Reading morphologically complex clause structures in Finnish

Jukka Hyönnä and Seppo Vainio
University of Turku, Finland

The study examined how morphologically complex clause constructions were processed during reading Finnish. Readers’ eye fixation patterns were recorded when they read two alternative versions of the same linguistic construction, a morphologically complex verb construction and its less complex subclause counterpart. The complexity of the verb construction is apparent in the construction being marked by less perceivable bound morphemes, which make the clause subject and predicate morphologically more complex and more dense in information. Experiment 1 showed that more complex verb constructions produced longer gaze durations than the length- and frequency-matched subclause constructions. Experiment 2 showed that the complexity effect is reversed when the more complex clause form was clearly more common in the language than its less complex counterpart. It is concluded that both structural complexity and structural frequency influence the ease with which linguistic expressions are processed during reading.

In his state-of-the-art review of human syntactic parsing, Mitchell (1994) concludes that the recent psycholinguistic research on parsing has been heavily preoccupied with (1) parsing of syntactically ambiguous materials (such as reduced relative clauses in English), (2) nonlinguistic issues, such as serial versus parallel nature and time constraints of on-line parsing, and (3) parsing of a single language, English. Thus, as suggestions for future directions, Mitchell (1994) calls for studies that examine parsing of unambiguous linguistic structures in languages most distantly related to English.

Requests for reprints should be addressed to J. Hyönnä, Department of Psychology, University of Turku, FIN-20014 Turku, Finland. Email: hyona@uttu.fi

We thank Matti Laine for giving us access to the newspaper corpus of Finnish words. We are also grateful to Päivi Laine for conducting us a corpus search in the LA syntax archives, Anneli Pajunen for her linguistic expertise, and to Don Mitchell, Wayne Murray, and an anonymous reviewer for their constructive criticism on an earlier version of this paper.

© 2001 Psychology Press Ltd
http://www.tandf.co.uk/journals/pp/09541446.html DOI:10.1080/0954144042000098
The study that is reported here is completely in line with Mitchell’s suggestions. In the study, we examined how morphologically complex, but syntactically unambiguous, clause constructions are processed during reading. The study was carried out in Finnish—a synthetic, non-Indo-European (Finno-Ugric) language that makes heavy use of morphological marking in expressing grammatical information. Although Finnish is psycholinguistically a highly interesting language for several reasons, it has received only limited attention so far (see Hyönä & Hujanen, 1997; Hyönä & Pollatsek, 1998; Laine, Vainio, & Hyönä, 1999). Consider, for example, its inflectional system that yields over 2000 possible forms for each noun and over 10,000 forms for each verb (Karlsson, 1983). Another feature that comes with the rich morphology is its highly flexible word order (see Hyönä & Hujanen, 1997).

In a morphologically rich language like Finnish, one of the key issues is to determine how morphological complexity influences syntactic parsing. In Finnish, information that is in many other languages conveyed as function words (such as prepositions and conjunctions) typically appears as bound morphemes attached to content words. For example, an English expression in the sauna reads in Finnish saunassa, where the inflection ssa is equivalent to the preposition in (note that Finnish lacks articles), in our sauna reads saunassamme, where the possessive suffix mme is equivalent to our, and also in our sauna reads saunassammekin, where the clitic kin is equivalent to also. Thus, the agglutinative feature of the Finnish language results in many lexical items being morphologically complex and high in information density.

Very little is known how morphological complexity influences on-line parsing. The following facts are known about on-line morphological processing of Finnish. Laine et al. (1999) showed that inflected nouns presented in isolation take longer to identify than monomorphemic nouns. Hyönä and Pollatsek (1998) reported evidence from a reading task consistent with the notion that compound words in Finnish are recognised via their constituent morphemes. Finally, Hyönä and Hujanen (1997) observed that case inflections are utilised in making immediate parsing decisions during sentence comprehension. All three studies strongly suggest that morphological information significantly influences both word identification and parsing in Finnish. However, no studies exist that have directly examined how morphological complexity affects sentence parsing. The present study was designed to fill in this gap.

The type of constructions we are interested in the present study are called conjunctional converbs (Nedjalkov, 1995; for a cross-linguistic study of converbs, see Haspelmath, 1995). In the linguistic literature, terms such as gerund, adverbial participle, and clause equivalent have previously been used for converb construction. What is interesting about the
conjunctional verb constructions is that they can be used interchangeably with conjunctional subclause forms. For example, a subclause *kun isä hinkkasi* (= when the father polished) can be expressed using a temporal verb structure, *isän hinkatessa*. In the verb construction, the *kun* (when/while) conjunction is left out and the subclause construction is denoted by a morphologically complex nominal verb form, in the earlier example, by the so-called second infinitive accompanied with a case inflection. Thus, the verb form, *hinkatessa*, consists of three morphemes, the stem *hinka*, the second infinitive marker *e*, and a case inflection *ssa* (so called inessive). The respective verb form in the full subclause, *hinkkasivat*, has two morphemes, the stem and a personal marker. The nominal verb construction is accompanied by the subject of the verb construction appearing in an inflected form (i.e., in the genitive case: *isän*) rather than in the base-form nominative case (*isä*), as in the full subclause. Syntactically, the main function of verb constructions is to mark adverbial subordination. In other words, they appear as non-obligatory adverbial modifiers (Haspelmath, 1995).

Historically, verb constructions appeared earlier in Finnish than full subclauses (Itkonen, 1966). It is claimed that full subclause forms appeared in Finnish under the influence of Germanic languages. Even today in some Finno-Ugric languages subclause structures are non-existent. Across languages, there appears to be an inverse relationship between the frequency of use of verbs versus conjunctions: “in languages that make an extensive use of conjunctions verbs play a minor role or are completely absent” (Nedjalkov, 1995, p. 100), and vice versa.

In modern Finnish, full subclause constructions dominate over verb constructions in colloquial speech, and verbs are not used very often. However, in written Finnish, verbs are still common (see description of materials in the Method section). Ikola (1986) distinguishes five types of verbs in Finnish (he calls them *lauseenvastike* = clause equivalent), of which we chose two for this study, the temporal verb (Experiments 1 and 2) and the final verb (Experiment 2). Temporal verb constructions can be used to replace a temporal subclause in a sentence like “When *X* happened/did something, *Y* happened/did something else”, while final verbs typically convey intentionality, “In order for *X* to happen, *X* needs something/something has to happen to *X*” (the final verb construction is used to replace the final subclause).

