Peak F0 downtrends in Central Catalan neutral declaratives

Eva Estebas-Vilaplana

Abstract

This paper examines peak fundamental frequency (F0) downtrends in Central Catalan sentences produced with a neutral declarative intonation. The tendency of an F0 decline over the course of an utterance is a well-known property of speech. However, the interpretation of such a downtrend varies in the literature. Whereas sometimes the downward trend of pitch has been analysed as a global effect, conceived as a component of the overall contour, at other times it has been treated as a local, phonologically controlled mechanism, which affects accents individually and in relation to previous accents. In order to determine the nature of F0 peak downtrends in Central Catalan, 192 sentences produced by four speakers were analysed. The results suggest that peak (or H(igh) accent) downtrends in Central Catalan are better explained as a linguistically controlled accent-by-accent downstep than as a global time-dependent declination. Thus, peak height in Central Catalan can be accurately predicted as a constant proportion of the height of the previous peak, except for the utterance-final H which undergoes a greater amount of lowering than that expected by the downstep rule. Final H values are better explained by means of a lowering constant, which is higher than the downstep ratio.

1 Introduction

One of the most controversial issues in the study of intonation is the interpretation or modelling of declination. Declination has been described as the tendency of fundamental frequency to gradually lower over the course of an utterance. Although this pitch downtrend[1] has been widely studied and recognised in many languages, researchers still disagree on its nature. Two major views have been followed in the description of F0 downtrends, referred to by among others Ladd (1983) and Nolan (1995) as the *Contour Interaction* model and the *Tone Sequence* model (henceforth CI

[1] Pitch is the perceptual correlate of fundamental frequency (F0), which is the acoustic correlate of the repetition rate of vocal fold vibration. In this paper, the terms "pitch" and "F0" are used interchangeably.

and TS respectively). Researchers within the CI tradition (Bruce, 1977; Cooper and Sorensen, 1981; Fujisaki, 1983, 1988; Gårding, 1983; Lieberman, 1967; Thorsen, 1980, 1983; Vaissière, 1983; among others) view declination as one of the components that interact to generate a given pitch contour. Within this proposal, the F0 contour of an utterance is specified as a number of separate domains that combine to create specific pitch configurations. At least two kinds of layer are specified in all accounts: 1) a global component or overall line that affects the whole phrase or utterance (declination component) and 2) a succession of local F0 movements or accent units (accent layer). This is illustrated in Figure 1 below (as represented in Nolan, 1995). The sloping line stands for the global declining component. Filled circles represent accented syllables and open circles unaccented syllables. The final F0 contour is a combination of the global component and the accent layer. In the final contour, each accent is much lower than the preceding one due to the effect of the global component.

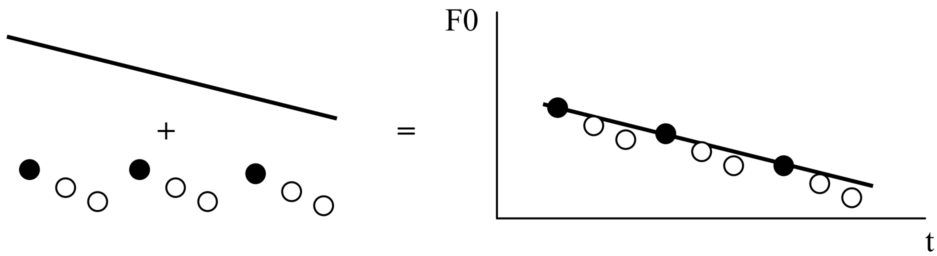

Figure 1. Schematic representation of the CI approach to modelling declination. The F0 contour (right panel) is the result of the interaction of two layer: 1) a global component (declination component) and 2) a succession of accent units.

Within the CI model, the temporal distance between two consecutive peaks is predicted to have an effect on the F0 value of the second peak, that is, the greater the distance between two consecutive peaks, the lower the F0 level of the second peak. Most researchers within the CI model (Prieto et al., 1996; Grabe, 1998; among others) measure temporal distance as the number of intervening syllables between accents. Thus, in this model, pitch drop can be analysed as an effect of the time interval (or number of syllables) between peaks. This is illustrated in Figure 2 below, where the F0 height of a peak is lower as the number of unaccented syllables between peaks increases. In the right panel of Figure 2, the second accent (second filled circle) is much lower than its equivalent in the left panel, due to a greater syllabic interval between accents.

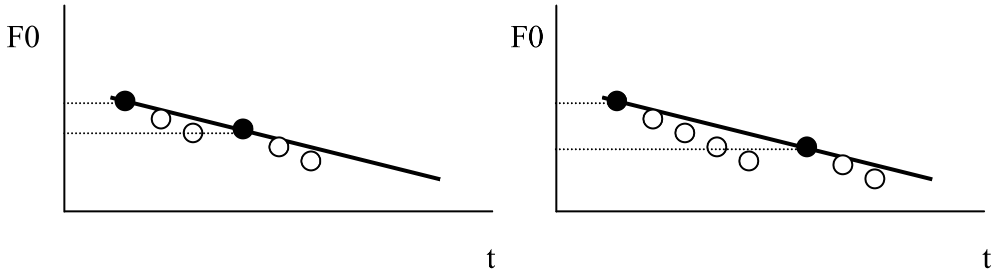

Figure 2. Expectations of the CI model: The number of intervening unaccented syllables between accents has an effect on the scaling of peaks. The greater the number of intervening unaccented syllables (open circles), the lower the scaling of an accent (filled circles).

The TS approach, on the other hand, analyses pitch downtrends as a linguistically controlled mechanism, which involves a deliberate use of step accents to attain specific tone targets. This view was originally proposed in Pierrehumbert (1980) and Liberman and Pierrehumbert (1984). They discovered that in American English descending contours, the value of F0 peaks was quite stable and that time-dependent lowering was almost absent in their data. This accent-by-accent decay was termed *downstep* or *catathesis* (see Pierrehumbert and Beckman 1988). In this model, no domain is superimposed on any other but the contour is specified as a sequence of local downstepping elements, each one lower than the preceding one. Within this proposal, the observed F0 downtrend originates from the repeating occurrence of downstep accents. This is schematised below after Nolan (1995).

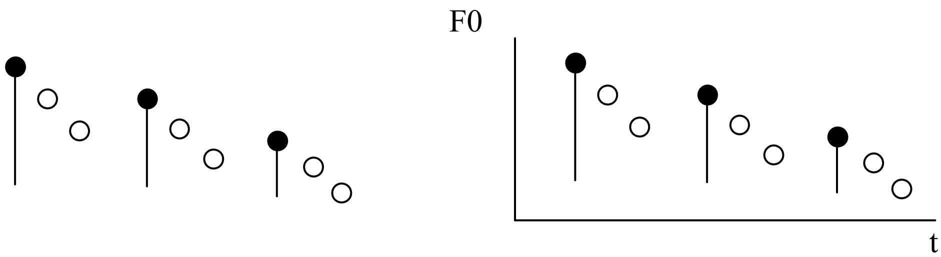

Figure 3. Schematic representation of the TS approach to modelling intonation. Each accent is lower than the preceding accent and this produces a downtrend over the whole F0 contour.

Given the fixed, phonologically-controlled F0 height of accents, no time-dependent effect on the scaling of peaks is expected. This is illustrated in Figure 4 below.

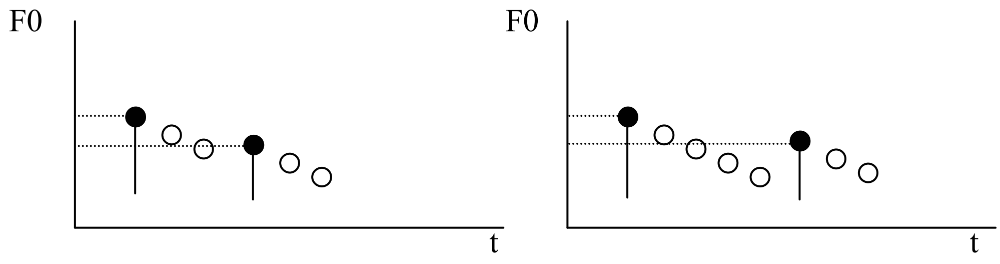

Figure 4. Expectations of the TS model: The number of intervening unaccented syllables between accents has no effect on the scaling of peaks.

