Word order and tonal shape in the production of focus in short Finnish utterances

Martti Vainio[1], Juhani Järvikivi[2], Stefan Werner[3]

[1]Department of Speech Sciences, University of Helsinki, Finland
[2]Department of Psychology, University of Turku
[3] Department of Foreign Languages, University of Joensuu

martti.vainio@helsinki.fi

Abstract

This paper presents results from a study on the production of Finnish prosody. The effect of word order and the tonal shape in the production of Finnish prosody was studied as produced by 8 native Finnish speakers. Predictions formulated with regard to results from an earlier study pertaining to the perception of prominence were tested. These predictions had to do with the tonal shape of the utterances in the form of a flat hat pattern and the effect of word order on the so called top-line declination within an adverbial phrase in the utterances. The results from the experiment give support to the following claims: the temporal domain of prosodic focus is the whole utterance, word order reversal from unmarked to marked has an effect on the production of prosody, and the production of the tonal aspects of focus in Finnish follows a basic flat hat pattern. That is the prominence of a word can be produced by an f_0 rise or a fall, depending on the location of the word in an utterance. The basic accentual shape of a Finnish word is then not a pointed rise/fall hat shape as claimed before since it can vary depending on the syllable structure and the position within an utterance.

Index Terms: focus, intonation, word order, Finnish.

1. Introduction

It has been shown that a later f_0 peak in an utterance has to be lower than the previous ones to be perceived as having an equally high pitch (see for instance [1] for English, [2] for Dutch as well as [3] and [4] for Finnish). Pierrehumbert [1] explains this by postulating a mental representation of declination which is used by the listener to normalize for physically conditioned declination of f_0. The peak height relations are directly related to the relative prominence of the words within an utterance and from the perceptual point of view it follows naturally that these phenomena should reflect the production of prosody. Prominence, on the other hand, is directly related to the accentuation of words; that is, a word has to be accented in some manner in order to gain prominence. The accentual shapes of Finnish words have usually thought to be the basic pointed hat shapes (see for instance [5] and the references therein). The possible shapes of pitch accents in Finnish and the fact that a mere fall in the f_0 curve could be used for accentuation have not, however, been attested before and no systematic studies exist.

Word order and prosody are the main means to mark the distribution of information within a sentence, i.e. the information structure. An important part of information structure has to do with the role of new (given) and old (inferred) information. The given or presupposed information is traditionally referred as the topic of the sentence. In contrast, focus is usually used to refer to new information, or, particularly information that is not within what is already pragmatically presupposed (e.g.,[6]). In addition to a particular referent being "old" or "new" information, often the relationship between a focused referent ("new" information) and what is pragmatically presupposed together make the focused referent informative. In Finnish, the flexible word order can be used to serve information structure. For example, in an unmarked case, such as "menimme laivalla Lemille" (we went by boat to Lemi), the canonical order of the two adverbs (manner + place) conforms to its default information structure, and the phrase as a whole is under so-called sentence focus [6] realized prosodically as broad focus. Consequently, no pragmatic presuppositions are evoked by the word order. In contrast, changing the word order to marked "menimme Lemille laivalla" presupposes that we did in fact go to Lemi, but now the word order is used to focus the fact that it was by boat we went to Lemi - not by a car - as if it were an answer to a question "how did you go to Lemi?" (For the pragmatic use of word order in Finnish, see, e.g., Hakulinen and Karlsson, [7], and Vilkuna, [8]). Apart from word order, however, prosody can be used to mark any constituent under the domain of focus even in the syntactically unmarked case by increasing the accent or stress on the part of an utterance that is intended to be brought into focus. Thus a Finnish speaker can say "*Manne* meni Lemille" ("Manne went to Lemi") as well as "Manne meni *Lemille*" ("Manne went to Lemi"; italics depict prosodic focus). An important question is, then, whether the two main means available – syntactic and prosodic – interact in production when of one or another part of an utterance has to be marked as focus

In earlier study on the perception of prominence in Finnish we found that word order reversal had an effect on the perceived prominence of words in a short Finnish utterance. On the other hand, we also found that the prominence of two nouns in the utterance followed a so called flat hat pattern; i.e., the prominence of the earlier word related to the f_0 rise and the prominence of the latter words was related to the f_0 fall with relative heights of the peaks being the most important factor.

