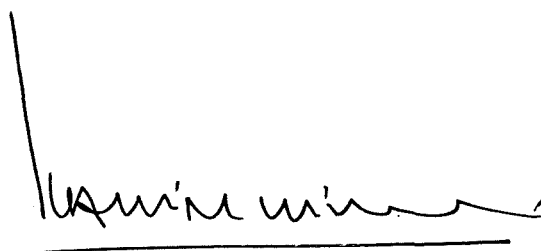Juan María Garrido Almiñana

Modelling Spanish Intonation for -Text-to-Speech Applications

Ph. D. Thesis

Tesis doctoral

Director: Dr. Joaquim Llisterri i Boix

Departament de Filologia Espanyola
Facultat de Lletres
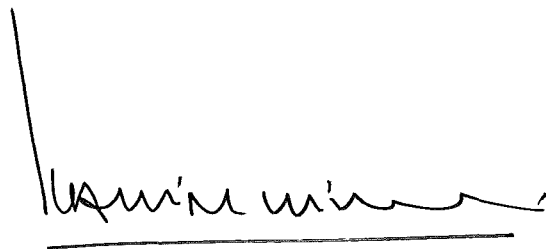*Universitat Autònoma de Barcelona*
Mayo 1996

Juan María Garrido Almiñana

Modelling Spanish Intonation for Text-to-Speech Applications

Ph. D. Thesis

Tesis doctoral

Director: Dr. Joaquim Llisterri i Boix

Departament de Filologia Espanyola
Facultat de Lletres
Universitat Autònoma de Barcelona
Mayo 1996

A todo aquel que encuentre
en estas páginas algo de utilidad

Agradecimientos

A Joaquim, el doctor Joaquim Llisterri. Por ser mi director de tesis, y por toda la ayuda que he recibido durante la realización de este trabajo. Pero sobre todo por la confianza depositada en mí desde el primer momento. Sin esa confianza ni esta tesis ni muchas otras cosas habrían sido posibles.

A Francesc, Francesc Gudayol. Por ser parcialmente 'culpable' de mi interés por la entonación, por haber tenido la idea del programa 'Pitch', y por tener la paciencia de llevar aquella primera idea hasta el final. Gracias también por los comentarios y las ideas tras leer la primera versión de esta tesis.

A Johanneke Caspers, Hugo Quené, Margreet Sanders, Vincent van Leuven y todas las personas que durante mi estancia en Holanda me ayudaron a perfilar el objetivo de esta tesis.

A Joan Torruella, por su ayuda en la selección de los locutores.

A los locutores, Antonio, Margarita y Mercedes, por la paciencia que demostraron durante horas de grabaciones y repeticiones. Y a Inmaculada Royo, de Radio Nacional de España en Barcelona, por las facilidades prestadas para conseguir las grabaciones de los diarios hablados.

A Eduardo Lleida, por las horas que perdió por mi culpa luchando hasta conseguir que las sílabas se contaran automáticamente.

A Alistair MacConkie y Pilar Prieto. Por la paciencia en leer tan detalladamente el primer manuscrito de esta tesis. Y por todos los comentarios, que me han ayudado a perfilar muchos detalles.

A los alumnos que me han tenido como profesor durante estos años. Por las horas que esta tesis le ha robado a la preparación de sus clases.

A mis dos Macintosh. Por no fallarme nunca. Sin ellos esta tesis tampoco habría sido posible. Y a mi equipo de música, por acompañarme en las muchas horas de soledad en las que esta tesis se ha ido haciendo.

A todas las personas que durante este tiempo han trabajado conmigo en el Laboratori de Fonètica de la Universitat Autònoma de Barcelona: Lourdes Aguilar, Marcel·la Andreu, Lorraine Baqué, Cristina Benesey, Mercè Bericat, Beatriz Blecua, Carme Carbó, Mònica Estruch, Helena Fernández, Josep Maria Fernández, Nati Fernández, Julia Giménez, Hanne Hesselager, José Luis Iglesias, María Jesús Machuca, Franca Mancini, Rafael Marín, Gemma Martínez, María José Navarro, Carme de la Mota, Dolors Poch, David Puigví, Jordi Renom, Montse Riera, Antonio Ríos y tantos otros. Gracias en unos casos por los

comentarios, las críticas, especialmente durante los seminarios del Grup de Fonètica, que me han sido muy útiles para el desarrollo de esta tesis. En otros, por el entusiasmo e ilusión demostrado en el trabajo de cada día, que en muchas ocasiones me han dado fuerzas para seguir trabajando.

Finalmente, a todos aquellos que habéis estado conmigo durante todo este tiempo, dándome vuestro apoyo y amistad. En vuestro caso, mi reconocimiento va mucho más allá de los agradecimientos de esta tesis. Gracias por estar ahí, a mi lado, aunque muchas veces no os haya tratado todo lo bien que os merecíais. Especialmente a mi familia, por los días de mal humor que han tenido que soportar durante este tiempo. Y a Carme, por el empujón final. Sin ese empujón esta tesis seguramente no existiría.

Resumen

*El objetivo de esta tesis es proponer un modelo para la entonación del español en lectura, orientado a la generación automática de contornos entonativos en sistemas de conversión de texto a habla. El término 'entonación' se ha considerado aquí equivalente a la evolución del contorno de frecuencia fundamental (F0), o **contorno entonativo**, a lo largo de un enunciado.*

*Para la consecución de este objetivo se ha llevado a cabo un análisis acústico de los contornos entonativos en un corpus de lectura de noticias en español. Este análisis se ha realizado en cuatro niveles entonativos diferentes: **grupo tónico, grupo de entonación** (identificado en este trabajo con el **grupo fónico**), **oración** y **párrafo**.*

Corpus de análisis

El corpus recogido consta de un conjunto de lecturas de artículos de diarios, grabadas por dos locutores no profesionales seleccionados. De este conjunto de grabaciones se han extraído cuatro sub-corpora, uno para cada nivel de análisis considerado.

Estilización

*Para la obtener una representación simplificada de los contornos entonativos del material analizado se ha empleado un procedimiento de **estilización** automática, integrado en el programa 'Pitch', un sistema de análisis, edición y síntesis de contornos entonativos desarrollado por Enginyeria La Salle de la Universitat Ramon Llull. Por medio de este sistema de estilización los contornos entonativos originales (obtenidos mediante un algoritmo de detección de F0) se han transformado en una serie de **puntos de inflexión***

*unidos mediante líneas rectas, o **contorno estilizado**, equivalente perceptivamente al contorno original.*

Los contornos estilizados mediante este sistema han sido posteriormente post-procesados para eliminar los errores de estimación de F0 y las variaciones micromelódicas. Las representaciones obtenidas han sido utilizadas en este trabajo para la fase de análisis acústico.

Análisis de los grupos tónicos

*De acuerdo con el **marco descriptivo** definido para el estudio del sub-corpus de grupos tónicos, los contornos estilizados de un enunciado pueden concebirse como una serie de **niveles** entonativos - 'pico' (**P**), 'valle' (**V**) o 'medio' (**M**) - asociados a determinados puntos de inflexión del contorno, que están unidos mediante diferentes **movimientos** entonativos - 'ascendente', 'descendente' o 'plano' -, representados por las líneas rectas del contorno.*

*El análisis de los contornos entonativos del sub-corpus utilizando este marco descriptivo ha permitido establecer un conjunto de **patrones entonativos locales**, definidos como concatenaciones de niveles y movimientos asociados a uno (o varios) grupos tónicos. Se han considerado tres grandes grupos de patrones: **iniciales**, **mediales** y **finales**. Los patrones mediales se han subdividido a su vez en patrones **post-iniciales** y patrones **interiores**. Los patrones finales también se han subdividido en tres grupos: patrones **descendentes**, patrones **ascendentes** y patrones **circunflejos**.*

*Se ha estudiado también la frecuencia de aparición de cada patrón en el sub-corpus analizado, así como el uso de estos patrones en función de diversos factores, unos fonéticos (la duración del grupo tónico), y otros lingüísticos (la posición de la sílaba acentuada en el grupo tónico, el límite sintáctico que aparece tras el grupo tónico, y la modalidad oracional). Los resultados obtenidos ponen en duda la validez del grupo tónico como unidad entonativa en español, y parecen favorecer la idea de que el **grupo acentual** (definido para otras lenguas como un conjunto de sílabas formado por una sílaba tónica y las sílabas átonas que la siguen hasta la siguiente tónica) es una unidad más adecuada para describir los fenómenos entonativos que se dan en el nivel local en español.*

Análisis de los grupos fónicos

*Para el estudio de los contornos entonativos en el sub-corpus de grupos fónicos (considerados aquí como una unidad equivalente al grupo de entonación) se ha empleado un marco descriptivo basado en la **hipótesis de***

*las **líneas superior e inferior** utilizada para otras lenguas. Se han definido aquí tres líneas de referencia, la **línea superior**, la **línea media**, y la **línea inferior**, que determinan el nivel teórico para los niveles P, M y V en cada instante de tiempo a lo largo de un grupo fónico.*

Utilizando este marco descriptivo, se han definido dos patrones globales básicos para los dos locutores analizados. Ambos patrones se caracterizan por el descenso progresivo de las tres líneas hacia el final del grupo fónico, y por tratarse de líneas convergentes, que presentan en general un rango mayor al principio que al final del grupo fónico.

También se han analizado las variaciones de este patrón básico en función de diferentes factores: la duración del grupo fónico, la posición del grupo fónico dentro de la oración (inicial, medial o final), y la modalidad de la oración en que está incluido el grupo fónico. Los resultados sugieren una relación entre la pendiente de las líneas y la duración del grupo fónico, y también que existe cierta relación entre la altura de F0 de los puntos iniciales y finales de las líneas y la posición del grupo fónico dentro de la oración.