The TS model's proposal to describe pitch downtrends as a localised occurrence of downstep derives from the behaviour observed in several African tone languages (Meyers, 1976; Clements, 1983). In these languages, the second H in a sequence HLH is realised at a much lower level than the first H[2]. The downstep of the second H is phonologically conditioned by the presence of the intervening L tone. Whenever there is no downstep trigger between the two H tones (that is, whenever there is no L between the Hs) as in the sequence HH, the second H does not show the same amount of lowering. Pierrehumbert (1980) showed that the behaviour observed in English declaratives was similar to that observed in tone languages. She claimed that any HLH sequence involves a downstepped second H due to the presence of the L tone. This L is not manifested as an F0 valley but it acts as a downstep trigger of the following H tone.

Liberman and Pierrehumbert (1984) proposed that in English downstepping contours the height of a given peak could be modelled by using a constant F0 reduction of the previous peak value. Thus, the amount of decay between peaks was calculated as a proportion of the second peak with respect to the first one, scaled above the reference line of the speaker. The reference line was described as an abstract line, which lies between the last peak of the utterance and the F0 minimum (see Liberman and Pierrehumbert, 1984 for more details). Additionally, Liberman and Pierrehumbert found that the final peak of a declarative sentence undergoes a more drastic lowering in F0 than that expected by the application of the downstep rule. They proposed to account for this final lowering by means of a lowering constant, defined as a fraction of the value of the peak predicted by the downstep rule. The modelling of F0

[2] H and L stand for High and Low tones respectively.

downstepping contours by means of a downstep ratio and a final lowering constant has been applied to several languages, such as, British English (Ladd, 1983, 1984), Japanese (Beckman and Pierrehumbert, 1986; Poser, 1984), Mexican Spanish (Prieto et al., 1996), German, (Möbius, 1993; Grabe, 1998, in press) and Dutch (van den Berg et al., 1992).

Some studies on intonation (Pierrehumbert and Beckman, 1988; Poser, 1984; Fujisaki, 1983, 1988) propose that the analysis of pitch downtrends as a sequence of localised phonological events does not exclude the existence of a global declination effect. Rather they suggest that both phenomena, declination and downstep, can coexist in the same utterance. Also Grabe (in press) suggested that what Liberman and Pierrehumbert (1984) interpreted as final lowering in American English sentences seems to be an effect of declination in British English utterances, since the greater F0 reduction observed on the utterance-final peak can be explained by a greater time interval between peaks. In the studies that claim the existence of the two mechanisms, declination is considered a residual downtrend after all other predictable lowering mechanisms have applied.

In this paper, we investigate whether the F0 downtrend observed in Central Catalan neutral declaratives responds to a global, time-dependent declination effect or to a local linguistically-controlled downstep mechanism. This is done by comparing the F0 maximum of consecutive peaks separated by different temporal distances (or number of intervening syllables). If declination is active, a greater amount of F0 lowering is expected as the temporal distance between Hs increases. Alternatively, if the height of peaks is stable and not affected by temporal distance, then the pitch downtrend would be the result of a controlled downstep or accent-by-accent decay. The results reported in this study will show that Central Catalan downtrends are better explained as a downstepping mechanism rather than as a declination effect. Given this evidence, we will also investigate to what extent the TS approach can be used to predict the location of peaks in Central Catalan downstepping contours.

2 Experimental design

The data used for the analysis of pitch downtrends in Central Catalan consist of 48 Subject-Verb-Object declarative sentences. The utterances were produced by four Central Catalan native speakers. The informants were asked to read the sentences as if they were responses to a "what happens?" question type, i.e. with a neutral declarative intonation. Overall, 192 sentences were recorded (48 sentences x 4 speakers)[3]. The whole list of sentences is included in the Appendix.

The number of stresses (and hence potential accents) per sentence was either three or four. There were 25 sentences with three stressed syllables and 23 with four. The number of unstressed syllables between successive stressed syllables was either one or two[4]. An example of each kind of sentence is illustrated in (1). Stressed syllables are underlined.

(1) a. *Three stresses*
La Mireia remena l'olla
"Mireia stirs the casserole"

b. *Four stresses*
Les noies volen rebre la reina
"The girls want to welcome the queen"

The four speakers had similar characteristics as far as age, geographical origin, social status and education are concerned. The informants were between 25 and 30 years old at the time of the recordings. Each of the subjects was born and lived in the same town, Ripoll, situated in the province of Girona. They were brought up in middle class Central Catalan-speaking families. Like all Catalan speakers, the subjects were

---

[3] These sentences were not specifically designed to analyse F0 downtrends in Central Catalan but were part of a bigger corpus of data, which was designed to examine the intonational properties of neutral declaratives in this language (see Estebas-Vilaplana 2000). F0 downtrends were analysed as part of the intonational characteristics of a neutral intonation.

[4] Even though in the data the number of intervening syllables between stresses was too small to allow an examination of temporal effects on the scaling of H accents, higher numbers of intervening syllables between accents appeared in the speakers' productions. This was due to the fact that not all stresses became accented (see section 4 for more details).

bilingual and spoke Spanish as a second language. However, Catalan was the language they always used at home, work and in normal daily conversations. The variety of Catalan they used was Central Catalan, which is the most widely spoken dialect of the whole Catalan linguistic domain and the best candidate to be treated as "standard". In the following sections, the four speakers are identified as CP, DV, MC and NG.

The recordings included two simultaneous but separate signals: *speech* and *laryngeal* (Lx) signals. For the speech signal, the instrumental tools used to obtain the data were a Marantz Superscope/CD 330 tape recorder and a Beyer Dynamic microphone on a stand. For the laryngeal signal, a portable laryngograph LX 12 was used. A Thandar portable DRO 26 oscilloscope was also used to check the quality of the laryngograph output. The recordings were conducted in a soundproof room of a private recording studio in Ripoll. The microphone for the recordings of the speech signal was set at about twenty centimetres from the speakers. To get the Lx signal a pair of electrodes was set on the speakers' neck by means of a collar band. The electrodes were connected to the portable laryngograph. Before the actual recording, subjects were asked to read a short passage to adjust the gain controls of the tape recorder and the laryngograph.

3 Analysis of the data

An acoustic analysis of the utterances was carried out by means of the Speech Filing System (SFS). SFS allows for a simultaneous inspection of several signals: in this case, speech waveform and F0 trace. To transfer the speech and laryngeal signals obtained in the recordings into the computer, the two signals were played on a Denon DN-770R tape recorder connected into a Sun Sparc-10 computer running SFS. Acquisition of the signals was done at a 16 KHz sampling rate, following the routines of the program. In order to obtain the F0 traces from the laryngeal signal, the VTX and FX programs were used. VTX converted the laryngeal waveform (Lx) into excitation period measurements (Tx). From VTX excitation period measurements, F0 traces were obtained by means of the FX program, which converts a Tx item into a F0 item. The extracted F0 was smoothed with a 5-point median filter.

Figure 5 illustrates the SFS display screen for the sentence *Les nenes haurien de donar l'enhorabona* ("the girls should send their congratulations") produced by speaker NG. The first screen contains the speech waveform. The second screen shows the F0 contour. The last screen is used to annotate relevant points in the utterance. In this particular example, the onset of each syllable is marked.

As observed in Figure 5, the F0 trace consists of a series of peaks. Each peak is lower than the preceding one. In order to determine whether the F0 descent observed in the utterances is a gradual time-dependent declination effect or a linguistically-controlled downstep, we will examine the scaling of F0 values as a function of the number of intervening syllables between H targets in sentences of different lengths (three stresses and four stresses).

