Using Prosogram to study final rises in South Swedish: Implications for the Scandinavian tone accent typology

Sara Myrberg
Centre for Languages and Literature, Lund University
sara.myrberg@nordlund.lu.se

Abstract – This paper presents a production study investigating the frequency of utterance-final rises in South Swedish. These rises are interesting from a dialect typological perspective. Recent findings suggest that South Swedish has two-peak accents utterance medially, whereas previous typologies have assumed only one-peak accents for South Swedish dialects. The South Swedish two-peak accents are similar to accents described from East Norwegian and West Swedish, and since these dialects are characterized by final rises in addition to their two-peak accents, better knowledge of final rises in South Swedish can be expected to contribute to the understanding of the South Swedish two-peakedness. Nine participants from three locations in South Sweden (Laholm, Hässloholm, Sjöbo) are recorded reading 1080 sentences. The f0 in the utterance final syllables are analyzed using Prosogram. The results indicate that the number of final rises increases in the more northerly locations, supporting a successive transition from one- to two-peak dialects rather than a binary distinction as predicted by previous work.

1. INTRODUCTION

In the traditional tone accent typology pf Scandinavia, dialects have been divided into two types, referred to as one-peaked and two-peaked dialects [3, 4, 8, 9, 18, 21]. The one- versus two-peakedness has been discussed in terms of a binary distinction, where dialects belong strictly to either one or the other group [3, 21] and South Swedish has been unanimously classified as a one-peak dialect.

One- and two-peaked dialects have been said to have different ways of expressing prominence. One-peak dialects have been said to use scaling of their single peak to boost accents that need to be extra prominent, e.g., because they signal information structural focus. Two-peak dialects, on the other hand, have been hypothesized to use the two-peak contour in accents that are prominent, whereas in less prominent accents, only the leftmost peak is realized, giving rise to two categorically distinct prominence contours [2, 3, 8, 15, 17, 21] (but cf. [12] for an alternative analysis of East Norwegian).

Recently, however, [16], described two-peak contours in South Swedish, in utterance-medial position. Earlier dialect typologies predict that such contours should not exist in South Sweden.

Two peak-contours have previously been described from Central and North Swedish, West Swedish and East Norwegian [2, 3, 5, 6, 7, 12, 15, 17, 21, 22]. These two-peak contours have been divided into two types: a late second peak that can appear outside the accentuated word, as illustrated in (1a), and an early second peak that appears directly after the first peak as in (1b).

(1) Two-peak contours in East Norwegian, Central Swedish and South Swedish. Representations adapted from Myrberg (to appear).

a. East Norwegian two-peak contour (late 2nd peak)

H*H
["Jen ta"]focus "vill"e "kunne" "cykla."
"The girl would be able to ride a bicycle."

c. South Swedish two-peak contour (late 2nd peak)

L*HL
["Flickan"]focus "vill"e "kunne" "cykla."
"The girl would be able to ride a bicycle."

In [16] it is shown that the two-peak contours in South Sweden exhibit drifting as in (1c), making them similar to the East Norwegian and West Swedish contours in (1a). In [16], therefore, it was therefore hypothesized that a) South Swedish is more closely related to West Swedish and East Norwegian than previous work has assumed, and b) two-peakedness is a gradient feature of the Scandinavian dialects rather than a binary one, so that a dialect cannot always be categorized into one or the other group, but may be best described as somewhere in between one- and two-peaked.

It is a previously well-described fact that East-Norwegian as well as West Swedish utterances often end with a rising
tone, in statements as well as questions, whereas Central Swedish questions and statements more often end with falling or low tone [2, 3, 5, 12, 11, 18, 21, 22].

The height of the final rise in East Norwegian reflects whether the utterance final word carries focus, so that a high final rise signals utterance-final focus, whereas a low rise does not. Because of this, the utterance-final rises in East Norwegian have been described as the second peak of two-peak contours [3, 5, 12, 21].

In South Swedish, the correlation between utterance-final rises and the two-peak accents appears to be less clear than in East Norwegian, and [16] therefore hypothesizes that South Swedish allows truncation in utterance-final position. As a result, some South Swedish speakers are predicted to have two-peak contours utterance medially, and one-peak contours utterance finally. However, a lot remains to be understood when it comes to the relation between two-peak accents and utterance final rises.