Análisis de las oraciones

*En el caso del análisis de los contornos estilizados en el sub-corpus de oraciones, se ha definido primero un marco descriptivo basado en el modelo de las líneas superior e inferior utilizado para los grupos fónicos. La forma de los contornos entonativos a lo largo de las oraciones se ha modelizado por medio de '**supra-líneas**', que determinan la altura relativa de las líneas superior, media e inferior de los diferentes grupos que componen la oración. Se han definido dos tipos diferentes de 'supra-líneas': las '**supra-líneas' altas**, que controlan la altura de F0 de los puntos iniciales de las líneas superior, media e inferior, y las '**supra-líneas' bajas**, que controlan la altura de F0 de los puntos finales de estas líneas.*

*El análisis de las oraciones seleccionadas ha puesto en duda la existencia de un sólo patrón de 'supra-líneas' en las oraciones, por lo que se ha estudiado la presencia de unidades entonativas de nivel intermedio entre el grupo fónico y la oración. Se ha propuesto la existencia de las llamadas '**cláusulas entonativas**', compuestas por varios grupos fónicos, que presentan un patrón entonativo propio, también representado por medio de 'supra-líneas'. Las diferentes cláusulas entonativas de una oración compleja se organizan formando una estructura jerárquica de unidades entonativas, que determinan la altura de F0 de los diferentes grupos fónicos dentro de la oración. Esta **estructura entonativa** puede ser representada mediante un **árbol prosódico**. Se ha observado también que los diferentes grupos fónicos de una oración tienden a organizarse en dos unidades, que coinciden con las*

llamadas *'rama tensiva'* y *'rama distensiva'* descritas en estudios anteriores sobre la entonación del español.

Finalmente, se han analizado algunos de los factores que pueden intervenir en la definición de la estructura entonativa de las oraciones, tales como su estructura sintáctica y su estructura pragmática.

Análisis de los párrafos

Para la modelización de los contornos entonativos en el sub-corpus de párrafos se ha utilizado el mismo marco descriptivo que para la modelización de las oraciones. La aplicación de este marco ha permitido comprobar que la definición de un único patrón global de 'supra-líneas' resulta muchas veces insuficiente para la modelización de los contornos entonativos en el nivel del párrafo. En función su duración, resulta en algunas ocasiones más adecuado dividir el párrafo en dos unidades entonativas diferentes, cada una con su propio patrón de 'supra-líneas'. Esto hace suponer que existe una *'preplanificación parcial'* en la generación de los contornos entonativos de los enunciados, que obligaría a iniciar una nueva unidad entonativa si la duración del párrafo excede el límite de preplanificación del hablante. Finalmente, se ha observado que los valores iniciales y finales de F0 en los párrafos son bastante constantes para cada locutor, por lo que pueden tomarse como puntos de referencia.

Definición del modelo

Los datos obtenidos en la fase de análisis han permitido elaborar una primera propuesta de modelo para la generación de los contornos entonativos del español en lectura. El modelo asume la existencia de una estructura entonativa para el texto de entrada, que define la posición de las pausas y la organización jerárquica de los diferentes grupos fónicos en unidades entonativas de nivel superior. Se trata por tanto de un modelo fonético, en el sentido de que no incluye ningún tipo de análisis lingüístico (sintáctico o fonológico), y se centra exclusivamente en la generación de los contornos melódicos. El modelo está organizado en dos sub-módulos, el **nivel global**, que se encarga de asignar los patrones globales ('supra-líneas' y líneas superiores, medias e inferiores) a la estructura entonativa de entrada, y el **nivel local**, que se encarga de superponer los patrones locales a las líneas superiores, medias e inferiores calculadas por el nivel global. La salida del modelo es un contorno entonativo estilizado, constituido por una cadena de puntos de inflexión. Cada punto de inflexión lleva asociado un valor de F0, e información acerca del segmento al que dicho punto va ligado en la cadena de elementos segmentales.

Table of contents

iv

INTRODUCTION

In this thesis, a first attempt has been made to define a phonetic model of Spanish intonation in read text, and to apply the obtained results to the automatic generation of intonation contours. Two main reasons can be argued to justify these goals:

1) The need of linguistic and phonetic models for Spanish intonation:

Several intonation models have already been proposed for other languages, but similar models are not yet available for Spanish. Partial descriptions of Spanish intonation can be found, but they still need to be completed to obtain a full model.

2) The need of formal descriptions of Spanish intonation to be used in the field of speech technology:

The problem of automatic generation of intonation is closely related to the development of text-to-speech (TTS) systems. The role of intonation in the intelligibility and naturalness of the output speech of these systems has been widely recognized. Developing automatic intonation generation modules for Spanish TTS systems, or improving the existing ones, requires formal, detailed and global descriptions of Spanish intonation. Although some progress has been made already in this direction, more accurate analyses, specially with a phonetic and linguistic approach, are still necessary.

The defined goals are too ambitious to be completely achieved in a work with the limitations of the one presented here. The present study is intended to be a first approach to the problem, presenting a general description of the main characteristics of Spanish intonation contours from an acoustic phonetics perspective. The detailed analysis of particular problems has not

been considered here, not only due to time constraints, but also because the selected material was not controlled enough to carry out this kind of studies.

A first proposal of generation model has been developed using the results obtained in the descriptive analysis. The proposed model is not complete, since it does not consider in detail all the generation process, but only the main steps, and it has not been perceptually validated yet. However, it may be used as a basic framework to be tested and refined.

The development of this work has been organized in three main parts:

1) A preliminary phase of theoretical work. This has included a review of the main characteristics of intonation in general, and of Spanish intonation in particular. A survey of the different approaches to intonation modelling for other languages has also been attempted. This first phase has led to an initial definition of the scope and main features of the model.

2) A second phase of phonetic analysis. This has included the analysis of Spanish intonation at different levels in a corpus of read texts. In this experimental part a formal descriptive framework has been applied to the obtained intonation contours in order to derive a quantitative description of Spanish intonation in read text.

3) A third phase of initial definition of the model, using the experimental data obtained in the second part.

The results of these three phases are described in the two volumes of this thesis. This first volume contains the theoretical considerations, the description and discussion of the experimental part, and a first definition of the model. A detailed presentation of the results (tables, figures, etc.) has been left for the appendix volume, in order to make this first volume more readable. Only the tables and figures necessary to support the conclusions and hypotheses have been kept in the first volume. For this reason, the appendix is necessary in many cases for a detailed analysis of the results, and to find out additional examples at some specific points of the discussion.

The contents of the present volume are organized as follows: the results of the first phase are presented in chapter 1; the results of the experimental part, including the definition of the analysis material and the description of the methodology are presented in the chapters 2-7; the results of the model definition phase are presented in chapter 8; the last chapter (9) contains some concluding remarks about the main implications that can be derived from the results of this work, and some proposals of further research.

The appendix volume contains a detailed presentation of the text used as reading material (appendix 1), the different defined sub-corpora (appendices 2, 4, 6 and 8), and a detailed presentation of the results of the different analyses carried out in the experimental part (appendices 3, 5, 7 and 9). Finally, an example of application of the proposed model to a sample paragraph is included in appendix 10.

Chapter 1

INTONATION AND INTONATION MODELS

1.1. The nature and structure of intonation

1.1.1. Intonation and fundamental frequency

A brief survey of the literature offers many different definitions of 'intonation':

- "the entire ensemble of pitch contours, pitch levels, and stress levels that occurs when a sentence is spoken". (Lieberman, 1965, p. 40)

- "I shall use the term *pitch* to refer to the perceptual correlates of frequency, *tone* to refer to the feature when it functions distinctively at word level, and *intonation* to refer to the feature when it functions at sentence level" (Lehiste, 1976, p. 229).

- "the distinctive use of patterns of pitch, or melody" (Crystal, 1980, p. 182).

- "no definition is completely satisfactory, but any attempt at a definition must recognise that the pitch of the voice plays the most important part" (Roach, 1983, p. 112).

- "INTONATION involves the occurrence of recurring pitch patterns, each of which is used with a set of relatively consistent meanings, either on single words or on groups of words of varying length" (Cruttenden, 1986, p. 9).

• "the salient pitch changes in the course of an utterance" ('t Hart *et al.*, 1990, p. 69).

• "*La línea melódica con que se se pronuncia un mensaje recibe el nombre de entonación*" (Alcina & Blecua, 1975, p. 452).

"*desde el punto de vista de su realización fonética, está formada, fundamentalmente, por el acento de intensidad, el tono y la transición final, aunque también contribuyen a la entonación, la cantidad y el ritmo*" (Alcina & Blecua, 1975, pp. 453-4).

• "*la entonación es la función lingüísticamente significativa, socialmente representativa e individualmente expresiva de la frecuencia fundamental en el nivel de la oración*" (Quilis, 1993, p. 410).

• "*La entonación es ante todo un fenómeno lingüístico, relacionado con la sensación perceptiva que produce la variación a lo largo de todo un enunciado de tres parámetros físicos -frecuencia del fundamental [...], amplitud y duración -, y que proporciona al receptor información de distintos tipos. A diferencia del acento, las relaciones y contrastes entre las variaciones de los tres parámetros se darían en unidades mayores que la palabra, normalmente sintagmas o frases*" (Garrido, 1991, p. 7).

These various definitions show some common points, that can be summarized as follows:

a) in general, these definitions relate intonation to the phonetic concepts of pitch and fundamental frequency (henceforth F0). Only Alcina & Blecua (1975) and Garrido (1991) include amplitude and duration as acoustic correlates of intonation.

b) many definitions (Lieberman, Lehiste, 't Hart *et al.*, Quilis, Garrido) point out that intonation is a phenomenon that occurs at a global domain (phrase, sentence or utterance). This distinguishes intonation from other prosodic phenomena, as tone or stress. These definitions thus exclude local pitch movements from intonational phenomena.

c) several definitions point out the role of intonation in the communication process (Cruttenden, Quilis).

In fact, it seems that these definitions refer to two different concepts of the term 'intonation':

1) A 'phonological' concept of intonation, that identifies intonation with a global domain in the utterances (phrases, sentences) which plays some

roles in the communication process. According to this definition, 'intonation' can be viewed as the prosodic phenomenon that occurs in a global domain (intonation units) and carries several types of information. The acoustic realization of these units is mainly pitch, but also duration and amplitude. With this meaning, intonation can be included within the classical 'phonological' **suprasegmental phenomena**, such as stress, tone or rhythm.

2) A 'phonetic' view of intonation, which considers intonation as the evolution over time of the F0 contour of an utterance. It this case, the word 'intonation' does not refer to any specific domain or unit, but only to the acoustic realization of these units in the F0 contours. In this sense, intonation can be grouped in the same category with duration, pauses or amplitude, as one of the possible **phonetic cues** of the suprasegmental phenomena mentioned above

Both concepts of intonation, the 'phonological' and the 'phonetic' one, are related in some way, but they also show important differences:

a) F0 contours are the main phonetic cue of 'phonological' intonation. However, other phonetic phenomena are also responsible for marking phonological intonation in the utterance: pauses, duration and amplitude. The analysis of F0 contours provides then only partial information about the phonetic realization of 'phonological' intonation.

b) Changes in the F0 contours are only partially conditioned by 'phonological' intonation. Other (linguistic or non linguistic) phenomena, such as stress or the segmental content of the utterance, define the shape of the F0 contours. The analysis of F0 contours involves then the consideration of other phenomena in addition to 'intonation' itself.