We chose these two constructions, because for them it was possible to create two alternative versions, a verb and a full subclause version, that are practically identical in meaning, but differ in morpho-syntactic complexity. There are two aspects in the relative complexity of the verb constructions. First, the verb construction is marked by short
bound morphemes, instead of a free morpheme in the full subclause version, which makes the convert verb construction less transparent and thus potentially less perceivable. Second, the subject and predicate contain more morphemes than the respective constituents in the full subclause version; also the total number of morphemes in the convert verb phrase is typically greater.

**EXPERIMENT 1**

In Experiment 1, subjects were exposed to both full subclause and convert verb forms of temporal constructions. On-line sentence processing was studied by recording readers’ eye fixation patterns when they read these sentences for comprehension. For the on-line study of language processing, the eye-tracking technique has now become the most preferred method for several reasons (for more details, see Rayner, 1998). The prediction we entertained in Experiment 1 is that convert verb structures as morphologically more complex and dense constructions will be more difficult to process, which will show up in longer fixation times on the convert verb than on the full subclause versions of the target clauses. This prediction is based on the view that morphological complexity brings about processing costs in the reader being required to decompose complex linguistic forms into their constituent morphemes to derive the structure and meaning of the expression. As the convert verb version contains morphologically more complex lexical items and its structure is less transparent and less perceivable, its processing is assumed to be more demanding.

Recording readers’ eye movements yields several processing measures (see, e.g., Murray, 2000), of which we chose gaze duration as our primary measure. For gaze duration, individual fixations landing on the target zone before exiting it are summed up, and it indexes the initial encounter of the target clause. Moreover, to get a more detailed understanding of the timing of the predicted complexity effect, gaze durations for pre- and post-target regions were also analysed.

**Method**

*Subjects.* Twenty-eight students from the introductory psychology course at the University of Turku participated in the experiment as a part of the course requirement. All subjects were native speakers of Finnish.

*Apparatus.* Eye movements were collected by the EYELINK eyetracker manufactured by SR Research Ltd (Canada). The eyetracker is an
infra-red video-based tracking system combined with hyperacuity image processing. There are two cameras mounted on a headband (one for each eye) including two infra-red LEDs for illuminating each eye. The headband weighs 450 g in total. The cameras sample pupil location and pupil size at the rate of 250 Hz. Recording is monocular and is performed for the selected eye by placing the camera and the two infra-red light sources 4–6 cm away from the eye. The resolution of eye position is 15 seconds of arc and the spatial accuracy is better than 0.5 degrees. Head position with respect to the computer screen is tracked with the help of a head-tracking camera mounted on the centre of the headband at the level of the forehead. Four LEDs are attached to the corners of the computer screen, which are viewed by the head-tracking camera, once the subject sits directly facing the screen. Possible head motion is detected as movements of the four LEDs and is compensated for on-line from the eye position records. The system allows free head motion within a 100 cm cube. The compensation is better than 1 degree over the acceptable range of head motion.

**Materials.** For the experiment two alternative versions, a full subclause and a deverbal construction, were created using the Finnish temporal clause construction. The temporal verb is commonly used in written Finnish; the newspaper corpus (22.7 million word tokens) by Laine and Virtanen (1999) lists 40,263 occurrences of this verb construction; the number of different verb stems taking the temporal form is 2288. A temporal clause construction can be expressed either as a full subclause including the conjunction *kun* (when) or with a deverbal construction, where the conjunction is eliminated and the subclause construction is denoted by a morphologically complex verb form. Next we provide an example sentence for both the full subclause and the deverbal version of the type of constructions employed in the experiment (see Appendix, for the complete set of stimulus materials).

**Temporal construction** (target phrase is shown in bold with morphological boundaries marked by slashes)

- **Full subclause**: Äiti ajoi nurmikkoa, *kun* isä hinkka*si* auton konepeltiä.
- **Deverbal construction**: Äiti ajoi nurmikkoa *isä/n hinkat/esse* auton konepeltiä.

“The mother mowed the lawn, while the father polished the car hood.”

The deverbal consists of three morphemes, the stem (*hinka*), a verb nominal form (*e*), and an inflectional ending (*ssa*). The nominal form is called the second infinitive, and the inflectional case accompanied with the nominal form is called the inessive. When using the full subclause
form with the conjunction, the verb has only two morphemes, the stem and a personal marker. The clause subject takes an inflected form (the genitive case) in the verb construction and the nominative case (i.e., the base-form without a case ending) in the full subclause construction. With few exceptions, the verb construction had one morpheme more than the subclause constructions. Syntactically, the temporal clause is an adverbial modifier of the main clause (i.e., a non-obligatory argument).

For each sentence pair, the target phrases were closely matched for length in number of characters (excluding the spaces between words). The mean lengths of the two versions of the target expressions were 20.4 and 20.7. By definition, as the verb version consisted of two words and the subclause version of three words, the individual words in the verb version were longer (particularly the verb).

The surface frequency of the noun and the verb were matched on the basis of the Laine and Virtanen (1999) corpus. The average surface frequency (per million) of the nominative form of the noun was 37.3 and that of the genitive form 25.7, and the average surface frequency of the indicative verb form was 0.3 and that of the verb 0.4.

All nouns selected for the target expressions were of high frequency; the average lemma frequency (the frequency of word stem including all inflectional variants) of the nouns was 168 per million. The average lemma frequencies of the verbs were lower, six per million. Note that lemma frequencies were perfectly matched, because the same word lemmas were used in the two alternative versions. We also computed average bigram frequencies for the target nouns and verbs on the basis of the Laine and Virtanen (1999) corpus. The average bigram frequency of the nominative noun was 7418 per million and that of the genitive noun 8206 per million; the average bigram frequency of the verb form was 7023 per million for the full subclause version and 7523 per million for the verb version of the temporal construction. What is noteworthy of the bigram frequencies is that, if anything, they are higher for the more complex constructions.

The sentences were presented in Courier font, one character space being 0.55 cm wide; with a viewing distance of about 60 cm one character extended approximately 0.5 degrees of visual angle horizontally.

