



Prosodic boundaries in subordinate syntactic constructions

Manon Lelandais, Gaëlle Ferré

LLING – Université de Nantes, FRANCE

manon.lelandais@univ-nantes.fr, gaelle.ferre@univ-nantes.fr

Abstract

Based on a video recording of conversational British English, this paper tests whether several different subordinate syntactic structures are evenly vocally integrated to their environment. "Secondary constructions" have been described in linguistics as dependent, subordinate forms elaborating on primary elements of discourse. Although their verbal and vocal characteristics have been deeply analysed, few studies have provided a qualified picture of their vocal integration. Beyond showing that secondary constructions are not evenly dependent on their environment, the results suggest that prosody demarcates secondary constructions more than it integrates them. The creation of a break preferentially takes place retrospectively, immediately after the subordinate structure through rhythmic features and/or pitch upsteps.

Index Terms: subordination, embedding, demarcation, boundaries.

1. Introduction

This article deals with subordination in spontaneous speech, more specifically on the sequences containing "secondary constructions" in English, operating at the syntactic level of modification.

In syntactic and discourse studies, "secondary material" refers to elements modifying or specifying some primary features, often described as additions associated to another propositional content in the host or embedding structure [1]. This paper focuses on adverbial clauses, appositive clauses, and restrictive relative clauses, as illustrated in examples (1-3). In (1), the adverbial clause specifies the circumstances in which *I tried driving once in her car* is valid as an utterance. The adverbial clause restricts the spatial scope in which the referential elements must be understood.

(1) **Adverbial clause (see appendix at the end of the paper for transcription conventions)**

Rhianna L i tried driving once in her car
SC **when we were on a # little road in the countryside #**
R and hem (swallows) she said turn left #

In (2), the appositive relative clause qualitatively evaluates *a place called Tropicana*, which can however be identified independently.

(2) **Appositive relative clause**

Beth L and then we went into # a place called Tropicana #
SC **which was horrible (laughs) #**
R it's on Saint Mary's street near the castle

Lastly, in (3), the restrictive relative clause increases the relevance of *the Spanish girls*, creating a subcategory for this referent.

(3) **Restrictive relative clause**

Joey L the Spanish girls
SC **that were there #**
R on our second one

Secondary constructions are generally defined as dependent on another predication [2]. However, the literature shows little consensus in defining clear scopes and boundaries for these forms [3], [4]. This study therefore questions whether they all express the same degree of dependence upon their context.

We investigate the production process of secondary constructions in English, focusing on demarcation at the segmental and suprasegmental levels. The main hypothesis, arising from the consensus in the previous results, is based on the capacity of these constructions to show distinct forms of autonomy in function of their syntactic type. Different degrees of prosodic demarcation are consequently identified from this new perspective, providing a qualified picture of their insertion in discourse.

2. Theoretical background

2.1. Syntactic subordination

In the traditional categorial division of clause complexes into two uneven and complementary syntactic subgroups, i.e. a main clause and a subordinate, modifiers are viewed as optional constituents functioning at a phrasal or clausal level [5], arising from the concept of minimal utterances: some elements of the message are deemed semantically useful without standing as constitutive elements. However, this acceptance has been reproved by a number of linguists (e.g. [6], [7]), described as imprecise for analysing spontaneous speech, especially regarding the nature of introductory elements.

2.2. Prosodic subordination

Prosodic subordination is essentially achieved through intonation [8]. Throughout a vocal paragraph, pitch height naturally declines in a progressive manner [9]. A subordinate unit is signalled through downwards changes in key (i.e. major levels in a speaker's pitch range) or in pitch height. Intonation can convey subordinating information that is not marked through verbal means. Syntactic units can prosodically be made autonomous or grouped depending on the speaker's semantic and/or pragmatic strategies [10], [11].

To integrate a prosodic unit to an adjacent segment, F0 (i.e. Fundamental Frequency = pitch) generally rises on the final syllable of the inserted segment, which indexes it as

prefacing further speech, continuing the paragraph and the ongoing point [12]. A downstepped tone compared to a preceding high tone corresponds to the general neutral relationship between two prosodic groups, often used to express seamless continuity [9].

