

# Workshop on Prosodic Labelling

## Stockholm, August 13, 1995

## Spanish

**Joaquim Llisterri**

**Departament de Filologia Espanyola, Edifici B, Universitat  
Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain.  
Fax: +34.3.581.16.86; E-mail: joaquim.llisterri@cc.uab.es**

## 0. Introduction

The aim of this report is: (a) to summarize the main approaches considered in the description and transcription of Spanish intonation; and (b) to provide some information about projects in this field, specially those which are oriented towards the development and exploitation of prosodic databases. Unfortunately lack of time has prevented from organizing surveys and coordination meetings at national level, and it was also difficult in practical terms to meet with groups working on American Spanish. Thus the materials presented here must be seen as preliminary; they reflect the information gathered by the author of the report, and by no means are to be considered exhaustive.

In the first part, the units used in the description of Spanish intonation by several authors are described. The second part addresses the transcription systems developed within the Spanish tradition. The final part presents information about some of the projects currently being carried out or documented in the literature available to the author. References and a contact electronic address for each project are provided.

## 1. Levels and units in the description of Spanish intonation

Before attempting to describe labelling systems for Spanish, it might be useful to survey the units that have been used in the description of Spanish intonation, since the theoretical conceptions of prosodic phenomena are closely linked to the different labelling system adopted. Two types of approaches are distinguished: those oriented towards a linguistic description, and those which are aimed at prosodic modelling for text-to-speech applications.

# 1.1. Linguistic approaches

## 1.1.1. Navarro Tomás (1945)

The description of Castilian Spanish intonation provided by Navarro Tomás in his *Manual de entonación española* published in 1945 is still the most exhaustive study of Spanish prosody available today. Most of the authors who have worked on Spanish intonation have used his descriptive framework for the characterization of Spanish intonation. The units adopted by Navarro Tomás (1945) are the following:

- **Breath-group ("grupo fónico")**

The breath-group is defined as the segment of an utterance comprised between two pauses.

- **Melodic unit or intonation group ("unidad melódica")**

The melodic unit - or intonation group - is considered by Navarro Tomás (1945) as the shortest possible segment with individual meaning and with a given melodic contour. Although melodic units tend to coincide with breath-groups, F0 movements and variations in intensity can also mark melodic unit boundaries.

Melodic units are divided into three parts:

- **Initial part**

- The initial part starts at the beginning of the group and ends in the first stressed syllable

- **Body**

- The body of the melodic unit comprises from the first stressed syllable to the last stressed syllable in the group

- **Final part**

- The final part starts at the last stressed syllable and ends at the end of the group. The final part of the melodic unit can show five different tonal movements, described by Navarro Tomás as:

- Anti-cadence ("anticadencia"): rising tone
    - Semi-anti-cadence ("semianticadencia"): half-rising tone
    - Level (suspensión)
    - Semi-cadence ("semicadencia"): half-falling tone
    - Cadence ("cadencia"): falling tone

- **Sentence**

Sentence is the highest level unit considered by Navarro Tomás (1945). It can be divided into a "rama tensiva" and a "rama distensiva". Each part or branch may be constituted by one or several melodic units.

- **Tonic group or stress group ("grupo rítmico-semántico", "grupo tónico")**

Navarro Tomás (1945) defines the tonic group - or stress group - as the segment of an utterance with a stress and with a coherent meaning that can not be divided into smaller units; however, he does not consider the tonic group as a unit for the analysis of Spanish intonation.

### **1.1.2. Quilis (1981, 1993)**

The description of Spanish intonation proposed by Quilis (1981, 1993) is based in two units:

- **Intonation group ("grupo de entonación")**

The intonation group is a segment of an utterance delimited by pauses or by F0 inflections.

- **Syllable**

The syllable is the unit related to stress placement. In the description proposed by Quilis (1981, 1993) each syllable has a pitch level associated to it.

### **1.1.3. Fant (1984)**

Although the work by Fant (1984) is not intended as an exhaustive description of Spanish intonation, it offers a framework based in two different types of units:

- **Tonic group or stress group ("grupo tónico")**

The tonic group is defined as in Navarro Tomás (1945)

- **Prosodic phrase ("frase prosódica")**

The prosodic phrase defined by Fant (1984) is equivalent to the melodic unit proposed by Navarro Tomás (1945).

