Focal F0 peak shape and sentence mode in Swedish

Ambrazaitis, Gilbert; Buanzur, Tuarik C.; Niebuhr, Oliver

Go to publication entry in University of Southern Denmark's Research Portal

FOCAL F0 PEAK SHAPE AND SENTENCE MODE IN SWEDISH

Gilbert Ambrazaitis[1], Tuarik C. Buanzur[2], Oliver Niebuhr[2,3]

[1]Centre for Languages and Literature, Lund University, Sweden; [2]Department of General Linguistics, ISFAS, Kiel University, Germany; [3]Dept. of Design and Communication, IRCA, University of Southern Denmark
Gilbert.Ambrazaitis@ling.lu.se, stu102801@mail.uni-kiel.de, olni@sdu.dk

ABSTRACT

Shape characteristics of rising-falling accentual F0 peaks of Stockholm Swedish Accent I words in narrow focus are studied in a corpus of 287 read sentences. The corpus includes statements and three types of polar questions. Results reveal a clear effect of sentence mode on the shape of the accentual rises: Statements are predominantly characterized by convex rises, questions by concave rises.

Keywords: prosody, question, convex, gender

1. INTRODUCTION

1.1. Questioning 'question intonation'

A conception called 'question intonation' has been postulated for many languages [13], commonly understood as one pole of a binary intonational distinction such as 'rising' vs. 'falling'. For instance, a falling intonation has been associated with statements and wh-questions in English and German, while word-order questions and syntactically unmarked questions have been claimed to rise at the end [12, 30]. These claims are challenged by a number of empirical studies showing that both rises *and* falls occur in different question types (e.g., for German [15, 16, 27]). Similarly, in an analysis of 200 wh-questions from spontaneous Stockholm Swedish, a considerable proportion of tokens (22%) showed a final rise instead of the expected fall [14].

Such inconsistencies led [14] and [15] to conclude, in accord with recent findings on English, Dutch [6], and Greek [2], that the rise—fall distinction is in the first place a matter of speaker attitudes or other social interaction signals, which just have a natural relationship with the functional categories of sentence mode. General support and some further refinement of this conclusion is provided by a recent investigation of 600 questions from spontaneous Swedish [28], suggesting that rising vs. falling pitch is determined by the decision whether the question is backward- or forward-looking in dialogue.

1.2. Multi-dimensional prosodic encoding of questions

In fact, the real indicators of sentence mode could be prosodic parameters, which, unlike the direction of the final pitch movement, come into play well before the end of the sentence [17, 18, 25]. For instance, a higher speech rate in questions compared to statements was found for German [20, 22], Cantonese, Neapolitan Italian, and Malaysian [7, 23, 29]. Sentence mode also involves differences in voice quality [22] and the number of prenuclear accents [25].

Recent evidence for German and Italian [24, 25] points to the *shape* of accentual F0 contours as a further correlate of sentence mode. More specifically, statements were found to have convex rise contours and concave falls, while questions are realized with concave rises and convex falls.

1.3. Stockholm Swedish

Swedish is well known for its lexical pitch accent distinction (Accent I vs. II). In the present study, only Accent I is involved, which can be modeled as a (H)L* pitch accent [5, 9]. Adding a focal accent (H-) and a low boundary tone results in a rising-falling L*H-L% sequence (see also [26]).

Final rises in Swedish are considered an optional add-on rather than an essential feature of question intonation [10, 11, 14]. Statements are associated with a falling baseline and a falling top-line of a tonal grid, while questions are associated with a raised topline and a widened F0 range [10, 11].

1.4. Research question and hypothesis

Peak shape characteristics have been discussed for Swedish [1], but neither strictly with respect to questions nor in terms of a systematic classification. Therefore, this study investigates if sentence mode is systematically linked with shape distinctions in Swedish. With reference to [24, 25], we test the following hypothesis: The rise of the rising-falling F0 peak in focal Accent I (L*H-L%) is predominantly characterized by a convex shape in statements and a concave shape in questions, while the fall is concave in statements and convex in questions.