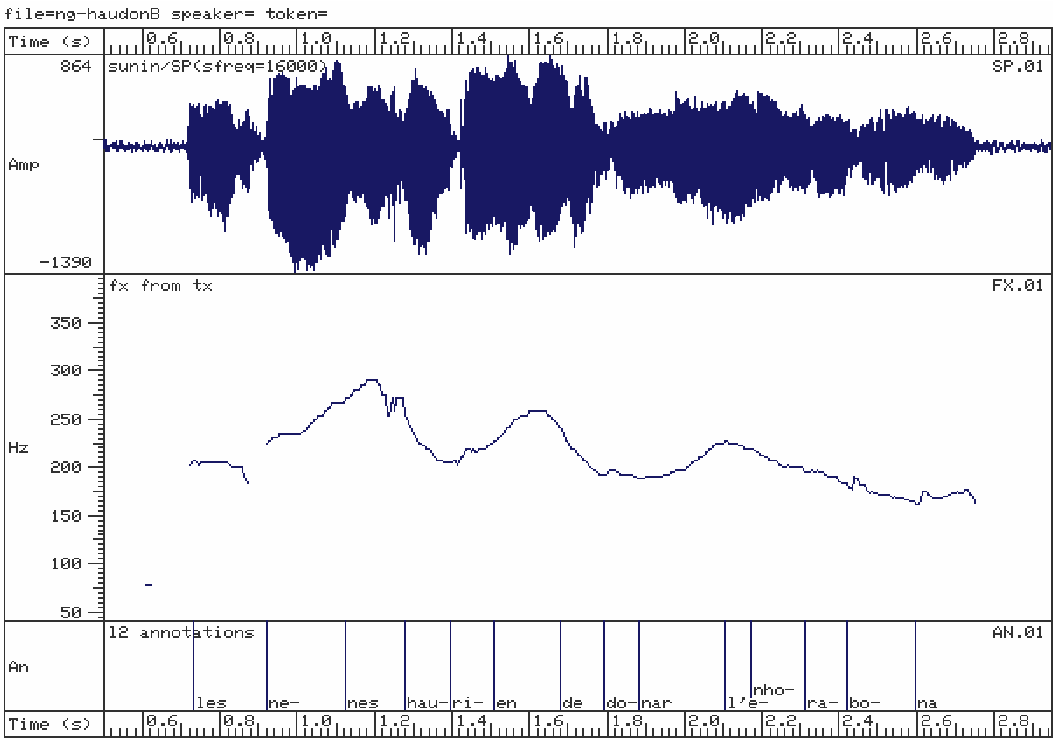

Figure 5. *Les nenes haurien de donar l'enhorabona* (speaker NG). "The girls should send their congratulations"[5].

[5]The alignment of some of the peaks beyond the limits of the accented syllable is explained by the fact that Central Catalan prenuclear accents in neutral declaratives are a combination of an L* pitch accent and an H word edge-tone anchored at the end of the word. See Estebas-Vilaplana (2000) for more details on this representation.

For each sentence, the following points were labelled in the F0 contours:

1. Phrase-initial F0 value
2. Highest F0 value for each pitch accent
3. Lowest F0 value between peaks
4. Phrase final F0 value[6].

4 Results

The mean schematised F0 contours of utterances with different numbers of stresses (and accents) are presented in Figure 6 for all speakers. In each graph, the number of stresses and the number of accents (in brackets) are specified. Thus, for example, "4(3)" stands for a sentence with four stresses but only three accented syllables. The L and H letters indicate the mean F0 minima and F0 maxima respectively and also the tonal interpretation proposed for those values. This explains why sometimes a quite low F0 value is marked as H (for more discussion on this issue see section 4.4). Finally, when a given F0 point is specified by means of two letters (as in L/H), the first letter corresponds to the sentence with more accents and the second letter to the sentence with fewer accents[7].

---

[6] Irregularities at the beginning and at the end of the F0 contour were ignored when identifying these points.

[7] Since this paper is intended to examine the scaling of H values, the rhythmic properties of H and L are not specified. For more information about the interpretation of H and L as pitch accents or edge tones, see Estebas-Vilaplana (2000).

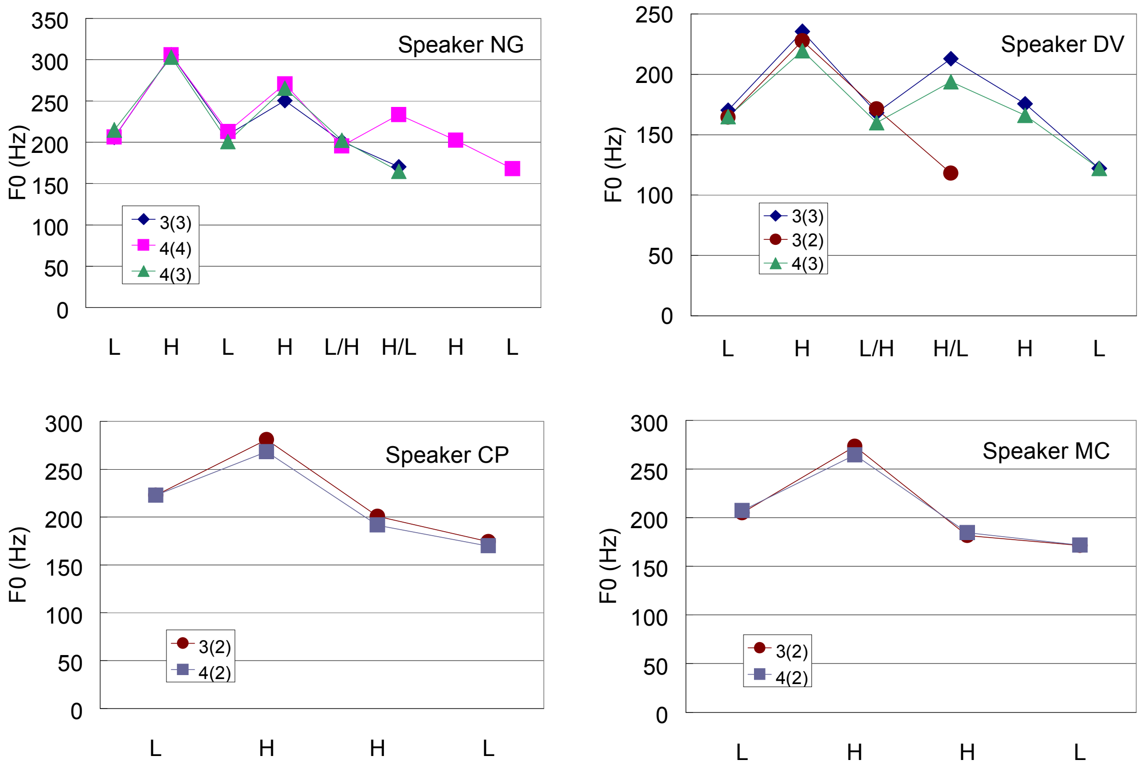

Figure 6. Mean schematised F0 contours for neutral declarative sentences with different number of stresses and accents (in brackets) for four speakers.

The results displayed in the graphs show inter-speaker variability as far as the number of accented syllables per sentence is concerned. Whereas speakers NG and DV tended to accent most of the potentially accentable syllables, speakers CP and MC only accented the first and last syllable of each utterance. Differences in the number of accents generated appropriate contexts to examine F0 downtrends since different numbers of intervening syllables between accents became available. For speakers NG and DV, the F0 descent observed in the schematised contours showed a gradual F0 lowering of peaks and a more constant F0 value for valleys, except for the last L which was lower than the others. For all speakers, the last H showed a very low frequency close to non-final L values. In the next sections, the nature of peak downtrends is examined as a function of the number of intervening syllables between H accents.

4.1 Utterance initial and final F0 values and pitch range

Before analysing the nature of Central Catalan F0 downtrends, it was considered necessary to see whether differences in sentence length had an effect on the utterance initial and final F0 values and on the pitch range with which sentences were produced. Whereas in some studies (Thorsen, 1980; Cooper and Sorensen, 1981), utterance length seems to have an effect on the F0 range of the first peak (and subsequently of all other peaks), in other studies (Liberman and Pierrehumbert, 1984; Prieto et al., 1996) sentence length does not influence the degree of prominence in sentences. Since in our data sentences with a different number of stresses (and hence with different length) are compared, we first need to prove that pitch accents are actually comparable and hence that utterance length does not affect pitch range. In order to do so, utterance-initial and utterance-final F0 values were calculated as well as the scope of the first F0 rise (measured as the difference in Hz from the lowest point at the beginning of a contour to the highest F0 value of the first peak). If the F0 excursion between the initial F0 value and the first peak is the same in sentences with different length, this will show that utterances with a different number of stresses are produced with the same pitch range and hence accents are comparable.

Figure 7 shows the mean values of utterance-initial and utterance-final F0 values for sentences of difference length (i.e. with 3 and 4 stresses). The results show that the initial and final F0 values are independent of the utterance length for all speakers. T-tests performed for each speaker showed no significant differences in utterance-initial F0 values for phrases of different lengths (where $p<0.01$): CP ($p=0.03$, $t=1.91$), DV ($p=0.23$, $t=0.72$), MC ($p=0.29$, $t=-0.55$) and NG ($p=0.49$, $t=-0.01$). Similarly, utterance-final F0 values were also rather constant and no significant differences were observed for any speaker: CP ($p=0.07$, $t=1.44$), DV ($p=0.18$, $t=-0.9$), MC ($p=0.47$, $t=-0.06$) and NG ($p=0.14$, $t=1.06$). These results are similar to those found by Liberman and Pierrehumbert (1984) for English and Prieto et al. (1996) for Mexican Spanish: utterance-initial and utterance-final F0 values were nearly constant for a given speaker despite differences in sentence length.