Thirty pairs of sentences were constructed. Two sentence sets were prepared by placing one member of the sentence pair in set A and the other member in set B. In both sets, 15 target expressions appeared in the full subclause form and 15 in the verb form. Each subject received either the materials of set A or B, and the two sets were counterbalanced across subjects. Thus, each subject read only one member of each sentence pair. The target sentences were mixed with 94 filler sentences. Two target sentence blocks were created with equal number of different
types of target sentences in each block. The order of blocks was counterbalanced across subjects, and the order of sentences within a block was randomised separately for each subject.

Procedure. Prior to the experiment, the eyetracker was calibrated using a nine-point calibration grid that extended over the entire computer screen. Each target sentence was presented one at a time roughly on the centre of the monitor. Prior to each sentence, the calibration was checked by presenting a fixation point on the top left corner of the screen; if needed, calibration was automatically corrected.

Subjects were instructed to read the sentences for comprehension at their own pace. They were further told that periodically they would be asked to paraphrase the last sentence they had just read to make sure that they attended to what they read. However, it was emphasised that the task was to comprehend, not to memorise the sentences. Subjects were asked to paraphrase a sentence approximately after every 10 sentences. The experimental session lasted a maximum of 30 minutes.

Results

Durations of individual fixations falling on the target phrase before exiting it were summed up. Following Murray’s (2000) suggestions, we refer to this measure as the gaze duration. A set of t tests were computed on the data using both subjects (t1) and items (t2) as the random variable. All analyses were done using paired samples tests (two-tailed). Means and standard deviations are given in Table 1.

Target phrase. Gaze duration was 45 ms longer for the converb than the full subclause forms, t1(27) = 2.41, p = .023, t2(29) = 2.01, p = .054. As predicted, these results show that morphological complexity affects the relative ease of processing these temporal clause constructions.

1Another option would have been to use a measure coined “regression path” (Konieczny, 1996) or “total pass reading time” (Murray, 2000), which takes into account all fixations after the eyes enter a text region until they exit it to the right. Thus, this measure also includes possible regressions to a previous text region. In the present set of data, this measure would have yielded a highly similar pattern of results, because there were hardly any regressions from the target region to the previous region.

2A separate analysis of the “sub-gazes” for the clause subject (i.e., the noun) and predicate (i.e., the verb) showed that both were associated with significantly longer subgazes in the converb construction (all ps < .001). However, these effects are not readily interpretable as such, because the subject and predicate forms differed in length between the conditions (i.e., the length was matched for the target phrase as a whole).
The effect did not show up in the number of fixations made before exiting the target clause \((ps > .25)\); if anything, verb constructions were read with slightly less fixations \((3.39)\) than subclause forms \((3.48)\). Thus, the effect is due to longer but not more fixations being needed to read the verb construction.

**Pre- and post-target regions.** To examine whether the effect is restricted only to the target segment, we analysed the gaze durations for two text segments preceding the target expression and the one following it. In the clause preceding the target phrase, we identified two regions, the final word of the preceding clause and rest of the preceding clause. The word prior to the target phrase differed in that in the subclause version the word was accompanied by a comma that was absent in the verb version. Other non-target regions were identical for the two conditions. None of the comparisons approached significance (see Table 1, all \(ps > .40\)). The fact that the two conditions did not differ in the post-target region is evidence for the lack of any spill-over effects in processing. Second, the finding that gaze durations were highly similar for the preceding word suggests that the presence of a comma did not result in processing differences. Finally, the small but clearly non-significant difference between the two conditions in gaze duration for the sentence-initial pre-target region (see Table 1) must be due to chance, as the conditions diverged only later in the sentence.

**Discussion**

The results of Experiment 1 are straightforward. The morphologically more complex and less transparent verb construction demanded more processing time than the less complex and more transparent subclause construction. This result suggests that for the verb construction the

**TABLE 1**

<table>
<thead>
<tr>
<th>Target phrase</th>
<th>Preceding clause (excl. final word)</th>
<th>Final word of the preceding clause</th>
<th>Following phrase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full subclause</td>
<td>665 (174)</td>
<td>733 (190)</td>
<td>340 (128)</td>
</tr>
<tr>
<td>Converb</td>
<td>710 (209)</td>
<td>745 (163)</td>
<td>336 (94)</td>
</tr>
<tr>
<td>Difference</td>
<td>45</td>
<td>12</td>
<td>-4</td>
</tr>
</tbody>
</table>
time saved by not having to process a conjunction was clearly less than the extra time needed to process the more complex subject and predicate forms. Thus, when the temporal clause construction was marked with a combination of bound morphemes incorporated as parts of a lexical item, the clause was more difficult to process than when the same information was conveyed as a separate word (i.e., a temporal conjunction). The fact that the converb construction typically had one additional morpheme may also have contributed to the effect.

It is possible that our results may partly be explainable as a “garden-path” effect (Frazier & Clifton, 1996). Converb structures are locally ambiguous at the subject: The genitive form (isän = father’s) may be taken either as a part of a noun phrase (isän auto = father’s car, isän kanssa = with the father), clause object (object often takes the genitive case), or the subject of a converb construction (isän hinkatessa).3 On the other hand, for full subclauses, there was no possibility that readers would have been “garden-pathed” to make a false parsing decision, as the subclause construction is already marked by the conjunction preceding the subject.

Although local ambiguity was in principle present in the converb constructions, in the sentences we employed the possibility for alternative interpretations was not particularly prevalent. First, interpreting the locally ambiguous noun phrase of the converb construction as the object of the preceding clause would have been possible only in 4 out of the 30 target sentences. Second, it is in principle possible to begin a clause with an object (see Hyönä & Hujanen, 1997) so that the ambiguous noun phrase could be interpreted as the clause-initial object of the second clause; however, this was not possible in the type of construction we used. On the other hand, it is a viable option when it appears sentence-initially (see Experiment 2). Third, the interpretation of the subject as part of an adverbial phrase (noun + postposition) was an option in the target sentences. We return to the issue of local ambiguity after presenting the results of Experiment 2.

Experiment 1 suggests that parsing clause structures in Finnish is influenced by the relative complexity and transparency of linguistic cues denoting the syntactic construction: When the cue was a combination of bound morphemes, the processing was made more difficult than when the structure was cued by a free morpheme. As noted previously, however, a second interpretation based on the notion of local ambiguity is also

3Note that syntactically the converb phrase is an adverbial modifier of the main clause. Another type of adverbial phrase is also a viable option: a noun in the genitive case followed by a postposition (isän kanssa = with the father).
plausible. Moreover, as only one single convverb structure was employed, the generalisability of the effect may also be questioned. Hence, Experiment 2 was conducted, where the final construction was also employed. In Experiment 2, the temporal construction was presented sentence-initially; if local ambiguity is the major source of the observed effect, the effect size should be significantly larger for the temporal construction in Experiment 2 than in Experiment 1.