2. METHODS

2.1. Speech materials and elicitation procedure

Four types of sentences are examined: (1) statements (henceforth, ASS = 'assertion'), (2) word-order ques-

tions (**QUES**), (3) echo questions (**QREP** = 'repetition question'), and a further type referred to as (4) 'disbelieving questions' (**QDIS**). All three types of questions are polar questions. Questions of the type **QDIS** are similar to those of the type **QREP**, but unlike the latter, **QDIS** questions involve an expression of the speaker's attitude.

Our corpus consists of read sentences, based on two target words – *vinet* 'the wine' and *bilen* 'the car' (both Accent I) – which were embedded in two similar variants of the same target sentence:

(1) A. Target sentence for ASS, QDIS, QREP
    De hade vinet i bilen./?
    'They had the wine inside the car./?'
    B. Target sentence for QUES:
    Hade de vinet i bilen?
    'Did they have the wine inside the car?'

The target sentence was elicited in 8 experimental conditions: 4 sentence types were cross-combined with 2 narrow-focus placements, i.e. a sentence-final focal accent on *bilen*, and a sentence-medial focal accents on *vinet*. This enlarges the database and supports the generalization of the findings.

The target sentence was presented to participants on a computer screen. *Statement* conditions (ASS) were elicited by a combination of a written situational context and a pre-recorded context question, played through loudspeakers. ASS' situational context was *Du pratar med en kompis om en tidningsartikel* 'You're talking to a friend about a newspaper article'. Focus was controlled by means of the auditory context question, see (2) for an example.

(2) Context question for medial focus in ASS:
    Vad var det nu som bovarna hade i bilen?
    'What did the villains have inside the car?'

The *question* conditions were elicited using a written context only. The context was displayed on the computer screen together with the target sentence. The question contexts included both explanations on the situational framework eliciting the type of question, and a semantic indication as to which word is to be put in focus. An example is shown in (3). Only the English translation is given for reasons of space. Bold face indicates the text to be read aloud.

(3) Context for QREP, medial focus:
    Your friend is telling you that some thieves had hidden something in the car. He also mentioned what it was, but you're not sure if you got it right. You want to be sure that it really was the wine that was hidden in the car: **Pardon? They had the wine inside the car?**

The task of the participants was to read the target sentences as naturally as possible given the provided written and/or auditory contexts. Three repetitions of all sentences were read by each speaker. Given twelve speakers (see 2.2), the corpus investigated in this study consists of 287 sentences (8 conditions, 3 repetitions, 12 speakers; 1 missing recording due to a technical problem).

2.2. Speakers and recording procedure

Twelve speakers were recorded for this study: six females, aged between 21 and 69 years (average age 35 years), and six males between 20 and 71 years (average age 42 years). All speakers represent the Standard variety of Swedish, as spoken in the urban area of Stockholm. Apart from one male speaker (M06), who was recorded at the first author's home, all speakers were recorded in an experimental studio at the Humanities Laboratory of Lund University. All recordings were made using a boundary microphone (IMG Stage Line ECM-302B) and a USB mic pre-amplifier (M Audio Duo) connected to a laptop computer outside the recording studio. Recordings were digitized at 44.1 kHz and 24 bit.

