Intonational sequences in Tuscan Italian

Judith Bishop, Marc Peake*, Dmitry Sityaev†*

*Appen Pty Limited, Level 6, North Tower, 1 Railway Street, Chatswood NSW, 2067, Australia
†Toshiba Research Europe Limited, St George House, 1 Guildhall St, Cambridge CB2 3NH, UK
jbishop@appen.com.au, mpeake@appen.com.au, dmitry.sityaev@crl.toshiba.co.uk

Abstract

Sequential distributions of intonational pitch accents in a hand-labelled, read speech corpus of Tuscan Italian are analysed in relation to (1) the global frequencies of accents; (2) evidence of consistent patterns of association between accents. A subset of syntactically controlled nominal phrases are examined to determine whether the fact of their constituting a coherent syntactic unit has any effect on their patterns of sequential tone distribution, i.e. whether they are associated with a given tune. Some evidence is found for patterns of association between accents, and between sequences of accents and our chosen type of nominal phrase.

1. Introduction

In this paper we present a preliminary investigation of the sequential distributions of intonational accents in a read speech corpus of Tuscan Italian. The annotation framework used is Tones and Break Indices (ToBI), which labels the intonation as a sequence of an initial boundary tone, pitch accent(s), intermediate phrase accent tone and intonation phrase boundary tone. We report only on the results of the tonal analysis here.

Following Pierrehumbert [1], research in the ToBI (Tones and Break Indices) framework has described the intonation phrase as a finite state machine: that is, as an independent sequence of equally possible choices (i.e. ToBI labels). However, the distributions of ToBI labels in American English, at least, are known to "occur with skewed frequencies". [2] For example, in English, H* accents are selected more frequently than L*+H accents, at least in declarative sentences. It appears that "strings of [H*] pitch accents are common primarily because H* makes up such a large fraction of all pitch accents".[3:106] The skewed global frequencies thus affect transition probabilities between accents. In the view of this paper, these two questions need to be distinguished: (a) which choices are absolutely more likely (global frequencies); and (b) given a particular choice has been made, in what way is the probability of the next choice affected by the preceding one? That is, are there actual dependencies between accent choices, which could point in the direction of "tune"-like sequences being chosen as a unit, or is the appearance of such dependencies due only to a confounding effect of frequency?

To some extent, the skewed frequencies reported in the literature must merely reflect the relative preponderance of declarative *vs* interrogative *vs* imperative sentence types in the data collected by studies to date; "rarer" tones seem to occur more rarely simply because the functions they serve are more rarely called upon in these particular corpora. And although useful, the amount of predictive information to be gained from broad statements as to the proportion of accents of different types in a given corpus is relatively limited. A more precise focus on the probability of recurring sequences of accents and/or accent/phrase accent/boundary tone sequences—bringing the ToBI model closer to the British "tunes" model of intonation which preceded it—might return greater dividends for speech applications such as automatic prosodic recognizers, a number of which are currently under development. Of particular interest, but still under-explored, is the relationship between intonation sequences and syntactic structures. In particular, it is still unclear as to whether the often remarked lack of clear alignment between syntax and intonation is because the syntax-intonation relationship is not direct, but mediated by information structure [e.g. 4], or also because we are looking for correspondences at the wrong level of syntactic structure (possibly too low: e.g. the rates of pitch accent and type of accent associated with nouns; or else too high: e.g. broadly 'declarative' sentences generally come with too wide a range of intonation sequences to be very useful for recognition). It may be that frequency effects—but at the level of the frequency of particular syntactic structures in a given language—play a part in determining the probability of certain intonation sequences being associated with them (i.e. the 'idiomaticity' of these intonation sequences, so to speak).

In this paper we will examine only declarative utterances, and only data from a single speaker. In the first half of the paper, we describe the global frequencies of accent types and transition probabilities between accents. In the second half of the paper, we examine a syntactically tightly controlled subset of the data—nominal phrases containing the particle *di* (of)—in order to determine whether these common syntactic structures are associated with any regularities in the intonation sequences applied to them.