On the contrary, a variation on the initial syllable signals a boundary. Likewise, a discourse segment featuring a low final syllable (termination contour) does not embed the following segment, and is autonomous regarding what ensues. Boundaries in speech can also be created with a variation of tempo [13]. Silent pauses conjointly participate in the segmentation of discourse [14].

3. Method

3.1. Corpus recording

The corpus used for this study, ENVID, is a collection of dialogues in British English. This collaborative corpus gathers video recordings realised between 2000 and 2012. Five dialogues were selected, making up a total of 2 hours and 10 minutes of interaction. Each interaction was recorded in a soundproof studio, guaranteeing its prosodic treatment. The participants are British people aged 20 to 23. Each participant had a lavalier microphone, providing two separate audio tracks. Two audio files corresponding to each microphone were created in a WAV format, facilitating the analysis of overlapping speech.

3.2. Corpus transcription

The corpus was first edited in Praat [15] for a standard orthographic transcription, in which secondary constructions were localised and coded on a separate track as SC. All the annotations made in Praat were then exported into Elan [16], an annotation tool facilitating data pooling and extraction.

3.2.1. Syntactic annotation

A total of 228 forms were annotated in the corpus, representing 4.82% of the total speaking time (i.e. 2.01 form/min): 88 adverbial clauses (1.88% of speaking time – 0.78 form/min), 81 restrictive relative clauses (1.71% of speaking time – 0.71 form/min), and 58 appositive relative clauses (1.23% of speaking time – 0.52 form/min)

The selection targeted forms without an interruption, surrounded with immediate left and right co-texts other than a single silent pause yielding the speaking turn.

The selected forms were classified according to their syntactic type in Praat (adverbial clause, appositive clause, restrictive relative clause). A second track delimitates their environment: the preceding tone-unit as described below was labelled L (left co-text), the subsequent one labelled R (right co-text).

3.2.2. Prosodic annotation

The corpus was segmented into tone-units, according to the British school of intonation [12], [13], based on dynamic pitch contours.

The Momel-Intsint algorithm [17], [18] was used for the automatic annotation of the F0 target points in the signal. Annotations are made in two respects: the algorithm notes pitch height (in Hz) on target syllables, which then allowed us to calculate mean F0 values for specific segments. The

algorithm also codes symbolic (relative) values of intonation, in which each measured F0 value is compared to preceding ones, i.e. significant changes in the F0 curve either regarding the speaker's pitch range (Top, Bottom) or regarding the neighbouring tones or sequences of tones (Upstep, Downstep, Same, Low, High). We are here particularly interested in values which indicate a significant pitch reset (Top, Bottom), or a significant change in pitch key (Upstep – change towards higher pitch range, Downstep – towards lower pitch range). We are also interested in the value "Same" which, if found in greater number in our sequences, would indicate that there is no break in between the different elements of the sequence.

Within each segment of the sequences under study, the nature of each nuclear contour was also coded manually (fall; fall-rise; rise; rise-fall; flat). Pitch key was then annotated in regards to each speaker's specific range (high; mid; low) on both the whole segments (L, SC, R) and the boundary (initial and final) syllables in these segments.

3.2.3. Working hypotheses

Based on the theoretical background defined by the literature and on our observations from the annotation, a specific list of vocal cues (including rhythmical and intonational patterns) is taken into account to survey different types of discourse boundaries. We mainly expect forms showing little autonomy.

If the constructions are not autonomous, they are expected to be integrated in the same tone-unit as the main clause, or to show continuation contours (final rises). They should be uttered in a low or mid-key, the usual declination line of the paragraph being followed without any break. These forms should not cause any important change in rhythm, featuring few pauses.

4. Results

This paper evaluates the prosodic autonomy of secondary constructions (integration vs. demarcation). We test whether these forms mainly create a break or whether they are preferentially integrated to their left and/or right co-text. After identifying and measuring the most relevant prosodic disjunctive cues drawing on our hypotheses, the three syntactic types can be placed on a continuum from dependence to autonomy. The particularities are detailed for each type from the most integrated form to the most disruptive.