This work is intended to be a description of the intonation of Spanish from a phonetic perspective. The goal of this work is to describe the organization and components of the Spanish F0 contours - henceforth, **intonation contours** -, and to propose a model that predicts their composition. The analysis of the phonological intonation is out of the scope of this work, although, given the relation between the two types of intonation described earlier, some conclusions can be extracted about it from the results presented here.

1.1.2. The structure of F0 contours: intonation as a hierarchical phenomenon

One of the features that can be observed in the intonation contours of any language is that they are made up of different-level components. In other

words, the intonation contours are the result of the interaction of several **intonational phenomena** covering different domains:

1) **Micromelodic phenomena**, which are realized in the intonation contours as small F0 movements covering parts of segments, or even a complete segment (Di Cristo, 1982, among others).

2) Above the segmental level, intonation contours show larger falls and rises ('pitch movements'), that may cover one or several syllables. These movements have been traditionally related to several phenomena, such as **'pitch accent'**, **'phrase accent'** or **'boundary tones'**, among others, depending on their position in the utterance (see Bolinger, 1958, Pierrehumbert, 1980, or Cruttenden, 1986, for example, for a more precise definition of these concepts).

3) An example of more global movements present in the intonation contours is the tonal lowering effect reported for the utterances of several languages. As 't Hart *et al.* (1990) state, "In languages like Dutch and English, the most important global attribute is the observed tendency of F0 to decrease slowly from beginning to end of an utterance" (p. 121). Although the question of whether this lowering effect can be considered as an universal phenomenon is still unsolved (t'Hart *et al.*, 1982; Umeda, 1982; Ladd, 1984), the fact that it has been widely reported for different languages provides some evidence of the influence of global phenomena on the intonation contours. There are also different hypothesis about the intonational phenomenon which is responsible of this tonal lowering. Pierrehumbert (Pierrehumbert, 1980; Beckman & Pierrehumbert, 1986) has suggested that it is caused by **downstep**, which relates the tonal lowering to a constant decrease of pitch height from one accented syllable to the next one, irrespective of the distance existing between the two syllables. In other cases, however, this lowering has been related to a constant **declination** effect, dependent on the length of the utterance ('t Hart *et al.*, 1990, among others). The domain of this lowering effect is different then depending on the accepted hypothesis: several accented syllables maximum in the case of downstep, up to the whole utterance in the case of declination.

4) Studies such as those presented by Thorsen (1980, 1985, 1986), among many others, have revealed the presence of even more global patterns in the intonation contours. Thorsen found tonal lowering effects similar to the ones reported previously in lower domains, covering even whole paragraphs. Thorsen related this to a **'supra-declination'** phenomenon having the paragraph as action domain. This effect has also been reported for Spanish in Garrido *et al.* (1993).

According to the way they are realized in the intonation contour, these different types of intonational phenomena can be classified in two main groups:

a) A first group of phenomena which are realized in the F0 contours in the form of specific F0 movements or patterns, that make up directly the shape of the intonation contours: micromelodic variations, pitch accents, and boundary tones. Since they share as common characteristic the fact that they tend to show a narrow domain (from one segment to several syllables), they are called here **local phenomena**.

b) A second group of phenomena which tend to define the relative height of these local movements depending on their position in the utterance: downstep, declination and 'supra-declination'. They do not have a 'visible' movement associated to them; their effects are rather recognizable in the realization of the local patterns. Due to the fact that they tend to have a large domain (from several pitch accents to a paragraph), and that they define in fact the global shape of the intonation contours, they are called here **global phenomena**.

It can be then hypothesized that the generation of the intonation contours is the result of the effect of these intonational phenomena operating at different domains. The question that arises next is whether these different phenomena act simultaneously or in several steps (cycles) to build the intonation contours. In other words, if the generation process of the intonation contours takes place in a linear or a hierarchical way. This problem has been discussed already in the literature - see, for example, the recent contributions of Ladd (1995), Grønnum (1995) and Beckman (1995) - but no definitive solutions have been offered at this time. Ladd (1988) summarizes the two main approaches to the question:

a) According to the **linear hypothesis**, intonation is the result of a sequential addition of F0 movements. Intonation contours are then generated from left to right, in only one step. This idea is present in the approaches followed by Ladd (1983) or Liberman & Pierrehumbert (1984), for example.

a) According to the **hierarchical hypothesis**, intonation contours are the result of the interaction of different-level modules that work one after another, in several cycles: global phenomena, such as declination, define first the global shape of the contours, from the beginning to the end, and once the global shape has been defined, the local phenomena add local variations to the global shape. The different-level intonation movements are then superimposed one over another, in a cyclical way. This hypothesis is underlying several descriptions of intonation, as those of

Thorsen (1979, 1985, 1986), for example, or Garding (1983), among others.

The acceptance of the hierarchical hypothesis implies the existence of **preplanning** in the generation of the intonation contours: the speaker knows in advance how long is going to be the utterance, and adapts the declination slope of the generated contour to its length. However, the linear hypothesis considers that intonation contours are not preplanned, that is, that the speaker does know in advance when he is going to stop speaking, and that the tonal lowering in not preplanned, but controlled locally. Ladd (1988) points out against the hierarchical hypothesis that some declination effects, as the declination reset, may be determined by factors different from a 'superordinated declination' effect, as the syntactic structure of the sentence; those effects would also be coherent with a linear model. The problem of preplanning is also considered by Toledo & Gurlekian (1990), using data from Spanish.

Although the question of the generation of intonation contours in fluent speech seems far from being solved, it seems clear at least that the structure of the contours can be described quite properly in a hierarchical way. For this reason, this work analyses and describes Spanish intonation assuming this hierarchical structure.

It is more difficult, however, to decide the number of domains or levels to be recognized in these structures. The literature does not agree at this point, as is discussed in the next section.

1.1.3. Intonation domains

In the previous section, it has been assumed that intonation contours are the result of different intonational phenomena, each one with its own domain of application. These intonational phenomena are recognized in the contours by the presence of special intonation movements (in the case of local phenomena), of by changes in the relative height of these local movements (in the case of global phenomena). The first consequence that is derived from the assumption of the hierarchical structure of intonation is the fact that intonation contours show different-level domains or units. The definition of intonation units has been a problem considered in many analysis of intonation, both from the field of Phonetics and Phonology. As a general definition, an **intonation domain** (or **intonational unit**) can be considered as the domain at which an intonational phenomenon takes place.

Following the general classification presented above, it is hypothesized here that at least two different levels, a local and a global one, can be recognized

in the intonation contours of any language. Accordingly, two main types of domains or units can be considered:

a) units within which local phenomena take place (**local units**): segments, syllables, stress groups.

b) units subject to global phenomena (**global units**): intonation groups, clauses, sentences, paragraphs.

In the following sections these two types of units are considered in more detail, although the question of the choice of the units for the analysis of Spanish is considered later, in section 1.2.

1.1.3.1. Local units: segments, syllables, stress groups, accentual phrases

The **segment** can be considered as the domain of the micromelodic variations. These variations appear usually as F0 movements covering part of the segment, or the whole segment. It has been demonstrated, for example, that the F0 level of vowels is affected by their height (intrinsic F0), as well as by their consonantal context (Lehiste & Peterson, 1961; Di Cristo, 1982).

The **syllable** is present as a basic intonation unit in many approaches to intonation (Pierrehumbert, 1980; Thorsen, 1978, among others). In these approaches, each syllable, or each accented syllable of a utterance, has an intonation 'level' associated to it. The interpolation of the different intonation levels along an utterance forms the complete intonation contour. Syllables would then be the domain of the pitch levels. A discussion about the validity of this type of units can be found in section 1.3.4.1.

The **stress group** has also been proposed as an intonation unit in several descriptions. It is defined in Thorsen (1978) as a unit formed by "a stressed syllable plus all succeeding unstressed syllables (within the same non-compound sentence), irrespective of intervening word- or morpheme boundaries" (p. 19). Thorsen argues that stress groups defined in this way show a special intonational pattern. Units similar to the stress group have also been defined from the field of Phonology, as in the case of the **foot** ("a string of one relatively strong and any number of relatively weak syllables dominated by a single node"; Nespor & Vogel 1986, p. 84). Stress groups, according to Thorsen's description of Danish intonation, would present a special intonation pattern that would be repeated along every stress group in a contour. However, other descriptions of intonation do not consider the presence of intonational patterns at this level, as is the case of Navarro's (1944) description of Spanish.

Finally, a similar unit has been proposed in Beckman & Pierrehumbert (1986), the **accentual phrase**, that can be defined as a unit containing a single pitch accent. According to their definition, however, it remains unclear what the scope of this unit is.

1.1.3.2. Global units: intermediate phrase, intonation group, sentence, paragraph

Beckman & Pierrehumbert (1986) defined the **intermediate phrase** as the prosodic constituent immediately below the intonational group. According to their definition, the scope of an intermediate phrase is a set of one or more accentual phrases, but rarely more than three. They also defined the intermediate phrase as the domain of the downstepping phenomena.

The **intonation group** is the most widely accepted unit of description in intonation. Intonation groups have been associated to many different names (sense-groups, breath-groups, tone-groups, tone-units, phonological clauses, intonational phrases), every one having its own definition, but all of them referring to a similar concept. Intonation groups can be defined as the domain of a basic (global) intonation pattern.

However, as Cruttenden (1986) points out, a precise definition of this unit has not yet been achieved: "almost all analysts operate with some notion of intonation-groups although most writers have no explicit discussion of how the division between intonation-groups is signalled" (p. 35). Several problems arise, in fact, when the recognition of intonation groups in an utterance is attempted.

The first problem is that, following the classification of Cruttenden (1986), it is sometimes difficult to recognize the internal and external markers of the presence of intonation groups. Internal markers, that is, global intonational patterns, are not always easily recognizable. And external markers, such as pauses, F0 movements and final-vowel lengthening, are not absolutely reliable cues for the location of an intonation boundary. It is important to note here that there is a tendency to identify the **phonic group** (part of a utterance between two pauses) with the intonation group. However, they are not always equivalent, because the presence of a pause is not always a cue of an intonation group boundaries, and inversely, these boundaries can also be indicated by cues different than a pause.