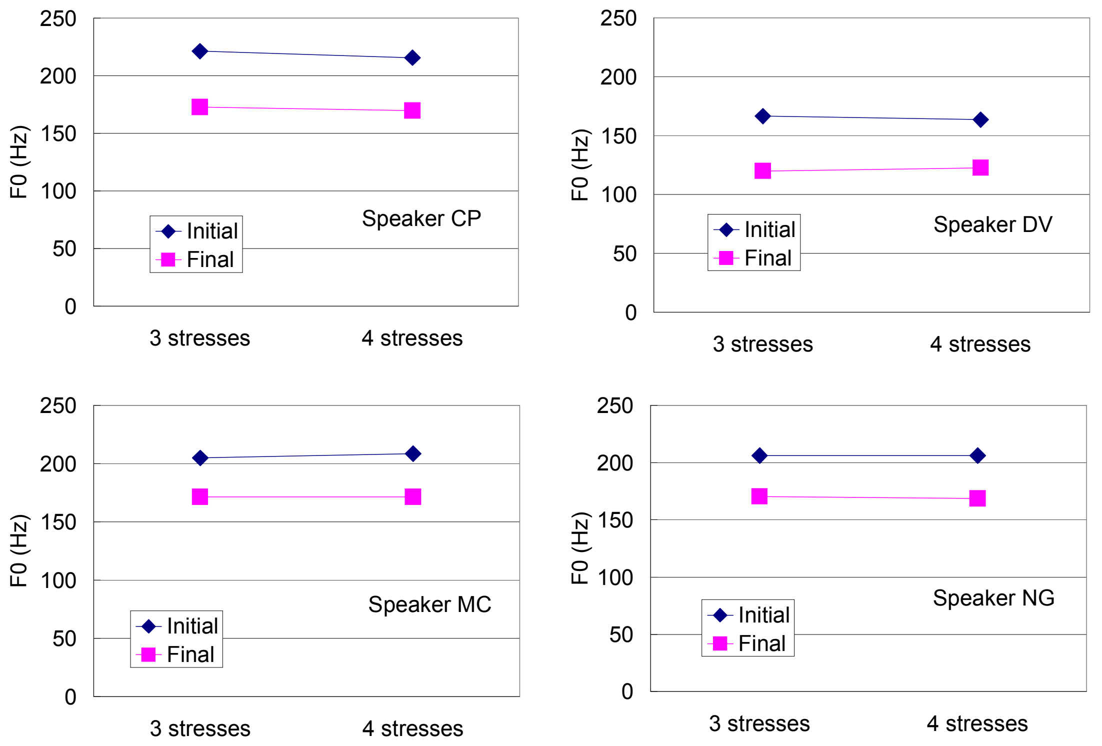

Figure 7. Mean values (in Hz) of utterance-initial and utterance-final F0 values of sentences of different lengths (3 to 4 stressed syllables).

Given the constant initial F0 values observed for all speakers, the F0 range was calculated as the difference in Hz from the lowest F0 value at the beginning of an utterance to the highest F0 point of the first peak, as in Prieto et al. (1996). The results are displayed in Table 1, which shows the mean F0 range (and SD) in utterances with three and four stresses and the results of t-tests comparing the F0 range values for all speakers.

	Speaker CP			Speaker DV			Speaker MC			Speaker NG		
Stresses	F0 range	p	t	F0 range	p	t	F0 range	p	t	F0 range	p	t
3	80.4 *17.45*	0.23	0.74	65.4 *18.5*	0.06	1.6	88.9 *13*	0.02	2.11	99.4 *20.1*	0.48	-0.03
4	75.9 *21.3*			56.1 *16.6*			80.5 *15.2*			99.6 *15.3*		

Table 1. Mean F0 range and SD (in italics) in utterances with three and four stresses and results of t-tests comparing the F0 range values for all speakers.

The results presented in Table 1 show that utterances were produced with a rather constant pitch range irrespective of the number of stresses per sentence. For all speakers, t-tests comparing the F0 range values in utterances with three and four stresses show that differences in pitch range are not significant (p>0.01 for all speakers). This means that H values in sentences with different lengths are comparable, since no pitch range variation is observed in the data.

4.2 Effects of temporal distance between peaks

One of the strategies used by some researchers (Prieto et al., 1996; Grabe, 1998; in press; among others) to detect whether declination is active in downtrend F0 contours is the analysis of the temporal distance (examined in terms of number of intervening syllables) between peaks. If declination is present, the F0 reduction is expected to be bigger as the time interval (or number of intervening syllables) between peaks increases. Otherwise, if there is no declination, the F0 reduction is expected to have a constant value irrespective of the time interval between peaks.

In our data, since the speakers tended not to accent all the potentially accentable syllables but showed a rather heterogeneous behaviour in the number of peaks, two appropriate contexts to examine declination appeared in the results.

Context 1

The first context is observed in the productions of speakers NG and DV. It involves the analysis of H_1 (first peak) and H_2 (second peak) in sentences where all stresses got an accent as opposed to those sentences where not all lexical stresses became accented. For example, sentences with four-lexically stressed words, as the one illustrated in (2), are sometimes produced with four pitch accents (as in (2a)) and sometimes with three (as in (2b)) for the same speaker. This means that the number of intervening syllables (and hence temporal distance) between H_1 and H_2 is bigger depending on whether a peak is observed at the end of *havia* or not. When all pitch accents are produced, the number of intervening syllables between H_1 and H_2 varies from one or two. However, when a pitch accent is not produced, the number of intervening syllables increases from three to five. This case seems to provide an appropriate context to analyse declination. If declination is present, there should be a larger F0 reduction of the second peak as the temporal distance between peaks increases. If there is no declination, the amount of lowering of the second peak is expected to remain constant despite differences in the time interval between peaks.

(2) La mare havia de menjar verdura

a. **H1** **H2** H3 H4

b. **H1** **H2** H3

"The mother should eat vegetables"

Similarly, since the first peak has been proved to have a similar scaling irrespective of the length of the sentence, another way to look at the same aspect is to compare three-stressed/three-accented sentences with four-stressed/three-accented sentences, as illustrated in (3). As before, the number of syllables (and temporal distance) between H1 and H2 is bigger in (3b) than in (3a) and hence the effects of temporal distance on F0 scaling can be tested.

(3) a. En Joan anava al museu

H1 **H2** H3

"Joan went to the museum"

b. La mare havia de menjar verdura

H1 **H2** H3

For speaker NG, we compared the peak height (mean F0) and amount of pitch drop (F0 mean difference) between H1 and H2 in the two conditions:

1. Four-stressed/four-accented sentences (henceforth (4,4)) vs. four-stressed/three-accented sentences (4,3)
2. Three-stressed/three-accented sentences (3,3) vs. four-stressed/three-accented sentences (4,3).

For speaker DV, only the second condition ((3,3) vs. (4,3)) could be examined since she did not produced any cases of four accented utterances.

Before analysing the effect of syllable distance on the scaling of two peaks, we wanted to confirm that differences in the number of syllables involved differences in real duration. Hence, the temporal distance between H1 and H2 was measured for the two speakers in all possible conditions. The mean duration (and SD) between H1 and H2 according to the number of intervening syllables and sentence conditions (4,4; 3,3; and 4,3) are presented in Table 2.

	Number of syllables (sentence condition)		
	1-2 (4,4)	1-2 (3,3)	3-5 (4,3)
NG	0.278 *0.056*	0.261 *0.071*	0.635 *0.083*
DV	-	0.279 *0.067*	0.574 *0.106*

Table 2. Mean duration (in seconds) between peaks with 1-2 intervening syllables and with 3-5 intervening syllables in different kinds of accented utterances. The SD values are presented in italics.

As expected, the bigger the number of intervening syllables between H1 and H2, the longer the temporal duration. T-tests comparing the results show that the mean duration is significantly different (NG: $p<0.001$, $t=10.9$ [for sentence condition (4,4) vs. (4,3)], $t=7.30$ [for (3,3) vs. (4,3)]; DV: $p<0.001$, $t=6.22$) between 1-2 syllables and 3-5 syllables for both speakers, indicating that the bigger the number of intervening syllables, the longer the time to produce them. In the same way, for speaker NG differences in the mean duration of 1-2 syllables in conditions (4,4) and (3,3) are non-significant ($p=0.3$, $t=0.5$). This proves that the analysis of declination with respect to the number of intervening syllables is reliable.