Our target sentences were elicited in combination with other unrelated speech material, which can be regarded as distractor material from the perspective of the present study. For this reason, a complete recording session took about 90-120 minutes per speaker, including instructions, and, in most of the sessions, two short breaks.
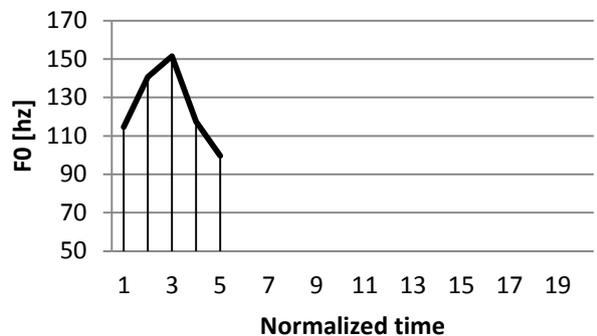
2.3. Classification of peak shapes

All focused words (i.e., *vinet* for medial focus, and *bilen* for final focus) were segmented into two syllables using Praat [4]. For each syllable, F0 was time-normalized by extracting 10 temporally equidistant F0 measurements using the Praat script ProsodyPro [31]. An example of a time-normalized F0 curve for an entire word (two syllables = 20 measurements) is plotted in Figure 1.

Based on all 287 representations of that kind, simplified versions of the plots were created by reducing the original plots to five F0 points, see Figure 2: (i) rise onset, (ii) peak maximum, (iii) fall offset, as well as two further intermediate points, one of them (iv) half-way in between rise onset and peak maximum, and the other one (v) half-way in between peak maximum and fall offset. Note that 'half-way' refers to the number of measurement points in the normalized time domain. For example, if points (i)-(iii) were determined at time points 3, 9, 17, then (iv) and (v) were defined as the F0 values at time points 6 and 13.

Figure 1: Time-normalized F0 course of a focally accented word (two syllables) with a concave rise.
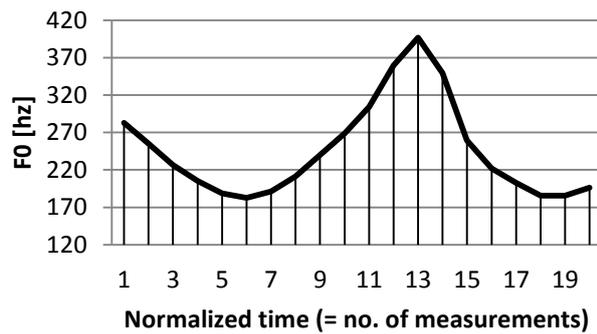
Figure 2: Simplified time-normalized F0 course of a focally accented word (two syllables) with a convex rise and a concave fall.

By means of visual classification of these simplified plots, the shapes of the rising and the falling parts of the accent contour were classified as either convex or concave (or undefined). A similar methodology has been applied in [8] in connection with phrase-final intonation patterns.

3. RESULTS

Table 1 summarizes the distribution of all four possible rising-falling peak shapes (such as concave rise + convex fall) among the different sentence types. Frequencies are pooled across all focus conditions and speakers. Based on this table, a Chi-squared test was conducted, in order to test whether the independent variable Sentence Type had a significant effect on the frequency of occurrence of the four different peak shapes. Two further Chi-squared tests determined possible effects of the independent variables Focus Condition and Speaker Gender.

Table 1 reveals a clear effect of Sentence Type, and the Chi-squared test confirms the significance of this effect (p<.01): The two peak shapes 'convex-concave' and 'convex-convex' are frequent in ASS, but infrequent in questions. By contrast, concave-concave peaks are frequent in questions, but infrequent in ASS. The peak shape 'concave-convex'

is the only one that occurs about equally often in both ASS and in questions.

Broken down to individual F0 slopes (i.e., rises and falls), Table 1 shows that *convex rises* occur by majority (17+20=47 of 62 or 76%) in ASS, while in all three types of question, *concave* rises are more frequent than *convex* rises (69-73% of the cases in Tab. 1). By contrast, the choice between convex and concave *falls* was hardly influenced by sentence mode. In the ASS condition, about two thirds of all falls took a convex shape. This bias towards convexity remained similar across all question types.

Table 1: Distribution of F0 peak shapes (e.g., cvx-ccv = convex rise plus concave fall) across the four sentence type conditions. Maximum total per line is 72 (2 focus positions x 3 repetitions x 12 speakers); missing cases occur due to rise or fall shapes classified as undefined.

