the one hand, it could mean explicit, verbalized knowledge of the speakers of a language to which they have some sort of conscious access. In the study of cognitive mechanisms, this assumption is rarely if ever made, for the obvious reason that a rather large amount of the knowledge of language (as well as knowledge of other cognitive domains) is totally unconscious for the informants. We do not have direct access to the neurocognitive mechanisms underlying linguistic communication, visual perception, feature extraction in auditory domain, and so on.

[3] A priori, many transitive verbs could involve other predicates besides causation, such as needing or wanting. According to an imaginable grammar which uses needing instead of causing, the transitive verb 'kill' would imply 'to need to die' instead of 'to cause to die'. This, however, is not how transitive verbs are crafted in known natural languages, so I will take it for granted that causation must have some special status in the way in which languages conceptualize and lexicalize events.

[4] Currently, there is ample evidence that the different types of causatives cannot be reduced to each other syntactically or semantically (Bouchard 1995:104-108, Comrie 1985, Fodor 1970, Gergely &

Bever 1986, Shibatani 2002, Salo 2003, Song 1995). In the domain of semantics, this is taken to mean that the expressions of the type T1 (say lexical causatives) are not synonymous with expressions of the type T2 (say analytical causatives). In the case of syntax, the non-reductionism means that the syntactic properties of the different types of causative constructions differ from each other. Many authors thus think that various causative constructions can be given a reduced representation in a suitable *conceptual or logical metalanguage*, while others assume that semantic and syntactic non-reductionism is more pervasive and applies to any linguistic level, whether at the level of the object language or at the metalanguage. Which way the chips will fall is immaterial to the present research question, however.

[5] My own view is that the problem of classification should be posed neither on the basis of form nor meaning; rather, the pairing between form and meaning for the relevant range of expressions should be deduced from the best linguistic theory available.

[6] Exactly the same question arises in other well-known cases. Thus, many transitive words involve causation; is that fact relevant to linguistic processing, or does it belong to cognitive psychology? Many words, especially verbs, describe events; do the combination of eventive semantics and a word form belong to linguistics or cognitive psychology?

[7] Since these techniques are somewhat uncommon in linguistics, it should be mentioned here that their purpose is only to check that the differences in the frequency of semantic interpretation between various complexity classes (1-5) are not produced by random fluctuations in the participants' responses, but rather are some systematic effect of the experimental manipulation. In the latter case, we say that the result is 'statistically significant'. In a participant-based analysis, the source of variation is constituted by the differences between subjects; in the item-based analysis, the source of variation is constituted by the stimulus items.

References

Achard, Michel. 2002. "Causation, constructions, and language ecology: An example from French". *The Grammar of Causation and Interpersonal Manipulation*, ed. by Masayoshi Shibatani, 127-156. Amsterdam: John Benjamins.

Aissen, Judith Lillian. 1974. *The Syntax of Causative Constructions*. New York: Garland Publishing.

Anderson, Stephen Robert. 1982. *A-Morphous Morphology*. Cambridge: Cambridge University Press.

Baker, Mark C. 1988. *Incorporation: A Theory of Grammatical Function Changing*. Chicago: University of Chicago Press.

Bouchard, D. 1995. *The Semantics of Syntax*. Chicago: University of Chicago Press.

Brattico, Pauli. 2005. "A category-free model of Finnish derivational morphology". *SKY Journal of Linguistics* 18: 7-45. http://www.ling.helsinki.fi/sky/julkaisut/sky2005.shtml

Brattico, Pauli, Alina Leinonen & Christina Krause. 2007. "On the limits of productive word formation: Experimental data from Finnish". *SKY Journal of Linguistics* 20: 109-139. http://www.ling.helsinki.fi/sky/julkaisut/sky2007.shtml

Chomsky, Noam. 1970. "Remarks on nominalization". *Readings in English Transformational Grammar*, ed. by Roderick A. Jacobs & Peter S. Rosenbaum, 184-221. Waltham, Mass.: Blaisdell.

Comrie, Bernard. 1976. "The syntax of causative constructions: Cross-language similarities and divergences". *Syntax and Semantics 6: The Grammar of Causative Constructions*, ed. by Masayoshi Shibatani, 261-312. New York: Academic Press.

Comrie, Bernard. 1985. "Causative verb formation and other verb deriving morphology". *Language Typology and Syntactic Description, Vol III: Grammatical Categories and the Lexicon*, ed. by Timothy Shopen, 91-151. Cambridge: Cambridge University Press.