4.1. Restrictive relative clauses

Restrictive relative clauses feature only two disruptive prosodic cues. They show a distinct duration pattern: SC is significantly the longest segment of the sequence in which it is inscribed (L: $F(54, 54) = 3.01, p < .0001$; R: $F(54, 54) = 4.48, p < .0001$). This type also features the highest number of intra-constituent silent pauses (44.6%, making a total of 10.26 sec). Most of these pauses are not accompanied with any significant increase or decrease in speech rate, and do not coincide with the syntactic boundaries; they are linked with displaced demarcation (i.e. a displaced pause making it possible for the speaker to keep the conversational turn while processing). Example (4) below illustrates these tendencies:

(4)	Michelle	L	oh she's that # woman	1.3 sec
		SC	that # looks after the Nottingham crowd (laughs)	3 sec
		R	that woman	0.5 sec

SC is the longest segment in the sequence. The conjunction *that* is separated from the clause it introduces with a silent pause, which is not correlated with any hesitation marker. While indexing a close link between the end of L and the beginning of SC, Michelle pauses outside syntactic boundaries to indicate she has not finished delivering the ongoing informational unit; her speaking turn is then secured.

Restrictive relative clauses are also the only type in which a majority of occurrences (55.6%) are directly bound to the left co-text under a same intonational unit, through flat contours on the left co-text's lexical items. In example (5) associated with Figure 1 below, SC shares a common prosodic and predicative unit with its left co-text, whose prosodic nucleus is in SC. An important break is however marked with R, through substantial extra-constituent silent pauses:

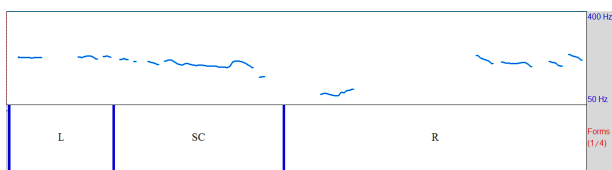


Figure 1: Praat pitch contour of example 5 (the transcription track shows segments – L, SC, R – in the sequence).

- (5) Zoe L and the daughter
 SC **that she's had all her life #**
 R the # cos there's the black one

Besides, they feature the highest distribution of integrative continuation contours (58% in L only; 16% both in L and SC as shown in (6) below; 9% in SC only).

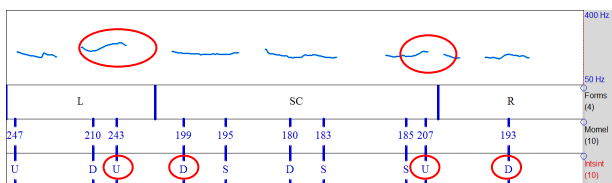


Figure 2: Two continuation rises in example 6 illustrated in the Praat curve and the Intsint codes (U = upstep, D = downstep).

- (6) Rhianna it's basically #
 L uh the school #
 SC **hem which is right # below**
 R that uh rents # the flat out

While L and SC display final rising contours, SC's and R's initial syllables are downstepped (Intsint "D" value). Although realised in distinct tone-units, these three segments are intonationally linked.

4.2. Adverbial clauses

Given their mobility in the macrostructure (i.e. they can be preposed to the main nuclear syntactic configuration or postponed), adverbial clauses show a diverse prosodic configuration with respect to boundedness and internal variation. However, some disjunctive markers are common to all forms.

Displaying three disruptive cues, adverbial clauses are demarcated thanks to intra-constituent intonational resources. These forms are both produced and followed with a greater pitch height variation within the neighbouring tones of the

same unit than the other types. Their immediate right co-text (i.e. R) features more demarcative tone values (hence pitch movement) than the other segments, as shown in Figure 3, associated with example (7):

- (7) Alex L (h) you know
 SC **when you're not allowed to laugh**
 R and then there's like a massive silence

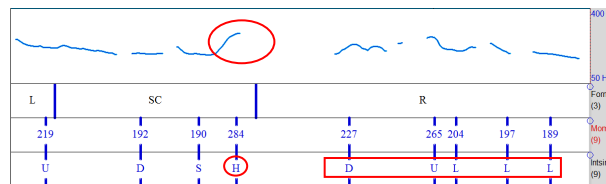


Figure 3: Final rise in example 7 illustrated in the Praat curve and the Intsint codes (H = high), followed by more pitch movement in R (L = low).