These findings then gave us a set of hypotheses to be tested with regard to the production of prosody. We therefore designed an acoustic prosodic experiment to test the following findings in the production experiment: 1) there is a clear perceptual bias caused

by word order changes in an adverbial phrase in a short Finnish utterance, 2) the relative difference of the two peaks in the f_0 contour is the most important single parameter in the perception of relative prominence within an utterance, and in addition to the relative heights 3) the rise of an earlier peak is mostly responsible for the prominence of that word, and 4) a fall after final peak its most important factor responsible for the prominence of that word. The rest of the paper describes the experiment and the results and ends with a discussion about possible further studies.

2. Materials

The list of sentences used in the experiment was similar to the ones used in the earlier perception experiment [4]; a simple declarative sentence starting with a verb and ending with an adverbial phrase whose word order could be reversed to mark the sentence for focus. The basic sentence "Menemme laivalla Jimille" ("We go - by boat - to Jimi's"), allows for three different focus conditions with regard to the nouns *laiva*, and *Jimi*; namely broad, narrow on the first noun, and narrow on the second noun. Two different words were used for the vehicle (laiva (boat) and juna (train)) and three proper nouns for the person to be visited (Jimi, Jani, and Lumi). With three different focus conditions and two different word order conditions a set of 36 different sentences was created. Accordingly, a set of prompt questions matching the intended three focus conditions was created as follows:

- Broad focus: Mitä teette tänään 'what do you do today'?

- Narrow focus on "laivalla": Millä menette Lumille 'With what/How do you go to Lumi's'?

- Narrow focus on "Lumille": Minne menette laivalla 'where do you go by boat'?

The question prompts were then recorded by a female speaker with two different emphases; neutral and emphatic. The 36 replies were paired with both emphatic and neutral questions which doubled the number of replies to 72. The emphasis in the questions was produced by strong emphasis on the question word of the utterance.

2.1. Participants and procedure

Eight participants (7 female) took part in the experiment. All of the participants were choir members living in the Helsinki area with similar backgrounds in eastern Finland. None of the participants were familiar with speech research and none reported any hearing problems. All of the speakers spoke with a neutral Helsinki area dialect/accent.

The 72 prompt-reply pairs were randomized for each participant and the he or she was given a sheet of paper with the corresponding replies. The focus was not indicated in any way on the paper as it was intended to be elicited by the type of question. The participants were not told of the nature of the experiment and were asked to speak lively.

The prompt questions were played to the participants through a high quality loudspeaker (Genelec 1029A) in a sound-treated recording studio at the Department of Speech Sciences in Helsinki. The prompts were spaced so that the participants had ample time to reply. The replies were recorded directly to a computer hard disk at 44.1 kHz sampling frequency and 16 bit quantization using a high quality analog-to-digital converter (Digi002 by Digidesign) and a high quality condenser microphone (AKG 4000B).

2.2. Results

Before data analyses each participant's responses were labeled and both intensity and f_0 were calculated. The utterances were automatically aligned with a speech recognizer and the results were manually corrected. Three points of interest for each word in the utterance were marked on the f_0 curve. These were meant to correspond to the basic point hat pattern mainly used for accentuation in Finnish. That is, the first point corresponds to the start of the f_0 rise, the second point to the peak, and the last point to the end of the f_0 fall. Both f_0 and intensity were measured at these points for subsequent statistical analyses. In addition to these points non-modal voice quality was marked. An example of the analysis points can be seen in Figure 1.
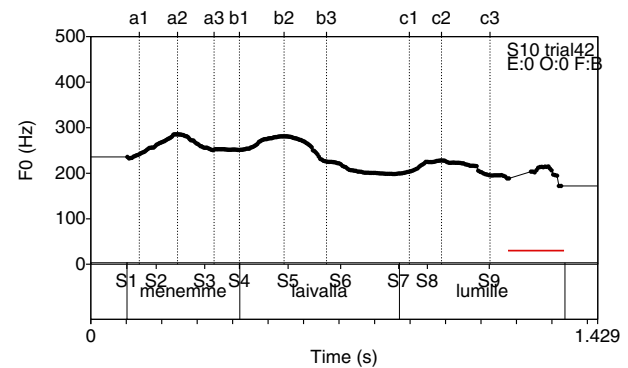

Figure 1: An example of an f_0 curve and the measured points of interest. The figure shows word level segmentation of an utterance with a broad focus; " mememme laivalla Lumille" (we go by boat to Lumi's). The horizontal (red) line at the end of the utterance depicts the span of vocal fry. The intervals marked with capital S stand for consecutive syllables.