II. QUESTIONS AND HYPOTHESES

The experiment presented here investigates utterance-final contours in three South Swedish varieties, in order to increase the understanding of the relation between two-peakness and utterance-final rises in a dialect typological perspective. The questions in (2) are targeted.

(2) Questions

a. Do dialects spoken in more northerly locations of South Sweden (geographically closer to West Swedish) exhibit more rises than more southerly locations (geographically farther from West Swedish)?

b. Is there a correlation between information structural focus and the number of rises, such that utterance-final rises are more frequent when the final word carries information structural focus?

As for (2a), Previous work, e.g., [3, 21] suggest that a relatively sharp isogloss could be identified, which separates one-peak dialects (with primarily low boundary tones) from two-peak dialects (with primarily high boundary tones). Under this hypothesis, one-peak dialects should not exhibit a correlation between information structural focus and utterance final rises, whereas two-peak dialects should exhibit a strong such correlation. The hypothesis from [16] that there is a gradient distinction between one- and two-peak dialects, however, indicates that there should be no sharp isogloss, but rather that the transition from low to high boundary tones spans a large geographical area.

As for (2b), [16] hypothesized that South Swedish is similar to East Norwegian and West Swedish in terms of having late second peaks (like in 1a, 1c), but is different from East Norwegian and West Swedish in that South Swedish frequently truncates the second peak in utterance final position, and thus exhibits fewer final rises.

A secondary aim of this study is to explore the possibility of using Prosogram [13, 14] for automatic annotation of Swedish prosody.

III. EXPERIMENT DESIGN

A read production experiment was conducted with native speakers of South Swedish. The experiment had a three-by-two design, in which lexical tone accent (1 vs. 2), focus placement (non-finally vs. finally) and sentence type (question vs. statement) were controlled for. There were five items, read with three repetitions. Nine speakers were recorded, which resulted in a corpus of 1080 read sentences (2*2*5*3*9=1080).

The nine speakers came from three different locations in the southern part of Sweden: Three speakers came from Laholm (Northwestern recording site, closest to West Swedish), three speakers came from Hässleholm (Central recording site), and three speakers came from Sjöbo (Southern recording site, farthest from West Swedish), cf. Figure 1.

![Figure 1. Map of the three recording sites.](image)

Participants were seated in front of a laptop computer screen that displayed question plus answer sentences to the participants in randomized order and interspersed with fillers. Participants were asked to read all text displayed on the screen out loud. They were recorded with an AKG C520 head-mounted condenser microphone, connected to a Zoom H5 Handy recorder. Sample sentences are provided in (3) and (4).

(3) Sample item, tone accent 1 (the raised 1 indicates a lexically stressed syllable with tone accent 1, see [19], [20] for a definition of stress in Swedish).

a. Question that triggers focus on subject

Vad är Vendelas? what is Vendela

‘What belongs to Vendela?’
b. Question that triggers focus in verb phrase
Vems är den e1norma 1bollen?
whose is the huge.DEF ball.DEF
‘To whom does the huge ball belong?’

c. Answer sentence
Den e’norma 1bollen är 1Vendela.GEN
the huge.DEF ball.DEF is Vendela.GEN
‘The huge ball belongs to Vendela.’

(4) Sample item, tone accent 2 (the raised 2 indicates a lexically stressed syllable with tone accent 1).

a. Question that triggers focus on subject
Vad 2kostar 2tio 2kronor?
what costs the tasty.DEF bun.DEF
ten crowns
‘What costs ten crowns?’

b. Question that triggers focus in verb phrase
Vad 2kostar den 2goda 2bullen?
what costs the tasty.DEF bun.DEF
‘What does the tasty bun cost?’

c. Answer sentence
Den 2goda 2bullen 2kostar 2tio 2kronor.
the tasty.DEF bun.DEF costs the tasty.DEF bun.DEF
ten crowns
‘The tasty bun costs ten crowns.