A second problem in the recognition of intonation groups is that they do not have to coincide with any specific linguistic unit. This problem has previously been addressed by Chomsky & Halle (1968), who pointed out the lack of mapping between syntactic domains and intonation groups. More recently, Nespor & Vogel (1986) have also analysed this question. They

define intonational phrases as "the domain of an intonation contour", and state that "the ends of intonational phrases coincide with the positions in which pauses may be introduced in a sentence" (p. 188). As far as linguistic scope is concerned, they consider that there are some syntactic units that may work as intonational units: "parenthetical expressions, nonrestrictive relative clauses, tag questions, vocatives, expletives, and certain moved elements"; sentences are candidates to be the domain of an intonation unit as well. But often intonational units do not coincide with these structures. To take this fact into account, they introduce the concept of "restructuring of intonational phrases". According to this, intonational phrases may be grouped, or split up into several smaller intonational phrases, depending on several factors, such as the length of the constituent, speaking style or tempo. Moreover, Selkirk (1984) considers that the constituents included in the same intonation group have to form a sense unit, and formulates the 'Sense Unit Condition' to predict the intonational grouping of syntactic constituents along utterances.

Despite these definition problems, it is generally accepted that intonation groups are units at which an intonational pattern appears, and that they tend to coincide with phonic groups, or the part of a utterance between two pauses.

There is some evidence of the existence of larger units above the intonation group. There is no agreement, however, about the nature of these units, their domain and organization. The finding of 'supra-declination' structures, pointed out in the works of Lehiste (1975), Sorensen & Cooper (Sorensen & Cooper, 1980; Cooper & Sorensen, 1981), Thorsen (1980, 1985, 1986), Bruce (1982b), Ladd (1988), Sluijter & Terken (1992, 1993) and Nicolas (1995) favours this hypothesis. F0 reset, considered in works as Collier (1985, 1986), for example, also supports this idea.

Several proposals have been made in this direction. Ladd (1986), for example, considers that 'major phrases' (MP), the equivalent of intonation groups in Ladd's description, can be grouped in larger MPs; he does not consider, however, the existence of a different higher-level intonational unit. Nespor & Vogel (1986), using phonological arguments, propose the existence of a higher level unit, the **phonological unit**, as a prosodic unit formed by one or more intonational groups, and whose scope would be a clause or sentence. As in the case of intonational groups, phonological utterances can be restructured, merging several units into a single one. These proposals, however, conflict with the 'Strict Layer Hypothesis' defined by Selkirk (1981), which defends the idea that a constituent of any level in the prosodic constituent structure would be formed by one or more units of the immediately lower level. To preserve this hypothesis, the existence of new higher level intonational units needs to be considered.

Sentence is also a unit traditionally considered as the domain of intonational phenomena. As reviewed in more detail in section 1.2., Navarro (1944) considered that sentences show special intonational patterns in Spanish. Also, the results of Cooper & Sorensen, Thorsen, Ladd and others provide phonetic evidence that sentences and **paragraphs** show some kind of intonational patterns. However, their consideration as intonational units is still being debated, at least from the phonological point of view.

1.1.3.3. Summary: a hierarchy of intonational units

The existence of a hierarchy of prosodic or intonational units has been recognized both from phonetic and phonological approaches. According to the state-of-the-art presented here, the local and global units that may be important in the generation of intonation contours could be arranged in the following layers:

1) segment
2) syllable
3) stress group / accentual phrase
4) intermediate phrase
5) intonation group
6) sentence
7) paragraph

In many cases, the precise definition of their domain and the type of intonational phenomena that occur within their scope is still in progress, as is the case of the accentual phrase. It is also possible that other units need to be defined, to take into account, for example, the fact that in some cases the intonational phrases can be grouped into larger units, as Ladd or Nespor & Vogel propose. These two questions are considered in the experimental part of this work.

1.1.4. Factors determining variations in intonation

A final question that is important to be considered in the description of intonation contours is how their shape can change according to a series of factors. It is well known that intonation contours are used in the communication process to transmit linguistic and non linguistic information. The shape of the contours tends to vary to indicate linguistic phenomena such as the syntactic structure, a focused element, or a stressed syllable, among others, but it also helps the listener to identify the geographical origin of the speaker, or his/her cultural level. There are several reviews in the literature that describe these possible sources of variation in the F0 contours, from purely physiological to stylistic, including different levels of

linguistic phenomena. See, for example, Rossi *et al.* (1981), Jassem & Demenko (1986) or Garrido (1991). These sources of variation can be grouped in the following way:

1) **Physiological and phonetic variations**. These types of variations do not add any relevant information to the message, and are related to the characteristics of the articulatory system of the speaker, or to the phonetic characteristics of the utterance. Examples of this type of variations are:

   a) F0 frequency of the speaker (related to sex, individual F0 variations and age).
   b) Pathological variations in the F0 of the speaker.
   c) Length of the utterance, that may be responsible, for example, of declination slope.
   d) Segmental components of the utterance, that determine, for example, micromelodic variations.

2) **Linguistic variations**. F0 contours are used in the communication process to transmit different types of linguistic information. The shape of F0 contours will be then varied accordingly to code this information in the phonetic realization of utterances. Using this information, listeners can decode the message contained in the utterance. This linguistic information can be also found at different levels:

   a) Lexical stress, used in languages such as Spanish as an additional identification cue for words, is marked - at least partially - in the phonetic form by movements of F0 contours (Lieberman, 1960).
   b) Word, phrase and sentence boundaries are also marked in several languages by special F0 movements at the limits of these units (Cooper & Sorensen, 1977; O'Shaughnessy, 1979).
   c) Sentence type is also marked by special F0 movements at the end of sentences, or by the global shape of the contours (Thorsen 1978, 1980, among others). However, sentence type may be indicated also by other language-dependent means, as the use of wh- particles in the case of wh- questions.
   d) Focus and emphasis may also be marked in the phonetic form by local movements (O'Shaughnessy, 1979; de la Mota, 1995). As in the rest of cases, they may be also marked by using other linguistic resources, such as topicalization.

3) **Emotional, sociolinguistic and stylistic variations**. These variables have in common that they are not used to decode the message, but add some type of information about the speaker or the situation in which the message has been uttered

a) The feelings and emotions of the speaker may also influence the shape of F0 contours, for example in the range of F0 variation (Léon, 1970, among others). In this case, it is difficult to establish if emotion is really an extralinguistic information in the utterance, or a part of the message that the speaker is intending to transmit. These may be called the expressive variations.

b) The shape of F0 contours may also be determined by the sociocultural background of the speaker (place of origin, mother tongue or dialect, social level). The range of the contours, for example, may depend of the cultural level of the speaker, and some particular movements may also depend of the his/her dialect. These may labelled as sociolinguistic variables. Quilis (1985, 1989) are examples of the analysis of the influence of sociolinguistic factors on intonation contours.

c) Finally, the type of communicative situation may also influence in some features of the shape of F0 contours. Variations related to rhythm, for instance, could be included in this group, because they can help to recognize the situation or context at which the utterance has been uttered. These may be considered as stylistic variations. Laan & van Bergem (1993) and Blaauw (1995) are examples of the present interest in the study of the influence of speaking style in intonation contours.

Only a subset of these variables has been considered for this work. The development of an intonation model specially oriented to its implementation in a TTS system has to take into account, at least at a first stage, mainly linguistic variables, since they are the ones that ensure the transmission of information in the message. The relation between three linguistic phenomena - stress, sentence type and syntactic structure - and the shape of the Spanish intonation contours is considered here. However, phonetic variables, such as the length of the different units, have also been taken into account.

1.1.5. Summary

The contents of the previous sections can be summarized in the following points:

a) The word 'intonation' is used to refer to two different phenomena, a phonological one and a phonetic one. In this work, intonation is considered in this second sense.
b) Intonation shows a hierarchical structure, with different phenomena operating at different domains. Two main domains can be considered in

intonation, namely, a local level and a global level. Each level includes its own intonation units, although their precise number and nature it is not yet well defined.

c) Intonation at all these different levels has different - linguistic and non-linguistic - functions. Only a limited number of the variables that determine the shape of an intonation contour have been considered for this analysis, linguistic in most of the cases .

1.2. Spanish intonation

The goal of this introductory section is to present the global descriptions of Peninsular Spanish intonation, that are closer to the aims of the work presented here. Global descriptions of non-Peninsular Spanish, such as the one by Sosa (1991), are excluded from this review. Partial descriptions are also reviewed later, when dealing with specific questions of Spanish intonation.

The first description that tries to offer a global view of Spanish intonation contours is the one by Navarro (1944, 1946). His work, based on kimographic and perceptual analyses of a corpus of literary recordings, is still the most exhaustive description available for Spanish intonation. It has obvious limitations, as the lack of a really formal description, due to the characteristics of the analysis procedure, and also to its scope, limited to 'literary' speech. However, it still remains as an obligatory reference point for any attempt of global study on Spanish intonation contours.

The description presented in Canellada & Kuhlmann (1987) is based in the work by Navarro. They accept Navarro's analysis, validating it with a more modern equipment, and using a more varied corpus. The description suggested by Alcina & Blecua (1975) is also based on the one of Navarro, with slight modifications.

Quilis (1981, 1993) presents a phonological description of Spanish intonation contours. The material used for this description is not specified, but it seems to cover simple utterances, main sentence types and different syntactic structures.

The work by Fant (1984) may also be included in this list of global descriptions of Spanish intonation contours. Although his goal is not the description of intonation contours and his results are not exhaustive, it is interesting because it represents an attempt to define a formal framework for the description of intonation contours.

More recently, some formal analyses have been carried out for the development of TTS systems for Spanish. It is possible to find references to

some of them in the literature (Olabe, 1983; Moreno *et al.*, 1989; Rodríguez *et al.*, 1993), but the results have not been fully published. The work by López (1993) can also be included in this group.

Garrido (1991) presented a formal description of the patterns of Spanish intonation corresponding to simple controlled sentences read within short dialogues. The work presented here tries to go beyond that first approach, not only in the definition of a more accurate methodology, but also in the presentation of a more complete set of intonation patterns.

The analysis of the works previously mentioned shows that the description of Spanish intonation contours is far from being complete. Although some global descriptions have been presented, there are still several questions that need further research. In the following sections, some of these questions, particularly those that are more related to the goals of this work, are reviewed. These questions are: the definition of the intonational units in Spanish; the definition of the intonational patterns for each type of unit; and the relation between intonational patterns and linguistic functions.

1.2.1. Intonation units in Spanish

Several hypotheses about the hierarchical structure of Spanish intonation contours have been proposed in the literature. They are summarized in the following sections.

