The Information Structure–Prosody Language Interface Revisited

Mónica Domínguez[1], *Mireia Farrús*[1], *Alicia Burga*[1], *Leo Wanner*[2,1]

[1]TALN Group, N-RAS Research Centre
Department of Information and Communication Technologies
Universitat Pompeu Fabra
[2]Catalan Institute for Research and Advanced Studies (ICREA)
{monica.dominguez|mireia.farrus|alicia.burga|leo.wanner}@upf.edu

Abstract

Several grammar theories relate information structure and prosody, highlighting a major correspondence between theme and rheme, and intonation patterns. Although these theories have been successfully exploited in some specific speech synthesis applications, they are mainly based on short default-order sentences, which limits their expressiveness for real discourse with longer sentences and complex structures. This paper revises these theories, identifying cases in which they are valid, and providing a new proposal for cases in which a more complex model is needed. Specifically, our experiments performed on real discourse from the Wall Street Journal corpus show that we need a model that: (1) foresees a hierarchical theme/rheme structure, and (2) introduces, apart from the traditional theme and rheme, a new element—the specifier.

Index Terms: information structure, thematicity, theme, rheme, prosody, ToBI.

1. Introduction

The influence of the information structure on intonation is widely reported in literature [1, 2, 3] under the heading of the "semantics-syntax-intonation language interface" [4, 5, 1, 6, 7]. Steedman [1] proposes a grammar theory that relates three different fields: syntax, semantics and intonation. Based on the theory stated by Beckman and Pierrehumbert [8] on intonation and information structure, Steedman establishes a main and recurrent correspondence between theme and rheme on the one side and intonation patterns on the other side. This correspondence has already been exploited experimentally in speech synthesis applications that serve as front ends in dialogue engines [9, 10]. However, it is not obvious that the experience gained in these applications can be transferred to, for instance, monologue generation: works such as [1] are based on rather short sentences with a simple structure and a default word order (SVO for English), which are not common for the genre of monologues. If we want to generate natural speech based on real discourse information, the hypotheses put forward in these works must be applicable to long sentences with complex syntactic structures as well. There are no descriptive studies on real data that provide evidence for their applicability or offer sound arguments that help revising these theories.

The aim of this paper is to test these theories at a small scale with a two-fold objective: (i) to validate them and, in case discrepancies between their proposed thematicity–intonation correlation and the observed correlation are identified, to determine when and why these discrepancies occur; and (ii) to propose a model that has the potential to capture better the thematicity–intonation correlation, especially in the case of complex linguistic constructions. Achieving this objective will also be instrumental for the establishment of a valid methodology for dealing with large corpora for the description of prosody at a deep structure level. To this end, this paper draws upon real discourse extracted from the Wall Street Journal and recorded in a professional setting.

The structure of this paper unfolds as follows. In Section 2, the theoretical background related to the annotation of both information structure and prosody is briefly explained. Section 3 presents the study in which we relate information structure and intonation patterns in the spirit of the "classical" theories, as well as the most outstanding findings in this respect. Section 4 suggests how at least some of the challenges encountered in the previous section can be met by drawing upon a more elaborated notion of information structure. The summary of our experiments and the conclusions we draw from them are sketched in Section 5.

2. Theoretical background

In what follows, we briefly introduce the notions of information structure and the information structure–prosody interface.

2.1. Information structure

The Information Structure (IS) (also known as Topic-Focus Articulation, TFA [11] in the Prague School [12], and Communicative Structure, CommStr, in the Meaning-Text Theory [6]) determines the "communicative" segmentation of the meaning of an utterance. This makes it central to the semantics–syntax–intonation interface [4, 5, 1, 6, 7] and therefore also to Natural Language Processing (NLP).

Steedman's work (which will be referred to in this paper as "the classical approach") is based on the interpretation of IS as a two-partite *thematicity* structure, with *theme* (that part of an utterance which connects it to the rest of the discourse) and *rheme* (what the utterance contributes to that theme) [1, p.655]. According to Steedman, it is also possible to have discontinuous themes and rhemes and all-rhematic sentences.[1] Consider an example of theme (Th)/rheme (Rh) distribution taken from [1]:[2]

(1) *Q: I know what Marcel SOLD to HARRY.*

[1]In fact, for Steedman, "the majority of themes in everyday utterances are null themes" [1, p.678], i.e., the majority of sentences are all–rhematic.