Once it had been confirmed that differences in the number of syllables involve differences in time, we compared the mean peak height of H2 preceded by 1-2 and 3-5 unaccented syllables as well as the amount of pitch drop (F0 mean difference) between H1 and H2. Table 3 presents the mean F0 values (and SD) of H2 according to the number of intervening syllables for the two speakers. Table 4 exhibits the mean F0 difference (and SD) between H1 and H2. Finally, Figure 8 shows the mean F0 values for H1 and H2 corresponding to an increase in the number of intervening syllables.

	1-2 (4,4)	3-5 (4,3)	t-test	1-2 (3,3)	3-5 (4,3)	t-test
NG	270.6 *23.5*	265.4	p=0.31 t=-0.48	250.3 *23.1*	265.4 *8.5*	p=0.08 t=1.4
DV	-	-	-	212.9 *16.2*	190.8 *12.5*	p<0.001 t=4.6

Table 3. Mean F0 values and SD (in italics) of H2 preceded by 1-2 or 3-5 unaccented syllables for the two speakers in different sentence conditions.

The results of Table 3 show that for speaker NG no significant differences were observed in the mean F0 value of the second peak between contexts having 1-2 or 3-5 intervening unaccented syllables. This seems to indicate that for this speaker declination is not present in her productions. Speaker DV, on the other hand, exhibits a significantly lower F0 value in H2 when the number of intervening syllables increases. We might first think that for this speaker declination is present in addition to downstep. However, if we look at Figure 8, we see that not only H2 but also H1 is lowered in her (4,3) sentences. This seems to indicate that the significant differences between the two H2s in Table 3 are not due to the differences in the number of intervening syllables, but to the fact that the whole pitch range of (4,3) sentences is a little lower than (3,3) sentences[8]. This is confirmed by the results displayed in Table 4, where the mean F0 difference between H1 and H2 with 1-2 and 3-5 intervening syllables is not significant for speaker DV. Similarly, no significant differences are observed for speaker NG. Furthermore, in condition (3,3) vs (4,3) speaker NG shows a higher amount of pitch drop when the number of intervening syllables is smaller than when it is bigger. This corroborates the idea that there is no declination effect in the data.

	1-2 (4,4)	3-5 (4,3)	t-test	1-2 (3,3)	3-5 (4,3)	t-test
NG	35.2 *11.8*	37.8 *15*	p=0.34 t=-0.4	55.3 *20.6*	37.8 *15*	p=0.04 t=1.78
DV	-	-	-	22.7 *12.5*	26.2 *10.1*	p=0.12 t=-1.27

Table 4. Mean F0 difference and SD (in italics) between H1 and H2 with 1-2 and 3-5 intervening syllables for the two speakers and in different sentence conditions.

[8] This was already observed in Table 1 where the F0 range between the utterance-initial F0 value and the first peak was lower in utterances with 4 stresses than in utterances with 3 stresses. This difference, however, was shown to be not significant (p=0.06).

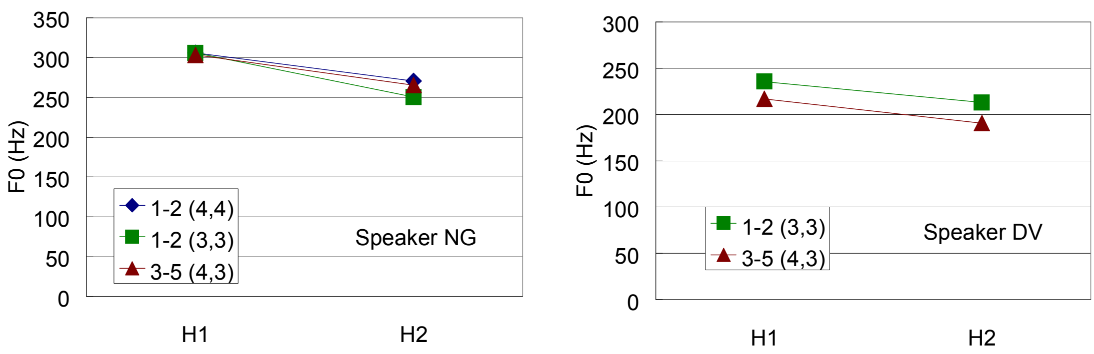

Figure 8. Mean F0 values for H1 and H2 corresponding to an increase in the number of intervening syllables (1-2 or 3-5) and in different sentence conditions for speakers NG and DV.

Context 2

The second context to analyse the presence/absence of declination is observed in those speakers (CP and MC) who produced only two pitch accents in sentences with three and four stresses. In this case, the number of syllables between H1 and H2 varies between 3 and 5 in three-stressed sentences and between 6 and 9 in four-stressed sentences. As before, the temporal distance between the first and last pitch accent is expected to be longer in four-stressed than in three-stressed utterances. This is confirmed in Table 5, where the mean duration distance (and SD) between H1 and H2 is significantly longer for the two speakers as the number of intervening syllables increases (CP: $p<0.001$, $t=8.9$; MC: $p<0.001$, $t=10.9$).

	Number of syllables (sentence condition)	
	3-5 (3,2)	6-9 (4,2)
CP	0.602 *0.129*	0.950 *0.132*
MC	0.622 *0.124*	1.028 *0.123*

Table 5. Mean duration (in seconds) between peaks with 3-5 intervening syllables and with 6-9 intervening syllables in sentences with three and four stresses but two pitch accents. The SD values are presented in italics.

As before, once it had been confirmed that differences in the number of syllables correspond to differences in time, the declination effect was tested in the data of these speakers by means of the same hypothesis: if the scaling of the last pitch accent decreases as the number of intervening stresses increases, declination is present. If, on the other hand, the last pitch accent shows no differences in F0 height or in the amount of pitch drop, declination is absent.

Table 6 presents the mean F0 values of H2 according to the number of preceding unaccented syllables for the two speakers. Table 7 exhibits the mean F0 difference (and SD) between H1 and H2. Finally, Figure 9 shows the mean F0 values (and SD) for H1 and H2 corresponding to an increase in the number of intervening syllables.

	3-5 (3,2)	6-9 (4,2)	t-test
CP	200.7 *9.78*	191.5 *8.55*	p=0.001 t=3.2
MC	181.6 *9.49*	184.7 *7.91*	p=0.06 t=-1.58

Table 6. Mean F0 values and SD (in italics) of H2 preceded by 3-5 or 6-9 unaccented syllables for the two speakers and in different sentence conditions.

	3-5 (3,2)	6-9 (4,2)	t-test
CP	80 *17.4*	76.6 *23.3*	p=0.23 t=0.74
MC	91.8 *13*	79.6 *15.2*	p=0.002 t=3.08

Table 7. Mean F0 difference and SD (in italics) between H1 and H2 with 1-2 and 3-5 intervening syllables for the two speakers and in different sentence conditions.

The results of Table 6 show that for speaker MC, no significant differences are observed in the mean F0 value of H2 according to the number of preceding syllables. This indicates the lack of declination in the productions of MC. For speaker CP, on the other hand, H2 is significantly lower when the number of preceding syllables is bigger. However, this does not indicate the presence of declination since, as displayed in Table 7, this speaker shows no significant differences between the mean F0 difference between H1 and H2 according to the number of intervening syllables. Even more, both speakers exhibit a tendency which contradicts the declination hypothesis. The pitch drop is slightly bigger with a shorter time interval (3-5 syllables) than with a longer duration (6-9 syllables). This corroborates the findings reported for speakers NG and DV. Thus, no declination effect is observed in Central Catalan neutral declaratives. This suggests that downtrend contours in this language originate from the repeating occurrence of downstep accents, rather than from a global declination component.