2. Material and method

The corpus used in this preliminary study comprised informational declarative sentences read by a single female speaker of Tuscan Italian. The sentences were recorded in a studio and digitized at 22kHz. Word and phone labels were time-aligned with the waveform and F0 contour in Wavesurfer, an Open Source speech analysis and manipulation program. The ESPS F0 extraction routine was selected. A ToBI annotation pane was added (see Figure 1). Transcription guidelines were adapted from the descriptions contained in [5] by the first author of this paper, an expert ToBI transcriber. This author also trained the ToBI transcribers, two native Italian speakers from Rome and Bologna, and thoroughly checked their transcriptions, in consultation with one of the transcribers. Few problems were encountered in matching the annotation symbols to the F0 contour, suggesting the general adequacy of the framework. Minor concerns arose only in relation to certain systematic

differences in types of *transition* between the starred tones of pitch accents. These will be discussed briefly below.

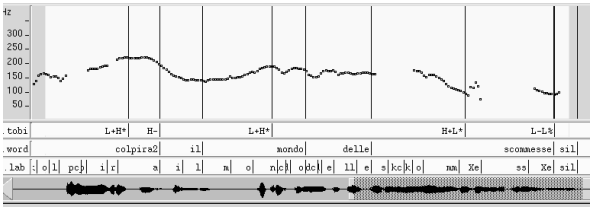

Figure 1: Sample intonation markup of a di phrase.

2.1. Intonation markup

The inventory of pitch accents was as follows: H*, L+H* (but see below), H+L*, L*, and L*+H, together with modifications for relative height of the high tones (upstep and downstep diacritics). Two phrase accents, H- and L-, and four boundary tones, L-H%, L-L%, H-L% and H-H% were annotated. An initial high boundary tone, %H, was also used quite frequently to annotate an unusually high intonation phrase onset (typically above 200Hz), *or else* a high prehead beginning at the same level as the initial accent. Pitch range modifications of high tones are ignored in the results reported, since our results are concerned only with categorical accent type, and not the local (extrinsic) pitch range with which each accent is realized.

In intonation phrase (IP)-initial position, only H* was annotated (not L+H*), since the onset of a new IP begins at a mid-low level by default, rendering the L+H*/H* distinction difficult to make. IP-medially, the distinction between H* and L+H* was maintained in the markup in order to annotate transitions between accents with greater phonetic accuracy. However, the dip described in the annotations by the L+H* accent frequently occurred much earlier than the syllable immediately preceding the stressed/accented syllable, where, canonically, it ought to occur. Indeed, it very often aligned close to the onset of the accented word, and frequently, when the particle *di* preceded the accented word, on the *i* vowel of that particle. In an analysis of tonal alignment in Bari and Pisa varieties of Italian, Gili Fivela conflates L+H* and H* and also the tones she annotates (L+)H*+L and H*+L. She observes, "…the best option seems to be to leave open the question as to whether the low target preceding the peak is a structural property of the peak accent…".[6] In the results below, L+H* and H* accents are also reported together as H*.

2.2. Analysis of *di* phrases

In the course of the intonation annotation, we observed that the corpus contained a considerable number of nominal phrases including the common syntactic particle *di*. We examined 240 such phrases, containing between 2 and 5 content words (i.e., accentable words), in utterance-initial (100) and utterance-final positions (140). We compared the intonational sequences in each position, in order to examine any positional effects on the tone sequences.

Table 1 describes the syntactic structures represented in the data. The structures can be summarised as (m) X of (m) (m) X (m) (m), where X stands for a noun and "m" stands for one of the following kinds of (optional) modifier – possessive adjective (e.g. nostre), demonstrative adjective (e.g. quelle), quantifier (e.g. tanti), number (e.g. due), or adjective. Sequences X X represent first and last names (e.g. Michele Giuttari). As there were very few instances in which either the X or the m were unaccented, and no obvious differences in terms of the accent types used on X *vs* m words, in reporting the results, we have conflated the sequences X of Xm, X of mX and X of XX on the one hand (to X of XX), and Xm of X, mX of X on the other (to XX of X).