### 1.1.4. Canellada & Kuhlmann Madsen (1987)

Canellada & Kuhlmann Madsen (1987) do not consider the tonic or stress group as an intonational unit, but they accept the melodic unit or intonation group and the sentence as prosodic units in the description of Spanish. They introduce a new level of analysis, namely the clause.

- **Clause ("cláusula")**

The clause is defined by Canellada & Kuhlmann Madsen (1987) as the segment of an utterance between two energy maxima.

## 1.2. Modelling for text-to-speech systems

### 1.2.1. López Gonzalo (1993)

The work by López Gonzalo (1993) is oriented towards the modelling of prosodic units for a Spanish text-to-speech system. The model takes as a starting point work on prosodic modelling at the CNET, and defines the following units:

- **Syllable**

- **Stress group ("grupo acentual) or prosodic word ("palabra prosódica)**

The stress group or prosodic word is defined as a group of consecutive syllables composed of unstressed words and a word containing a syllable bearing the lexical accent.

- **Intonative group or breath group**

The definition of the breath group - or intonative group - coincides with the one given by Navarro Tomás (1945)

## 1.3. Summary

The following units can then be considered for the description of Spanish intonation:

- **Sentence**

This is the highest unit of prosodic analysis, accepted by all authors who have described Spanish intonation.

- **Breath-group ("grupo fónico")**

The definition of breath group is also uncontroversial, and this unit is used by the majority of the authors.

- **Intonation group ("grupo entonativo"), melodic unit ("unidad melódica"), prosodic phrase ("frase prosódica")**

This unit is considered as different from the breath group by Navarro Tomás (1945) and Quilis (1981, 1993); however, López Gonzalo (1993) considers that it is equivalent to the breath group. Fant (1984) uses the denomination of prosodic phrase to refer to the intonation group.

- **Clause ("cláusula")**

Canellada & Kuhlmann Madsen (1987) are the only authors that consider the clause, defined as a function of changes in the energy contour.

- **Stress group ("grupo acentual"), tonic group ("grupo rítmico-semántico", "grupo tónico"), prosodic word ("palabra prosódica")**

Although Navarro Tomás (1945) and Canellada & Kuhlmann Madsen (1987) do not consider it a unit for the description of Spanish intonation, Fant (1984) and López Gonzalo (1993) consider this level of analysis in their description or in their modelling.

- **Syllable**

This is the lowest level of analysis considered by all the authors.

## 2. Transcription systems for Spanish

A standard transcription system for Spanish intonation has not been yet developed. In traditional linguistic descriptive studies the systems proposed by Navarro Tomás (1918, 1945), Alcina & Bleca (1975) and Quilis (1981, 1993) are used. In work more oriented towards speech technology applications, systems such as ToBI and INTSINT are being applied in current projects.

### 2.1. Navarro Tomás (1918)

In the prosodic transcriptions provided by Navarro Tomás (1918) in his *Manual de pronunciación española* the following symbols can be found:

## • Pauses

- ||| pause (less than 1 s.)
- || pause (less than 0.5 s.)
- | pause (less than 0.25 s.)
- ' very brief, almost unnoticeable pause

## • Pitch movements

→ normal tone

up and down arrows are used to show tonal inflections; when the inflection is lower than the normal tone the arrow is under the text; the direction of the arrow indicates the tone of the syllable and of the next ones until a change does occur

## 2.2. Navarro Tomás (1945)

The following set of symbols is used in the prosodic transcriptions published by Navarro Tomás (1945) in the *Manual de entonación española*:

- || anti-cadence (raising tone)
- ] semi-anti-cadence (half-raising tone)
- ] semi-cadencia (half-falling tone)
- | suspensión (level tone)
- ; cadencia (falling tone)
- ||; absolute question
- ] ; reiterative question
- ] ; pronominal and assertive question
- | relative question
- | continuation in interrogative sentences

## 2.3. Alcina & Blecua (1975)

Alcina & Blecua (1975) provide three classes of transcription symbols: those delimiting intonation groups or melodic units, those describing the pitch movement at the end of the intonation group and those describing pitch level in syllables