Dixon, R.M.W. 2000. "A typology of causatives: Form, syntax and meaning". *Changing Valency: Case Studies in Transitivity*, ed. by R.M.W. Dixon & Alexandra Y. Aikhenvald, 30-83. Cambridge: Cambridge University Press.

Dowty, David R. 1979. *Word Meaning and Montague Grammar*. Dordrecht: Kluwer.

Fodor, J.A. 1970. "Three reasons for not deriving 'kill' from 'cause to die'". *Linguistic Inquiry* 1: 429-438.

Gergely, G. & T.G. Bever. 1986. "Relatedness intuitions and the mental representation of causative verbs in adults and children". *Cognition* 23: 211-277.

Hakulinen, Auli, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen & Irja Alho. 2004. *Iso suomen kielioppi*. (A Grammar Book of Finnish.) Helsinki: Suomalaisen Kirjallisuuden Seura.

Hankamer, Jorge. 1989. "Morphological parsing and the lexicon". *Lexical Representation and Process*, ed. by William Marslen-Wilson, 392-408. Cambridge, Mass.: MIT Press.

Julien, Marit. 2002. *Syntactic Heads and Word Formation*. Oxford: Oxford University Press.

Karlsson, Fred. 1983. *Suomen kielen äänne- ja muotorakenne*. (The Sounds and Forms of Finnish.) Juva: WSOY.

Kayne, Richard S. 1975. *French Syntax: The Transformation Cycle*. Cambridge, Mass.: MIT Press.

Kemmer, Suzanne & Arie Verhagen. 1994. "The grammar of causatives and the conceptual structure of events". *Cognitive Linguistics* 5: 115-156. https://openaccess.leidenuniv.nl/handle/1887/2393

Kytömäki, Leena. 1977. *Suomen verbijohdosten generointia*. (Generating Finnish Derived Verbs.) Licentiate thesis, Department of Finnish and General Linguistics, University of Turku.

Kytömäki, Leena. 1992. *Suomen verbiderivaation kuvaaminen 1600-luvulta nykypäiviin*. (A Description of Finnish Verbal Derivation from the Seventeenth Century to the Present.) PhD thesis. Turku: University of Turku.

Lasnik, Howard & Mamoru Saito. 1993. *Move α: Conditions on Its Application and Output*. Cambridge, Mass.: MIT Press.

Lehtonen, M., V.A. Vorobyev, K. Hugdahl, T. Tuokkola & M. Laine. 2006. "Neural correlates of morphological decomposition in a morphologically rich language: An fMRI study". *Brain and Language* 98(2): 182-193. doi:10.1016/j.bandl.2006.04.011

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 own lexicon". *Proceedings of the 21st Annual Penn Linguistic Colloquium*, ed. by Alexis Dimitriadis, Laura Siegel, Clarissa Surek-Clark & Alexander Williams, 201-225. (= *University of Pennsylvania Working Papers in Linguistics*, 4.2.) Philadelphia: Penn Linguistics Club. http://ling.upenn.edu/papers/v4.2-contents.html

Niemi, Jussi, Matti Laine & Juhani Tuominen. 1994. "Cognitive morphology in Finnish: Foundations of a new model". *Language and Cognitive Processes* 9: 423-446. doi:10.1080/01690969408402126

Sagart, Laurent. 2001. "Vestiges of archaic Chinese derivational affixes in modern Chinese dialects". *Sinitic Grammar. Synchronic and Diachronic Perspectives*, ed. by Hillary Chappell, 123-142. Oxford: Oxford University Press.

Salo, P. 2003. Causatives and the Empty Lexicon: A Minimalist Perspective. Ph.D. dissertation, University of Helsinki. http://urn.fi/URN:ISBN:952-10-1477-6

Selkirk, Elisabeth. 1982. *The Syntax of Words*. Cambridge, Mass.: MIT Press.

Shibatani, Masayoshi, ed. 1976. *The Grammar of Causative Constructions*. (= *Syntax and Semantics*, 6.) New York: Academic Press.

Shibatani, Masayoshi. 2002. "Some basic issues in the grammar of causation". *The Grammar of Causation and Interpersonal Manipulation*, ed. by Masayoshi Shibatani, 1-22. Amsterdam: John Benjamins.

Song, Jae Jung. 1995. *Causatives and Causation*. London: Longman.

Ullman, Michael T. 2001. "A neurocognitive perspective on language: The declarative/procedural model". *Nature Reviews Neuroscience* 2: 717-726. doi:10.1038/35094573

Zubizarreta, Maria L. 1985. "The relation between morphophonology and morphosyntax: The case of Romance causatives". *Linguistic Inquiry* 16: 247-289.