**EXPERIMENT 2**

In Experiment 2, we examined the processing of two types of convverb constructions, temporal and final constructions. The final convverb is used to denote intention or goal. The final convverb construction can be used to replace the conjunction *jotta* (~ in order to) in a clause like: *jotta uhri toipuisi* (= in order for the victim to recover). Using the convverb form the same clause reads *toipuakseen uhrin*, where the verb form contains four morphemes, the verb stem *toipu*, the first infinitive marker *a*, a trans-lative case ending *kse*, and *en* as a possessive suffix to mark the third person singular (a possessive suffix is also obligatory in the construction). The final construction has the characteristic that the convverb and the main clause share the same subject (see Nedjalkov, 1995), as in *Toipuakseen uhrin on saatava pikaisesti lisää verta* (= in order for the victim to recover, he quickly needs to get more blood), where *uhrin* (= victim) is both the subject of the convverb as well as that of the main clause. In the corresponding subclause construction, a personal pronoun *hän* is added in the main clause (not unlike in the English equivalent), as in *Jotta uhri toipuisi, hän on saatava pikaisesti lisää verta*.

The two convverb structures differ in the frequency of usage, the temporal construction being more frequent than the final construction. The newspaper corpus by Laine and Virtanen (1999) lists almost 1800 occurrences of the temporal verb construction per million words, while the final verb construction occurs less commonly, about 300 occurrences per million words. More importantly, not only does the frequency of occurrence vary between the different constructions, but so also does the relative frequency of the alternative forms, the more transparent subclause and the more complex convverb construction. A corpus search of written Finnish (the Lauseopin arkisto syntax archives) revealed that for the temporal construction the full subclause and the convverb constructions are about equally common (with a slight bias favouring the full subclause form), whereas for the final construction, the convverb construction was clearly more common than the full subclause construction (a ratio of 9 to 1). These frequency estimates are based on all instances of these
constructions found in the corpus irrespective of their sentence position (i.e., whether or not they were preposed).

In Experiment 2, we were interested in seeing whether the results obtained in Experiment 1 would generalise to another clause structure that has the interesting feature that the more complex clause construction is more common than the less complex construction. If a complexity effect could be established also for this new clause construction, it would clearly support the generalisability of the effect. However, it is also possible that the relative frequency with which the two alternative forms appear in written text may play an important role. The idea that structural frequencies are potentially relevant is somewhat analogous to the word frequency effect (i.e., low-frequency words are harder to access in the mental lexicon than high-frequency words): Frequent exposure to a given linguistic structure would result in the corresponding mental structure becoming more easily accessible and its on-line parsing thus being easier to compute.

In Experiment 2, the target phrases appeared sentence-initially, unlike in Experiment 1, where they appeared in the second clause. In the final construction, the sentence is initiated with the converb followed by the subject, whereas in the temporal form the converb is preceded by the subject. All final converb constructions were unambiguous, whereas temporal converb constructions were locally ambiguous at the subject. As we argued earlier, the subject of the temporal converb is clearly more ambiguous when it occupies the sentence-initial (Experiment 2) rather than a sentence-medial position (Experiment 1). Without any prior constraining context, all the three alternative interpretations we mentioned above (i.e., part of a noun phrase, clause object, or subject of a converb) are possible. Thus, if local ambiguity significantly contributes to the processing difficulty effect, the effect for the temporal constructions should be greater in Experiment 2 than in Experiment 1.

Method

Subjects. Twenty-seven university students served as subjects as a part of a course requirement. All subjects were native Finnish speakers.

Apparatus. The same eyetracker was used as in Experiment 1.

Materials. In addition to the temporal construction, the final construction was also employed. The final construction is used to denote intention or goal. The newspaper corpus (22.7 million word tokens) by Laine and Virtanen (1999) lists 7030 occurrences of the final converb; the number of different verb stems taking the final form is 810. The subclause
structure is marked by the conjunction jotta, which translates to so that or in order to (see the next example). The converb is marked by the first infinitive transative verb form (e.g., toipu/a/kse/en = in order to recover), which consists of the verb stem (toipu), the first infinitive marker (a), the case ending denoting the transative (kse), and the possessive suffix (en). As may be noted from the example (see Appendix for the complete list of target sentences of Experiment 2), the converb construction contains two morphemes more than the subclause version. This was true for all the final construction sentences used in the experiment.

**Final construction** (target phrase is shown in bold with morphological boundaries marked by slashes)

**Full subclause:** Jotta uhri toipu/isi, hänen on saatava pikaisesti lisää verta.

**Converb construction:** Toipu/a/kse/en uhrijn on saatava pikaisesti lisää verta.

“In order for the victim to recover, he quickly needs to get more blood.”

A new set of temporal constructions were created, where the target clause appeared as the first clause of the sentence to make them comparable to the final construction sentences, for which the two alternative versions are only possible at the sentence-initial position. A set of 22 temporal sentence pairs were constructed, for which the target clause could naturally occur as the first clause (see the next example), and these were compared to a set of 22 pairs of final construction sentences, such as the previous one. Counterbalancing of sentences was performed as described earlier for Experiment 1.

**Temporal construction** (target phrase is shown in bold with morphological boundaries marked by slashes)

**Full subclause:** Kun myyjä punnit/si tomatteja, ostaja kaivoi kolikkoja tuskusta.

**Converb construction:** Myyjä/n punnit/e/ssa tomatteja ostaja kaivoi kolikkoja tuskusta.

“When the shopkeeper weighed the tomatoes, the customer dug up coins from his pocket.”