S-Type	ccv-cvx	cvx-ccv	ccv-ccv	cvx-cvx	Total
ASS	22	17	3	20	62
QDIS	27	9	18	7	61
QREP	25	6	18	10	59
QUES	26	7	15	11	59

Figure 3: Distribution of contour shapes of the focal rises across the two focus conditions (medial, final) and four sentence types; 18 tokens per conditions. (a) female, (b) male speakers.

(a) Female speakers
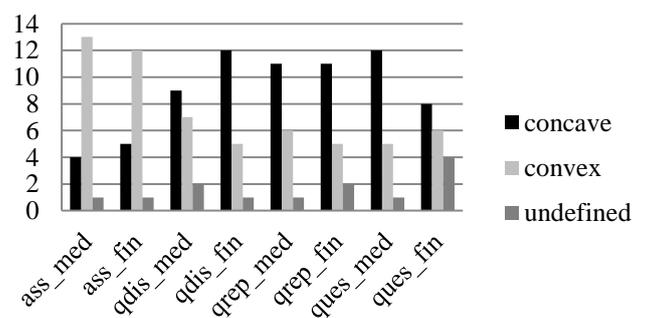
(b) Male speakers
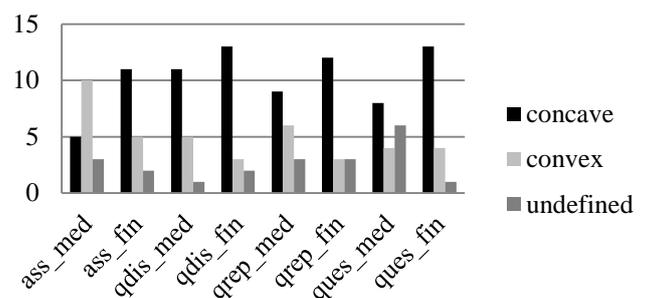
The results yielded no significant effect of Focus Condition (p=.1571), thus suggesting that the effects

of sentence type just described apply to the medial word *vinet* as well as to the final word *bilen*.

Figure 3 offers a more detailed account of the distribution of the *rising* contours, divided into female and male speakers. It displays, once again, that concave rises are preferred in all three question types (and across both focus conditions), whereas convex falls are preferred in ASS. However, the figure also reveals an exception from this rule: statements with final focus of *male* speakers.

That is, there is an effect of Gender on the distribution of the peak shapes, which proved marginally significant (p=.0615) in the corresponding Chi-squared test: female speakers used convex rises more frequently than male speakers.

In summary, speakers preferred concave over convex focal rises in questions, and vice versa in statements. However, male speakers have applied this peak shape distinction first and foremost in sentence-medial target words, while in final focal accents, they have preferred concave over convex rises, irrespective of sentence mode.

4. DISCUSSION AND OUTLOOK

The results show that there is systematic variation in the F0 peak shape of focal accents, and suggest that this variation plays a role in the prosodic encoding of sentence mode in Stockholm Swedish: *Convex focal rises* were clearly preferred in statements, *concave focal rises* in polar questions. No differences of that kind were observed between the three types of polar question involved in this study, which supports our conclusion that focal peak shape is a correlate of sentence mode proper, rather than of, for example, speaker attitude (see 1.1., 1.2.). Nevertheless, further question types, not least wh-questions, should be tested for peak shape effects.

In addition, the results suggest a gender-specific behavior in the application of peak shape variation. Gender effects in the prosodic encoding of questions have also been reported for German [21] and Greek [3], which are, however, not directly comparable to those in the present study. Future research should explicitly address various kinds of gender effects. Are they physiological in nature, or rather social?

Results of a complementary study (in progress) show that, compared with statements, questions are also characterized by a higher F0 maximum and a larger F0 range in the nuclear section of the sentence (i.e., the section comprising the focal accent). That is, in addition to the present finding that questions have concave rises, these concave rises also reach higher than their convex counterparts in statements, suggesting that the shape of the rise is part of a bundle of multiple F0 features. However, whether or not one of these features is more important for sentence mode identification still has to be determined.