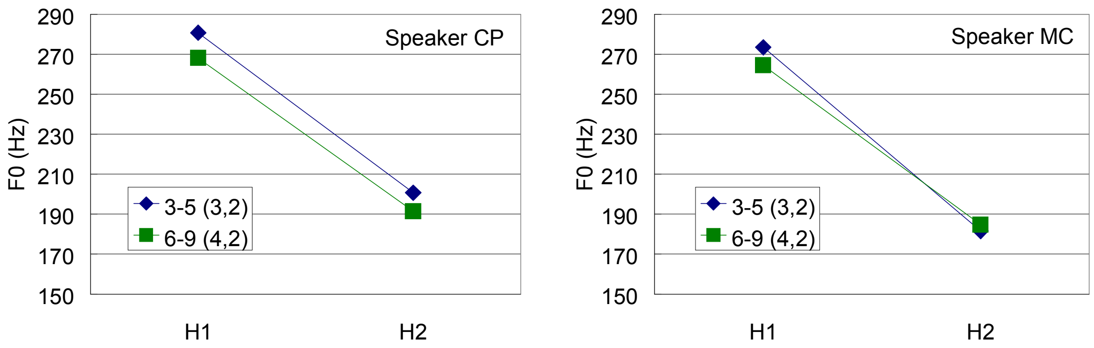

Figure 9. Mean F0 values for H1 and H2 corresponding to an increase in the number of intervening syllables (3-5 or 6-9) and in different sentence conditions for speakers CP and MC.

4.3 Predicting F0 values

The results observed so far show that Central Catalan downtrend patterns have to be explained by a linguistically controlled downstep rather than by a gradual declination effect. This supports the proposal of the TS view, which analyses pitch downtrends as the result of an accent-by-accent decay. Given these results, the next aim is to find out whether the model proposed by Liberman and Pierrehumbert (1984) to predict downstep in American English can be used to determine the scaling of peaks in Central Catalan downstepping sequences. Liberman and Pierrehumbert's model is examined since it has successfully predicted the scaling of downstepped accents in several languages, namely, Dutch (van den Berg et al., 1992), Mexican Spanish (Prieto et al., 1996), Japanese (Beckman and Pierrehumbert, 1986; Poser, 1984), and German, (Möbius, 1993; Grabe, in press). The ability of Liberman and Pierrehumbert's model to predict the F0 height of downstepped peaks in Central Catalan can only be tested in the utterances of speakers NG and DV, who were the ones that produced sequences of downstepped accents in neutral declaratives. As observed in Figure 6, speaker DV produced sentences with 3 peaks and speaker NG with 3 and 4 peaks. In this section, only the prediction of peaks in non-final utterance position will be analysed since the last peak of a sentence exhibited a much more abrupt F0 decay, which will be dealt with in section 4.4.

Liberman and Pierrehumbert (1984) noticed that in American English downstepping contours the height of a given peak was dependent on the F0 value of the preceding peak. That is, a correlation was observed between the height of two adjacent peaks: in general, the higher the first peak, the higher the second peak. Given this evidence, they decided to model downstep as an exponential decay to a constant nonzero asymptote, which they called the *reference line* of the speaker. The reference line (R) is a value lying between the last peak and the speaker's F0 minimum (i.e. final F0 value). Liberman and Pierrehumbert decided to introduce the notion of the reference line in their model to avoid predicting peaks that fall below the speaker's F0 minimum. The reference line and the F0 minimum of an asymptotically downstepping contour are illustrated in Figure 10.

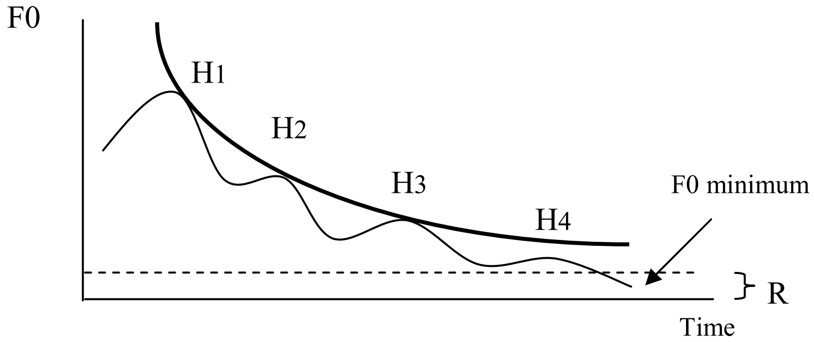

Figure 10. Adaptation of Liberman and Pierrehumbert's (1984) modelling of downstep as an exponential decay to a constant nonzero asymptote. R stands for the reference line.

Then, the ratio of decay between adjacent peaks (or downstep constant) is calculated by using the following equation:

$$\text{Downstep ratio } (r) = (P(x+1)-R)/(P(x)-R)$$

where P(x) = peak height of a peak in a position x, and R = the reference line value.

For each speaker, the reference line was calculated as the mid value between the mean F0 of the last "peak"[9] and the mean F0 minimum for all sentences. For speaker NG, R=184Hz and for speaker DV, R=145Hz. Once the reference values were obtained, the downstep ratios (r) between peaks were calculated for each sentence using the

[9] In final H accents, the highest F0 value was considered to be at the onset of the accented syllable, since, as pointed out in the next section, the last H of neutral declaratives does not show a clear peak.

equation above. The means of downstep ratios are 0.64 for speaker NG and 0.7 for speaker DV.

Having obtained these values, the F0 height of a given peak was calculated as a constant fraction (downstep ratio) of the previous one, scaled above the reference line of each speaker. This was obtained by means of the following equation:

$$P(x+1) = R + r \cdot (Px-R)$$

where P(x) = peak height of a peak in position x, r = downstep ratio, and R = reference line.

In Figure 11, the F0 values predicted by the model for a given peak are plotted against the observed F0 values for both speakers. For speaker NG, the predicted vs. the observed values of peak 2 and 3 are represented. For speaker DV, the values of peak 2 are shown. In each graph, the dotted line is a reference line which indicates the values when x = y (predicted = observed). The solid line is the regression line. R^2 values are included in each graph.

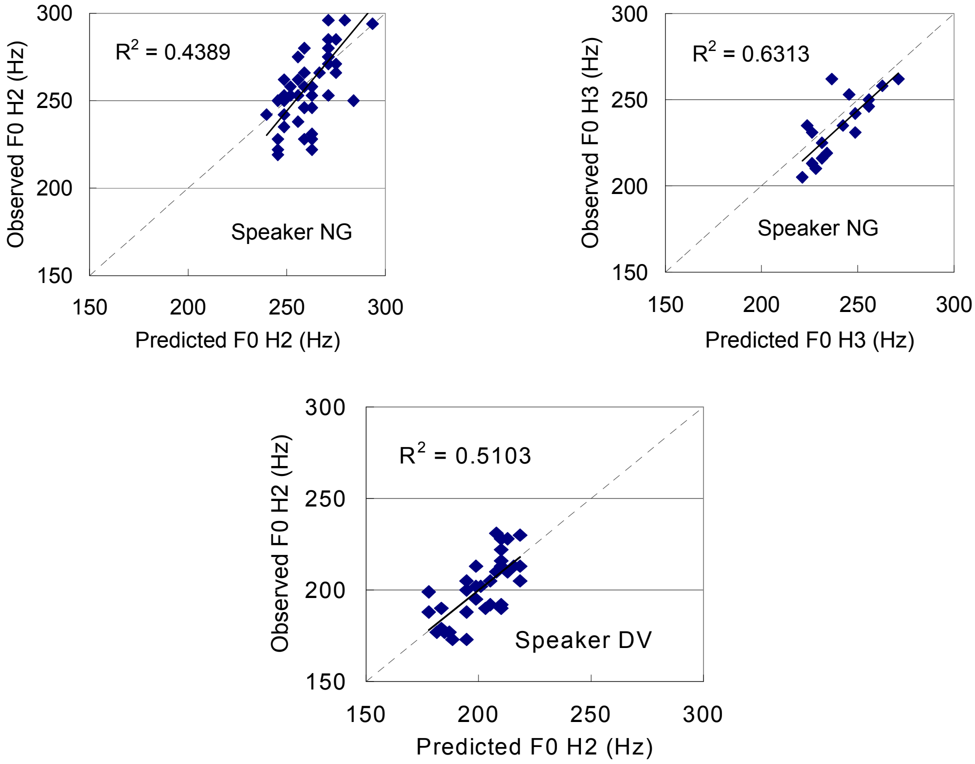

Figure 11. Predicted F0 values of a given peak (H2 or H3) as opposed to the observed F0 values of the same peak for speakers NG and DV. Dotted lines indicate x = y and solid lines are the regression lines.

The results observed in the graphs of Figure 11 showed that the model predicts the F0 height of a given peak quite successfully since most observations are located quite close to the x = y line. This indicates that the proposal of modelling downstep as a fixed proportion of the previous peak relative to the speaker's reference line seems to work in Central Catalan. These results support the proposal of the TS approach (Pierrehumbert, 1980; Liberman and Pierrehumbert, 1984) to analyse F0 contours as a series of H (and L) tones that are controlled at a local level.