Before the statistical analyses were conducted the first author went through the utterances and marked problematic ones by visually inspecting the f_0 curves. The problematic 36 utterances – together with a same number of filler utterances – were then played to a set of 20 naive listeners who judged the focus condition of the utterances. The ones whose focus was judged to be the intended one by zero to three listeners were removed from most of the subsequent analyses as outliers. All in all, 25 items were rejected this way (4% of the data). The rises, falls, and peak height differences used in the statistical analyses were calculated in semitones.

2.2.1. Tonal shape – the flat-hat pattern

The following predictions concerning the flat hat pattern of accentuation were tested: 1) the most important feature responsible for the focus conditions is the difference in peak heights of the two accented words, 2) the rise of the first peak is more important than the fall, and 3) the fall of the latter peak is more important than its rise. The relative importance of the features were tested with logistic regression with the different pitch related features as predictors and the given focus condition as the dependent variable. We only analyzed the two narrow focus conditions as there were no predictions with regard to the broad focus condition.

A logistic regression clearly showed that the in the narrow fo-

cus on the first word condition the difference between the peaks to be highly significant ($\chi^2(1) = 63.94, p < .0001$). The f_0 rise was also significant ($\chi^2(1) = 7.76, p < .0053$), but the subsequent fall failed to reach significance ($\chi^2(1) = 2.27, p < .109$). Naturally, the fall and the peak difference are highly correlated ($r = 0.8664225$) making the local fall redundant in the analysis. The non-linearities of the parameters was also tested using *restricted cubic splines* ([9]), but they all failed to reach significance.

The latter peak turned out to be more complicated than the the previous peak with regard to the analyses. First of all the rise of the peak was significant ($\chi^2(4) = 19.23, p < .0007$) with a significant non-linearity ($\chi^2(3) = 8.87, p < .03$); the fall of the peak was also significant ($\chi^2(4) = 12.49, p < .014$) with a significant non-linearity ($\chi^2(3) = 11.54, p < .009$), and the difference between the two peaks was, again, highly significant ($\chi^2(1) = 77.47, p < .0001$).

Why then, was the rise of the last peak significant. Figure 2 shows the *log odds* of the rise parameter as a function of the last peak rise in the logistic model. It is obvious from the figure that the amount of rise (depicted on x-axis with negative values as the the amount was calculated by subtracting the peak value from the rise start) does not affect the probability of the narrow focus category until there is no peak left; or when the peak is, in fact, negative. The confidence of the model is lost after that point (see the 95% confidence intervals [dotted lines] in the figure). This can be seen as an elbow around zero semitones in the figure. It can therefore be inferred that the mere presence of a peak is required. The presence of the peak, on the other hand, is probably conditioned by the local tonal shape of the words. In order to accent the word (prerequisite for prominence) either a H tone or an F (depending on the moraic structure of the accented syllable) has to be produced. Both are characterized by a local rise in pitch (see [10] for Finnish and [11] in general). On the other hand, the presence of the fall of the accent is also obligatory in the sense that Finnish statements end with a fall; very often to a level where a vocal fry is introduced. The creaky voice is used to signal finality and turn yielding in Finnish; a suitable function with regard to the design of the experiment. The whole of the utterance is, thus, used also in production to signal focus related prosodic prominences. Moreover, there are clearly cases where a person can produce a narrow focus on the last word without producing a conspicuous local rise. Figure 3 depicts an f_0 curve with a narrow focus on the latter word. The accent is clearly produced with a falling accent. The accented syllable in this case is bimoraic which are produced by a falling dynamic tonal target in Finnish (see [10] for more information). Additionally, the valley after the first peak is due to a low target on the second syllable of the word. That is, the disyllabic basic Finnish word structure is characterized by low target on the second, unstressed syllable of the word. The other syllables are probably tonally neutral, which accounts for the gradual rises towards the latter peak often seen in the data.