A corpus analysis was conducted using the Lauseopin arkisto syntax archives of oral and written Finnish to determine the frequency of usage of the two alternative clause forms for the temporal and final constructions. The examined corpus consisted of 36,165 clauses that stem from a wide variety of different written genres (excluding spoken discourse). For both the temporal and final construction, we counted the number of occurrences in the corpus of the respective subclause and converb constructions. Only those sentences were considered for which the
subclause and the converb structure were both possible. For the temporal construction, there were in the corpus 602 such subclause structures and 487 converb sentences; for the final construction the respective occurrences were 9 and 84. Thus, for the final construction the subclause form is clearly less frequent, whereas for the temporal construction the two forms were fairly evenly balanced with the subclause construction appearing somewhat more frequently than the respective converb construction. Overall, the temporal construction is more common than the final construction.

Separately for each sentence pair, the target phrases were closely matched for length. The average length of the subclause and the converb phrase of the temporal construction was 19.7 characters; for the final construction the respective lengths were 21.1 and 20.6. Moreover, the words in the target phrases were matched for lemma and surface frequency. The word frequencies are based on the Laine and Virtanen (1999) corpus. The average lemma frequency of the subject of the temporal construction was 57 per million and that of the final construction 51 per million; the respective average lemma frequencies of the predicate were 128 and 40 per million. The average surface frequency of the noun in the temporal construction was 7.0 for the non-inflected form and 8.6 for the inflected form; the respective surface frequencies for the final construction were 10.1 and 9.6. The average surface frequency of the verb in the temporal construction was 10.0 for the indicative verb form and 0.5 for the converb; the respective frequencies for the final construction were 0.5 and 0.3. We also computed the average bigram frequencies (per million). For the subject in the temporal construction the bigram frequency was 8541 for the non-inflected and 9040 for the inflected form; the respective frequencies in the final construction were 6826 and 8416. The average bigram frequency of the verb of the temporal construction was 7247 for the indicative form and 7464 for the converb; the respective frequencies of the final construction were 9093 and 8449. Thus, there was a tendency for the bigram frequencies to be greater for the converb version.

**Procedure.** The procedure was identical to that of Experiment 1.

**Results**

As in Experiment 1, gaze duration was used as the processing measure.\(^4\) A 2 (type of clause: temporal vs final) × 2 (form of clause: converb

---

\(^4\)As the target phrase appeared sentence-initially, gaze duration is here equivalent to the regression path measure (Konieczny, 1996).
construction vs full subclause) ANOVA was computed on these data with both variables being within-subject variables in the subject analysis and type of clause being a between-item and form of clause a within-item variable in the item analysis. The means and standard deviations are presented in Table 2.

**Target phrase.** There was a main effect of type of clause, $F_1(1, 26) = 24.58, MSe = 7277, p < .001; F_2(1, 42) = 5.89, MSe = 25069, p = .02$. Final clauses obtained longer gaze durations than temporal clauses. More importantly, however, the type of clause $\times$ form of clause interaction proved significant, $F_1(1, 26) = 11.04, MSe = 3806, p = .003; F_2(1, 42) = 5.08, MSe = 6910, p = .03$. For the temporal construction, the converb version was read with longer gazes than the full subclause version (a difference of 34 ms), thus replicating the main result of Experiment 1. For the final construction the opposite pattern emerged; the morphologically more complex converb structure was read with shorter gazes than the full subclause structure (a difference of 45 ms).

The interaction was also apparent in the number of fixations made before exiting the target clause, $F_1(1, 26) = 10.25, MSe = .12, p = .004; F_2(1, 42) = 8.50, MSe = .119, p = .006$. For the final construction the means were 4.56 and 4.17 for the subclause and converb constructions, respectively; the respective means for the temporal construction were 3.93 and 3.97.

**Post-target regions.** As is apparent from the list of target sentences (see Appendix), the main clause following the target phrase was identical in the two forms of clause except for one additional word in the converb form of the final construction. To test for possible spill-over effects in processing, we compared gaze durations for the identical part of the post-

<table>
<thead>
<tr>
<th>TABLE 2</th>
<th>Mean gaze durations (ms) and their standard deviations (in parentheses) for the two versions of the temporal and final construction in Experiment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Temporal construction</strong></td>
<td><strong>Final construction</strong></td>
</tr>
<tr>
<td><strong>Target phrase</strong></td>
<td><strong>Following phrase</strong></td>
</tr>
<tr>
<td>Full subclause</td>
<td>763 (180)</td>
</tr>
<tr>
<td>Converb</td>
<td>797 (233)</td>
</tr>
<tr>
<td>Difference</td>
<td>34</td>
</tr>
</tbody>
</table>
target region separately for temporal and final constructions. No signs of spill-over effects were observed (all Fs < 1). For the temporal construction, the means were 1553 ms and 1578 ms, and for the final construction 1659 ms and 1645 ms for the subclause and verb versions, respectively.

In the majority of the temporal constructions, the target phrase was immediately followed by a clause final adverbial phrase (see Appendix). These adverbial phrases were identical in the two clause forms except that in the subclause form the phrase was followed by a comma. A separate analysis was conducted for this text region as a further test of possible spill-over effects; moreover, it also allowed us to see whether the presence of a comma may affect processing time (see Hill & Murray, 2000). Gaze durations proved very similar between the two clause forms (ts < 1, the means were 510 ms and 513 ms for the subclause and the verb versions, respectively). This null effect suggests (1) that there were no spill-over effects and (2) that the presence of a comma did not seem to lengthen the reading time of the clause final phrase.

Discussion

Experiment 2 demonstrated that the structural complexity of linguistic expressions does not hamper on-line parsing when the more complex construction occurs more frequently in the language, but does so when the more and less complex versions are equally frequent. As regards the processing of the two equally frequent temporal constructions, the two experiments yielded a complexity effect of similar magnitude, 43 ms in Experiment 1 and 34 ms in Experiment 2. We reasoned earlier that if local ambiguity at the subject would be responsible for the processing time difference, the difference should have been larger in Experiment 2 than in Experiment 1. This is because in Experiment 2 the target phrase appeared sentence-initially with no prior linguistic context to narrow down the scope of alternative interpretations. As the size of the effect did not appear to be greater in Experiment 2 (if anything, it was slightly smaller), we reason that the effect is not explainable as an ambiguity effect. Nor can structural frequencies account for the effect, because the two alternative structures were reasonably similar in frequency. Instead, we argue that the observed difference in processing the two types of temporal construction has to do with the relative transparency and complexity of linguistic cues denoting a linguistic structure. When the structure is marked by bound morphemes, it is less transparent and thus more difficult to process than an analogous linguistic construction where the structure is signalled by free morphemes. In addition, the temporal verb construction contained one morpheme more than the corresponding subclause construction, which may also have contributed to the processing difficulty.
The competition model of Bates and MacWhinney (1989) ascribes special relevance to the transparency of linguistics cues that mark a given structure. According to this model, sentences are interpreted by direct form to function mappings, where the reliability, availability, assignability, and perceivability of a linguistic cue all contribute to the strength a cue has in suggesting a particular interpretation. By exposure to language, the language user takes advantage of the correlations that exist between forms and functions. In terms of the competition model, linguistic cues (i.e., bound morphemes) denoting a verb are reliable and available, and an interpretation can be assigned to the form immediately when the cues are encountered. Although available, as bound morphemes they are less easily perceivable, compared to the full subclause form, where the construction is marked by a visually salient free morpheme (i.e., a conjunction). In this sense, the verb constructions are less transparent than the full subclause constructions. Another difference between the two types of clauses is that for full subclauses a form-function mapping can be carried out already when reading the conjunction, whereas with temporal verb constructions the mapping has to be deferred until the predicate is encountered, making the processing under certain circumstances more demanding.

