# QUESTION INTONATION IN NON-SCRIPTED DANISH DIALOGUES

*Nina Grønnum and John Tøndering**

Linguistics Laboratory, Department of Scandinavian Studies and Linguistics,
University of Copenhagen, Denmark
{ninag,johnt}@hum.ku.dk

## ABSTRACT

Global intonation contour slopes in Danish read speech have been found to vary systematically according to utterance type. Statements have the steepest gradients, wh-question contours are slightly less steep, questions with word order inversion less steeply falling again, and declarative questions have no gradient at all, i.e. their global contour is level. Furthermore, in all but the very shortest utterances onset and offset of the global intonation contour appear relatively constant in the frequency range across varying utterance length. This paper is a first exploration into Danish question intonation in non-scripted speech.

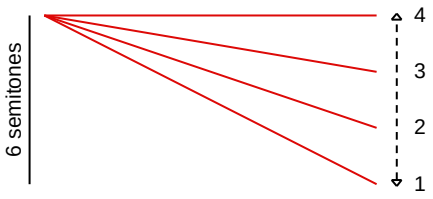
**Keywords:** Intonation; questions; Danish; non-scripted dialogues.

## 1. INTRODUCTION

Only the briefest summary of Standard Danish intonation is presented here for background. The reader is referred to [2] and her references for details.

Danish intonation is a hierarchical structure where components of larger structural/temporal scope carry—and scale the manifestation of—components of lesser structural/temporal scope. Thus, the stress group pattern—i.e. the fundamental frequency pattern associated with a stressed and all succeeding unstressed syllables—sits upon, and is subject to variation determined by, the prosodic phrase contour; in its turn the prosodic phrase contour rides upon and is constrained by the more global utterance contour. Furthermore, phrase and utterance contours are characterized solely by their stressed syllables. Two facts are responsible for this state of affairs: (1) The local pitch movements (rise-falls) depend for their existence upon the presence of unstressed syllables in the stress group: in a succession of stressed syllables—with no intervening unstressed ones—there is no upwards deflection of the $F_0$ course between them. (2) The stressed syllables appear to be frequency scaled in relation only to each other, without regard to the presence or not of any local 'highs' between two stressed 'lows.' In other words, the local deflections have no independent role in the shaping of grosser trends in phrase and utterance contours.

**Figure 1:** Intonation contour slopes in: declaratives (1); wh-questions (2); questions with word order inversion (3); declarative questions (4)




For a given intonation type the range spanned by the contour is constant: In utterances of up to five stress groups (above which an utterance is typically decomposed into a succession of prosodic phrases), the stressed syllables are equidistantly spaced in frequency, in frequency intervals which are inversely proportional to their number. In other words, longer utterances exhibit smaller frequency intervals between the stressed syllables and accordingly less steep gradients than shorter ones, ceteris paribus.

The cue to utterance modality lies, not in any local final pitch movement or any final high or low tone, but in the global development of the utterance contour from the onset of its first stressed syllable to the offset of the last stressed syllable. For a given utterance length, this global contour is most steeply declining in statements, and level in declarative questions, i.e. questions without any overt lexical or syntactic cues to their interrogative function. In between—and with a certain degree of variability—are found wh-questions and questions with word order inversion, cf. Fig. 1. Accordingly, the end of the contour (as defined by its last stressed syllable) is lowest in statements, next come wh-questions, then questions with word order inversion and at the top we find declarative questions with their level contours. This state of affairs may be seen as a trade-off between syntax and prosody: the less syntactic information an utterance contains about its interrogative function, the stronger are the prosodic cues (the

*Authors in alphabetical order.

less statement-like is the global contour).

All of this holds for read speech materials. But what about non-scripted speech?

## 2. MATERIAL

From the map task dialogues in the publicly available DanPASS corpus (http://www.danpass.dk), we chose six pairs of speakers who navigate through four sets of maps, i.e. 24 dialogues, lasting a total of 3 hours and 15 minutes, including pauses. An utterance was classified as a question if it evoked a reaction from the partner, either in terms of a confirmation or a negation, and/or a clarification/expansion/explanation. This is a fairly broad understanding of the concept 'question' and one with obviously fuzzy boundaries and accordingly a few indeterminate cases inevitably exist. The numerous utterances which evoked a reaction, typically a *yes*, which we took to mean *I understand, I follow you* etc., were not, of course, classified as questions. 444 utterances fell into our question category. 71 of those contained syllables with extra prominence (emphasis). Extra prominence disrupts the smooth declination of the phrase or utterance contour, i.e. it introduces an uncertainty in the determination of the contour slope. Accordingly, these questions were excluded from further analysis. The remaining 373 were further reduced by the exclusion of a variety of utterances which occurred in insufficient numbers to allow us to establish any trend in their behaviour. Thus, some questions with tags (*xxx, right?*), a few more or less complete polarity questions (*xxx or ....? xxx or what?*), a few long questions whose contours were broken down into two or three prosodic phrases, and a couple of other marginal instances.

