

Identifying Existing and Novel Compound Words in Reading Finnish: An Eye Movement Study

Jukka Hyönä

Department of Psychology
University of Turku
Turku, Finland
hyona@utu.fi

Minna Koski

Department of Psychology
University of Turku
Turku, Finland
mmjkos@utu.fi

Alexander Pollatsek

Department of Psychology
University of Massachusetts
Amherst, MA, USA
pollatsek@psych.umass.edu

1 Introduction

According to the dual-route race model of compound word identification (Pollatsek, Hyönä, & Bertram, 2000), the holistic route and the morphological decomposition route operate in tandem. Bertram and Hyönä (2003) posited that word length modulates the interplay between the two access routes. When a compound word is sufficiently short so that all or most of its letters fall on the foveal region when fixating it during reading, the holistic route gets a head start and completes faster than the morphological route and thus the word is more likely to be identified as a whole. On the other hand, when a compound word is so long that a subset of letters is beyond foveal reach, the identification is initiated by first recognizing the initial constituent followed by the recognition of the second constituent and that of the whole word.

In their study examining the processing of novel compound words, Pollatsek et al. (2011) demonstrated that the decomposition route played even a more prominent role in processing novel than lexicalized compound words. Pollatsek et al. (2011) compared the processing of novel and existing Finnish compound words by manipulating the frequency of first constituent as an independent word, separately for long (average length of 13 letters) existing and novel compound words. The length of the first constituent as well as the frequency of the second constituent was matched across conditions. For first fixation duration, which indexes early effects in word processing, an effect of first-constituent frequency was observed that was similar in size for existing and novel compound words. For gaze duration (i.e. the summed duration of fixations made on the word before exiting to the right or left) first-constituent frequency was greater for novel than existing compound words. For the

latest stages of processing during the first-pass reading, indexed by fixation time spent on the target word after fixating away from the first constituent but before exiting the word, only a main effect of novelty was observed. As regards to the processing of long novel compound words, the pattern of results was taken to suggest a two-stage process. During the first stage, lexical access is achieved for the compound word constituents. During the second stage, the meaning of the novel compound word is composed out of the constituent meanings. The second stage is assumed to take longer when the frequency of the first constituent is low, because the prototypical relationships that the low-frequency first constituent would be engaged in compounding are not firmly established.

In the present study, we further investigated the processing of novel and lexicalized Finnish two-noun compound words. This time we manipulated the frequency of the second constituent (the compound head). It was done separately for existing and novel compound words. Moreover, we also manipulated the length of the compound words. If indeed word length strongly determines the interplay between the holistic and decomposition route in compound word identification, as argued by Bertram and Hyönä (2003), the manipulation of the second-constituent frequency tapping into the decomposition process should result in different types of processing especially for short existing versus novel compound words. Short existing compound words are more likely to be identified by the holistic route, whereas short novel compound words have to be processed via the morphological decomposition route. For long compound words, on the other hand, the manipulation of the second-constituent frequency should lead to less dramatic differences between existing and novel compounds, as the decomposition

Copyright © by the paper's authors. Copying permitted for private and academic purposes.

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at <http://ceur-ws.org>

route is assumed to be in operation for both word types.

Adult readers read sentences silently for comprehension while their eye movements were registered. The target compound words were embedded somewhere in the middle of the sentences. The frequency of the second constituent as a separate word was manipulated for short (7-9 letters) lexicalized (e.g., *savukala* = smoked fish) and novel (e.g. *hymykisa* = smile contest) compounds as well as for long (12-16 letters) lexicalized (e.g., *hiekkapaperi* = sand paper) and novel (e.g., *skandaalivaali* = scandal election) compound words. Thus, the experimental design was a 2 (low vs. high frequency second constituent) x 2 word type (existing vs. novel) x 2 word length (short vs. long) within-participants design. Comprehensibility of the novel compound words was secured by a rating test conducted prior to the experiment proper. Only novel compound words whose meaning could be computed without providing any linguistic context were chosen for the study. The frequency of the first constituent was matched across the conditions, as was the frequency of the short and long existing compound words.

2 Results

Several eye fixation measures were used to tap into the time course of compound word processing. The earliest effects were measured by first fixation duration. Early, but less immediate effects were measured by second fixation duration and gaze duration. Still later effects were measured by total fixation time, which is the sum of all fixations, both first-pass and second-pass, made on the target word.

First fixation duration: In the earliest stages of foveal word processing, indexed by first fixation duration, no effects of novelty or second-constituent frequency were observed.

Second fixation duration: A bit later in the processing timeline, main effects of novelty and second-constituent frequency were obtained. These effects were modified by interactions involving word length, including the three-way interaction. This interaction was broken down by computing a separate 2x2 ANOVA for short and long compounds, respectively. These analyses revealed no effect of novelty or second-constituent frequency for long compounds, whereas for short compounds both main effects and their interaction proved significant. The interaction reflected the fact that the second-

constituent frequency effect was only observed for short novel compounds.