[2]For theme/rheme determination in a sentence, Steedman pictures, as is common in the field, the sentence in question as an answer to a hypothetical question.

But what did he GIVE to FRED?
A: (Marcel GAVE)$_{Th}$ (a BOOK)$_{Rh}$ (to FRED.)$_{Th}$

Mel'čuk [6] proposes a more complex thematicity structure. Thus, on the one hand, he distinguishes, apart from the traditional theme and rheme (whose definition, in general terms, coincides with Steedman's), a specifier element (SP), which sets up the context of the utterance; and, on the other hand, he defines thematicity over propositions rather than over sentences. The second feature implies that thematicity is *per se* a hierarchical structure: if a proposition is embedded, its thematicity partition will be embedded as well. In sentences containing coordinated propositions, there is a parallel thematicity structure (one partition by proposition)[3].

In further contrast to Steedman, Mel'čuk assumes that apart from existential and zero-argument propositions, which are allrhematic, any proposition has at least theme and rheme. Consider an example of theme (T)/rheme (R)/specifier (SP) distribution in the sense of Mel'čuk:

(2) $\{$*[Years ago]$_{SP}$, [he]$_{T}$ [collaborated with the new music gurus Peter Serkin and Fred Sherry in the very countercultural chamber group Tashi,* $\{$*[which]$_{T(P2)}$ [won audiences over to dreaded contemporary scores like Messiaen's Quartet for the End of Time]$_{R(P2)}\}_{P2}]_{R}\}_{P1}$

2.2. The classic information structure–prosody interface

Although all three elements of prosody (intonation, rhythm and stress) are equally important from a theoretical point of view, intonation is the most relevant feature for speech synthesis applications, since its correct prediction helps to obtain naturalness and variability in the generated speech [13]. As has been argued by many authors, among them by Beckman and Pierrehumbert [8] and Steedman [1], intonation is also directly correlated with IS. Beckman and Pierrehumbert identified six pitch accents and classified them as theme-rheme markers (see Table 1).

Table 1: *Pitch markers of theme and rheme (stated by [8]).*

	patterns
theme	L+H*, L*+H
rheme	H*, L*, H*+L, H+L*

As far as pitch accents are concerned, Beckman and Pierrehumbert suggest that the characteristic bitonals for theme and rheme are L*+H and H+L* respectively. Steedman [1] builds upon this theory and hypothezises on complete pitch accent/boundary tone (PABT) patterns, claiming that:

> *the intonational phrase L+H* LH%* [a clearly increasing Low-High pattern] *(among others) is associated with the theme, whereas the H*L% and H*LL%* [clearly decreasing High-Low] *tunes (among others) are associated with the rheme*; cf. [15, p.275, 16, p.28, 17].

Accordingly, Steedman correlates the theme/rheme in example (1) with intonation patterns (IP) as follows:

(3) *Q: I know what Marcel SOLD to HARRY.*
But what did he GIVE to FRED?
A: (Marcel GAVE) (a BOOK) (to FRED.)
L+H LH% H* L L+H* LH%*

In order to test Steedman's theme/rheme-IP correlation hypotheses, we have grouped together ToBI patterns resulting from automatic labeling and a further reduction process into three cathegories according to their final intonation curve typology, i.e. 'falling', 'rising' and 'flat'. We assume that our automatic labeling [20], based upon one main pitch accent within each intonational phrase, may not always coincide with detailed manual labeling accounting for all intonation events occurring along the pitch contour line, such that a need for broader categories arises in order to establish a comparable ground between our automatic labeling and traditional theories. Table 2 summarizes the collection of patterns we used in this experiment.