**Table 1:** Number of questions of each type against number of stressed syllables. See text

| stress syll | 1 | 2 | 3 | 4 | 5 | 5-10 |
|---|---|---|---|---|---|---|
| decl_quest-a | **6** | **36** | **29** | **8** | **8** | 11 |
| word ord inv | **13** | **51** | **28** | **10** | **8** | 4 |
| wh-quest | 4 | **19** | **11** | **10** | 2 | 1 |
| decl_quest-b | 2 | **13** | **10** | **15** | 1 | |
| total | 25 | 119 | 78 | 43 | 19 | 16 |

300 questions remained for analysis: 114 have word order inversion, 47 are wh-questions, and the remaining 139 have no lexical or syntactic markers of their interrogative function, they are declarative questions. Note that this category is subdivided in Table 1; explanation follows in section 4. The majority of the questions analysed contain from 1 to 5 stressed syllables, 2 syllables being the absolutely most frequent. But note that there are relatively more short questions with word order inversion and relative more long declarative questions. Numbers in bold face in Table 1 are those whose average contours are depicted in Fig. 2.

## 3. DATA

The intonation contours are represented symbolically in the corpus, cf. [3]. The pitch of the stressed syllables was annotated on a coarse scale of high (*h*), mid (*m*) or low (*l*). A finer gradation exists, however, within a succession of stressed syllables within a given range (between high and mid, high and low, or mid and low). Thus, e.g., *h_>_>_m* designates a succession of four stressed syllables descending gradually from high to mid.

We did not want a priori to exclude the possibility that the movement from stressed to post-tonic in the last stress group might have more of a role to play in non-scripted speech than in read speech, where it is always rising and where its magnitude follows from the contour it rides upon: it is larger on contours that terminate higher up in the range. In the dialogues, falls were however also encountered, constituting about 20% of all final movements, but not in any readily apparent correlation with utterance type, except that there are hardly any final falls among the wh-questions. The level of the post-tonic relative to the preceding stressed syllable is graded in the corpus in 7 steps from the highest rise, *H/*, through *H*, *h*, to equal pitch, =, and through *l*, *L* to the lowest fall, *L*. Given the approximately even distribution of falls across utterance types, and given the fact that rises are by far the most common, we have simply calculated averages over all post-tonics per length per type. Granted the proportion of rises vs. falls, the average movement of course comes out as a rise and appears in Fig. 2 as a thinly dotted line. Note, however, that approximately one half of the utterances ended without a post-tonic at all. Among the 326 utterances in Fig. 2, this is true of 173 items.

The symbolic representation was converted to pseudo-interval scales: *h*, *m*, and *l* were assigned the (arbitrary) values 3, 2, and 1, respectively, with interpolation of the finer gradations within an interval. E.g., the *h_>_>_m* sequence above translates into 3_2.667_2.333_2. *H/*, *H*, *h*, =, *l*, *L*, *L* were assigned the equally arbitrary values 1.5, 1.0, 0.5, 0, –0.5, –1.0, –1.5 added to the value of the preceding stressed syllable. E.g., *m_H/* translates into 2_3.5.

## 4. RESULTS

If speaker strategy carries over from read to non-scripted speech, we shall expect wh-questions to

**Table 2:** Onset, offset and difference between onset and offset of the contours depicted in Fig. 2. See text

| stress syll | 2 | | | 3 | | | 4 | | | 5 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | on | off | grad | on | off | grad | on | off | grad | on | off | grad |
| decl_quest-a | 2.4 | 2.1 | 0.3 | 2.2 | 1.9 | 0.2 | 2.6 | 2.0 | 0.2 | 2.5 | 1.9 | 0.2 |
| word ord inv | 2.4 | 2.4 | 0.0 | 2.2 | 2.3 | 0.0 | 2.4 | 2.3 | 0.0 | 2.5 | 2.3 | 0.1 |
| wh-quest | 2.2 | 1.7 | 0.5 | 2.4 | 1.9 | 0.2 | 2.1 | 1.3 | 0.3 | | | |
| decl_quest-b | 2.2 | 1.2 | 1.1 | 2.4 | 1.2 | 0.6 | 2.4 | 1.1 | 0.4 | | | |
| declar | 2.3 | 1.4 | 0.9 | 2.7 | 1.5 | 0.6 | 2.5 | 1.2 | 0.5 | 2.7 | 1.1 | 0.4 |

have the steepest contours and offset furthest down in the frequency range, questions with word order inversion should decline less, i.e. offset somewhat higher up, and declarative questions should have high level contours.