The second major finding is that the above discussed complexity effect is reversed when the more complex construction is more frequent in the written language than the less complex construction, as is the case for the final constructions. It is noteworthy that this “reversed complexity effect” came about when the more complex structure not only was less transparent but also had two morphemes more than the less complex structure. In the case of these final constructions, no local ambiguity is present. The only confounding factor is the presence of a comma attached to the verb in the subclause structure, which is absent in the verb structure. Hill and Murray (2000) have reported evidence that during processing of locally ambiguous sentences, there is a tendency for increased fixation time on the word followed by a comma. Although we cannot completely rule out this possible confound, we consider it quite unlikely. Namely, in both experiments we conducted separate analyses for the words that were identical except for the presence or absence of a comma after it, and we observed no signs of a “comma effect”.

Instead, we would like to interpret the results for the final construction as a structural frequency effect. The more complex verb construction was nine times more frequent than the less complex subclause construction. According to this view, frequent exposure to a given linguistic structure results in the corresponding mental structure becoming more easily accessible, which in turn will speed up the on-line processing of these structures when they are encountered in text. To parse and comprehend these
converbs, the reader has to mentally represent the separate morphemes the converb consists of. However, this is not sufficient, but a mental representation is also needed of the morpheme combination. This is because it is only the proper combination of morphemes that signals the linguistic structure, not the individual morphemes as such. Thus, a language user learns and represents these structures only by being exposed to them directly, it does not suffice to be exposed to the individual morphemes in other structures. With frequent exposure, the mental structure is strengthened and its processing facilitated, to the extent that the effect of structural complexity is overridden by the effect of structural frequency.

Although the complexity effect observed for the temporal conjunction is consistent with the competition model, the finding that structural frequency is capable of reversing the complexity effect appears to be at odds with it. Instead, it suggests there are also other factors, such as frequency that can influence the ease with which the mapping between linguistic form and function takes place. It is important to differentiate the notion of structural frequencies outlined above from lexically based exposure models of sentence parsing (see, e.g., Trueswell, Tanenhaus, & Kello, 1993), which are tuned to capture effects dependent on lexical preferences (e.g., the frequency with which a given verb takes a direct object). It should also be differentiated from morphological processing models of word recognition (see, e.g., Frauenfelder & Schreuder, 1992), which assume that the surface frequency of single lexical items (together with some other factors) determines if a word is recognised via the direct route or via the morphological parsing route. Lexical statistics are not at stake here, because the lexical items used in the two alternative forms (1) had identical stems and (2) had comparable surface frequencies. Instead, we are dealing here with frequencies of specific linguistic structures as such, without recourse to the lexical level.

CONCLUSIONS

The study showed that when a clause construction is marked by less perceivable bound morphemes, it is more effortful to process than when the same construction is denoted by a visually distinct free morpheme. This was true only when the complex and less complex structures were equally frequent. The complexity effect was reversed when the more complex structure was more frequent in the language—a finding in compliance with the view that structural frequencies also play a relevant role in parsing and comprehending morphologically complex structures during reading.

Manuscript received March 1999
Revised manuscript received February 2000
REFERENCES


APPENDIX

Target sentences in Experiment 1

(target phrases in bold; a = subclause, b = converb construction)

