

A Galician Syntactic Corpus with Application to Intonation Modeling

Montserrat Arza¹, José M. García Miguel², Francisco Campillo³,
Miguel Cuevas-Alonso⁴

¹ Centro Ramón Piñeiro para a Investigación en Humanidades. Xunta de Galicia.

Estrada Santiago - Noia, Km. 3 - A Barcia 15896 - Santiago de Compostela - Galicia (Spain)

² Dpto. Traducción y Lingüística. Universidad de Vigo. Campus Universitario - Lagoas-Marcosende s/n 36200 -Vigo (Spain)

³ Dpto. Teoría de la Señal y Comunicaciones. Universidad de Vigo.

⁴ Dpto. Lengua Española. Universidad de Vigo.

{marza, campillo}@gts.uvigo.es, gallego@uvigo.es, mcuevasalonso@gmail.com

Abstract

This paper will present the design of a Galician syntactic corpus with application to intonation modeling. Our paper paid attention to phonetics and syntactic phenomena to achieve a comparison between its appearances. A corpus of around 3000 sentences was designed with variation in the syntactic structure and the number of accent groups, and recorded by a professional speaker to study the influence on the prosodic structure.

Keywords: Syntax, intonation, corpus.

1. Introduction

Intonation modeling is acknowledged to be one of the most relevant stages in speech synthesis, since languages use intonation variations to mark important parts of the discourse and dependencies between the different phrases (Pierrehumbert and Hirschberg, 1990; Selkirk, 1984). One of the functions of intonation is to divide up sentences into sequences of chunks or phrases (Ladd, 1996; Gussenhoven, 2004). Called prosodic structure, this plays a determining role both in naturalness and intelligibility (Ostendorf and Veilleux, 2004).

In a previous work (Campillo et al., 2009) syntactic and morphosyntactic information were used to model intonation, achieving a great improvement in the quality of synthetic speech. However, given the limitations of the available linguistic module, only shallow parsing could be used, with no information regarding the function of the phrases in the sentence, which probably degraded naturalness.

This paper is a part of an ambitious project in which the influence on intonation modeling of a complete syntactic analysis will be studied. It was proposed to study whether the syntax have an effect on the intonation (Selkirk, 1984; Nespor and Vogel, 1986; Campillo et al., 2009). Given the severe scarcity of resources in a minority language such as Galician, in this work a corpus of 3000 sentences paying attention to syntactic variability was designed and recorded by a professional speaker. Then, the prosodic structure of the sentences was manually labeled, in order to extract information that could be useful in next stages of the project. Note that the purpose of this work is not to draw general conclusions about the Galician language, but to develop a methodology to design a corpus that can be useful to study the relation between syntax and intonation. In any case, this corpus is available for further research, if someone is interested in extending the number of speakers.

The outline of this paper is as follows: section 2. is dedicated to the design of the corpus, taking into account the

variables that we want to study and the different factors that might affect the analysis in a random way; section 4. and section 3. describe the process of tagging, and the recording of the corpus, respectively. Finally, the conclusions and discussion are in section 5.

2. The Corpus

A speech corpus is needed to perform a prosodic analysis. Unfortunately, Galician is a minority language, with a severe lack of linguistic resources, so we decided to develop our own corpus, paying special attention to the different features involved to avoid other factors affecting our subject of study.

On the one hand, the corpus is constrained in several aspects: syntactic and phonetic. First, one constraint is that we choose the declarative modality sentence because it is the most abundant and their structure is more variable than interrogative or exclamatory. Second, the position of the accent within the word: the corpus is limited to paroxytone words, that is a word with a stronger phonetic accent on the next to the last syllable. Paroxytone words allow us to avoid two consecutive accented syllables in the boundaries between syntactic constituents.

On the other hand, we are going to study the influence of three variables on intonation: the number of accent groups (AG) where AG is a sequence of unaccented words ending in an accented one; the number of syntactic constituents and the order of syntactic constituents.

Taking into account these variables, the final corpus comprises 3000 sentences, including fifteen examples of each specific case.

The next subsections are dedicated to explaining the three variables that are considered in our study.