Table 2: *Intonation patterns classified according to their final intonation curve type.*

	patterns
falling	H*L%, L*+HL%, H*+LL%
rising	L*H%, L*+HH%, H*+LH%, H+L*H%
flat	L*L%, H*H%, H+L*L%

3. Validating the classic interface

In order to validate the classic information structure–prosody interface defined in terms of the correlation between theme/rheme of a sentence and its prosodic patterns, we carried out a number of experiments. In the context of these experiments, a non-expert native speaker of standard American English was instructed to read 109 sentences from the corpus annotated with information structure (composed of around 450 sentences from the Penn TreeBank).[4] The selection of those 109 sentences was based on the variation and complexity of their deep-linguistic (and thus also information) structures, making them prosodically interesting and useful for our study. The speaker was recorded in a professional recording studio to guarantee the quality of the sound signal. In a first stage, those sentences were annotated prosodically. In a second stage, the prosodic and theme/rheme patterns of these 109 utterances were assessed and contrasted with the patterns as prognosticated by Steedman's proposal.

3.1. Prosody annotation

Among several prosody annotations models, ToBI (Tone and Break Indices) [17] is the most widely used for annotating and adapting a markup language for open-source speech synthesizers such as Festival [18]. In our experiments, intonation patterns have also been annotated following the ToBI annotation convention. ToBI labels account for pitch accents (PA) within the intonational phrase (IP) and significant boundary tones (BT) within the sentence, which perfectly suits the purpose of these experiments. However, in contrast to most annotation exercises, which use a manual ToBI annotation,[5] we developed an own au-

[3]The hierarchical relations in a given thematicity segmentation are in practice controlled by parentheses (e.g. 'T(T)' will stand for "theme within the theme" and 'R(T)' for "rheme within theme". In coordinations and subordinations, each proposition is pointed out, e.g. T(P2) stands for the theme of the second proposition.

[4]The Penn Treebank was chosen as corpus base because it already contains the annotation at different linguistic levels (semantic and syntactic). Thus, we just added the information structure level to obtain all the information we need.

[5]The manual ToBI annotation has the advantage of being reliable and highly descriptive, but on the other hand it is subjective and very time consuming.

tomatic annotation interface that is based on AuToBi as initial labeling stage [19]. Since AuToBI only labels sentences word by word and our aim is to describe intonation patterns within intonational phrases (see [20] for details), we have processed the data automatically in a second stage to obtain a single ToBI pattern that is *a posteriori* manually validated and matched to the corresponding IP in the utterance. The advantage of working at the IP level is that we can correlate it with other layers, especially with IS.

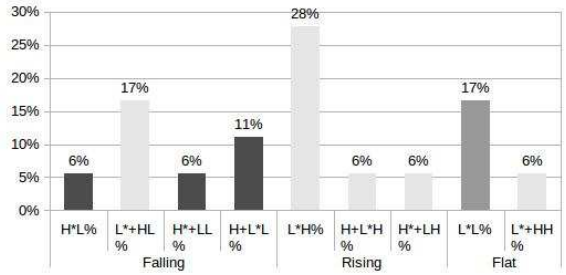
3.2. Assessment of the IS–prosody correlation

As already mentioned above, in the scope of our IS–prosody correlation assessment, we aimed to observe to what extent the classical proposals on the information structure–prosody interface can be applied to general (monologue) discourse with its rather complex sentential (and thus also thematic and prosodic) structures.

We analyzed pitch accents within the intonational phrase and complete intonation patterns (PABT — the combination of a pitch accent and a boundary tone). The analyzed themes include the set of patterns shown in Figure 1. All IPs are taken into account for calculating the percentages, regardless their position. It can be observed that rising PABT patterns together with L*+H tones represent a reasonable amount of the total (63%).[6]

(4) shows an example where the short theme, "Mister Kuehn", matches the pattern L*+H as expected by [1]; the same can be observed in (5), where the non-subjectival theme, "for some players", also matches the L*+H pattern.

Figure 1: *Theme ToBI Patterns.*



That is, as Steedman [1] claims there is indeed a tendency in themes to contain a rising pattern, at least in our recording, which includes only one speaker performing a reading task—although somewhat less than 40% of themes do not reflect the intonation pattern suggested in [1].