1.2.1.1. Navarro's levels of description. Related approaches

Navarro (1918, 1944, 1946) establishes two levels in the intonation contours of Spanish:

a) **'Melodic unit'** (*Unidad melódica*):

According to Navarro (1946, p. 77), *"La unidad melódica es la porción más corta del discurso con sentido propio y con forma musical determinada"*. This melodic unit tends to coincide with the phonic group, that is, a portion of utterance located between two pauses: *"Los límites de la unidad melódica coinciden con los del grupo fónico"* (Navarro, 1946 ,p. 77). However, Navarro also remarks that melodic units are not always delimited by pauses; they may also be marked by changes in energy and/or duration, or by special F0 movements: *"Las divisiones entre estos grupos o unidades no van siempre marcadas por verdaderas pausas. Con frecuencia el paso de una unidad a otra se manifiesta solamente por la depresión de la intensidad, por el retardamiento de la articulación y por el cambio más o menos brusco de la altura musical, sin que ocurra real y*

efectiva interrupción de las vibraciones vocálicas" (Navarro, 1944, p. 31).

This melodic unit coincides with the intonation group described in chapter 1.

Navarro also states that the intonation contours of the melodic units may be divided into three different parts:

a) the initial part, from the beginning of the group to the first stressed syllable.

b) the body of the contour, from the first stressed syllable to the last stressed syllable in the group.

c) the final part, from the last stressed syllable to the end of the group.

These parts, as will be considered later, will determine the shape of the intonational pattern associated to the melodic units in Spanish.

b) **Sentence**:

The sentence is the highest-level unit considered in the description of Navarro (Navarro, 1944, p. 31). Sentences are organized, according to their intonational structure, in two parts (Navarro, 1944, p. 40), the 'tension part' (*'rama tensiva'*) and the 'distension part' (*'rama distensiva'*). Each part may be formed by one or more melodic units.

Navarro mentions a third type of unit, the **'tonic group'**, (*'grupo rítmico-semántico'* or *'grupo tónico'*), defined as *"la parte del discurso que tiene por base prosódica un sólo acento espiratorio y por contenido ideológico un núcleo de significación no susceptible de divisiones más pequeñas"* (Navarro, 1944, p. 29). However, he considers that it is not a real unit for intonation analysis: *"el grupo rítmico-semántico, como la palabra misma, no figura con individualidad propia en la entonación castellana"* (Navarro, 1944, p. 30). This unit is similar to the stress group discussed in a previous section, although its definition is slightly different: the stress group always starts with a stressed syllable, but this is not necessarily the case of the tonic group.

Some other studies have borrowed this proposal, introducing certain variations in some cases. This is the case of Alcina & Blecua (1975), Fant (1984), and Canellada & Kuhlmann (1987). Their inventory of units is summarized in the following lines.

Alcina & Blecua (1975) define three levels of analysis which coincide with the units defined by Navarro:

1) The rhythmic unit (*unidad rítmica*), equivalent to Navarro's tonic group.
2) The melodic unit.
3) The utterance, that coincides with the sentence in Navarro's description.

Fant (1984) also established a set of intonational units for Spanish similar to those proposed by Navarro. He defined two main levels of description:

1) the 'tonic group' (*grupo tónico*), equivalent to the one defined by Navarro.
2) the 'prosodic phrase' (*frase prosódica*) equivalent to the 'melodic unit' proposed by Navarro.

Finally, in the case of Canellada & Kuhlmann (1987), the only difference with Navarro's model is the consideration of a new unit, called *cláusula*, that is defined as a portion of utterance between two energy maxima: *"Cada altura en la línea de intensidad marca el comienzo de una unidad que llamamos cláusula. [...] Esta cláusula se extiende hasta que, tras un mínimo, una nueva altura dé comienzo a otra unidad. La lengua castellana no admite más de dos sílabas seguidas sin acento"* (p. 84). However, this unit is not strictly speaking an intonation unit in the sense considered here, because its definition is based in the presence of 'energy accents', not pitch accents. They do not consider the tonic group as an intonational unit, but they accept the intonation group and the sentence as higher-level intonation units.

1.2.1.2. Quilis' levels of description

A different proposal is the one by Quilis (1975, 1981, 1993). He only considers two levels of description, namely the intonation group and the syllable:

a) **Intonation group** (*grupo de entonación*):

Quilis defines the intonation group as the *"porción de discurso comprendida entre dos pausas, entre pausa e inflexión del fundamental, entre inflexión del fundamental y pausa, o entre dos inflexiones del fundamental, que configura una unidad sintáctica más o menos larga o compleja (sintagma, cláusula, oración)"* (Quilis, 1993, p. 419). As in the case of Navarro's description, the intonation groups tend to coincide with the phonic groups, although in some cases their boundaries can be marked by cues other than pauses.

b) **Syllable**:

The syllable is the unit related to stress placement in Quilis' description. As is discussed later, each syllable has a pitch level associated to it.

1.2.1.3. Mixed approaches: López (1993)

The work by López (1993) is an example of hybrid description, taking units from both approaches. López analyses Spanish intonation in order to define a model specially oriented to a TTS system, the 'Hispavoc' of CNET (France Telecom). He establishes three levels for his analysis:

a) Syllable.
b) Stress group:

López defines the stress group as the *"conjunto de sílabas consecutivas con un único acento léxico principal. Está formado por aquel conjunto de palabras no acentuadas seguida de una palabra acentuada"* (p. 39). This definition, in fact, refers to the same type of unit that the tonic group of Navarro's description.

c) Intonational group, which always coincides with phonic group.

1.2.1.4. Summary

As can be observed from the previous paragraphs, the intonational units defined for Spanish coincide more or less with those defined for other languages. At a local level, the syllable has been used in some cases, and the notion of tonic group has been widely considered, although its real status as an intonation unit is far from being demonstrated. It has to be noted, however, that the tonic group considered here does not refer to the same type of unit as the stress group defined in section 1.1.3.1.

At a global level, the melodic unit or intonation group has been considered the basic domain for intonational phenomena, as in other languages. Above the intonation group, only Navarro considers one more unit, the sentence.

A hierarchy of proposed units can also be established, similar to the one discussed in section 1.1.3.3.

Spanish   **General**

	• segment
• syllable	• syllable
• tonic group	• stress group / accentual phrase
	• intermediate phrase
• melodic unit / intonation group	• intonation group
• sentence	• sentence
	• paragraph

In this work, the tonic group, the intonation group and the sentence, three widely used domains in the classical phonetic analyses of Spanish intonation, will be considered as analysis units. They have been chosen to determine to what extent they can be really accepted as intonation units in Spanish. Also, the paragraph has been added to these three levels, taking into account the results obtained for other languages, and for Spanish, as will be reviewed later. Lowest-level units, as syllables and segments, have been excluded from this analysis.

1.2.2. Intonation patterns in Spanish

Another problem in the studies of Spanish intonation published so far is the lack of information concerning the evolution of F0 contours at each level of description. In the following sections the patterns defined in the literature are presented, and the main gaps in the descriptions are outlined. In the following sections, the patterns that have been defined for each one of the considered levels of analysis - tonic group, intonation group, sentence and paragraph - are presented.

1.2.2.1. Intonation patterns in the tonic group domain

Following the criteria suggested by Fant (1984), the Spanish tonic groups may be classified into four types, depending on their position in the intonation group, : initial, internal, final and initial-final. In fact, this proposal partially agrees with the three different parts of the intonation groups - initial, medial, and final - defined by Navarro (1944). The evolution of F0 movements in each type has been only partially analysed: there are not global descriptions of the evolution of the intonation contours at tonic groups, but only descriptions focused in the stressed syllable. Only the movements in the final tonic group have been described in some detail.

1.2.2.1.1. Initial tonic groups

Fant (1984, p. 30) considers that the evolution of F0 along initial tonic groups goes from a low F0 level to an extra-high level (A+, according to his notation; see section 1.3.4.1.2. for a description of his proposal) at the end of the syllable following the stressed one, where the peak is located. Navarro (1944) also mentions that the F0 peak tends to delay to the following (unstressed) syllable at the first stressed syllable of an intonation group: *"Es frecuente que la sílaba débil que sigue inmediatamente a la primera acentuada resulte algo más alta que ésta, con diferencia de uno o dos semitonos"* (p. 49).

1.2.2.1.2. Internal tonic groups

For internal tonic groups, Fant (1984) proposes an F0 pattern characterized by a medium level (M) at the beginning of the stressed syllable, and a high level (A) at the end of the stressed syllable. This displacement of the peak to the end of the stressed syllable or even to the next (unstressed) syllable has also been reported in Garrido *et al.* (1993) and Prieto *et al.* (1994).

1.2.2.1.3. Final tonic groups

The interest of all researchers who have described Spanish intonation has been focused in the movements at the end of the intonation groups. Navarro (1944) distinguished 5 different types of F0 movements that can appear at the end of a melodic unit:

   1) falling (*cadencia*).
   2) half-falling (*semicadencia*).
   3) level (*suspensión*).
   4) half-rising (*semianticadencia*).
   5) rising (*anticadencia*).

According to Navarro, these final movements contain the main intonational information, and they are called 'tonemes' (*tonemas*). Alcina & Blecua label these movements as 'final transitions' (*transiciones finales*), and Stockwell *et al.* (1956) and Quilis (1981, 1993) prefer to use the term 'terminal junctures' (*junturas terminales*) to refer to them.

Navarro (1944) also considers the rise-fall pattern (*tonema circunflejo*) as a cue for special types of sentences (relative questions, for example). This pattern, according to Navarro's description, would start to rise in the last stressed syllable, and to fall at the end of the same syllable or at the following one.

Finally, Fant (1984) considers only two possible final movements or terminal junctures: rising and falling. The typical pattern for falling movements is, according to Fant's description, a medium (M) level along the stressed syllable, and a low (B) level in the syllable following the stressed one. Quilis (1981, 1993) establishes a similar distinction in two types of terminal junctures, falling and rising.

1.2.2.1.4. Initial-final groups

Fant (1984) is the only study that gives some information about the F0 movements in intonation groups formed by just one tonic group. According to his description, the peak (F0 maximum) appears before the end of the stressed syllable, and reaches a high (A) or extra-high (A+) level. Then the contour falls until a low (B) level at the end of the utterance.