Our hypothesis was confirmed with respect to the rising part (L*H-), but not for the falling part of the focal L*H-L% pattern. With reference to [24, 25], convex falls were expected for questions and concave falls for statements, but convex falls dominated across all sentence conditions.

This result could indicate that it is primarily the rise of the focal or nuclear accent which is subject to sentence-mode related shape variation. Variation in the shape of the fall could be optional, or a (bio-mechanical) epi-phenomenon of the preceding shape variation in the rising slope; or it could even be related to the signalling of other communicative functions. In the case of our medial focal accents, the fall could indeed be associated with either of two different functions, which we did not control for: Assuming that a L% boundary tone is optional after medial focus [19], the fall could in some of our recordings represent a L% boundary tone, while in others, it could be identical with the word-accent fall of the final target word (L*H- HL*).

Note that our assumption may even hold cross-linguistically and is not inconsistent with the fact that previous studies on Neapolitan Italian and German [24, 25] found sentence-mode related shape variation for falling accent slopes. These falls were associated with prenuclear accents, whereas the present study addressed nuclear accents. That is, the factor prenuclear vs. nuclear may also determine which slope of the accent peak is involved in signalling sentence mode. This sounds even more reasonable, if we take into account that the final *fall* of nuclear accents already has quite a high functional load, and a similar assumption can be made for the prenuclear *rise*.

So, do we find peak shape variation in *non-focal* (prenuclear) accent contours in Swedish, where – at least in the case of Accent II – lexical tone comes into play [5, 26]? Can we also find peak shape variation in *focal* Accent II-words, where the focal accent is realized in a post-stress syllable? Is it the focal accent or the stress-aligned lexical tone in Accent II that varies with respect to sentence mode? These basic questions show that we have only just begun to discover the relevance of F0 peak shape variation in communication. Progress in methodology and automatic F0 processing will be of great help in dealing with this level of phonetic detail in the future.

5. ACKNOWLEDGEMENTS

6. REFERENCES

[1] Ambrazaitis, G., Frid, J. 2014. F0 peak timing, height, and shape as independent features. *Proc. 4th International Symposium on Tonal Aspects of Languages*, Nijmegen, 138-142.

[2] Arvaniti, A., Baltazani, M., Gryllia, S. 2014. The pragmatic interpretation of intonation in Greek wh-questions. *Proc. 7th Speech Prosody,* Dublin, 1144-1148.

[3] Arvaniti, A., Ladd, D.R. 2009. Greek wh-questions and the phonology of intonation. *Phonology* 26, 43-74.

[4] Boersma, P., Weenink, D. 2014. *Praat: doing phonetics by computer [Computer program].* Version 5.3.49 http://www.praat.org/ (accessed 13 May 2013)

[5] Bruce, G. 1977. *Swedish Word Accents in Sentence Perspective*. Lund: Gleerup.

[6] Chen, A., Rietveld, T., Gussenhoven, C. 2001. Language-specific effects of pitch range on the perception of universal intonational meaning. *Proc. Eurospeech 2001,* Aalborg, 1403–1406.

[7] D'Imperio, M. 2000. *The Role of Perception in Defining Tonal Targets and their Alignment*. Ph.D. Dissertation, The Ohio State University, OH, USA.

[8] Dombrowski, E., Niebuhr, O. 2010. Shaping phrase-final rising intonation in German. *Proc. 5th Speech Prosody,* Chicago, 1-4.

[9] Engstrand, O. 1995. Phonetic interpretation of the word accent contrast in Swedish. *Phonetica* 52, 171–179.

[10] Gårding, E. 1979. Sentence intonation in Swedish. *Phonetica* 36, 207-215.

[11] Gårding, E. 1989. Intonation in Swedish. *Working Papers, Lund University, Department of Linguistics* 35, 63-88.