## • Intonation groups

- | melodic unit boundary

## • Final pitch movements

- ↑ anti-cadence (rising tone)

	semi-anti-cadence (half-rising tone)
→	level
⊥	semi-cadence (half-falling tone)
↓	cadence (falling tone)

### • Pitch levels

/1/	pitch level 1
/2/	pitch level 2
/3/	pitch level 3

## 2.4. Quilis (1981, 1993)

The system proposed by Quilis (1991, 1993) is very similar to the one described by Alcina & Blecua (1975). It is oriented towards the transcription of phenomena that are relevant from a phonological point of view.

### • Final pitch movements ("junturas terminales")

↓	falling tone
↑	rising tone
	level tone

### • Pitch levels

/1/ low	- it occurs after a pause or at a terminal juncture
/2/ mid	- it occurs in weak syllables after the last stressed syllable in an utterance
/3/ high	- it can be found in all the syllables with strong stress in the utterance

### • Stress

// strong	
/~/ weak	- weak stresses are not normally transcribed

## 2.5. Fant (1984)

Fant (1984) transcribes five different pitch levels that can be assigned to Spanish syllables:

M medium	- it typically occurs at the boundary between a pre-tonic and a stressed syllable
B low	- is found at the end of a declarative utterance
A high	- corresponding to the end of a stressed syllable in medial position
A+ extra high	- can be found at the end of the first stressed syllable of an utterance

## 2.6. Canellada & Kuhlmann Madsen (1987)

The proposal by Canellada & Kuhlmann Madsen covers the transcription of intonation group boundaries, of boundaries within the intonation group and of the pitch movements at the end of an intonation group:

### • Intonation groups

- / intonation group boundary
- // separation between the "rama tensiva" and the "rama distensiva" within an intonation group

### • Final pitch movements

- a semi-anti-cadence (half-rising tone)
- c semi-cadence (rising tone)
- A anticadence (rising tone)
- C cadence (falling tone)
- s level tone

## 2.7. Garrido (in preparation)

The aim of Garrido's work is to provide a formal description of Spanish intonation that could be applied to speech technology developments (Garrido 1991 a, b; Garrido et al., 1993). The method of study closely follows the IPO approach.

Intonative contours are viewed as a series of inflection points linked by straight lines and described in terms of combinations of levels and movements. Inflection points are defined as points where a significant change in the slope of the contour is found; after perceptual tests, the threshold of change has been fixed at 8%.

### • Levels

The intonative contours are made up of a series of inflection points labelled with one of the three values: P, V, M

- P peak - inflection points in the F0 contour corresponding to "full" accent peaks
- V valley - inflection points appearing at full lowerings or minima in the intonative contours
- M medium - inflection points coinciding with "half" accent peaks or half lowerings

### • Movements

The intonative movements are defined in terms of pairs of symbols that define a single movement in the contour:

V-P	- full rising
V-M, M-P	- half rising
P-V	- full falling
P-M, M-V	- half falling
V-V, M-M, P-P	- high, medium and low level movements

### **3. Aspects of prosody covered by the different transcription systems and phenomena which have been transcribed**

The survey presented above shows that the majority of transcription proposals cover three different types of phenomena: intonational unit boundaries, pitch movements at the end of an intonation unit and pitch levels associated to syllables. The following aspects and phenomena are then covered:

- **Intonation unit boundaries**

Boundaries between intonation units are usually transcribed by all authors

- **Pitch movements at the end of an intonation unit**

Since several authors distinguish five types of pitch movements (Navarro Tomás, 1918, 1946; Alcina & Blecua, 1975; Canellada & Kuhlmann Madsen, 1987) Quilis argues that only two of them (rising and falling) are phonologically relevant, although he provides for a symbol for the level tone.

- **Pitch level in syllables**

Three different pitch levels - high, low and mid - are distinguished in traditional descriptions. Fant (1984) introduces an extra-high level.

- **Pauses**

Navarro Tomás (1918) distinguishes four different types of pauses in the transcriptions published in his *Manual*.