1.2.2.2. Intonation patterns in the intonation group domain

As stated before, the description presented in Navarro (1944, pp. 48-49) established three different components in the intonation groups: an initial part, until the first stressed syllable; a medial part, from the first to the last stressed syllable; and a final part, from the last stressed syllable to the end of the group. Following this scheme, he offers a general pattern for the intonation contours in Spanish intonation groups comprising three parts:

a) an initial rising, until the first stressed syllable or the syllable following it.
b) a more or less uniform tone in the central section.
c) depending on the type of intonation group, the final part will show a different final movement, or toneme, as described above (1.2.2.1.3).

The tonal lowering effect discussed in chapter 1 has also been reported for Spanish intonation groups. Fant (1984, p. 40) considers this possibility, although he states that it is only present in the internal stressed syllables. Garrido (1991) also reports this effect in one-intonation-group declarative sentences. This downtrend tendency in Spanish has recently been considered in terms of downstepping phenomena in Prieto *et al.* (1995a), using a corpus of Mexican Spanish.

1.2.2.3. Intonation patterns in the sentence domain

As has been discussed earlier (section 1.2.1.1.), Navarro (1944) considers two parts in the intonation structure of any sentence, namely a 'tension' and

a 'distension' part. According to this hypothesis, the different intonation groups of a sentence are included in one of these two parts. Navarro also proposes a general hypothesis about the organization of intonation contours in Spanish in the sentence domain. The idea is that the contours of the intonation groups of a sentence show rising or falling final patterns depending of whether they appear in the 'tension' or the 'distension' part, according to the following general rules:

1) the final group of the 'tension part' presents a rising final F0 movement.
2) the final group of the 'distension part' shows a falling final F0 movement.
3) the non-final groups of the 'tension part' tend to finish with a half-falling movement.
4) the non-final groups of the 'distension part' tend to finish with a half-rising movement.

Navarro indicates that this rule presents some exceptions in the case of the non-final groups, depending on the type of syntactic boundary appearing at the end of the group.

This hypothesis is illustrated in the case of Navarro (1944) and Canellada & Kuhlmann (1987) with examples of declarative sentences of relatively low syntactic complexity. It is not clear, however, if this model makes adequate predictions in the case of more complex sentences.

1.2.2.4. Intonation patterns in the paragraph domain

As has been reported before for other languages, the above mentioned work by Garrido *et al.* (1993) showed that Spanish intonation also shows 'supra-declination' phenomena at paragraph level. However, the results of this work are based in data taken from 'laboratory' speech. However, the results of the work presented in Garrido (1993), based on less controlled speech (readings of news by professional and non-professional speakers), seem to indicate that this 'supra-declination' phenomenon does not always cover the entire paragraph.

1.2.2.5. Summary

This review shows that the description of intonation patterns in Spanish leaves some unanswered questions:

a) the analysis of the non-final movements in the tonic group domain has not been considered in detail yet.

b) the analysis of intonational patterns can also be attempted in more detail: the basic pattern proposed by Navarro (1944) does not take into account, for example, other global phenomena more recently defined, as declination or downstep. In addition, this pattern does not explain how local movements are integrated into the global shape.

c) in the sentence domain, the Navarro's model establishes some predictions that seem to be valid in the case of short sentences, but it is still necessary to check to what extent this model can be applied to longer sentences.

d) higher level phenomena seem to exist in Spanish, at least in 'controlled' material. The formal description of these phenomena and their integration in a global model are still pending.

1.2.3. Intonation patterns in Spanish and the expression of linguistic information

A third set of problems related to the description of Spanish intonation is the definition of the relationship between linguistic functions and F0 movements. This section reviews the different studies of the intonational cues that are used in Spanish to mark different linguistic phenomena. Only the three ones considered in this work are reviewed, namely, lexical stress, syntactic structure, and sentence type.

1.2.3.1. The relationship between F0 shape and lexical stress

There is still some debate about the role of F0 in the phonetic realization of lexical stress in Spanish. This phonetic realization has been traditionally related to cues different than F0, that is, amplitude and duration. Navarro (1944, p. 21 and ff.), for example, argues that amplitude and pitch are the main correlates of stress when a word is spoken in isolation; however, in the case of words embedded in longer utterances, Navarro considers that amplitude is the main contributor to stress perception, while pitch is used for more global functions: "*En español es el acento de intensidad el que mantiene la estructura prosódica de las palabras, mientras que el tono, por su parte, rige la organización melódica del periodo*". This question has also been debated in Bolinger & Hodapp (1961), Contreras (1963, 1964) and Monroy (1972). Canellada & Kuhlmann (1987) consider that duration is the most important acoustic cue involved in the perception of stress. More recently, perceptual experiments (Enríquez *et al.*, 1989) have demonstrated the relevance of F0 in the identification of stressed syllables in Spanish,

although this does not imply that F0 is really the main phonetic realization of stress in Spanish.

Anyway, some relation between F0 peaks and the presence of a stressed syllable seems to exist. Navarro (1944, p. 21 & ff.) reports that in general F0 tends to rise in stressed syllables and to fall in the unstressed ones: "*De un modo general, en la serie sucesiva de los núcleos silábicos de una frase, el tono en español se eleva más o menos sensiblemente en las sílabas fuertes y desciende en las inacentuadas*".

However, an exact correlation between between stressed syllable and F0 peak may not be found. The F0 peak may be delayed to the next syllable in some cases, as Navarro (1944, p. 39) reported, specially in the first stressed syllable of a phonic group: "*Es frecuente que la sílaba débil que sigue inmediatamente a la primera acentuada resulte algo más alta que ésta, con diferencia de uno o dos semitonos*" . Fant (1984) also reports this fact, and it has also been found in Garrido (1991), at least for the first stressed syllable of the utterances. For non-initial stressed syllables, this effect has also been found in Garrido *et al.* (1993). Llisterri *et al.* (1995) have analysed some of the factors affecting this peak displacement, finding some variations of the peak displacement patterns if a pause is located after the word containing the stressed syllable. They also report some effects of the type of syntactic boundary following the word on the location of the peak. The problem has also been considered in Prieto *et al.* (1994, 1995b), analyzing data from Mexican Spanish. Their results indicate that peak delay is related to the duration of the syllable onsets, the vicinity of a prosodic boundary (intermediate phrase, intonational phrase), and the distance to the next stressed syllable. The results of these studies are reviewed in more detail in section 4.7.1.

As can be observed, most research interest has been focused in the location of the peaks in relation with the stressed syllable. Much less attention, however, has been paid to the location of the start of the rising part in relation to the stressed syllable, which can also be an important cue for their identification. Prieto *et al.* (1995b, p. 449) have recently provided some experimental data about this point: "A description of the accent gesture in the different prosodic contexts revealed that, on the one hand, the location of the start of the rise was stably anchored in the onset of the [stressed] syllable".

According to these data, it can be hypothesized that the presence of an F0 peak in a syllable does not seem to be an absolutely necessary cue to be perceived as stressed, but rather a rising movement within the stressed syllable.

1.2.3.2. The relationship between F0 shape and syntactic structure

The analysis of the relationship between intonation contours and syntactic structure in Spanish has focused in the study of F0 movements at the end of phonic groups, depending on the type of syntactic boundary that coincides with the pause. The general conclusion that can be extracted from the study of these boundaries, as Fant (1984, p. 42) suggests, is that, apart from some special constructions, the use of a particular terminal juncture is not directly linked to a specific syntactic boundary: "*La ocurrencia de una y otra [juntura terminal] normalmente no es deducible de la estructura sintáctica*". The most studied syntactic boundaries have been, accordingly, those where there is a tendency to insert a pause. Examples of these syntactic boundaries can be found in the following sentences:

1) Fronted elements:

Entonces / sale a pasear al campo.
Cuando hace buen tiempo y la temperatura es agradable / sale a pasear.
Sobre el caballo, con el sombrero en la mano, / cabalga el jinete.

2) Parenthetical elements:

Los caballeros, / vestidos con sus armas, / se fueron al combate

3) Adjective clauses:

    a) non-restrictive:

    Los alumnos, / que viven lejos, / llegan tarde.

    b) restrictive:

    Los alumnos que viven lejos / llegan tarde.

4) Asyndetic coordinations:

    a) complete: with a connector ('*y*', '*o*') linking the last two elements of the asyndetic coordination.

    El caballo es fuerte, / rápido / y negro.

    b) incomplete: the last two elements of the asyndetic coordination are not linked by a connector.

    El piso desciende en un declive suave, / resbaladizo, / bombeado.

5) Coordinate phrases or clauses:

   a) 'first degree' coordinations: those showing an '*y*', '*o*' connector (*copulativas, disyuntivas*).

   Dijo que venía / y nos quedamos esperando.

   b) 'second degree' coordinations: those showing a '*pero*' connector (*adversativas*).

   No podía hacer nada, / pero se estaba quieto
   Viene todas las tardes, / a pesar del frío

6) Adverbial subordinate clauses:

   Si mañana no recibo noticias, / le escribiré de nuevo

A brief review of the specific F0 movements associated with these syntactic boundaries is presented in table 1.1.

Syntactic boundary	Intonational cue		
	Movement before the boundary	Movement at the end of the element	Other intonational cues
Fronted phrases	Rising, half-falling (Navarro, 1944)		
Embedded phrases Parenthetical elements	Level or half-falling if it appears in the 'tension part' Half-rising if the group is in the 'distension part'. (Navarro, 1944; Canellada & Kuhlmann, 1987; Quilis, 1993)		General F0 level lower that in the preceding and following groups
Restrictive adjective clauses	Rising (Garro & Parker, 1983)	Falling (Garro & Parker, 1983)	
Non-restrictive adjective clauses	Falling (Garro & Parker, 1983)	Level (Quilis, 1981) Half-falling (Canellada & Kuhlmann, 1987) Rising (Garro & Parker, 1983)	Pause
Complete asyndetic coordinations		Half-rising in the penultimate group Falling in the rest of groups (Navarro, 1944; Quilis, 1993)	
Incomplete asyndetic coordinations		Falling (Navarro, 1944; Quilis, 1993)	
'First-degree' coordinate phrases or clauses	Half-rising (Navarro, 1944)		
'Second -degree' coordinate phrases or clauses	Half-falling (Navarro, 1944)		
Adverbial subordinate clauses	Rising (Navarro, 1944)		

Table 1.1: Described F0 cues for some types of syntactic boundaries in Spanish

The division of sentences into 'tension' and 'distension' parts has also been related to syntactic factors. Canellada & Kuhlmann (1987) present a list of

boundaries that are candidates to the placement of the boundary between the 'tension' and the 'distension' parts:

1) Subject/Predicate:

>   *La niña de la caseta // tenía un vestido rojo.*

2) Fronted PP/rest of the sentence:

>   *En torno // había mesas y bancos rústicos*

3) Adverbial subordinate clause/main clause:

>   *Cuando nadie lo miraba // venía haciendo equilibrios por cima de un raíl.*

4) Fronted element/rest of the sentence:

>   *La caída // era fácil que sucediese.*

5) Fronted Object/rest of the sentence:

>   *A nosotros // no nos regalaba nadie nada.*

6) Other fronted elements:

>   *A la pipa // se le iba el humo por todas partes*

7) Absolute participle, gerundive or infinitive clause/main clause

>   *Al venir nosotros // se aguó la fiesta*

8) Participle or gerund clause, predicative/main clause:

>   *Oculto, hundido entre los rebaños // discurría el Jarama*

9) Subordinate subject clause/main clause

>   *Que tú pagues la merienda // es imperdonable*

10) Two '*y*' coordinate clauses or sentences

>   *La niña lo vio venir // y se paró a mirarlo*

11) Two '*pero*' coordinate clauses

De esta sanamos los tres // o nos volvemos de remate

12) Main clause/subordinate relative clause

Había zonas cubiertas de madreselva y vid americana // que avanzaban por los alambres

13) Beginning asyndetic coordination/last element asyndetic coordination

Mauricio la miró un momento, abrió el cajón // y sacó las pesetas.