1a Yleisön joukossa vallitsi jännittynyt tunnelma, **kun autot ampaisivat liikkeelle.**
1b Yleisön joukossa vallitsi jännittynyt tunnelma **autojenampaistessa liikkeelle.**
2a Pappi keskusteli hääparin kanssa, **kun sukulaiset ate rioivat ruokasalissa.**
2b Pappi keskusteli hääparin kanssa **sukulaistenaterioidessa ruokasalissa.**
3a Laiskottelijat löhöävät sohvalla, **kun kuntoilijat hölkkäävät pururadalla.**
3b Laiskottelijat löhöävät sohvalla **kuntoilijoidenhölktäessä pururadalla.**
4a Tuomari pyyhki hikeä otsaltaan, **kun valmentajat jututtivat poikia aikaisän aikana.**
4b Tuomari pyyhki hikeä otsaltaan **valmentajienjututtaessa poikia aikaisän aikana.**
5a Seppo siivosi paniikinomaisesti, **kun ystävätkoputtivat ovelle.**
5b Seppo siivosi paniikinomaisesti **ystävien koputtaessa ovelle.**
6a Henkilövahingoilta välttytiin, **kun mopot kolaroi vat Hameentielä.**
6b Henkilövahingoilta välttytiin **mopojenkolaroidessa Hameentielä.**
7a Opiskelijat pelasivat jätkänsakkia, **kun professorit lu ennoivat esittelyluennolla.**
7b Opiskelijat pelasivat jätkänsakkia **professorienluen noidessa esittely- luennolla.**
8a Kokoneet vaeltajat lepälivät jo nuotionääressä, **kun retkeilijät patikoivat vielä jyrkkäärinnettäylös.**
8b Kokoneet vaeltajat lepälivät jo nuotionääressä **retkeilijöiden patikoidessa vielä jyrkkää rinnettäylös.**
9a Vauvan nukuttaminen on hankalaa, **kun kärpsäset surraavat ympärillä.**
9b Vauvan nukuttaminen on hankalaa **kärpästen sur ratessa ympärillä.**
10a Ala-asteen pihapiirissä syntyv vilskettä, **kun kellowpirisevät välitunnin loppumisen merkiksi.**
10b Ala-asteen pihapiirissä syntyv vilskettä **kelojoenpirstessä välitunnin loppumisen merkiksi.**
11a Luottamusmiespitääharjakaispuhetta, **kun työntekijät hörppivät kahvia tai olutta.**
11b Luottamusmies pitää harjakaispuhetta työntekijöiden hörppiessä kahvia tai olutta.
12a Saattaa miltei aistia entisaikojen olemassaolon, kun purjelaivat ankkurivat Aurajoen rannoille.
12b Saattaa miltei aistia entisaikojen olemassaolon purjelaivojen ankkuroissa Aurajoen rannoille.
13a Vahtikoira haukkuu taukoamatta, kun harakat hännävät sitä ja sen pentua.
13b Vahtikoira haukkuu taukoamatta harakoiden hännätessä sitä ja sen pentua.
14a Moraaliarvot on arvioitava uudelleen, kun normit höltyvät ja lainsäädäntö monimutkaistuu.
14b Moraaliarvot on arvioitava uudelleen normien höltyessä ja lainsäädännön monimutkaistuessa.
15a Jäänmurtajilla on sesonkiaika, kun tankkerit juuttuvat ahtojäihin.
15b Jäänmurtajilla on sesonkiaika tankkerien juuttuessä ahtojäihin.
16a Poika pelasi jalkapalloa, kun tyytö askarteli lastenhuoneessa.
16b Poika pelasi jalkapalloa tytön askarrellessa lastenhuoneessa.
17a Eläimiä kuoli runsaasti, kun puro ehtyi kokonaan.
17b Eläimiä kuoli runsaasti puron ehtyessä kokonaan.
18a Juhlija laahusti kotiin, kun ilta häämärti ja sataa tihutti.
18b Juhlija laahusti kotiin illan häämättyessä ja sateen tihuttaessa.
19a Lomalaiset nauttivat loistavasta lomasäästää, kun aurinko helotti kirkkaalta taivaalta.
19b Lomalaiset nauttivat loistavasta lomasäästää auringon helotteessa kirkkaalta taivaalta.
20a Äiti ajoi nurmikkoa, kun isä hinkkasi auton konepeltiä.
20b Äiti ajoi nurmikkoa isän hinkatessa auton konepeltiä.
21a Palovahti seurasi valppaana, kun asentaja hitsasi teräsputkia.
21b Palovahti seurasi valppaana asentajan hitsatessa teräsputkia.
22a Huoltimies leikkasi pensasaitaa, kun kodinhoitaja kokkasi keittiössä.
22b Huoltomies leikkasi pensasaitaa kodinhoitajan kokatessa keittiössä.
23a Toinen onkija vielä kalasti, kun toinen jo perkasi kaloon.
23b Toinen onkija vielä kalasti toisen jo peratessa kaloon.
24a Apumies rakensi telineitä, kun muurari rappasi seinää.
24b Apumies rakensi telineitä muurarina rapatessa seinää.
25a Sähkö kulutus väheni, kun pakkanen lauhtui maaliskuun lopulla.
25b Sähkö kulutus väheni pakkasen lauhtuessä maaliskuun lopulla.
26a Kotiavustajien tarve kasvaa, kun väestö vanhenee ja kylät autioituvat.
26b Kotiavustajien tarve kasvaa väestön vanhetessa ja kylien autioituessa.
27a Ruuan varastaminen ei liene vakava rikos, kun nälkä kurnii vatsassa.
27b Ruuan varastaminen ei liene vakava rikos nälän kurniessa vatsassa.
28a Öljyn hinta nosee, kun kauppasota laajentuu Eurooppaan.
(a) Final constructions
1a Jotta pannukahvi selkiintyisi, sen on annettava seistä muutama minuutti.
1b Selkiintyaikseen pannukahvin on annettava seistä muutama minuutti.
2a Jotta putkimies työlistyisi, hänen on oltava itse vähän aktiivisempi.
2b Työlistyaikseen putkimiehen on oltava itse vähän aktiivisempi.
3a Jotta taimi juurtuisi, sen on annettava kasvaa rauhassa ainakin pari viikkoa.
3b Juurtuakseen taimen on annettava kasvaa rauhassa ainakin pari viikkoa.
4a Jotta neste haihtuisi, sen täytyy kohota yli 120 asteeseen.
4b Haihtuakseen nesteen täytyy kohota yli 120 asteeseen.
5a Jotta solmu aukeaisi, sen on oltava oikein tehty.
5b Auetakseen solmun on oltava oikein tehty.
6a Jotta uhri toipuisi, hänen on saatava pikaisesti lisää verta menetetyn tilalle.
6b Toipuakseen uhrin on saatava pikaisesti lisää verta menetetyn tilalle.
7a Jotta maalaiskunta laajenisi, sen on ostettava valtion maita kunnan rajan pohjoispuolelta.
7b Laajentuakseen maalaiskunnan on ostettava valtion maita kunnan rajan pohjoispuolelta.
8a Jotta vanhus parantuisi, hänen on voimisteltava joka päivä kahdesti.
8b Parantuakseen vanhuksen on voimisteltava joka päivä kahdesti.
9a Jotta nuotio syttyisi, sen on oltava riittävän kuiva.
9b Syttyäkseen nuotion on oltava riittävän kuiva.
10a Jotta sadetakki suojaaisi, sen on oltava riittävän pitkä.
10b Suojatakseen sadetakin on oltava riittävän pitkä.
11a Jotta katsoja viihtyisi, hänen on kyettävä samaistumaan johonkin näytelmän henkilöistä.
11b Vihtiöykseen katsojan on kyettävä samaistumaan johonkin näytelmän henkilöistä.