2.1. Syntactic variation

Table 1 shows the basic syntactic constituents that are used in the design of the corpus. To consider a broader range of

S	Subject	A choiva inunda o túnel (<i>The rain floods the tunnel</i>)
V	Verb	A choiva <i>inunda</i> o túnel (<i>The rain floods the tunnel</i>)
O	Object	A choiva inunda o <i>túnel</i> (<i>The rain floods the tunnel</i>)
X	Adjunt	A choiva inunda o túnel <i>hoxe</i> (<i>The rain floods the tunnel today</i>)

Table 1: *Basic syntactic constituents.*

cases, extended versions of the subject and the object with a prepositional phrase or an adjective phrase are included, like in, for example, “A Casa de Usher foi destruída” (*The House of Usher was destroyed*) or, “O neno comía mazás verdes” (*The boy ate green apples*). Moreover, the presence of the object as a clitic of the verb is considered as well, like in “O neno cómeas” (*The boy eats+them*).

From the basic SVO structure and the extended versions, which involve a minimum of syntactic and phonological constituents, we vary the order of the different functions. So we got series of combinations similar to the ones listed below:

- S-O-V - A choiva o túnel inunda (*The rain the tunnel floods+it**)
- S-V-O - A choiva inunda o túnel (*The rain floods the tunnel*)
- S-V-X - A choiva inundao hoxe (*A choiva floods+it today**)
- O-V-S - O túnel inúndao a choiva (*The tunnel floods+it the rain**)
- S-V-Oa - A choiva inunda o túnel antigo (*The rain floods the old tunnel*)
- Op-V-X - O túnel da estrada inúndao hoxe (*The tunnel of the road floods+it today**)
- S-V-O-X - A choiva inunda o túnel hoxe (*The rain flood the tunnel today*)

Combinations found to be unnatural in Galician were discarded, like long sentences ending in a verb:

- S-O-X-Vo - A choiva o túnel hoxe inúndao (*The rain the tunnel today floods+it**)

2.2. Number of accent groups

The number of AG is important in our study because sentences are organized phonologically according to the accents. In this case, sentences are limited to having from three to five AG, like the ones shown in the examples below, where accented syllables are represented in bold type and AG are divided with “-”:

- 3AG - A **choiva** - **inunda** - o **túnel** (*The rain - floods - the tunnel*)
- 4AG - O **túnel** - a **choiva** - **inúndao** - **hoxe** (*The tunnel - of the road - floods+it - today**)
- 5AG - A **choiva** - o **túnel** - da **estrada** - **inúndao** - **hoxe** (*The rain - the tunnel - of the road - floods+it - today**)

2.3. Number of syntactic constituents

Depending on the number of the accent groups and the functions in Table 1, the number of syntactic constituents will be between two and four. Therefore, we can have: two syntactic functions and three accentual groups; three syntactic functions and three accentual groups; three syntactic functions and four accentual groups; four syntactic functions and five accentual groups; etc...

- Sa-V: two syntactic functions and three accentual groups - A **forte choiva inúndao** (*The strong rain floods+it**)
- S-V-O: three syntactic functions and three accentual groups - A **choiva inunda o túnel** (*The rain floods the tunnel*)
- S-V-Op-X: four syntactic functions and five accentual groups - A **choiva inunda o túnel da estrada hoxe** (*The rain floods the tunnel of the road today*)

The syntactic functions do not appear in every examples to compare different functions in the same position:

- S-V-O: A **choiva inunda o túnel** (*The rain floods the tunnel*)
- S-V-X: A **choiva inunda hoxe** (*The rain floods today*)

In the last example, we can compare if the object (O) and adjunct (X) have the same prosodic structure.

We may relate AG to syntactic variation and number of syntactic constituents. For example, it is important that the extended subject and the object, by adjective and prepositional phrase, represent two accentual groups.

- 4AG - O **túnel** - a **choiva** - **inúndao** - **hoxe** (*The tunnel - the rain - floods+it - today**)
- 5AG - O **túnel** - da **estrada** - a **choiva** - **inúndao** - **hoxe** (*The tunnel - of the road - the rain - floods+it - today**)

2.4. Control of phonemic effects

Due to the fact that the aim of our study is to examine the potential effects of syntactic limits in the phonological and intermediate intonation phrases, sentences were carefully designed paying attention to different phonetic and contextual factors:

1. The syllables in a syntactic limit are always open and the consonants selected in onset position are voiced to reduce the micro-prosodic movements and to observe clearly the evolution of the fundamental frequency.

- (A roupa) (**d**enígrea) (o lodo vermello) (*The clothes fade the red mud*)

The verb begin with the sound /D/, thus the boundary between subject and verb contains voiced consonant.

- (O dote querido) (**r**ecíbeo) (a dona **d**e Lugo) (*The expect present receive+it* the girl of Lugo*)

The verb begin with the sound /R/, thus the boundary between object and verb contains voiced consonant. And, we suspect that in the modification of the subject, could make a intonation rupture for this we use the “de” preposition.