14) Main clause/gerund clause

Carmen jugaba con los brazos en alto // trenzando los dedos.

In all these examples, the double bold lines indicate the boundary between 'tension' (left) and 'distension' (right) parts of the sentence. According to Navarro's hypothesis, the end of the 'tension part' is indicated in all these cases with a final rising intonation movement. It can be considered then that the selection of this final movement is also determined by syntactic factors.

Finally, the use of intonation to indicate the presence of syntactic boundaries in internal position of intonation group has been much less analysed. Some preliminary data about this question can be found in Garrido *et al.* (1995).

1.2.3.3. The relationship between F0 shape and sentence type

The relationship between sentence type and intonation contours is the best known aspect of Spanish intonation. Four types of sentence types have been usually considered in the studies of Spanish intonation:

  a) Declarative
  b) Interrogative
    • 'Yes-No' questions
    • 'Wh-' questions
    • Relative questions
  c) Exclamative
  d) Imperative

Sentence type has been traditionally associated in Spanish to the terminal juncture appearing at the end of the sentence. However, other intonational cues have also been proposed. Table 1.2 summarizes the different cues reported in the literature for each sentence type:

	Intonational cue		
	Final F0 movement	Global shape	Other intonational cues
Declarative	Falling (Navarro, 1944; Canellada & Kuhlmann, 1987)	Steady central part (Navarro, 1944)	
'Yes-No' questions	Rising (Navarro, 1944; Canellada & Kuhlmann, 1987; Quilis, 1987, 1993)	Continuous fall along the central part of the contour until the beginning of the final rise. (Navarro, 1944) Higher mean F0 than the declarative sentences (Canellada & Kuhlmann, 1987)	Initial peak 3-4 semitones higher than in declaratives (Navarro, 1944; Garrido, 1991)
'Wh-' questions	Falling Rising (Navarro, 1944; Quilis, 1987, 1993; Garrido, 1991)		Initial peak 3-4 semitones higher than in declaratives (Navarro, 1944; Garrido, 1991)
Relative questions	Rise-fall (Navarro, 1944) Falling (Canellada & Kuhlmann, 1987)		Initial peak 3-4 semitones higher than in declaratives (Navarro, 1944; Garrido, 1991)
Exclamative	Falling Rising Rise-fall (Navarro, 1944)	'Wave-like' contour (Navarro, 1944; Canellada & Kuhlmann, 1987; Garrido, 1991) Higher F0 range than in declaratives Garrido (1991)	
Imperative	Falling Rise-fall (Navarro, 1944)		

Table 1.2: Described F0 cues to mark sentence type in Spanish

1.2.3.4. Summary

As far as the relation between stress and F0 movements is concerned, the main questions seem to be related to the location of the beginning and end of the rising movement traditionally associated with stressed syllables. A more formal description of these movements is needed.

The description of the intonational marking of syntactic structure in Spanish is far from being complete. First, as Fant (1984) suggested, the selection of final movements at the end of non sentence-final phonic groups does not seem to be directly related to syntactic factors. The selection of these movements according to other phonological or pragmatic factors, for example, seems an interesting research question.

Moreover, some other intonational cues, such as F0 reset, whose relevance for the description of the influence of the syntactic structure in the intonation contours has been outlined for other languages (Cooper & Sorensen, 1977, 1980; Collier, 1985, 1987), could also be applied to the description of Spanish, as was already attempted in Garrido (1993). In the case of the studies for other languages, the analysis has been focused on major syntactic boundaries, that is, the boundary between clauses (coordinate or subordinate) within a sentence, or between sentences (Thorsen, 1985, 1986; Ladd, 1988). According to the results of these works, F0 reset seems to be a way of marking hierarchical dependencies between sentences and clauses in speech.

Finally, the study of the relationship between sentence type and intonation is perhaps the most completely described aspect of intonation in Spanish. The relationship between sentence-final movements and sentence type has been widely described, specially for the most frequent types, such as declarative and interrogative. In the case of exclamative or imperative sentences, however, more research needs to be carried out.

1.2.4. General summary

The review of the studies of Spanish intonation presented here leads to the following conclusions:

a) The intonation units proposed for Spanish are in general similar to those of other languages: syllables, tonic groups, melodic units and sentences. However, it is still unclear in some cases (the tonic group is the most evident one) that these units are the domain of some kind of intonational phenomena. The validity of these units should be then revised. In addition, recent studies on Spanish intonation seem to indicate that more global units, such as the paragraph, should be considered in a

model that intends to be a global description of Spanish intonation contours. Taking into account these considerations, four analysis levels have been defined for this work: tonic group, intonation group, sentence and paragraph.

b) More research is necessary to obtain a complete inventory of intonational patterns at each level of description. More or less informal descriptions have been offered for tonic groups, intonation groups, sentences and paragraphs. However, if the development of a formal model is attempted, more accurate definitions are needed.

c) The relationship between the shape of the intonation contours and the main linguistic functions is still incomplete. In addition, nothing has been said about the influence of non linguistic factors, such as duration of utterances, on the shape of the contours of Spanish. This type of information, however, is also important for the development of a formal model of intonation, specially if its application to TTS systems is being considered.

1.3. Intonation description and modelling

This third section deals with those aspects of intonation analysis that involve the description and modelling of the intonation contours, or, in other words, with the methodological aspects of intonation analysis.

1.3.1. Phonetic vs. phonological approaches to the study of intonation

As Cutler and Ladd stated, there are two possible approaches to the study of intonation:

> "There are two broad traditions in the study of prosody that may be characterized - or caricatured - by their methodological preferences for one or the other of the scientific activities mentioned in the title: making measurements and constructing models. On one side of the dichotomy stand instrumental and experimental studies that seek to quantify acoustic features and investigate perceptual responses. On the other are descriptive and theoretical studies of prosodic structure and its relation to other aspects of grammar and phonology. In a great deal of past work these two traditions have simply ignored one another" (Cutler & Ladd, 1983, p. 1).

The first one may be called the phonetic, or 'bottom-up' approach, and tries to study intonation as is manifested in the acoustic or perceptual properties of the speech signal. As Cutler and Ladd pointed out:

> "for many of those who take the measurer's approach, the primary concern is not representation, but *realization*. The question being asked is: What are the physical correlates of this or that prosodic message? To the extent that such investigators have

> constructed explicit models of prosodic representation, they have tended to think in terms not of linguistic categories, but of interacting parameters; their models assign acoustic correlates to individual functions such as word stress, sentence stress, sentence modality, affective use of pitch range, and so on, and attempt to specify the interaction of all these effects on individual parameters like fundamental frequency" (Cutler & Ladd, 1983, p. 5)

The second approach could be called 'phonological', or 'top-down', in the sense that it starts from the linguistic structures, and tries to find out in the speech how these structures are realized:

> "The model-builder is interested in establishing an inventory of abstract categories - a formal *representation* - of prosodic function and prosodic form. The goal of the model-builder's enterprise is to describe the systematic structure underlying prosodic distinctions; the basic assumption is that there are well-defined abstract levels of representation that mediate between specific prosodic functions like "phrase boundary" and specific acoustic traits" (Cutler & Ladd, 1983, p. 5).

The approach followed here may be defined as phonetic, in the sense that it starts by analyzing the acoustic properties of the phonetic output, and tries to obtain more abstract intonation patterns from the results of this acoustic description.

1.3.2. Description and modelling of intonation

When attempting a phonetic approach to the analysis of intonation, the first problem that arises is the interpretation of the F0 contour itself. First, the F0 contours obtained from F0 trackers usually include detection errors, that may be easily recognized in some cases, but not in others. The recognition of intonation patterns in F0 contours may also be difficult due to the presence of micromelodic movements, which 'hide' the real shape of a contour. For these reasons, almost every phonetic study of intonation makes use of some type of simplification of the intonation contours, prior to the analysis and modelling task.

Once the intonation contours have been simplified, the second step involves the description of the simplified intonation contours. This description involves different subtasks:

a) first, a **formal framework** has to be chosen for the description of the intonation contours;

b) second, the different **intonation patterns** present in the analysed intonation contours have to be identified and defined, using the chosen descriptive framework;

c) third, a study of the relation between the different intonational patterns and the different **functions** of intonation has to be performed.