Given the theoretically dubious nature of calculating an average intonation contour (for each type of question at each length) from values on a pseudo-interval scale, cf. above, we exhibit average contours only in so far as there are at least 5 observations behind the average. Accordingly, there are no questions longer than 5 stressed syllables in Fig. 2, cf. the bold faced cells in Table 1.

In the course of classifying utterances in the 24 dialogues we noticed that certain utterances, which—according to our criteria—functioned as questions, were indistinguishable from statements when listened to in isolation. Those are the 38 "decl_quest-b" utterances in Fig. 2 and Tables 1 and 2. We return to them in 4.1.2. To bring out differences or similarities with utterances which functioned as statements in the dialogues we selected, at random, 51 such statements, around 10 at each length from 1 to 5 stressed syllables, denoted "declar" in Fig. 2 and Table 2, bottom panel/row.

The contours in Fig. 2 are lined up to the last stressed syllable, to bring out the difference in the location of the offset of the contour and in the rise to the post-tonic in the last stress group.

Table 2 lists utterance onsets, offsets and the interval between the two. A pseudo-gradient results by dividing the pitch interval with the number of stressed syllables minus one.

If we keep in mind that the variability in the data is very considerable, we can nevertheless distill certain trends from Fig. 2 and Table 2.

### 4.1. Gradients, onsets and offsets

#### *4.1.1. True declaratives, wh-questions and inversion*

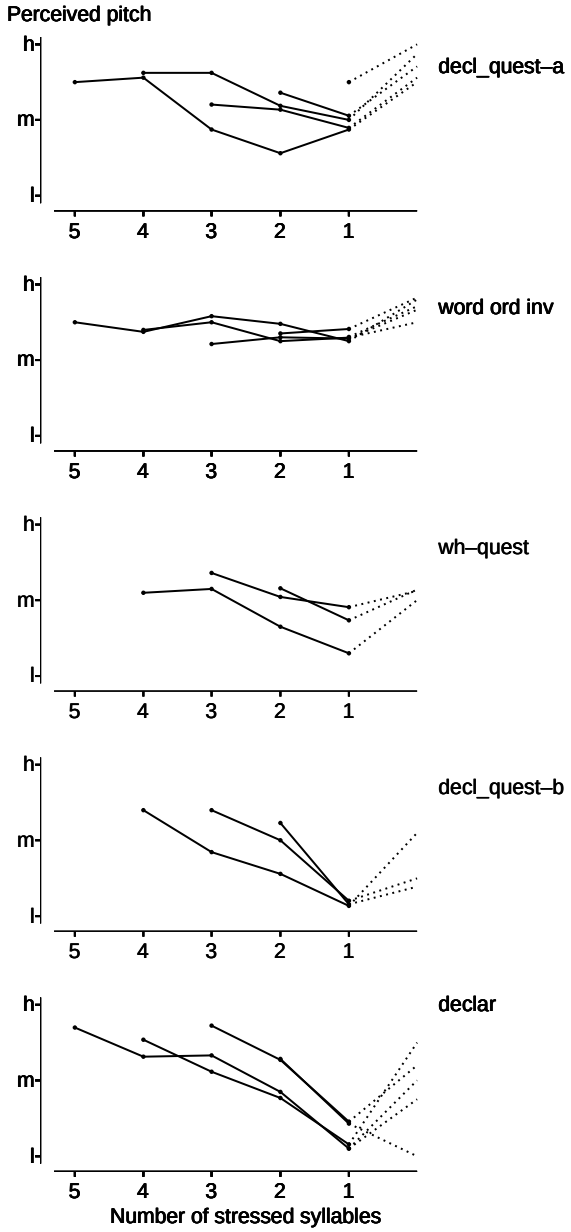
By and large, the predictions from read speech hold true: For any given length, declaratives have the steepest slopes and the lowest offsets. Wh-questions are somewhat less steep and terminate a little higher. Questions with inversion hardly decline at all and terminate still higher up. Note also that it is true of declaratives and wh-questions that the longer utterances have less steep slopes than the shorter ones. We can find no discernible pattern in contour onsets. They seem to vary randomly between 2.7 and 2.1 units on the perceived pitch scale. Note also in Fig. 2 that, in so far as a gradient is noticeable, i.e. in declaratives and wh-questions, the stressed syllables are fairly nearly equidistantly spaced on the pitch axis, again as would be predicted from read speech data.