2. Special attention has been paid to the number of potential accent across the carrier sentences; the utterances contain three, four and five pitch accents, this variation has been strictly controlled and covariates with the number of sentence constituents.

- (A roupa) (denígreá) (o lodo vermello) (*The clothes fade the red mud*)

3. The distance between stressed syllables and constituent or virtual intonation boundaries was controlled to avoid the coincidence between them and to offer sufficient phonetic space to the realization of the intonation pattern, keeping in mind that such coincidence could produce truncation phenomena; the accentual pattern chosen for the words that compose the sentences is always paroxytone.

- (A roupa) (denígreá) (o lodo vermello) (*The clothes fade the red mud*)

4. The distance between stressed syllables has been fixed -2 syllables- in order to pass up the syllabic clash.

- (A roupa) (denígreá) (o lodo) (vermello) (*The clothes fade the red mud*)

For example the patro belows are unstressed (A), stressed(T) and unstressed syllables and other time the same patron.

- (A roupa) (denígreá) (o lodo) (vermello)
- (ATA) (ATA) (ATA) (ATA)

3. Recording of the corpus

The corpus was recorded by a female professional speaker, with five years of experience working on the radio, in the sound studio of the Group of Multimedia Technologies of the University of Vigo. The studio was partially sound-proofed to avoid problems with exterior sounds.

The task was organized in three sessions of five hours each, taking short breaks every hour and every time the speaker needed it. The order of the presentation of the prompts was randomized to avoid effects of focalization. Two people paid attention to technical recording issues and errors in the pronunciation in the one room, and while the speaker speak over the microphone, in the other room.

One person paid attention of the technical aspects of the record, putting boundaries beetwen phrases and marking the disposables sentences. Other person paid attention of the unexpected phonetics aspects, or when the fail was a intonation for the effects of focalizations, stopping the recording to repeat other time the complete sentence.

Every sentence found to be wrong, either for being mispronounced or for a spontaneous focalization not related to the syntactic structure, was re-recorded in a different session.

Apart from that, the recording was listen at the finish of the sessions with the aim of repeat that sentences in the following day. This repetitions and the items was been wrong, we kept to compare beetwen the right sentences in the other steps of the project.

4. Tagging

Recording of the corpus, the next stage are to label the syntactic and intonative boundaries on the sentences. To compare syntactic boundaries and intonation boundaries the sentences were syntactically and morphosyntactically labelled. Morphosyntactically labelled was made automatically with a program design for it, it result are:

- A = DETERMINANTE, FEMININO, SINGULAR (*Determiner, feminine, singular*) (a)
- forte = ADXECTIVO, FEMININO, SINGULAR (*Adjective, feminine, singular*) (forte)
- choiva = SUBSTANTIVO, FEMININO, SINGULAR (*Substantive, feminine, singular*) (choiva)
- inunda = VERBO, 3ª DO SINGULAR, PRESENTE DO INDICATIVO, 1º CONXUGACIÓN (*Verb, 3rd singular, present, 1st conjugation*) (inundar)
- o = DETERMINANTE, MASCULINO, SINGULAR (*Determiner, masculine, singular*) (o)
- túnel = SUBSTANTIVO, MASCULINO, SINGULAR (*Substantive, masculine, singular*) (túnel)

As Table 3 (Seijo Pereiro, 2009) below shows, twelve main categories are distinguished in our typology: verb, noun, pronoun, adjective, determiner, article, adverb, preposition, conjunction, interjection, residual and punctuation (see . This proposal not only is based on the EAGLES group's one but also aims to follow it closely and be compatible with it.

EAGLES GROUP	OUR TAGGER
Nouns	Nouns
Verbs	Verbs
Adjectives	Adjectives
Pronouns and determiners	Pronouns
Articles	Determiners
Adverbs	Articles
Adpositions	Adverbs
Conjunctions	Prepositions
Numerals	Conjunctions
Interjections	Interjections
Unique/unassigned	
Residual	Residual
Punctuations marks	Punctuation marks

Table 2: Main categories recognised by the taggers

And the syntactic labelled was made for each sentence an we had used as Table 3 below shows, four categories are distinguished in our typology: Subject, verb, object and adjunct:

Syntactic function	Our tagger	Categories
Subject	S	Nominal Phrase
Verb	V	Verbal Phrase
Object	O	Nominal Phrase
Adjunt	A	Adverb

Table 3: Main functions recognised by the taggers

The phrases can be modified by the adjectival phrases or by prepositional phrases:

- A forte choiva(*The strong rain*)= Adjectival phrase
- A choiva de inverno(*The rain of the winter*)= Prepositional phrase