IV. Annotation Procedure

The prosodic pattern of the final syllable of each target sentence was extracted using Prosogram [13, 14]. Prosogram is implemented as a Praat script [1], which attempts to create a stylization of the f0 contour that corresponds to human perception of pitch [13, 14]. Prosogram categorizes f0 movements in terms of flat, rising, or falling tone, based on a threshold for perception of pitch movements (glissando threshold). Three options for the glissando threshold are provided, and in the current analysis, the most sensitive of these options was used (cf., also [10]).

For the purposes of the current experiment, the Prosogram segmentation was based on segmentation of the speech into phonemes and syllables. This was carried out using an automatic aligner recently developed for Stockholm Swedish by [23]. Some amount of manual correction was necessary before the output from the aligner could be used as input to the Prosogram analysis.

A plot showing the output of the Prosogram analysis together with the output from the automatic aligner, after manual correction, for the target sentence in (3c) is provided in Figure 2.

Figure 2. Prosogram plot of one target sentence. Cf. Mertens (2020).

V. Results and Discussion

Beginning with the question from (2a), the results show that the speakers from the three recording sites exhibit different frequency of rises, in questions as well as statements, as seen in Figures 3 and 4. Laholm speakers have the most rises, followed by Hässleholm speakers. Sjöbo speakers have the smallest number of rises in both questions and statements. As can be expected, questions exhibit a higher degree of final rises than statements.

This indicates that the frequency of final rises varies between South Swedish dialects and increases successively in geographical areas that are closer to West Swedish dialects. Assuming a connection between the frequency of final rises and two-peakedness in a dialect, as in the East Norwegian literature [5, 6, 7, 12] this result can be said to support the idea that the distinction between one- and two-peaked dialects should be seen as a scale on which dialects place themselves, rather than a binary distinction between two dialect types.

As for the question from (2b), the pattern is more complex. In the answer sentences (3c, 4c, Figure 3), the utterance-final word carries the nuclear accent in the focus condition (following the questions in 3a and 4a), and a postfocal accent in the given condition (following the questions in 3b and 4b). Under the assumption that the type of two-peakedness found in South Swedish is similar to that in East Norwegian and West Swedish as in [16], a more or less strong correlation between final rises and information structural focus is expected. Such a correlation is found in Laholm but is not significant in Hässleholm or Sjöbo (Figure 3).

However, the interpretation of this result is complicated by the fact that the same pattern between the focus and given conditions is seen in the questions (3a-b, 4a-b, Figure 4) from Laholm as well as Hässleholm. In the questions, it is not the case that the nuclear accent is final in the focus condition and non-final in the given condition. Therefore, the hypothesis that the presence of a nuclear accent is the source of the larger number of rises in the focus condition does not apply to the questions in this dataset. Instead, the cause of this difference must be sought elsewhere. Future research will thus be necessary to further elucidate the function of final rises in South Swedish.
This paper has presented a production experiment with speakers from three locations in South Sweden, with the aim of better understanding the nature of utterance-final rises in South Sweden and their relation to the two-peaked contours that have recently been described in utterance-medial accents of some South Swedish speakers [16]. The results show that speakers from the southmost recording location (Sjöbo) have fewer rises than speakers from the northmost recording location (Laholm). Assuming that the frequency of final rises correlates to some extent with where a dialect positions itself on a scale between one- and two-peekedness as outlined in [16], this result has been taken as an indication that the Sjöbo dialect is more one-peaked than the Laholm or Hässleholm dialects.

However, the results also suggest that there is no simple correlation between one- versus two-peekedness and the frequency of utterance final rising contours. The two different types of questions in this dataset exhibited different prevalence for rising tones in Laholm and Hässleholm. While this may not be surprising in a wider typological perspective, question intonation has, to my knowledge, not been studied in detail for the Scandinavian dialects (some of which primarily have falling contours in questions). Future research will thus be necessary to reach a fuller understanding of the function of utterance-final rises in Scandinavian dialects, in relation to information structural focus and other pragmatic and information structural categories.

ACKNOWLEDGEMENTS

I want to express my sincere gratitude to Anders Myrberg for performing the programming tasks necessary to conduct the present study, to Nathan Young for sharing his automatic aligner with me, to Mattias Heldner and Shin Ishihara for helpful discussion, and to Agnes Wallner for doing recordings. This research was funded by The Swedish Research Council, grant number 2016-01536.

REFERENCES