## 4. Further resources available

Although it has been impossible to carry out a proper survey of existing prosodically labelled databases for Spanish, in a general survey of spoken and written language resources for Spanish carried out last year at the Instituto Cervantes (Arrarte & Llisterri, 1994) the existence of such databases developed in Spain was not detected. The following paragraphs intend to give some information about projects or groups known by the author, but are not at all exhaustive.

### 4.1. MULTEXT

One of the tasks of the LRE-MULTEXT project is to provide a prosodic coding of part of the EUROM.1 database in Spanish (together with, at least, coding for French and English). Initially the EUROM.1 40 passages read by one speaker will be analyzed and an automatic symbolic coding of F0 target points using INTSINT will be provided. The work for Spanish is carried out at the Universitat Autònoma de Barcelona in close collaboration with Daniel Hirst (Aix-en-Provence), and is to end by October 1995.

References: Hirst, Ide & Véronis (1994)

Contact: Joaquim Llisterri <joaquim.listerri@cc.uab.es>

### 4.2. AT&T Bell Laboratories

AT&T Bell Labs (Pilar Prieto & Julia Hirschberg) has recorded a small corpus with one Mexican Spanish speaker in order to study pitch accents. The corpus will be labelled using the ToBI system and adapting it to Spanish if necessary.

References: Prieto, van Santen & Hirschberg (1994)

Contact: Pilar Prieto <prieto@research.att.com>

### 4.3. UQAM, Université du Québec à Montréal, Department of Linguistics

A corpus of spontaneous speech from eight speakers from Panamá representing two socioeconomic levels and two age groups has been analyzed by Henrietta Jonas-Cedergreen at the UQAM (Université du Québec à Montréal). Syllabic boundaries, rhythmical group boundaries and intonational phrases have been auditorily delimited; the utterances have also been labelled at the segmental level with time alignment between the digital spectrogram and the waveform. This corpus is part of a larger research project, within which a similar corpus from 16 French speakers from Québec has been already prepared.

Contact: Henrietta Jonas-Cedergreen <r14730@er.uqam.ca>

## 4.4. Corpus of Spoken Spanish for prosodic studies, Department of Linguistics and Phonetics, University of Leeds

A paper by Cid Uribe and Roach (1990) describes a corpus of spoken Spanish build up for prosodic studies at the University of Leeds. Seven native speakers of Spanish were recorded and the material was complemented by recordings of TV news and comments and a recording of a special TV overseas report. The materials obtained from the speakers comprise spontaneous dialogue, spontaneous autobiography, spontaneous anecdote, poetry, narrative and descriptive reading. A total of 25.250 words was obtained.

The corpus was orthographically transcribed and syllabic boundaries were also introduced. The prosodic transcription followed the conventions developed for the Spoken English Corpus, marking major and minor tone unit boundaries, eight different tones and features of connected speech such as hesitations, compressions, elisions and lengthenings. Transcriptions were stored in machine-readable files.

References: Cid Uribe & Roach (1990)

## 4.5. UPM-SSR, Universidad Politécnica de Madrid, Departamento de Señales, Sistemas y Radiocomunicaciones

A database for the modelling of Spanish prosody has been build up within a larger project aimed at the development of a Spanish text-to-speech system in the Departamento de Señales, Sistemas y Radiocomunicaciones at the Universidad Politécnica de Madrid. The corpus is made up of 144 sentences from one single speaker, and consists on 306 intonation groups and 818 stress groups, classified in several types. Stylized F0 contours together with F0 and duration information for each vowel are stored in the database used for the modelling.

References: López Gonzalo (1993); López-Gonzalo & Hernández-Gómez (1993); López Gonzalo, Álvarez Cercadillo & Hernández Gómez (1994)  
Contact: Eduardo López Gonzalo <eduardo@gaps.ssr.upm.es>

## 4.6. UPM-DIE/UNED, Universidad Politécnica de Madrid, Departamento de Ingeniería Electrónica / Universidad Nacional de Educación a Distancia, Departamento de Lengua Española

A prosodic corpus has been build up and analyzed in order to improve intonation in a Spanish text-to-speech synthesis system developed by the Departamento de Ingeniería Electrónica at the Universidad Politécnica de Madrid in collaboration with the Departamento de Lengua Española at the Universidad Nacional de Educación a Distancia.