4.4 H with final lowering or L target?

One of the questions that arises from the data is whether the last pitch accent in neutral declaratives is a strongly downstepped H (i.e. with final lowering) or an L target, as suggested in Prieto (1995) and in Bonet (1984) for Central Catalan, or in Sosa (1999) for Spanish. The main problem in classifying the last pitch accent as H derives from the fact that there is no evident peak in the F0 contour. Instead the last accent is realised as a progressively falling slope (see Figure 5 for an example). Based on perceptual and acoustic evidence, studies such as Prieto (1995) propose that the last pitch accent of a Central Catalan multi-accented declarative is L. On the other hand, the accent of a single accented declarative is H, since an F0 peak is observed in the F0 contour. Even though this proposal is supported by phonetic detail, it presents some phonological problems since two tonal entities (H and L) are used to describe a non-contrastive intonation (that is, a neutral declarative intonation). This problem may be overcome by interpreting the last pitch accent of multi-accented declaratives in the same way as that of a single accented declarative (i.e. H) and by explaining that in multi-accented sentences the H has a low realisation due to a final lowering effect. This view is investigated in this study on acoustic and theoretical grounds.

If we look at the F0 values corresponding to what has been classified as the last H of a declarative sentence in Figure 6, we observe that the mean F0 of the last H is in fact very similar to the mean F0 of the previous L targets. However, the last H is much higher than the following L. If the last pitch accent of a declarative was L, then the expected transition between an L pitch accent and the L edge tone would be low and level, but not falling[10]. In the data of all speakers, however, there is a falling movement from the last pitch accent to the L edge tone, indicating that the last pitch accent is a high (although strongly lowered) target. The falling transition from H to L is illustrated in Figure 12 for the data of all speakers.

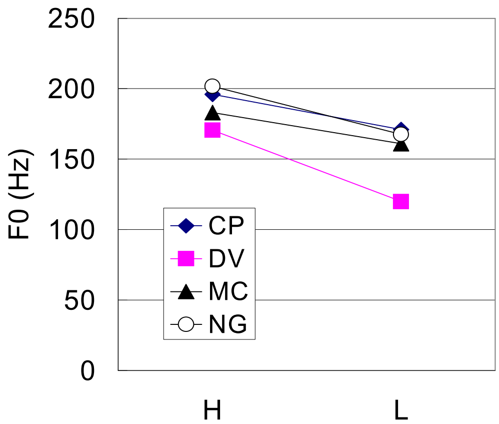

Figure 12. Mean F0 values of the last H (measured as the F0 height at the onset of the accented syllable) and the last L of a neutral declarative contour.

[10] Pierrehumbert (1980) and Grice (1995) claim that the expected F0 transition between two L tones is low and level and that between H and L involves an F0 fall.

In Mexican Spanish downstepping contours, Prieto et al. (1996) reported different behaviour as far as the last peak was concerned. Whereas two speakers produced a clear final peak, one speaker realised the final accent as a continuously falling slope, similar to that observed in the data of Central Catalan speakers. Prieto et al. showed that the final fall behaved as the final peaks: even though there was no F0 target in the F0 contour, a model which assumed a target value better predicted the F0 value than other models. In the following lines, the final fall observed in Central Catalan neutral declaratives is modelled following Liberman and Pierrehumbert's (1984) lowering constant. We assume that if the highest F0 value of the last pitch accent is accurately predicted by the lowering constant, the interpretation of this accent as H will be supported.

The fact that the last peak of a contour shows a more extreme F0 descent is observed in many languages (American English: Liberman and Pierrehumbert, 1984; Mexican Spanish: Prieto et al., 1996). This final lowering involves a more abrupt F0 lowering of the pitch in the final peak of the utterance, which falls below the values predicted by the downstep rule. This means that the F0 value of the last peak in an utterance cannot be modelled using the downstep ratio as for the non-final peaks. Thus, Liberman and Pierrehumbert (1984) proposed that utterance final peaks should be modelled by using a lowering constant, which is defined to be a fraction of the value of the peak predicted by the downstep rule above the reference line. The F0 value of the last peak is calculated as follows:

$$P = R + l \cdot (P_{(down)} - R)$$

where P = peak height of the last peak of the utterance, $P_{(down)}$ = peak height of the last peak predicted by the downstep rule, R = the reference line and l = the final lowering constant. The l was obtained by dividing the distance between the reference line and the observed final peak value by the distance between the reference line and the predicted downstepped peak value as below:

$$l = (P_{(obs)} - R)/(P_{(down)} - R)$$

In languages where a clear peak is observed in the F0 contour, P(obs) equals the observed peak height. However, since in Central Catalan there is no surface F0 peak on the last pitch accent of a declarative, we assume that P(obs) corresponds to the F0 height at the onset of the accented syllable (as in Prieto et al., 1996). According to this, the mean *l* value for the two speakers is 0.85 for speaker NG and 0.62 for speaker DV.

Once the *l* value was obtained for the two speakers, the predicted height of the last "peak" was calculated for all sentences and compared to the observed or real last peak value. The results are displayed in Figure 13, which shows predicted against observed F0 values for the last H. As before, the dotted line indicates the values when x = y (predicted = observed). The solid line is the regression line. R^2 values are included in each graph. Even though the R^2 values are not very high, the plots show a quite close association between predicted and real values since most observations are clustered fairly equally round the diagonal. These results suggest that the predictability of the final pitch accent highest F0 value by means of the lowering constant is quite accurate and thus the interpretation of this accent as a strongly lowered H is favoured.

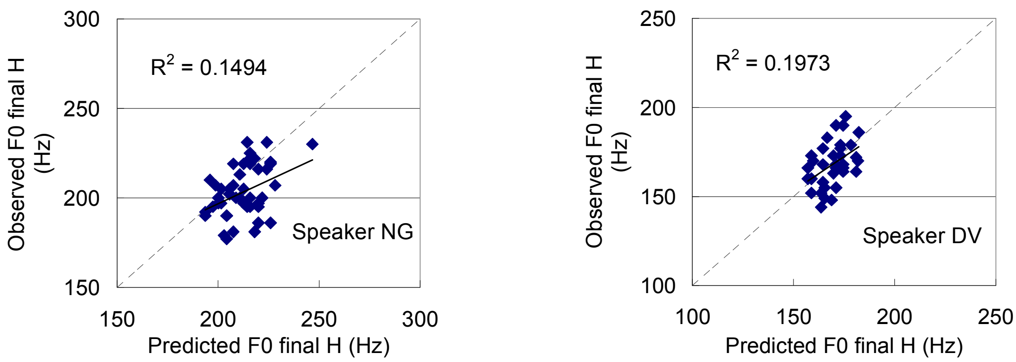

Figure 13. Predicted F0 values for the last H of an utterance as opposed to the real F0 values observed in the data for speakers NG and DV. Dotted lines indicate x = y and solid lines are the regression lines.

5 Discussion

The results of the scaling of F0 peak height in Central Catalan declaratives have shown that downtrend patterns are better explained as a linguistically controlled downstep than as a global declination component. This has been postulated due to the absence of a significant effect of temporal distance between peaks on the F0 value of a given peak. The F0 height of consecutive H values has proved to be independent of the number of intervening syllables (or temporal distance) between peaks, indicating that time-dependent declination is not active. This behaviour is consistent with the models that claim that F0 downtrends are better explained as a deliberate use of step accents (Pierrehumbert, 1980; Liberman and Pierrehumbert, 1984; Pierrehumbert and Beckman, 1988). Thus, the data support a linear interpretation of F0 contours, as claimed by the TS models, rather than a superpositional interpretation, as in the CI tradition.

Consequently, in line with Liberman and Pierrehumbert's proposal, downstepping H values in Central Catalan were predicted as a constant fraction (downstep ratio) of the previous peak, scaled above the reference line of each speaker. This model has successfully accounted for downstepping patterns in several languages (English: Pierrehumbert, 1980; Liberman and Pierrehumbert, 1984; Ladd, 1983, 1984, 1993; Japanese: Pierrehumbert and Beckman, 1988; Mexican Spanish: Prieto et al., 1996; or Dutch: van den Berg et al., 1992). In all these languages, descending H patterns could be explained as the repeated application of the downstep rule.