(4) *"Mister Kuehn, the company said, will retain the rest of the current management team".*

T	**Mister Kuehn**			
	H+L*H%			
R	**the company said**	**will retain the rest**	**of the current**	**management team**
	L*L%	L*+HL%	L*L%	L*L%

(5) *"For some players the lure is money up to fifteen thousand dollars a month".*

T	**For some players**			
	L*+HL%			
R	**the lure**	**is money**	**up to fifteen thousand**	**dollars a month**
	L*H%	L*H%	L*+HL%	L*L%

[6]It obviously remains to be proved that this tendency is kept in a big recorded corpus with several speakers and different registers, such as spontaneous speech.

However, in both (4) and (5), the rhemes do not show the expected pattern. In (4), there is no explanation for the L*L% pattern of the IP "the company said" and in (5), there is no explanation for any of the rising patterns found in the rheme span. Consequently, these two examples suggest that [1]'s approach to include everything, apart from theme, into a flat rheme span lacks the prediction accuracy we would need for speech synthesis applications.

4. Towards a more accurate IS–prosody interface

[1] is based on a linear dimension of thematicity. However, the study of our recorded material suggests that if we apply a hierarchical three-partite thematicity structure in the sense of Mel'čuk, we may be able (i) to find a justification for the discrepancies we saw in (4) and (5) between the prognosticated and the observed rheme patterns; (ii) propose a more accurate modelization of the intonation–thematicity correlation for the about 40% of non-coincident patterns within the theme span captured in Figure 1. Consider in (6) and (7) the sentences already cited in the examples in (4, 5), with a thematicity annotation as suggested by Mel'čuk.

(6) *"Mister Kuehn, the company said, will retain the rest of the current management team".*

T1	**Mister Kuehn**			
	H+L*H%			
SP1	**the company said**			
	L*L%			
R1	**will retain the rest**	**of the current**	**management team**	
	L*+HL%	L*L%	L*L%	

(7) *"For some players the lure is money up to fifteen thousand dollars a month".*

T1	**For some players**		
	L*+HL%		
T1(R1)	**the lure**		
	L*H%		
R1(R1)	**is money**	**up to fifteen thousand**	**dollars a month**
	L*H%	L*+HL%	L*L%

As already pointed out in Section 2.1, the notion of thematicity in the sense of Mel'čuk's includes, apart from theme and rheme, specifier elements. We have observed in our corpus that the identification of specifier elements provides very stable intonation patterns, being L*L% the commonest pattern, especially in reported speech. The IP "the company said" is a specifier (SP) and indeed carries this pattern. Another example in (8) also shows that the specifier element introduced in our information structure-prosody interface can be intonationally characterized as a separate entity from theme and rheme. This observation reinforces the tripartite division proposed by Mel'čuk.

(8) *"There is a large market out there hungry for hybrid seeds, he said".*

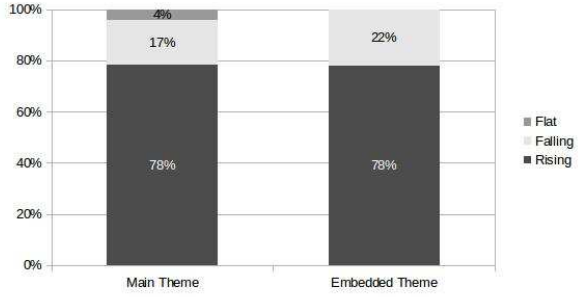
R1	**There is a large market out there**	**hungry**	**for hybrid seeds**
	L*+HL%	L*H%	L*L%
SP1	**he said**		
	L*L%		

In (7), the rheme element of (5) is hierarchically decomposed into an embedded theme/rheme structure. "The lure" becomes an embedded theme of the rheme and "is money up to fifteen thousand dollars a month" becomes the rheme of the rheme. As a consequence, we can at least explain why "the lure" carries a rising pattern: it is theme.