12a Jotta vene uppoaisi, sen täytyy täyttyä ääriään myöten vedellä.
12b Upotakseen veneen täytyy täyttyä ääriään myöten vedellä.
13a Jotta jälkikasvu varttuisi, sen on saatava mahdollisimman monipuolistaa ravintoa.
13b Varttuakseen jälkikasvun on saatava mahdollisimman monipuolistaa ravintoa.
14a Jotta Ilves toipuisi, sen on saatava puolustuspelinsä kontrolli paremmaksi.
14b Toipuakseen Ilveksen on saatava puolustuspelinsä kontrolli paremmaksi.
15a Jotta työtön selviytyisi, hänen on otettava aloite omiin käsiinsä.
15b Selviytyökseen työttömen on otettava aloite omiin käsiinsä.
16a Jotta päiäma karttuisi, sen on annettava kasvaa korkoa vielä useita vuosia.
16b Karttuakseen pääoman on annettava kasvaa korkoa vielä useita vuosia.
17a Jotta kaupankäynti kiihtyisi, sen tulee saada apua ruplan kurssin paranemisesta.
17b Kiihtyökseen kaupankäynnin tulee saada apua ruplan kurssin paranemisesta.
18a Jotta tiele kukostaaisi, sen täytyy saada kehittyä omilla ehdollaan.
18b Kukoistaakseen tieteen täytyy saada kehittyä omilla ehdollaan.
19a Jotta maissi kypsyisi, sen tulee saada riittävästi lämpöä.
19b Kypsyökseen maissin tulee saada riittävästi lämpöä.
20a Jotta myrsky laantuisi, sen on syytää pian näyttää selviä heikkemisen merkkejä.
20b Laantuakseen myrskyn on syytää pian näyttää selviä heikkemisen merkkejä.
21a Jotta yhteispele sujuisi, se vaatii vielä paljon harjoitusta.
21b Sujuakseen yhteispele vaatii vielä paljon harjoitusta.
22a Jotta laavavirta pysähyy, se tarvitsee tielleen ison ja kuumuutta kestävän esteen.
22b Pysähtyökseen laavavirta tarvitsee tielleen ison ja kuumuutta kestävän esteen.

(b) Temporal constructions

1a Kun myyjä punnitsi tomaatteja, ostaja kaivoi kolikkoja taskustaan.
1b Myyjän punnitessa tomaatteja ostaja kaivoi kolikkoja taskustaan.
2a Kun merivesi länpäti, Ruissalon rannoilla alko näkyä taas uimareita.
2b Meriveden lämmetessä Ruissalon rannoilla alko näkyä taas uimareita.
3a Kun humalainen horjui kotiin, talonmies oli aloittanut jo aamutoimensa.
3b Humalaisen horjuessa kotiin talonmies oli aloittanut jo aamutoimensa.
4a Kun nuorukainen siivosi huonettaan, vanhemmat valmistivat päivällistä.
4b Nuorukaisen siivotessa huonettaan vanhemmat valmistivat päivällistä.
5a Kun velka pienentyi, aloin mieltää uuden auton hankkimista.
5b Velan pienentyessä aloin mieltää uuden auton hankkimista.
6a Kun hevonen ikääntyi, sillä ei enää voinut ratsastaa useita tunteja päivässä.
6b Hevosen ikääntyessä sillä ei enää voinut ratsastaa useita tunteja päivässä.
7a Kun eläkeläinen kuivasi tiskiä, kotiavustaja imuroi makuuhuoneen.
7b Eläkeläisen kuivatessa tiskiä kotiavustaja imuroi makuuhuoneen.
8a Kun apulainen lastasi tavaarolta lavalle, kuljettaja jututti kaupan myyjää.
8b Apulaisen lastatessa tavaarolta lavalle kuljettaja jututti kaupan myyjää.
9a Kun pakkanen paukkui ulkona, sisällä oli miellyttävää takkatulen ääressä.
9b Pakkasen paukuessa ulkona sisällä oli miellyttävää takkatulen ääressä.
10a Kun Aurajoki kuohui valtoimenaan, ihmiset pelkäsivät veneiden iartoavan kiinnityksistään.
10b Aurajoen kuohuessa valtoimenaan ihmiset pelkäsivät veneiden iartoavan kiinnityksistään.
11a Kun karkulainen sieppasi avaimet, toiset vangit kiinnittivät vartijan huomion puoleensa.
11b Karkulaisen siepatessa avaimet toiset vangit kiinnittivät vartijan huomion puoleensa.
12a Kun isäntä pilkoi saunapuita, naapurit huolehtivat saunaveden kantamisesta.
12b Isännän pilkoessa saunapuita naapurit huolehtivat saunaveden kantamisesta.
13a Kun mustalainen vähätteli syylisyyttään, poliisit alkoivat menettää kärsivällisyyttään.
13b Mustalaisen vähätellessä syylisyyttään poliisit alkoivat menettää kärsivällisyyttään.
14a Kun koululainen uskaltautui avaamaan suunsa, opettaja palkitsi sen välittömästi kehuminisella.
14b Koululaisen uskaltautuessä avaamaan suunsa opettaja palkitsi sen välittömästi kehuminisella.
15a Kun tanssimusiikkii taukosi, nuoriso siirtyi tuota pikaa kadulle.
15b Tanssimusiikkin tauotessa nuoriso siirtyi tuota pikaa kadulle.
16a Kun talouslaman väisty, autokauppa elpyi nopeasti.
16b Talouslaman väistyessä autokauppa elpyi nopeasti.
17a Kun velallinen erehtyi kertomaan todellisen rahatilanteensa, pankki irtisanoi lainan välittömästi.
17b Velallisen erehtyessä kertomaan todellisen rahatilanteensa pankki irtisanoi lainan välittömästi.
18a Kun professorit luennoivat esittelyluennolla, opiskelijat pelasivat jätänsakkia.
18b Professorien luennoidessa esittelyluennolla opiskelijat pelasivat jätänsakkia.
19a Kun kellot pirisevät välitunnin loppumisen merkiksi, ala-asteen pihapiirissä syntyy vilskettä.
19b Kellojen piristessä välitunnin loppumisen merkiksi ala-asteen pihapiirissä syntyy vilskettä.
20a Kun tankkerit juuttuvat ahtoäihin, jäänmurtajilla alkaa varsinainen sesonkiaika.
20b Tankkerien juuttuessa ahtoäihin jäänmurtajilla alkaa varsinainen sesonkiaika.
21a Kun väestö vanhenee ja kylät autioituvat, kotiavustajien tarve kasvaa.
21b Väestön vanhetessa ja kylien autioituessa kotiavustajien tarve kasvaa.
22a Kun kauppasota laajentuu Eurooppaan, öllyn hinta tulee nousemaan.
22b Kauppasodan laajentuessa Eurooppaan öllyn hinta tulee nousemaan.