2.2.2. Word order

First, to investigate whether emphasizing the prompt affected the production of focus 2 x 2 x 3 analyses of variance (ANOVAs) were done using peak difference in semitones as the dependent measure averaged over subjects (F1) and items (F2). Emphasis (neutral, emphasized), Word Order (unmarked, marked) and Focus (broad =B , noun 1 = N1, noun 2 = N2) were within-variables in the subject analyses. In the item analyses Word Order was a between-
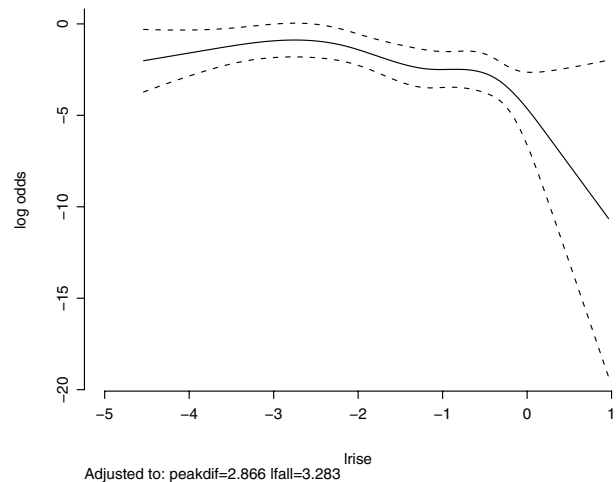

Figure 2: Log odds as a function of final peak rise in the logistic model for narrow focus on latter word.

item factor. The results showed a significant main effect of Focus, $F1(2, 14) = 70.28, p < .001; F2(2, 20) = 521.22, p < .001$, (average peak differences in semitones: Broad, 3.396; N1, 6.962; N2, -0.845) and a main effect of Emphasis that was significant by subjects, $F1(1, 7) = 8.97, p < .001$, (average peak differences: neutral, 3.020; emphasized, 3.322) but failed to reach significance in the item analyses $[F1(1, 10) = 1.49, p > .1]$. There was no effect of Word Order ($F's < 1$). In addition none of the interactions reached statistical significance (all $p's > .09$).

Expectedly, the type of focus was a significant factor affecting the production results. Contrary to our expectations, however, the interaction between word order and focus type did not quite reach statistical significance ($p's > .09$), although the results are suggestive to the expected direction. On the one hand, it is possible that due to the number of subjects and items there was simply not enough statistical power for it to reach significance. On the other hand, it is also possible that putting emphasis on the prompt question created unnecessary noise in the data. The results showed that the emphasis on the question prompt had an effect across all focus types and both word orders. However, as the emphasis was realized in stressing the question word in each case, i.e., having a narrow focus on the question word, it is possible that it may have affected the broad focus condition differently than the other two focus conditions. In particular, it is possible that emphasis may have produced noise in the data, because there is no such obvious way to interpret the (narrow focus) emphasis on the question word in the case of broad focus as there is in the case of narrow focus. Thus, it is possible that subjects interpreted the prompts in such cases in an item-by-item fashion, as suggested by the non-significant item analyses. In order to check whether this was indeed the case, separate ANOVAs were carried out for the neutral and emphasized conditions.

Emphasis on the prompt question: ANOVAs showed a significant main effect of Focus, $F1(2, 14) = 59.98, p < .001; F2(2, 20) = 208.70, p < .001$, but again no effect or
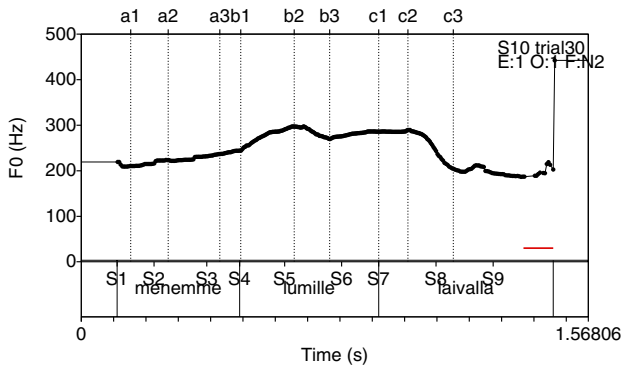
Figure 3: An example of a flat hat pattern and a final fall without a local rise in pitch. The intended focus condition in the utterance is a narrow focus on the last word.