While sharing a single prosodic contour with L, SC's final syllable is higher than the initial one (284 Hz vs. 219 Hz) and does not match R's beginning, which is downstepped (Intsint "D" value). This preposed SC forms a landmark for what follows, embedding R, which nonetheless comprises more demarcative tone values.

Adverbial clauses are also demarcated thanks to rhythmic strategies. Although they do not feature any distinctive duration or speech rate, they display the highest distribution of filled pauses (19.7%), mostly occurring immediately after their production, i.e. in the right co-text. The most frequent hesitation marker is *hem*, as shown in R in (8) further below.

While L and SC are indexed as in a common cognitive unit through their fluent delivery and a structuring silent pause at the end of SC, R is indexed as requiring a greater processing load. Kate combines the discourse marker *like*, both suggesting approximation and functioning as a filler, with the hesitation marker *hem*.

- (8) Kate L that's why i haven't any cutlery or anything
 SC **when i came came back to uni last year #**
 R but we should have like hem (noise) a dinner #

Figure 4 shows once again more demarcative tone values in R. This postponed SC is realised in a distinct tone-unit. Yet, contrary to what could be expected, SC shows an initial upstep (Intsint "U" value) despite its lack of internal variation.

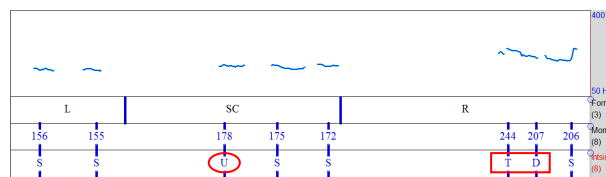


Figure 4: Initial upstep in example 8 followed by more pitch movement in R, illustrated in the Praat curve and the Intsint codes (U = upstep, T = top).

4.3. Appositive clauses

The prosodic signals indexing independence in appositive clauses are more numerous and varied than in the other forms we treated so far. They display the greatest number of vocal boundary cues with eight segmental and suprasegmental markers.

Appositive clauses are the shortest and fastest forms; both in their embedding sequence (L: $F(54, 54) = 2, p < .05$; R:

$F(54, 54) = 1.9, p < .05$) and compared to the other types (duration restrictives: $F(54, 54) = 3.5, p < .0001$, adverbials: $p > .05$; speech rate: $p > .05$). This host sequence also reveals the highest distribution of extra-constituent silent pauses (36.5%, making up a total of 30.8 sec; restrictives: $F(57, 44) = 3.8, p < .0001$; adverbials: $F(57, 36) = 2.40, p < .005$), 61% of which are immediately following the subordinate segment (i.e. in between SC and R).

In addition, appositive clauses show a relatively high distribution of filled pauses (15.8%), most of them located after their production. They then show a very irregular rhythm. Example (9) illustrates these tendencies:

(9)	Tim	L	and they played this Irish #	1.7 sec
			tune	4.12 syll/sec
		Sc	which was awesome #	0.8 sec
				5 syll/sec
		R	this guy with <u>hem</u> # a	1.5 sec
			whistle	4.6 syll/sec

From the point of view of intonation, 98.3% of occurrences are uttered in a separate tone-unit. These forms are then vastly made autonomous. They also show less integrative rising contours: 79 % of the sequences containing them do not feature any rising tone, as shown by example (10) and Figure 5:

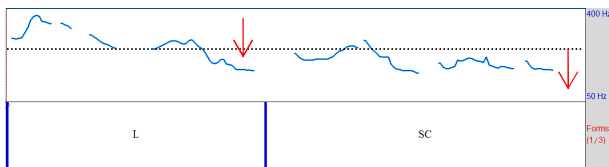


Figure 5: Two similar contours in example 10 with a lower pitch key in SC illustrated in the Praat curve.

(10)	Rhianna	L	even compared to Easyjet
		SC	which is another low cost company #
		R	yeah i hate Ryanair

If the absence of a pause between L and SC suggests their proximity at a propositional level, the definitive falling contours indicate they make two distinct discursive moves. While L supports the main theme (the staff is very rude), SC turns this new argument into a concession, with a change in point of view. While the pitch key in SC is lower, the similar contour signals similarity to L.