Going back to the results presented in Figure 1, a deeper analysis has been carried out to shed some light on the 40% of patterns within the theme span that do not coincide with the L*+H pattern in that they are either flat or falling and thus subsumed by the category Steedman called "among others". These diverging patterns are mostly found in complex structures, i.e., long themes containing more than one IP, coordinated sentences, subordinated clauses, etc. Consequently, an information structure model that can cater for complex utterances may be suitable to target this linguistic reality.

In order to test whether Mel'čuk's proposal of thematicity can be used as such a model, another experiment has been carried out. Thematicity, in accordance with Mel'čuk's definition, has been labeled in our selected corpus of 109 sentences following the guidelines outlined in [21]. Taking into account only theme spans which are prosodically marked up, we have classified their intonation patterns into rising, falling and flat, as shown in Table 2. The pattern L*+H L% is included into the rising pattern classification due to the fact that the state of the art considers this rising PA as characteristic of theme tunes. As a result, we have found that both main and embedded themes contain rising intonation patterns at equal rates of 78%. Consequently, we can affirm that embedded themes behave as main themes in terms of intonation when they are prosodically marked up. If we take into account the total of spans analyzed, main themes containing a rising pattern will characterize 34%, and when we add embedded themes we reach 50%. Therefore, we can conclude that a hierarchical approach to thematicity has the potential to provide more clues than traditional approaches when attempting to predict intonation patterns under complex communicative conditions.

Figure 2: *Comparison between main and embedded theme spans.*



In Section 3, we have already shown several examples where both main and embedded themes are characterized by rising tunes. The fact that data from a broader selection of utterances at an intra-speaker level also show this tendency sets a sound ground for further insight into the characterization of diverging intonation patterns based upon a hierarchical IS-prosody interface.

However, despite these advances in the explanation of the IS-prosody correlation there is still a substantial amount of cases that call for further investigation. These are, first of all, cases where a whole theme is not intonationally marked: 19% of primary themes and 49% of embedded themes. The example in (9) shows a sentence where the theme is a deaccented subject pronoun ("he") that, intonationally, forms part of the first IP of the rheme.

(9) *"Nevertheless he said he is negotiating with plant genetic to acquire the technology to try breeding hybrid cotton".*

SP1	**Nevertheless**				
	H*L%				
SP2	**he said**				
	L*L%				
T1	**he**				
R1	**is negotiating**	**with plant genetic**	**to acquire the technology**	**to try breeding**	**hybrid cotton**
	L*+HL%	H*+LL%	H+L*H%	H+L*H%	L*L%

The characterization of this kind of thematicity is also worthwhile to be born in mind, and further experiments will aim to find out if there is a characteristic IP in those cases. So far, we have observed that IPs containing a deaccented theme tend to bear a rising pattern, but the tendency is not so clear, and it seems that more layers interact at this level. Therefore, further research needs to be carried out, also including more speakers of the same dialect in order to test inter-speaker consistency and thus be able to draw more definite conclusions.

5. Conclusions

We have presented results of a descriptive study on a limited set of sentences from a wider corpus, attempting to determine which intonation patterns better characterize thematicity in real utterances, with the ultimate goal to build a model suitable for use in speech synthesis applications. We have observed that classical theories on IS–prosody interface are partially applicable in that themes of a specific (simple) nature have been proved to be characterized by a rising tune. On the other hand, the flat theme/rheme interpretation prevailing in these theories fails to explain complex linguistic structures. Drawing upon more advanced proposals on information structure, we have shown that further descriptive work needs to be done in order to accurately and concisely describe the IS–prosody correlation. Complex and hierarchical thematicity structures as well as the introduction of specifiers into the thematicity structure are bound to render positive results. Furthermore, the tri-partite division and the possibility of hierarchy are features of thematicity that facilitate a fine-grained communicative partition of complex utterances and thus a more detailed projection between the different layers of the semantics–syntax–intonation interface (or, more especifically, a more accurate description of the prosodic patterns related to each span). Our work can also be considered as detailing Steedman's proposal where he remains vague, stating that the patterns he identifies are only a few "among others".

The ultimate goal of developing a model combining prosodic and communicative structures for speech synthesis requires a deeper insight into descriptive studies on how these two linguistic layers interact. A good understanding of the structure and sequence of intonation patterns, as well as rare and/or exceptional cases will hopefully provide clues to more efficient NLP.