[12] Halliday, M.A.K. 1970. *A Course of Spoken English: Intonation*. London: Oxford University Press.

[13] Hirst, D.J., Di Cristo, A. 1998. *Intonation Systems. A survey of Twenty Languages*. Cambridge: Cambridge University Press.

[14] House, D. 2005. Phrase-final rises as a prosodic feature in wh-questions in Swedish human–machine dialogue. *Speech Communication* 46, 268-283.

[15] Kohler, K. 2004. Pragmatic and attitudinal meanings of pitch patterns in German syntactically marked questions. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung (AIPUK)* 35a, 125-142.

[16] Kügler, F. 2014. Do we know the answer? Variation in yes-no-question intonation. *Experimental studies in linguistics* 1, 9-29.

[17] Liu, F., Surendran, D., Xu, Y. 2006. Classification of statement and question intonations in Mandarin. *Proc. 3rd Speech Prosody,* Dresden, 603-606.

[18] Liu, F., Xu, Y. 2007. Question intonation as affected by word stress and focus in English. *Proc. 16th ICPhS,* Saarbrücken, 1189-1192.

[19] Myrberg, S. 2013. Focus type effects on focal accents and boundary tones. *Proc.26th Annual Swedish Phonetics Conference (FONETIK 2013)*, Linköping, 53-56.

[20] Niebuhr, O. 2012. Das ist (k)eine Frage – Phonetische Merkmale in der Identifikation standarddeutscher Deklarativfragen. In Anderwald, L. (ed.), *Sprachmythen – Fiktion oder Wirklichkeit.* Frankfurt/New York: Peter Lang, 203-222.

[21] Niebuhr, O. Gender differences in the prosody of German questions. *Proc. 18th ICPhS*, Glasgow.

[22] Niebuhr, O., Bergherr, J., Huth, S., Lill, C., Neuschulz, J. 2010. Intonationsfragen hinterfragt - Die Vielschichtigkeit der prosodischen Unterschiede zwischen Aussage- und Fragesätzen mit deklarativer Syntax. *Zeitschrift für Dialektologie und Linguistik* 77, 304-346.

[23] Petrone, C. 2008. *Le rôle de la variabilité phonetique dans la représentation des contours intonatifs et de leur sense*. Ph.D. Dissertation, Université de Provence, France.

[24] Petrone, C., D'Imperio, M. 2011. From tones to tunes: Effects of the f0 prenuclear region in the perception of Neapolitan statements and questions. In Frota, S., Elordieta, G, Prieto, P. (eds.), *Prosodic categories: production, perception and comprehension*. Berlin: Springer, 207-230.

[25] Petrone, C., Niebuhr, O. 2013. On the intonation in German intonation questions: The role of the prenuclear region. *Language and Speech* 57, 108-146.

[26] Riad, T. 2006. Scandinavian accent typology. *Sprachtypol. Univ. Forsch. (STUF)* 59, 36-55.

[27] Selting, M. 1995. *Prosodie im Gespräch. Aspekte einer interaktionalen Phonologie der Konversation*. Tübingen: Niemeyer.

[28] Strömbergsson, S., Edlund, J., House, D. 2012. Prosodic measurements and question types in the Spontal corpus of Swedish dialogues. *Proc. Interspeech 2012,* Portland, Oregon, 839-842.

[29] Van Heuven, V., van Zanten, E. 2005. Speech rate as a secondary prosodic characteristic of polarity questions in three languages. *Speech Communication* 47, 87-99.

[30] von Essen, O. 1966. *Allgemeine und angewandte Phonetik*, 4th edn. Berlin: Akademie Verlag.

[31] Xu, Y. 2013. ProsodyPro — A tool for large-scale systematic prosody analysis. *Proc. Tools and Resources for the Analysis of Speech Prosody (TRASP 2013),* Aix-en-Provence, 7-10.