This lower key is another prototypical means for achieving disruption in these forms. While pitch discontinuity marks the segment following the subordinate form in the other types, 60% of speakers produce appositive clauses with a simultaneous change in pitch height (>20 Hz; $p < .05$ for these speakers). Some forms are then realised with a genuine "parenthetical" intonation (i.e. lower F0 and no modulation [19]–[22]). The following segment is globally uttered with a higher key (1.48 in L; 1.31 in Sc; 1.40 in R – pitch key is normalised according to each speaker's pitch range; however, $p > .05$). An initial upstep on the first syllable of the right co-text corroborates a non-neutral interval between SC and R.

5. Discussion

Our analysis confirms that the different syntactic types can be distinguished in their degree of autonomy. They can be positioned on a continuum, from integration (restrictive relative clauses) to autonomy (appositive clauses), going

through intermediate combinatory strategies (adverbial clauses).

Restrictive relative clauses are the most intonationally integrated forms to their (left) co-text. However, they rhythmically stand out (duration, filled pauses).

Adverbial clauses are freer, reflecting their pragmatic ability to project an interpretative frame for several consecutive segments or to close a unit. Their intonation conveys change, indicating an alteration in the interpretative/informational networks rather than segmenting adverbial clauses as isolated units. The postponed forms present variations from the prototypical afterthought realisation [23]. Speakers also index more processing difficulties after the production of this type. However, these subsequent segments are more characterised with *hem*, signalling a local interruption, than with *uh*, signalling a strong boundary [24]–[26].

Appositive clauses are finally the most independent forms, mainly showing a total prosodic autonomy. They are the only forms displaying such an array of disjunctive cues, both at the segmental and suprasegmental levels. Disruption is then more perceptible, with the highest significant gap in pitch, and the weakest distribution of continuation contours.

Disjunction is mostly expressed by speakers through rhythmical resources, in all types of secondary constructions. Duration and extra-constituent silent pauses are especially drawn on to create a break. As far as intonation is concerned, initial pitch upsteps are widely used. The presence of these resources suggests that prosody demarcates secondary constructions more than it integrates them; their location indicates that prosodic boundaries are preferentially marked retrospectively, immediately after the secondary construction.

6. Conclusions

When analysing how semiotic units form larger sequences of action in discourse and conversation, spontaneous speech presents both complex chains of structures embedded in one another, and disruptions in which the discourse parts no longer follow one another. Secondary constructions introduce a break when they establish a different assertive position from the preceding utterance [22]. While this break can directly be created through syntactic or discursive means, prosody creates a break immediately afterwards through rhythmic features or pitch upsteps, signalling the previous elements have to be recontextualised.

This study aimed at demonstrating that a wide survey of prosodic resources for demarcation sheds new light on subordination, deriving from numerous interactions between segmental and suprasegmental cues. One way to widen the picture of prosodic boundaries would be to take final syllable lengthening into account. However, there is little consensus on the array of thresholds to be taken into account regarding the sampler types and weights of syllables; we are currently working to provide the most qualified and reliable criteria.

Appendix: transcription conventions

(h)	audible inbreath	#	pause
(...)	vocal activity (laughs, swallowing, sighs)	L	left co-text
SC	secondary construction	R	right co-text