Of the 240 phrases, only 12 did not fall into one of the three syntactic structure types X of X, X of XX and XX of X. These 12 phrases had the structures X of XXX and XX of XX. For simplicity we have excluded these phrases from the data set. This brings the number of utterance-initial *di*-phrases analysed to 96, and the number of utterance-final phrases to 132.

Syntactic structure	Utterance-initial (n = 100)	Utterance-final (n = 140)
X of X	67 (67%)	83 (59.5%)
X of X m	12 (12%)	15 (11%)
X of m X	10 (10%)	8 (5.5%)
m X of X	6 (6%)	10 (7%)
X m of X	0 (0%)	11 (8%)
X of X X	1 (1%)	5 (3.5%)
Other (X of XXX, XX of XX)	4 (4%)	8 (5.5%)

Table 1: Syntactic structures of \phrases with particle *di*

3. Results and discussion

3.1. Frequencies of accent types in the complete corpus

The proportions of each accent type were determined for intermediate phrases in two prosodic positions, non-IP-final and IP-final, together comprising the entire corpus. The significance of the differences in the relative proportions of (L+)H*, L*, L*+H and H+L* accents in non-IP-final intermediate phrases *vs* IP-final intermediate phrases was measured. A two-sample t-test of proportions shows that the proportions of L* and H+L* accents, relative to the total number of accents, are statistically significantly greater in IP-final phrases (z = -10.61 and -4.99, respectively; p=0) than in non-IP-final, while the proportion of (L+)H* accents is significantly greater in non-IP-final phrases (z =12.62; p=0). The proportion of L*+H accents is not significantly different in the two positions.

A chi-square statistical method was used to analyse the observed *vs* expected counts of two-accent (bigram) sequences in non-IP-final intermediate phrases, given the different frequencies of different accent types (Table 2). None of the differences between observed and expected counts were significant in this position. There was only 1 L*+H followed by another accent in the data; the corresponding row was therefore omitted from the table. The lack of a significant difference between observed and expected values indicates that any differences in transition probabilities in two-accent sequences in non-IP-final intermediate phrases would be due only to the absolute differences in the frequency of each accent type, and not to any dependencies between the accent types, i.e. preferred sequences of accent type.

Table 3 displays the same data for IP-final intermediate phrases. This time, there are a few statistically significant associations between accents in sequence (all at p<0.001 level; the relevant cells are shaded). First, more H+L* accents are

followed by another H+L* accent than expected. We might speculate that the sequence H+L* H+L* has a certain recognizable, tune-like quality associated with IP-final position. This would be supported by the fact that the sequence H+L* H+L* is, proportional to each sample size, considerably more frequent in IP-final phrases than non-IP-final phrases. More L*+H accents are also followed by H+L* accents than expected. Again, this sequence is considerably more frequent, proportionally, in IP-final position than in non-IP-final. This sequence also notably comprises a 'hat pattern': a pattern of rise, plateau, fall, which is easily recognizable when encountered in the data, and may again constitute a tune-like sequence for speakers, associated with IP-final position. More L* accents are followed by L* than expected; impressionistically, it seems likely that these L* L* sequences correspond to right-detached phrases, which characteristically carry low, level intonation and are always complete IPs. Finally, there are fewer L* accents after H+L* accents than expected; the reason for this is unclear as yet.

Non-IP-final					
Observed	(L+)H*	L*	L*+H	H+L*	Total
(L+)H*	459	85	5	95	644
L*	16	4	0	5	25
H+L*	54	4	2	14	74
Expected					
(L+)H*	458.5	80.6	6.1	98.8	
L*	17.8	3.1	0.2	3.8	
H+L*	52.7	9.3	0.7	11.3	

Table 2: Observed *vs* expected frequencies of bigrams of accent types, for non-IP-final intermediate phrases

IP-final					
Observed	(L+)H*	L*	L*+H	H+L*	Total
(L+)H*	1719	1196	44	1132	4091
L*	65	105	6	40	216
L*+H	9	3	2	47	61
H+L*	168	42	1	173	384
Expected					
(L+)H*	1688	1158.8	45.6	1198.4	
L*	89	61	2.4	63	
L*+H	25	17	0.7	18	
H+L*	158.5	108.8	4	112.5	