Gaze duration: In gaze duration, summing up all fixations made during the first-pass reading, the main effect of word type, word length and second-constituent frequency were all significant. Gaze duration was significantly longer for novel than existing words, longer for long than short words, and longer for compounds containing a low-frequency than high-frequency second constituent. Similarly to second fixation duration, gaze duration also revealed a reliable three-way interaction between the manipulated factors. In order to examine in more detail the interaction, it was broken down into two separate 2x2 ANOVAs, one for the short and another for the long compound words.

For the long compound words, there was a main effect of word type and second constituent frequency, but no reliable interaction between them, suggesting that the second-constituent frequency effect was of similar magnitude for existing (an effect size of 91 ms) and novel (an effect size of 111 ms) compound words. However, for short words, the Word Type x Second-Constituent Frequency proved significant. This interaction reflected the fact that the second-constituent frequency effect was considerably greater for novel (an effect size of 155 ms) than for existing (an effect size of 42 ms) compound words.

Total fixation time: We also analyzed the total fixation time spent reading the target words. This measure indexes late effects; it sums up the duration of all fixations made on the word during the first-pass and second-pass reading. In this measure, the three-way interaction obtained for second fixation duration and gaze duration was no longer significant. However, the interaction between word type and second-constituent frequency was almost significant. This interaction reflects the fact that in total fixation time the effect of second constituent frequency was greater for novel than existing compound words, regardless of word length. The size of the second-constituent frequency effect was 51 ms for the existing compounds and 151 ms for the novel compounds.

Summary of results: The following picture emerges from the pattern of results presented above. In the earliest stages of word processing, no signs of either novelty or second-constituent frequency were seen, which suggests that these effects took some time to develop during compound word identification. For long compounds,

these effects were still absent in second fixation duration but emerged in gaze duration. For short compounds, the effects were already visible in second fixation duration. Finally, the measure indexing second-pass reading demonstrated a greater second-constituent effect for novel than existing compounds. All in all, the pattern of results suggests that meaning composition takes place with more delay for long than short compound words.

3 Conclusions

The present study provided further evidence for the view (Bertram and Hyönä, 2003, 2013), according to which word length modifies the relative role of the holistic versus the morphological decomposition route in compound word identification. The decomposition route is an integral part in identifying long compound words, because holistic processing is not viable due to visual acuity constraints. This became apparent in the effect of the second constituent frequency indexing access via morphological constituents being similar in magnitude for the novel and lexicalized compound words. On the other hand, when lexical access via the holistic route is a viable option, as is the case with short existing compound words that fit in the foveal area of the eyes when the word is fixated, the novelty effect emerged relatively early (during second fixation) and the second-constituent frequency effect was considerably smaller for existing than novel compound words during first-pass reading. Finally, the second-pass reading measure demonstrated a greater effect of constituent frequency for novel than lexicalized compounds. This may be taken to suggest that meaning composition takes longer when the frequency of the second constituent is low, since the typical relationships low-frequency constituents are engaged in compounds are less firmly established.

There are also two findings that are not completely in line with the visual acuity principle proposed by Bertram and Hyönä (2003). One is the absence of an early novelty effect for short compound words. If the holistic route is immediately activated when making the first fixation on the word, there should have been a novelty effect in first fixation duration. Second, there was a 42 ms effect of second constituent frequency in gaze duration even for existing short compound words, suggesting that the decomposition route

also becomes active when identifying short lexicalized compounds.

A possible theoretical framework that can account for the obtained pattern of results is a dual route cascade model assuming that identification always starts out with the decomposition route with the process quickly cascading into the holistic access in the case of short compounds and with some delay in the case of long compounds. When a compound word is short, its constituents are also short and may be accessed rapidly. On the other hand, when the word is long, not only the constituents are likely to be longer and may thus lengthen their access, but the morphological segmentation process may also need additional time. Hence, the holistic route is activated with some delay after the decomposition route is activated. The suggested model may be further tested by replicating the present study by manipulating the frequency of first constituent separately for long and short, novel and lexicalized compound words.

References

- Bertram, R., & Hyönä, J. (2003). The length of a complex word modifies the role of morphological structure: Evidence from eye movements when reading short and long Finnish compounds. *Journal of Memory and Language*, 48, 615-634.
- Bertram, R., & Hyönä, J. (2013). The role of hyphen at the constituent boundary in compound word identification: Facilitative for long, detrimental for short compound words. *Experimental Psychology*, 60, 157-163.
- Pollatsek, A., Bertram, R., & Hyönä, J. (2011). Processing novel and lexicalised Finnish compound words. *Journal of Cognitive Psychology*, 23, 795-810.
- Pollatsek, A., Hyönä, J., & Bertram (2000). The role of morphological constituents in reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 820-833.