The corpus consists of a 5 minutes text read by 8 Spanish speakers, comprising a total of 1127 breath groups. A database has been build up with pauses and their duration, duration of breath groups, number of stressed and unstressed syllables, pitch values and syntactic patterns of each breath group.

References: Pardo, Martínez, Quilis & Muñoz (1987); Moreno, Martínez, Pardo & Vallejo (1989)

Contact: José Manuel Pardo <pardo@die.upm.es>

## 4.7. URL-ESEEI, Universitat Ramon Llull, Escola Superior d'Enginyeria Electrònica i Informàtica

A corpus aimed at modelling prosody for text-to-speech in Spanish has been build up and analyzed at the Universitat Ramon Llull. It consists of a 21s. text read by 8 Spanish speakers (320 stress groups). Pauses, breath groups, stress groups, vowel duration, vowel F0 and declination have been measured.

References: Martí & Gudayol (1993, 1994)

Contact: Josep Martí <elsmarti@perseus.url.es>

## 4.8. UAB-DFE, Universitat Autònoma de Barcelona, Departament de Filologia Espanyola

A corpus for the analysis of Spanish intonation has been build up by Juan M. Garrido at the Departament de Filologia Espanyola, Universitat Autònoma de Barcelona. The corpus consists of news bulletins and news comments from Spanish radio and TV and newspaper articles dealing with the same topics; the materials have been recorded from the broadcasts and have also been read by two non-professional speakers.

Several databases have been build up with information concerning paragraphs (number of syllables and number of sentences), sentences (number of syllables and sentence type), intonation groups (number of syllables, sentence type and position within the sentence) and stress groups (number of syllables, position of the stressed syllables, sentence type, position in the intonation group, position in the sentence). A total number of 1426 stress groups has been analyzed, corresponding to 400 intonation groups and 582 sentences.

References: Garrido (1993)

Contact: Juan M. Garrido <juanma@liceu.uab.es>

## 5. Acknowledgments

The author wishes to acknowledge the information received from Professor Henrietta Jonas-Cedergreen (UQAM) and Pilar Prieto (AT&T). Section 1 has been prepared with the helpful collaboration of Juan M. Garrido (UAB).

## 6. References

ALCINA FRANCH, J- BLECUA, J.M. (1975) *Gramática española*. Barcelona: Ariel ( Letras e Ideas, Instrumenta 10).

ARRARTE, G.- LLISTERRI, J. (1994) *Informe sobre recursos lingüísticos para el español (I): Corpus escritos y orales disponibles y en desarrollo en España*. Madrid: Instituto Cervantes (also available in English; contact <g.arrarte@cervantes.es> or <joaquim.llisterri@cervantes.es>)

CANELLADA, M.J. - KUHLMAN MADSEN, J. (1987) *Pronunciación del español. Lengua hablada y literaria*. Madrid: Castalia.

CID URIBE, M.- ROACH, P. (1990) " Spanish intonation: Design and implementation of a machine-readable corpus ", *Journal of the International Phonetic Association* 20,2: 1-8

FANT, L. (1984) *Estructura informativa en español. Estudio sintáctico y entonativo*, Upsala: Acta Universitatis Upsaliensis.

GARRIDO, J.M. (1991a) "Estilización de patrones melódicos del español para sistemas de conversión texto-habla", *Procesamiento del Lenguaje Natural, Boletín nº 11* : 209-220

GARRIDO, J.M. (1991b) *Modelización de patrones melódicos del español para la síntesis y el reconocimiento*. Bellaterra: Departament de Filologia Espanyola, Universitat Autònoma de Barcelona

GARRIDO, J.M. (1993) "Analysis of global pitch contour domains at paragraph level in Spanish reading texts", in HOUSE, D.- TOUATI, P. (Eds) *Proceedings of an ESCA Workshop on Prosody*. September 27-29, 1993, Lund, Sweden. Lund University Department of Linguistics and Phonetics, Working Papers 41. pp. 104-107.