Table 3: Observed *vs* expected frequencies of bigrams of accent types, for IP-final intermediate phrases; shading indicates a statistically significant discrepancy

3.2 Analysis of *di* phrases

Syntactically, all of the utterance-initial *di* phrases were subject NPs. Note that these phrases were also informationally highly loaded: both the content words before and after the particle *di* are likely to constitute 'new' information in the 'out of the blue' context of read sentences. Perhaps unsurprisingly therefore, of the 100 phrases, 86 were followed immediately by an intonation phrase break. 56% of the intonation boundary tones were of type L-H%, 30.5% of type L-L% and 13.5% of type H-L%. 7 of the phrases were followed immediately by an intermediate phrase break (H-: 2, L-: 5). 3 were followed by one further accent before an intonation phrase boundary (L*: 2, H+L*: 1). All of the utterance-final phrases were followed immediately by an L-L% boundary.

Syntactic structure	**$(H*)^nL*$**	**$(H*)^nH*$**	**$(H*)^n$ H+L***	Other	Total
X of X	43	17	1	6	67
X of XX	18	1	1	3	23
XX of X	3	0	2	1	6
Total	64	18	4	10	96
% of total	**67**	**19**	**4**	**10**	**100**

Table 4: Results for utterance-initial nominal phrases with particle *di* (n=96). *Note*: $(H*)^n$: n >0

Syntactic structure	**$(H*)^nL*$**	**$(H*)^nH*$**	**$(H*)^n$ H+L***	Other	Total
X of X	30	1	38	14	83
X of XX	7	0	12	9	28
XX of X	4	0	4	13	21
Total	41	1	54	36	132
% of total	**31**	**0.8**	**41**	**27.2**	**100**

Table 5: Results for utterance-final nominal phrases with particle *di* (n=132). *Note*: $(H*)^n$: n>0

The total in the 'other' tone sequence column for utterance-final phrases is particularly high because the types of pre-final accents were more mixed in this context. Table 6 shows the counts and proportions of only the *final* accent in both utterance-initial and utterance-final *di*-phrases. The figures include the final accent in the 'other' tone sequences in both cases, but, in the case of the utterance-initial *di*-phrases, they exclude the phrases which end in an intermediate phrase boundary or a further accent rather than an IP boundary (10 phrases in total). For comparison, counts and proportions of final accents in all utterance-initial IP phrases, excluding the set of *di*-phrases, and all utterance-final IP phrases, excluding the set of *di*-phrases, are also included in the table.

di-phrases	(L+)H*	L*	L*+H	H+L*	Total
U-initial IP	13	64	0	9	86
% of total	15	74.5	0	10.5	
Non- di-phrases					
U-initial IP	601	212	2	130	945
% of total	63.60	22.43	0.21	13.76	
Significance	Z = -8.80 p=0	Z=10.44 p=0	n.s.	Z=0.85 p>0.05 (n.s.)	
di-phrases	(L+)H*	L*	L*+H	H+L*	
U-final IP	2	45	0	85	132
% of total	1.5	34.1	0	64.4	
Non-di-phrases					
U-final IP	848	2	0	567	1417
% of total	59.84	0.14	0	40.01	
Significance	Z=-12.88 p=0	Z=6.07 p=0	n.s.	Z=9.96 p=0	

Table 6: Frequencies of types of final accent in Utterance-initial and Utterance-final IPs.

Using a two-sample t-test of proportions, the comparison between *di*- and non-*di*-phrases shows a number of strongly statistically significant differences in the distribution of final accent types in *di*-phrases compared with non-*di*-phrases, and also in utterance-initial *vs* utterance-final phrases. The far greater proportion of H+L* accents in utterance-final position in both *di*- and non-*di*-phrases may indicate a role for this accent in cueing utterance-finality, especially if realized in a lowered pitch range. Turning to utterance-initial phrases, we find that far more utterance-initial *di*-phrases end in the low pitch accent (L*), rather than a high pitch accent ((L+)H*), compared with non-*di*-phrases. The same result is obtained in utterance-final *di*-phrases. In utterance-final position, the accent H+L* appears to largely replace L* in non-*di*-phrases, and also, but to a lesser extent, in *di*-phrases. To put the results in another perspective, 80% of final accents in utterance-initial *di*-phrases have low tone on the stressed syllable (the most perceptually salient part of the accent); in utterance-final position, the percentage is even higher (98.5%). The final accent falls on a modifier 12.5% of the time in utterance-initial position, and 11% of the time in utterance-final position, in the syntactic structures of the *di*-phrases under consideration here; the remaining accents fall on the noun. A number of hypotheses are possible regarding this disproportion of low final accents. Could it be that the low tone acts as a sort of terminal cue for the end of the syntactic unit? Could there be a correspondence between a low accent on the final noun and the syntactic subordination of the phrase headed by *di* in relation to the higher-level NP in which it is included?