Word Order ($F's < 1$) and no significant interaction ($p's > .3$). No emphasis on the prompt question: ANOVAs showed again a significant main effect of Focus, $F1(2, 14) = 71.46, p < .001; F2(2, 20) = 429.79, p < .001$, but again no effect or Word Order ($F's < 1$). In contrast to the above, however, there was a significant interaction between Word Order and Focus, $F1(2, 14) = 3.48, p = .059; F2(2, 20) = 4.17, p = .031$.

Table 1: Mean f_0 peak difference by focus and markedness (word order) in semitones.

	marked	unmarked
Broad B	4.063	2.830
Narrow N1	6.647	7.005
Narrow N2	-1.008	-1.420

That mean values for each focus condition broken down by markedness (word order) are depicted in Table 1. Pairwise comparisons across markedness showed that the difference between the unmarked and marked broad focus conditions was indeed significant ($t1(7) = 2.51, p < .05; t2(10) = 1.78, p = .05$). The tendencies to compensate for the word order reversal can be seen in the broad condition as well the N2 condition. That is, in the broad focus case the top-line difference is raised to compensate for the extra prominence due to the marked word order. Similarly the difference is decreased in the N2 condition.

3. Discussion

The results displayed in this paper reveal a complicated phenomenon relating to the production of focus in Finnish. In general, the results are in consonance with similar results on perception of prominence. Most importantly the production of focus is not localized to the prominent or emphasized word but relates to the time domain of the whole utterance or at least to the part of it where the relative prominences are relevant; a whole adverbial phrase in our case. The results are, moreover, in accordance with a somewhat similar study by Xu and Xu on the realization of focus in English [12].

The experiment described in this paper was designed before we became aware of the tonal differences of Finnish stressed syllables [10] and the effect of the different tonal structures on the realization of focus should be systematically studied. It could well be the case that local prominences are in large part dependent on the different tones. That is, an H tone could in general be related to a rise, whereas an F tone could be, respectively related to a fall in both production and perception. This is not a confound regarding this study, but is calls for further systematic studies regarding tone and intonation in Finnish.

5. References

[1] Janet Pierrehumbert, "The perception of fundamental frequency declination," *Journal of the Acoustical Society of America*, vol. 66, pp. 363–369, 1979.

[2] C. Gussenhoven, B. H. Repp, A. Rietveld, H. H. Rump, and J. Terken, "The perceptual prominence of fundamental frequency peaks," *Journal of the Acoustical Society of America*, vol. 102, no. 5, pp. 3009–3022, 1997.

[3] Martti Vainio, Hansjörg Mixdorff, and Juhani Järvikivi, "Perception and production of focus in Finnish," in *Proceedings of the 15th International Congress of Phonetic Sciences*, M. J. Solé, D. Recasens, and J. Romero, Eds., Barcelona, Spain, 2003, pp. 1831–1834.

[4] Martti Vainio and Juhani Järvikivi, "Tonal features, intensity, and word order in the perception of prominence," In press: *Journal of Phonetics*.

[5] Kari Suomi, Juhani Toivanen, and Riikka Ylitalo, "Durational and tonal correlates of accent in Finnish," *Journal of Phonetics*, vol. 31, pp. 113–138, 2003.

[6] R. Van Valin and R. La Polla, *Syntax: Structure, meaning and function*, Cambridge University Press, 1997.

[7] Auli Hakulinen and Fred Karlsson, *Nykysuomen lauseoppia [Finnish Syntax]*, Suomalaisen Kirjallisuuden Seura, Helsinki, 1979.

[8] Maria Vilkuna, *Free Word Order in Finnish: Its Syntax and Discourse Functions*, Suomalaisen Kirjallisuuden Seura, Helsinki, 1989.

[9] Frank E. Harrell, *Regression modeling strategies*, Springer, 2001.

[10] Martti Vainio, Juhani Järvikivi, and Daniel Aalto, "Quantity and tone in Finnish lexically stressed syllables," In proceedings of TAL 2006, La Rochelle, France, 2006.

[11] Yi Xu, "Effects of tone and focus on the formation and alignment of f0 contours," *Journal of Phonetics*, vol. 27, no. 1, pp. 55–105, 1999.

[12] Yi Xu and Ching X. Xu, "Phonetic realization of focus in english declarative intonation," *Journal of Phonetics*, vol. 33, no. 2, pp. 159–197, 2005.