7. References

- [1] R. Huddleston and G. K. Pullum, *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press, 2002.
- [2] C. Lehmann, "Towards a Typology of Clause Linkage," in J. Haiman & S. A. Thompson (Eds.) *Clause Combining in Grammar and Discourse*, Amsterdam and Philadelphia: John Benjamins, pp. 181–225, 1988.
- [3] H. Tao and M. McCarthy, "Understanding Non-restrictive Which-clauses in Spoken English, Which is Not an Easy Thing," *Language Sciences*, vol. 23, pp. 651–677, 2001.
- [4] E. Couper-Kuhlen, "Intonation and Grammar," in *An Introduction to English Prosody*, Tübingen: Max Niemeyer Verlag, pp. 139–157, 1986.
- [5] M. A. K. Halliday, *An Introduction to Functional Grammar*. London: Edward Arnold, 1985.
- [6] J. Haiman and S. A. Thompson, "'Subordination' in Universal Grammar," in *Proceedings of the Tenth Annual Meeting of the Berkeley Linguistics Society*, pp. 510–523, 1984.
- [7] W. Chafe, "Linking Intonation Units in Spoken English," in J. Haiman & S. A. Thompson (Eds.) *Clause Combining in Grammar and Discourse*, Amsterdam and Philadelphia: John Benjamins, pp. 1–27, 1988.
- [8] D. Bolinger, "Intonational Signals of Subordination," in *Proceedings of the Tenth Annual Meeting of the Berkeley Linguistics Society*, pp. 401–413, 1984.
- [9] A. Wennerstrom, *The Music of Everyday Speech. Prosody and Discourse Analysis*. Oxford: Oxford University Press, 2001.
- [10] P. Mertens, "Intonational Grouping, Boundaries, and Syntactic Structure in French," in *Proceedings of the ESCA Workshop on Prosody, September 27–29, Lund, Sweden*, pp. 156–159, 1993.
- [11] Y. Lee, "On the Role of Intonation and Pauses in the Interpretation of English Parenthetical Adverbs," *Linguistic Association of Korea Journal*, vol. 15, pp. 85–107, 2007.
- [12] J. C. Wells, *English Intonation: An Introduction*. Cambridge: Cambridge University Press, 2006.
- [13] D. Crystal, *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press, 1969.
- [14] J. Local, "Continuing and Restarting," in P. Auer & A. Di Luzio (Eds.) *The Contextualization of Language*, Amsterdam and Philadelphia: John Benjamins, pp. 273–296, 1992.
- [15] P. Boersma and D. Weenink, "Praat: Doing Phonetics by Computer." [Online]. Available: <http://www.fon.hum.uva.nl/praat/>. [Accessed: 30-Jan-2013].
- [16] H. Sloetjes and P. Wittenburg, "Annotation by Category: ELAN and ISO DCR," in *Proceedings of the 6th International Conference on Language Resources and Evaluation*, 2008. <http://www.lat-mpi.eu/tools/elan/>.
- [17] D. J. Hirst, "A Praat Plugin for Momel and INTSINT with Improved Algorithms for Modelling and Coding Intonation," in *Proceedings of the XVIth International Conference of Phonetic Sciences*, pp. 1233–1236, 2007.
- [18] B. Bigi, "SPPAS: a Tool for the Phonetic Segmentation of Speech," in *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2012), Istanbul (Turkey)*, pp. 1748–1755, 2012.
- [19] J. Hirschberg and B. Grosz, "Intonational Features of Local and Global Discourse Structure," in *Proceedings of the Workshop on Speech and Natural Language*, NY: Harriman, pp. 441–446, 1992.
- [20] A. Wichmann, "Spoken Parentheticals," *Gothenburg Studies in English*, no. 81, pp. 177–193, 2001.
- [21] J. Local, "Phonetic Detail and the Organisation of Talk-in-interaction," in *Proceedings of the 16th ICPhS, August 6–10, Saarbrücken, Germany*, 2007. <http://www.icphs2007.de/>
- [22] M. Lelandais and G. Ferré, "Multimodal Analysis of Parentheticals in Conversational Speech," *Multimodal Communication*, vol. 3, no. 2, pp. 197–217, 2014.
- [23] W. Chafe, "How People Use Adverbial Clauses," in *Proceedings of the Tenth Annual Meeting of the Berkeley Linguistics Society*, California, pp. 437–449, 1984.
- [24] E. E. Shriberg and R. J. Lickley, "Intonation of Clause-internal Filled Pauses," *Phonetica*, vol. 50, no. 3, pp. 172–179, 1993.
- [25] M. Swerts, "Filled Pauses as Markers of Discourse Structure," *Journal of Pragmatics*, vol. 30, no. 4, pp. 485–496, 1998.
- [26] H. H. Clark and J. E. Fox Tree, "Using 'uh' and 'um' in Spontaneous Speaking," *Cognition*, vol. 84, no. 1, pp. 73–111, 2002.