6. Acknowledgements

Parts of this work have been funded by a grant from the European Commission under the contract number FP7-ICT-610411. The second author is partially funded by a grant from the Spanish Ministry of Economy and Competitivity in the framework of the Juan de la Cierva fellowship program (JCI-2012-12272).

7. References

[1] Steedman, M., "Information structure and the syntax-phonology interface", Linguistic Inquiry, 4(31):649–685, 2000.

[2] von Heusinger, K., "Intonation and information structure". Habilitation dissertation, University of Konstanz, 2007.

[3] Büring, D., "Semantics, intonation and information structure". The Oxford Handbook of Linguistic Interfaces, ed. by Gillian Ramchad and Charles Reiss. Oxford University Press, 2007.

[4] Lambrecht, K., "Information structure and sentence form: Topic, focus and the mental representations of discourse referents". Cambridge University Press, Cambridge, 1994.

[5] Hajičová, E., Partee, B. and Sgall, P., "Topic-Foucs Articulation, Tripartite Structures, and Semantic Content. Kluwer Academic Publishers, Dordrecht, 1998.

[6] Mel'čuk, I. A., "Communicative Organization in Natural Language: The semantic-communicative structure of sentences". Benjamins Academic Publishers, Amsterdam, 2001.

[7] Erteschik-Shir, N., "Information Structure: The Syntax-Discourse Interface". Oxford University Press, Oxford, 2007.

[8] Beckman, M. and Pierrehumbert, J., "Intonational structure in Japanese and English". Phonology Yearbook, 3: 255–310, 1986.

[9] Moore, J.D., Foster, M.E., Lemon, O. and White, M., "Generating tailored, comparative descriptions in spoken dialogue". In Proceedings of FLAIRS-04, 917–922, Miami Beach, USA, 2004.

[10] White, M., Clark, R.A.J. and Moore, J.D., "Generating tailored, comparative descriptions with contextually appropriate intonation". Computational Linguistics, 36(2): 159–201, 2010.

[11] Sgall, P., "Functional sentence perspective in a generative description of language". Prague Studies in Mathematical Linguistics, 2: 203–224, 1967.

[12] Daneš, F., "Zur linguistischen Analyse der Textstruktur". Folia Linguistica, 4: 72–78, 1970.

[13] Llisterri, J., Carbó, C., Machuca, M.J., De la Mota, C., Riera, M. and Ríos, A., "El papel de la lingüística en el desarrollo de las tecnologías del habla". VII Jornadas de Lingüística, 137–191, 2003.

[14] Steedman, M., "Structure and Intonation". Language, 68: 260–296, 1991.

[15] Steedman, M., "Surface Structure, Intonation, and Focus". In Ewan Klein and Frank Veltman (eds.), Natural Language and Speech, Proceedings of the ESPRIT Symposium, Brussels, 21–38, 260–296. Dordrecht: Kluwer, 1991.

[16] Steedman, M., "The Syntactic Process", MIT Press, Cambridge MA, 2000.

[17] Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J. and Hirschberg, J., "ToBI: a standard for labelling English prosody". Proceedings of the ICSLP, vol. 2, 867-870, Sydney, Australia, 1992.

[18] Steedman, M., "Using APML to specify intonation". Magicster Project Deliverable 2.5. University of Edinburgh, 2005. Available at http://www.ltg.ed.ac.uk/magicster/deliverables/annex2.5/apml-howto.pdf

[19] Rosenberg, A., "AutoBI - a tool for automatic ToBI annotation". Proceedings of Interspeech, 146–149, 2010.

[20] Domínguez, M., Farrús, M., Burga, A. and Wanner, L., "Automatic extraction of prosodic patterns for speech synthesis applications". Submitted to Speech Prosody Conference 2014.

[21] Bohnet, B., Burga, A. and Wanner, L., "Towards the Annotation of Penn TreeBank with Information Structure". Proceedings of the Sixth International Joint Conference on Natural Language Processing, 1250–1256, Nagoya, Japan, 2013.