#### *4.1.2. Declarative questions*

As noted above, 38 declarative questions (*decl_quest-b*) behaved in what we found to be an odd manner: in isolation they sounded like statements. To corroborate our impression we put together an exploratory test with 51 utterances, mixing 25 of these suspicious utterances with another 14 declarative questions which to us indeed sounded interrogative and 12 true declaratives for control. Eight linguists and phoneticians took the test on-line and were simply asked to choose between two alternatives: *does the utterance sound interrogative or not*. They varied a good deal in their propensity for perceiving an utterance to be interrogative; but by and large our impressions were confirmed. We therefore felt confident that we could pool the remaining 13 utterances of which we were sole judges with those utterances which received 4 or more responses for *does not sound interrogative* in the test.

It is no surprise that those 38 utterances were mainly perceived as being not interrogative, cf. Fig. 2 and Table 2: They even tend to terminate slightly lower in the range than the true declaratives. This is curious: One would expect that an utterance which has no overt lexical or syntactic markers of its interrogative function would have to have a prosodic cue, a non-declarative-statement-like intonation contour, to be perceived as a ques-

**Figure 2:** Intonation contours in 3 types of questions and in declaratives of varying length. See text




tion. But apparently, under the proper circumstances this is not a prerogative. What those circumstances are is not clear: we have found no systematic contextual differences between the 38 questions in *decl_quest-b* and the more prosodically conventional 87 *decl_quest-a* in Fig. 2. Furthermore, trying to render a few random items among the 38 with a less sloping intonation contour did not make them sound odd or out of place.

The 87 declarative questions, in the upper panel in Fig. 2, are not, in fact, the upper end of the intonation contour continuum as established in read speech. Their slopes and offsets are located somewhere between wh-questions and questions with word order inversion. This does not jeopardize their identification as interrogative utterances, given that their contour slopes are, ceteris paribus, less steeply declining than the true declaratives.

*4.1.3. The last rise*

The rise from the last stressed syllable to the first post-tonic holds another surprise. From read speech we would expect this rise to be smaller on more steeply declining contours, but in fact the contrary appears in Fig. 2: the highest rises are found in the true declaratives, and the smallest in the questions with word order inversion. If in Danish non-scripted speech this final rise were to have independent cue value (as found for other languages, cf. e.g. [1]), a reasonable expectation would be higher rises in questions than in statements. But so is not the case, on the contrary. This is indicative of the same state of affairs as reigns in read speech: Utterance modality is distributed across the whole utterance, by the global course of the intonation contour, not by any local, final pitch movement (unless, rather perversely, higher rises should cue statements and smaller rises should cue questions). This is in line also with what was found in [4].

## 5. SUMMARY

We set out to confront data from read speech with data from non-scripted speech. Our expectations were mainly confirmed: intonation contour slopes vary in an apparent trade-off with lexical/syntactic information. But a subgroup of declarative questions exhibited a rather unexpected behaviour, having intonation contours similar to true declaratives. The final rise from stressed 'low' to post-tonic 'high' turned out higher on steeper contours, and smaller on lesser gradients, exactly contrary to the read speech norm.

## 6. REFERENCES

[1] Grice, M., Benzmüller, R., Savino, M., Andreeva, B. 1995. The intonation of queries and checks across languages: Data from map task dialogues. *Proc. XIIIth ICPhS* Stockholm. Vol. III, 648–651.

[2] Grønnum, N. 1995. Superposition and subordination in intonation - a non-linear approach. *Proc. XIIIth ICPhS* Stockholm. Vol. II, 124–131.

[3] Grønnum, N. Submitted. Danpass - a Danish Phonetically Annotated Spontaneous Speech Corpus. *Speech Communication*.

[4] Tøndering, J. 2003. Intonation contours in Danish spontaneous speech. *Proc. 15th ICPhS* Barcelona. 1241–1244.