Similar to other languages, utterance-final peaks in Central Catalan underwent a much greater F0 lowering than that predicted by the downstep rule. In the data of all speakers, this final lowering was not realised as a compressed final peak but as continuously falling slope. This behaviour was also observed in Mexican Spanish (Prieto et al., 1996). Despite the lack of a clear peak, final accents in Central Catalan declaratives have been analysed as H targets and have been modelled by means of a lowering constant, which applies to the output of the downstep rule. Following Prieto et al., the F0 value at the syllable onset was taken as the target "peak" value.

The interpretation of the final accent in declarative sentences as a strongly lowered H accent has some advantages over those studies that interpret this final accent as an L tone (Prieto, 1995; or Bonet, 1984). One of the most important implications is that the same phonological entity (H) can be used to describe the final accent of both single and multi-accented declaratives. In Prieto (1995), whenever a declarative sentence consisted of only one accent, this was interpreted as H since an F0 rise was observed in the F0 contour. On the other hand, whenever a declarative contained more than one accent, the last one was described as L due to the lack of an F0 peak in the F0 contour. This interpretation triggers the undesirable consequence that two different phonological entities are used to describe the final accent of neutral declaratives, depending on the number of accents per sentence. Even though the acoustic evidence does show a different pattern in the F0 contour of single and multi-accented utterances, the two patterns do not trigger a contrastive opposition. This problem is sorted out by postulating the same entity (H) for both single and multi-accented sentences and by explaining the lack of F0 peak in multi-accented sentences as the result of final lowering.

Overall, the results obtained in the data support the main tenets of the TS approach that F0 contours can be analysed as a series of primitives controlled at a local level (as in Pierrehumbert 1980 and Liberman and Pierrehumbert 1984).

6 Conclusion

This paper has examined the nature of peak F0 values in Central Catalan neutral declaratives. The results showed no effect of temporal distance (or number of intervening syllables) between peaks on the scaling of H values. This indicates that peak downtrends in Central Catalan are better explained as a linguistically controlled accent-by-accent downstep than as a global time-dependent declination. As in English (Liberman and Pierrehumbert, 1984), Mexican Spanish (Prieto et al., 1996) and Dutch (van den Berg et al., 1992), peak height in Central Catalan can be accurately predicted as a constant proportion of the height of the previous peak. Similarly to these languages, utterance final peaks (or progressively falling slopes) undergo a greater amount of lowering than that expected by the downstep rule. Hence, the final H is better explained by means of a particular ratio of decay (lowering constant), which is higher than the downstep ratio.

References

Beckman, M. E. and J. Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology Yearbook* 3: 255-310.

Bonet, E. 1984. *Aproximació a l'entonació del català.* Tesi de llicenciatura. Universitat Autònoma de Barcelona.

Bruce, G. 1977. *Swedish word accents in sentence perspective.* Lund: Gleerup.

Clements, G. 1983. The hierarchical representation of tone features. In I. Dihoff (ed.). *Current approaches to African linguistics, Vol 1.* Dordrecht: Foris.

Cooper, W. F. and J. M. Sorensen. 1981. *Fundamental frequency in sentence production.* Heidelberg: Springer.

Estebas-Vilaplana, E. 2000. The use and realisation of accentual focus in Central Catalan with a comparison to English. PhD dissertation. University College London.

Fujisaki, H. 1983. Dynamic characteristics of voice fundamental frequency of speech and singing. In P. F. MacNeilage (ed.). *The production of speech.* New York and Berlin: Springer-Verlag. 39-55.

1988. A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour. In O. Fujimura (ed.). *Vocal physiology: voice production, mechanisms and functions.* New York: Raven. 347-355.

Gårding, E. 1983. A generative model of intonation. In A. Cutler and D. R. Ladd (eds.). *Prosody: Models and measurements.* Berlin: Springer. 11-26.

Grabe, E. 1998. *Comparative Intonational Phonology: English and German.* Wageningen: Ponsen and Looijen bv.

In press. Fundamental frequency downtrends in English and German. *Language and Speech.*

Grice, M. 1995. Leading tones and downstep in English. *Phonology* 12: 183-234.

Ladd, D. R. 1983. Peak features and overall slope. In A. Cutler and D. R. Ladd (eds.). *Prosody: Models and measurements.* Berlin: Springer. 39-52.

1984. Declination: a review and some hypotheses. *Phonology Yearbook* 1: 53-74.

1993. In defense of a metrical theory of intonational downstep. In H. van der Hulst and K. Snider (eds.). *The phonology of tone: the representation of tonal register.* Berlin: Mouton de Gruyter.

Liberman, M. and J. Pierrehumbert. 1984. Intonational invariance under changes in pitch range and length. In M. Aronoff and R. Oerhle (eds.). *Language sound structure.* Cambridge, MA: MIT Press. 157-233.

Lieberman, P. 1967. *Intonation, perception, and language.* Cambridge, MA: MIT Press.

Meyers, L. 1976. Aspects of Hausa tone. *Working Papers in Phonetics* 32. University of California, Los Angeles.

Möbius, B. 1993. *Ein quantitatives Modell der deutschen Intonation. Analyse und Synthese von Grundfrequenzverlaeufe.* Niemeyer: Tübingen.

Nolan, F. 1995. The effect of emphasis on declination in English intonation. In J. Windsor Lewis (ed.). *Studies in general and English Phonetics. Essays in honour of Professor J. D. O'Connor.* London: Routledge. 241-254.

Pierrehumbert, J. 1980. *The Phonology and Phonetics of English Intonation.* PhD dissertation. Bloomington: Indiana University Press.

Pierrehumbert, J. and M. E. Beckman. 1988. *Japanese Tone Structure.* Cambridge, MA: MIT Press.

Poser, W. J. 1984. The phonetics and phonology of tone and intonation in Japanese. PhD dissertation. MIT.

Prieto, P. 1995. Aproximació als contorns tonals del català central. *Caplletra* 19: 161-186.

Prieto, P., C. Shih, and H. Nibert. 1996. Pitch downtrend in Spanish. *Journal of Phonetics* 24: 445-473.

Sosa, J. M. 1999. *La entonción del español.* Barcelona: Cátedra.

Thorsen, N. 1980. Intonation contours and stress group patterns in declarative sentences of varying length in ASC Danish. *Annual report of the Institute of Phonetics* 15: 13-47. University of Copenhagen.

1983. Two issues in the prosody of Standard Danish. In A. Cutler and D. R. Ladd (eds.). *Prosody: Models and measurements.* Berlin: Springer. 27-38.

Vaissière, J. 1983. Language-independent prosodic features. In A. Cutler and D. R. Ladd (eds.). *Prosody: Models and measurements.* Berlin: Springer. 53-66.

van den Berg, R., C. Gussenhoven and T. Rietveld. 1992. Downstep in Dutch: implications for a model. In G. J. Docherty and D. R. Ladd (eds.). *Papers in Laboratory Phonology II.* Cambridge: Cambridge University Press. 335-358.

Appendix

Sentences with three stresses

Jo menjava melmelada
La Rosa llegia llibres
L'ou bullia a l'olla
La mare menja arengades
L'au venia de l'illa
L'Emília vol amanida
Ell anava a Girona
L'Anna vivia a Vilavella
La núvia rentava la faldilla
En Juli diu bajanades
La Mila nega la maionesa
Els homes llimaven l'armari
En Joan domina l'àlgebra
La Glòria ve de Vilanova
La Rosa rega els geranis
En Joan anava al museu
La Mireia remena l'olla
En Jaume mou la galleda
En Jordi mira les mones
La Neus anima els alumnes
En Jordi guardava medalles
La Marina beu llimonada
L'Emili mana els obrers
La Remei mima les nenes
En Ramon olora la benzina

Sentences with four stresses

L'Eva guarda monedes romanes
La Núria vol una germana menuda
L'illa brillava amb la llum de l'albada
L'Isidre du una gavardina vermella
L'home venia llimones madures
La Mari du una sivella daurada
L'Amèlia duia robes negres
Les noies vénen de veure l'examen
Les ovelles volen jeure a l'herba
En Raimon volia arruïnar la germana
La Lluïsa havia d'arribar diumenge
Les noies volen rebre la reina
Les mares veuen riure les nenes
En Juli hauria de demanar l'hora
Les nenes haurien de donar l'enhorabona
Els joves volen viure la vida
L'àvia va a raure a la masia
La Laura volia dividir l´herència

La mare havia de menjar verdura
L’Elena mira de moure el mobiliari
Els homes volen beure aiguardent
L’Elena volia munyir l’ovella
La Ramona volia mullar la roba