GARRIDO, J.M.- GUDAYOL, F. (1991) " Modelization of intonation patterns in Spanish for automatic recognition", in *Actes du XIIème Congrès International des Sciences Phonétiques*. 19-24 août 1991, Aix-en-Provence, France. 5 vols. Aix-en-Provence: Université de Provence, Service des Publications. Vol 2 pp. 246-249

GARRIDO, J.M.- LLISTERRI, J.- de la MOTA, C.- RÍOS, A. (1993) " Prosodic differences in reading style: Isolated vs. Contextualized Sentences" in *Eurospeech 93. 3rd European Conference on Speech Communication and Technology*. Berlin, Germany, 21-23 September 1993. Vol 1 pp. 573-576.

HIRST, D.J.- IDE, N. - VÉRONIS, J. (1994) "Coding fundamental frequency patterns for multi-lingual synthesis with INTSINT in the MULTEXT project", in *Conference Proceedings of the Second ESCA/IEEE Workshop on Speech Synthesis*. September 12-15, 1994. Mohonk Mountain House, New Paltz, New York, USA. pp. 77-80

LÓPEZ GONZALO, E. (1993) *Técnicas de procesamiento lingüístico-prosódico y acústico para conversión texto-voz mediante concatenación de unidades*. Tesis doctoral. Universidad Politécnica de Madrid.

LÓPEZ GONZALO, E.- ÁLVAREZ CERCADILLO, J.- HERNÁNDEZ GÓMEZ, L. (1994) "Metodología para el modelado prosódico de un sistema de conversión de texto a habla en castellano", *Actas del X Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural*, Córdoba, 20-22 de julio de 1994.

LÓPEZ-GONZALO, E.- HERNÁNDEZ-GÓMEZ, L.A. (1993) "Prosodic Modelling for a Text-to-Speech System in Spanish", in HOUSE, D.- TOUATI, P. (Eds) *Proceedings of an ESCA Workshop on Prosody*. September 27-29, 1993, Lund, Sweden. Lund University Department of Linguistics and Phonetics, Working Papers 41. pp. 74-77

MARTÍ, J.- GUDAYOL, F. (1993) "El ritmo y la entonación en la lectura del castellano", *Input 3* (octubre): 5-15

MARTÍ, J.- GUDAYOL, F. (1994) "Ritmo y entonación en la lectura del castellano", *Boletín de la Sociedad Española para el Procesamiento del Lenguaje Natural* 14: 273-290

MORENO, P.J. - MARTÍNEZ, M.- PARDO, J.M.- VALLEJO, J.A. (1989) "Improving naturalness in a text-to-speech system with a new fundamental frequency algorithm", in TUBACH, J.P.- MARIANI, J.J. (Eds) *Eurospeech 89. European Conference on Speech Communication and Technology*. Paris, September 1989. Edinburgh: CEP Consultants Ltd. vol 1 pp. 360-363

NAVARRO TOMÁS, T. (1918) *Manual de pronunciación española*. Consejo Superior de Investigaciones Científicas: Madrid, Instituto Miguel de Cervantes ( Publicaciones de la Revista de Filología Española, III ). 21ª edición, 1982.

NAVARRO TOMÁS, T. (1945) *Manual de entonación española*. New York: Hispanic Institute. Cuarta edición: Madrid: Guadarrama (Punto Omega, 175), 1974.

PARDO, JM.- MARTÍNEZ, M.- QUILIS, A.- MUÑOZ, E. (1987) "Improving Text to Speech Conversion in Spanish: Linguistic Analysis and Prosody " in LAVER, J.- JACK, M.A. (Eds) *European Conference on Speech Technology*. Edinburgh, September 1987. Edinburgh: CEP Consultants Ltd. pp. 173-178.

PRIETO, P.- van SANTEN, J.P.H.- HIRSCHBERG, J. (1994) "Patterns of Fo peak placement in Mexican Spanish", in *Conference Proceedings of the Second ESCA/IEEE Workshop on Speech Synthesis*. September 12-15, 1994. Mohonk Mountain House, New Paltz, New York, USA. pp. 33-36

QUILIS, A. (1981) *Fonética acústica de la lengua española*. Madrid: Gredos (Biblioteca Románica Hispánica, Manuales, 49)

QUILIS, A. (1993) *Tratado de fonología y fonética españolas*. Madrid: Gredos (Biblioteca Románica Hispánica, Manuales, 74)