Whatever the explanation, these differences do suggest that in terms of accent distribution, the *di*-phrases comprise a distinct population, i.e. that there are strong regularities in the accent sequences associated with these phrases, which are quite distinct from the regularities present in the same prosodic positions in the remainder of the corpus. This finding suggests that there may be important predictive patterns to be found by studies combining a fine-grained syntactic structure analysis and intonational sequence analysis. Such predictive patterns might be used to improve automatic prosody recognition.

4. Conclusion

This study has focused on frequencies of different accent types and sequences of accent. It has found that true dependencies between accent types are relatively rare; apparent dependencies are likely to be the result of a confounding of the absolute frequency of different accent types with actual associations between the different accents. We found a significant association only between certain accents in intonation-phrase-final position; it may be that the perceptual salience of this position is in some way related to the use of distinctive tone sequences there, such as H+L* H+L* and the hat pattern L*+H H+L*. In examining the intonation of *di*-phrases and non-*di*-phrases, we focused principally on phrase-final accents, since it was evident from a cursory overview of the data that most variation in the intonation of the *di*-phrases, at least, was concentrated in this position. The results indicated a much greater proportion of final low accents (L* and H+L*) in both utterance-initial and utterance-final *di*-phrases, relative to other utterance-initial and –final intonation phrases.

One of our hypotheses is that relatively frequent syntactic structures in a given language—such as *di*-phrases in Italian—may be more likely to have certain intonational sequences strongly associated with them. We have concentrated on accents because it is our impression that rather less is generally known about the motivations for accent choice than the choice of boundary tone and phrase accent (but see [7] in relation to English). Numerous studies have attempted to characterise 'declarative intonation' with interesting, but limited, results; it may be that this is simply too broad-brush a syntactic level at which to make the observations. Although the results presented in this study are very preliminary, they point to interesting possibilities for studies combining a fine-grained analysis of syntactic structures with an equally fine-grained intonational analysis. Finally, it is well-known that read speech may differ in the frequency and types of intonational events represented. It would be interesting to compare the results obtained in this study with a spontaneous speech sample.

5. Acknowledgements

The authors would like to thank Toshiba Research Europe Ltd. for permission to use their Italian speech corpus.

6. References

[1] Pierrehumbert, J. The Phonology and Phonetics of English Intonation. PhD thesis, MIT, 1980.

[2] Pitrelli, J. F., "ToBI Prosodic Analysis of a Professional Speaker of American English", *Proc. of Speech Prosody 2004*, Nara, Japan, March 23-26, 2004.

[3] Dainora, A. "An empirically based probabilistic model of intonation in American English." PhD thesis, University of Chicago, 2001.

[4] Steedman, M. "Structure and Intonation", *Language*, 68, 260-296, 1991.

[5] Grice, M., D'Imperio, M., Savino, M. and Avesani, C. "Strategies for intonation labelling across varieties of Italian." *Prosodic Typology:The Phonology of Intonation and Phrasing*. S.-A. Jun, ed., O.U.P., 2005.

[6] Gili Fivela, B. "Tonal Alignment in Two Pisa Italian Peak Accents", *Proc. of Speech Prosody 2002, Aix-en-Provence,* 11-13 April, 2002.

[7] Hirschberg, J. and Pierrehumbert, J. "Intonational structuring of discourse", *Proc. of the Association for Computational Linguistics of New York*, 134-44, 1986.