The syntactic tagging was made manually:

- A forte choiva= **S** (*The strong rain*)
- inunda= **V** (*flood*)
- o túnel= **O** (*the tunnel*)
- hoxe= **X** (*today*)

In parallel, the minor phrase breaks were labelled using an ad hoc program developed by the research group that allowed the user to listen to the recording and see the text. The mark we used to tagging the minor phrase was #R-E#, we can see in the next example:

- A forte choiva **#R-E#** inunda o túnel **#R-E#** hoxe (*The strong rain #R-E# floods the tunnel #R-E# today*)

The final tagging will be similar to the ones listed below:

- (A forte choiva(S)) **#R-E#** (inunda(V)) (o túnel(O)) **#R-E#** (hoxe(X)) (*The strong rain(S) #R-E# floods(V) (the tunnel(O)) #R-E# today(X)*)

Note that only minor phrase boundaries are tagged, since it is the information we need for the further steps of the global project. Therefore, we are not following any of the well known models like ToBI or other ones, like AC or AP (see (Hidalgo-Navarro, 2006) for further information about these models). We use de Praat (<http://www.fon.hum.uva.nl/praat/>) to estimate the intonation contour, and to compare with manual tagging(See Figure ??).

5. Conclusions

Intonation modeling is acknowledged to be one of the most relevant stages in speech synthesis and we expect that this study help us. When I began to study the intonation parse, I noted that the syntax maybe had an influence on the intonation. The best manner to prove this hypothesis is observe

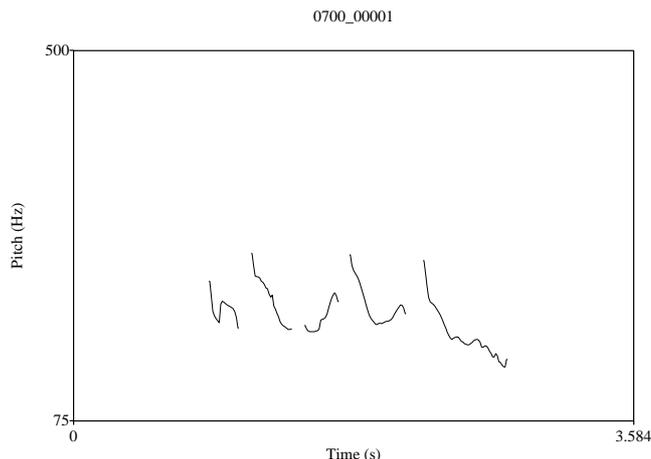


Figure 1: Pitch

the corpus and remark in it. For this, in this paper the design of a syntactic Galician corpus with application to intonation modelling was presented. We needed a corpus that contain the expected situations. The sentences were chosen taking into account variation of syntactic constituents, both in ordering and number, and also in the number of accent groups. Moreover, special attention was paid to avoid segmental effects in the boundaries between consecutive constituents that could affect the shape of the intonation contour. Finally, the sentences were recorded taking care of voiding spontaneous focalizations that would affect intonation and masquerade the subject of study. In the future, our plan will be to proof that our hypothesis can be right, although every authors don't share this opinion.

6. Acknowledgements

The work reported here was partially supported by the Spanish government and ERDF funds under the project TEC2009-14094-C04-04 "Búsqueda de Información en Contenidos Audiovisuales plurilingües", Xunta de Galicia "Isidro Parga Pondal" research programme and *Centro Ramón Piñeiro para a Investigación en Humanidades*.

7. References

- Campillo, F., van Santen, J., and Banga, E. R. (2009). Integrating phrasing and intonation modelling using syntactic and morphosyntactic information. *Speech Communication*, 51(5):207–214.
- Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.
- Hidalgo-Navarro, A. (2006). *Aspectos de la entonación española: viejos y nuevos enfoques*. Arco Libros.
- Ladd, R. (1996). *Intonational phonology*. Cambridge University Press.
- Nespor, M. and Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris Publications.
- Ostendorf, M. and Veilleux, N. (2004). A hierarchical stochastic model for automatic prediction of prosodic boundary location. *LREC Proceedings*, pages 1760–1762.

- Pierrehumbert, J. and Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. *Intentions in Communication*, pages 271–311.
- Seijo Pereiro, Lorena, A. M. Í. F. M. P. F. C. D. E. R. B. (2009). A galician textual corpus for morphosyntactic tagging with application to text-to-speech synthesis. *Speech Communication*, 51(5):207–214.
- Selkirk, E. (1984). *Phonology and Syntax. The relation between Sound and Structure*. Cambridge. MIT Press.