The scheme presented in figure 1.1 shows this procedure:

```
┌─────────────────────────────────────────────────┐
│   ┌─────────────────────────────────────┐        │
│   │                                     │        │
│   │             F0 extraction           │        │
│   │                                     │        │
│   └─────────────────────────────────────┘        │
│                      │  Intonation               │
│                      │  contour                  │
│                      ▼                            │
│   ┌─────────────────────────────────────┐        │
│   │          Simplification             │        │
│   │     (Stylization/Transcription)     │        │
│   └─────────────────────────────────────┘        │
│                      │  Simplified intonation contour │
│                      ▼                            │
│   ┌─────────────────────────────────────┐        │
│   │          Standardization            │        │
│   │      (Descriptive framework         │        │
│   │        + Pattern definition)        │        │
│   └─────────────────────────────────────┘        │
│                      │  Intonation patterns      │
│                      ▼                            │
│   ┌─────────────────────────────────────┐        │
│   │            Analysis of              │        │
│   │        intonation functions         │        │
│   └─────────────────────────────────────┘        │
│                      │                            │
│                      ▼                            │
└─────────────────────────────────────────────────┘
```
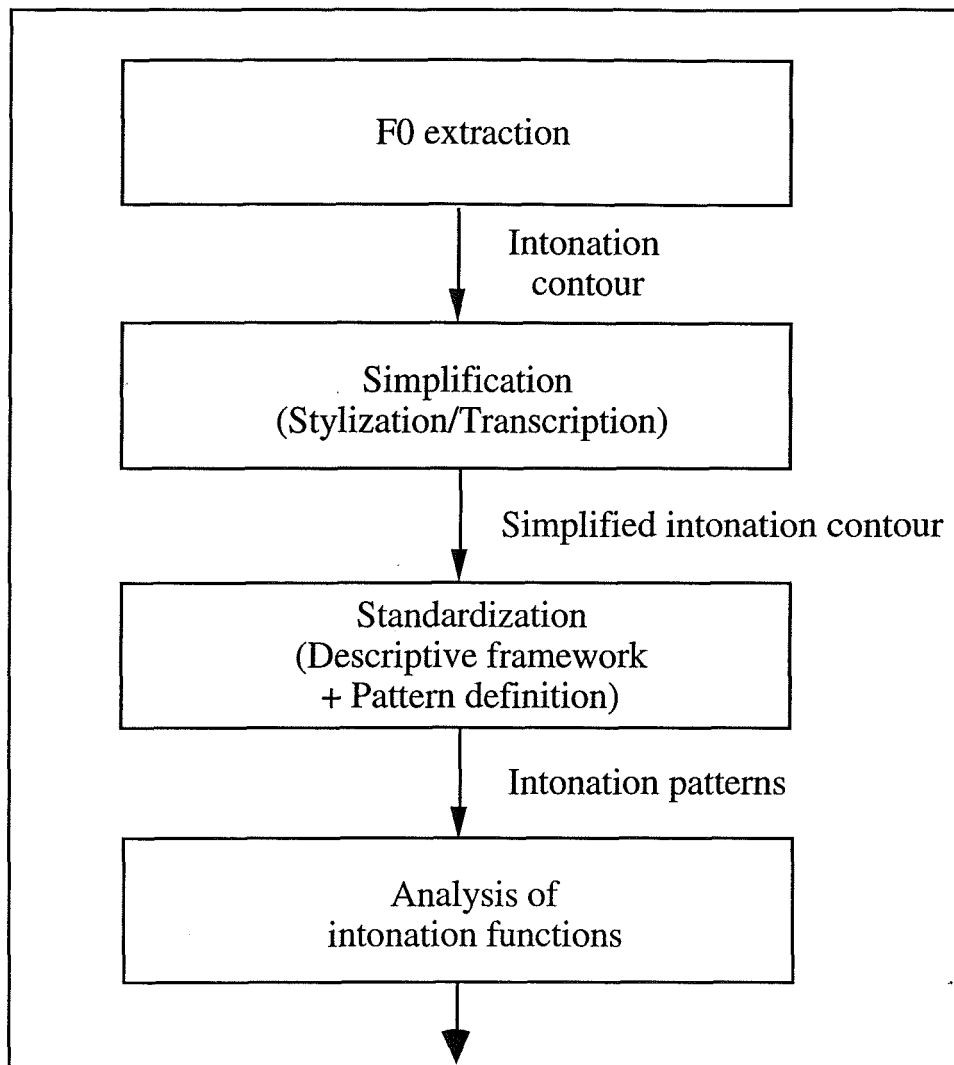
Figure 1.1. Schematic representation of a phonetic approach to the analysis of intonation

Phonetic descriptions of intonation usually include, in a more or less systematic manner, these three tasks. But the definition of an intonation model involves not only a description, but also a prediction of how the intonation patterns are combined to form the intonation contours. These combinatory constraints may be determined in part by the linguistic uses of the intonation contours, but also by other factors. The development of this predictive component can be properly called **modelling** of intonation, and should involve, for example, the definition of a **grammar** of intonation, that will predict how intonation patterns have to be combined according to a series of linguistic and non linguistic parameters.

A good example of phonetic approach to intonation modelling is the one proposed by the IPO for Dutch ('t Hart *et al.*, 1990). The IPO approach considers two main steps:

1) a first step of **stylization**, that allows to eliminate the unnecessary variations from the intonation contours. The result of this stylization process is a 'close-copy' stylization of the analysed contour.

2) a second step of **standardization**, which involves the definition of a formal framework, and the definition of patterns from the stylized contours.

The work presented here follows closely the methodology used in the IPO approach, that is, a first step of stylization and a second step of standardization have been differentiated. More details about the procedures of stylization and standardization used here are given in chapter 3. In the following sections, however, a general overview of these procedures is given.

1.3.3. Stylization and transcription of intonation contours

As has been stated before, raw F0 contours as they are calculated by a pitch extractor are difficult to analyse. They contain errors and variations that are not relevant. In order to facilitate its analysis, it is useful to simplify these F0 contours. To avoid this problem, **stylization** and **transcription** systems are used. Both types of procedures are different, but they share the same goal: to provide a representation of the intonation contours that can be used for analysis and description.

1.3.3.1. Stylization

Stylization procedures simplify F0 contours to obtain schematic representations. In general, these procedures assume that F0 contours may be decomposed in series of **turning points** linked with straight lines or parabolas. These methods may be classified according to different criteria.

A first classification may be proposed according to whether they need some previous linguistic information to carry out the stylization procedure

a) Stylization procedures based on linguistic criteria: they need some previous linguistic information to carry out the stylization procedure. Rossi *et al.*'s method (Rossi *et al.*, 1985), Thorsen's method (Thorsen, 1983) and Mertens & d'Alessandro's method (Mertens & d'Alessandro, 1995), for example, need a previous segmentation of the speech signal into syllables or allophones.

b) Stylization procedures not based on linguistic criteria: these systems take the raw F0 contour and apply directly the stylization procedure. This

is the case, for example, of the IPO method ('t Hart *et al.*, 1990), or the MOMEL system (Hirst *et al.*, 1991; Hirst & Espesser, 1993).

A second classification can be established depending on the type of phonetic information used by the stylization methods:

a) Stylization procedures based on acoustic criteria only use acoustic properties of the F0 contour to derive the stylized representation. Examples of this type are the Rossi *et al.* method, the Thorsen's method, and the MOMEL system

b) Other stylization procedures use perceptual criteria to keep only the perceptually relevant movements in the resulting stylized contour. The IPO stylization method and the Mertens' method fall within this group.

Finally, stylization methods may also be classified into manual and automatic:

a) Manual systems involve the operator making decisions about the stylization process. The IPO method, Thorsen's method and the Rossi *et al.*'s method are included in this type.

b) Automatic methods extract the turning points or values automatically. These automatic methods are faster than the manual ones, but usually need a 'post-processing' step, to eliminate errors in the stylization procedure. The MOMEL method and Mertens & d'Alessandro's methods are examples of this type.

A detailed review of each type of stylization system is not attempted here. However, a short description of the IPO procedure and the MOMEL systems is offered, due to their similarity with the procedure used in this study.

The goal of the IPO stylization system is to obtain 'close-copy stylizations'. A 'close-copy stylization' is an stylized version of an original intonation contour, made up of a series of straight lines, that has to meet two conditions (t'Hart *et al.*, 1990, p. 42):

a) the stylized version should be auditorily indistinguishable from the (resynthesized) original.
b) it must contain the smallest possible number of straight-line segments with which the desired perceptual equality can be achieved.

The MOMEL system is an automatic procedure of stylization that converts the original contour in "a sequence of target-points <ms.; Hz>. These target-points can subsequently be used to generate a quadratic spline

function which is then directly usable as input for PSOLA resynthesis". The target points are obtained by means of "a dissymetric version of robust regression to provide an optimal fit for a sequence of parabolas" (Hirst *et al.*, 1991, p. 234).

1.3.3.2. Transcription

In transcription, the utterances are annotated with predefined symbols to mark the presence of some F0 levels or movements. This is the case, for example, of the ToBI (Beckman & Hirschberg, unpublished; Silverman *et al.*, 1992; Beckman & Ayers, 1994; Pitrelli *et al., 1994*), or the INTSINT system (Hirst, 1991; Hirst *et al.*, 1993; Nicolas, 1995).

The ToBI transcription system uses two main types of symbols (Beckman & Hirschberg, unpublished):

1) A set of 'Break Indices', used to express "the degree of juncture perceived between each pair of words and between the final word and the silence at the end of the utterance".

2) Two tone symbols, H (high) and L (low), that are used to label the "pitch events associated with intonational boundaries (phrasal tones) and pitch events associated with accented syllables (pitch accents)".

This transcription method is based on the description of American English intonation presented in Pierrehumbert (1980).

The INTSINT method considers also two types of symbols, but both representing different kinds of tonal height:

1) Absolute tones: T (top), M (mid) and B (bottom).
2) Relative tones:
    a) non-iterative: H (higher), L (lower), and S (same).
    b) iterative: U (up) and D (down).

A complete review of different transcription systems can be found in Llisterri (1994).

1.3.3.3. Summary

Transcription procedures are more related to phonological approaches to intonation, while stylization procedures seem to be more appropriate for phonetic analysis. However, transcription and stylization seem to be two compatible methods: stylization can be viewed as the first (phonetic) step in

the procedure of simplification of intonation contours, and transcription as the second (phonological) step. This idea is present in the stylization (MOMEL) + transcription (INTSINT) procedure proposed by Hirst.

For this work, an automatic stylization procedure has been used to obtain a first simplified representation of the intonation contours. The reasons for this choice have been the following:

a) a stylization approach seems more adequate for a phonetic study of intonation than a transcription-based approach.
b) the use of an automatic system has two main advantages: first, the stylization criteria can be applied in a more accurate and consistent way; and second, it allows the processing of a large corpus in a reasonable time.

This procedure is described in detail in chapter 3. However, a transcription procedure has also been used to annotate - by hand, in this case - the inflection points of the obtained stylized contours. This transcription system is presented in chapter 4.

1.3.4. Descriptive frameworks for the intonational phenomena

As has been outlined before, a second problem in a phonetic analysis of intonation is the definition of the descriptive framework that will be used to describe the intonation contours. In fact, it is sometimes difficult to distinguish the end of the stylization/transcription procedures and the beginning of the definition of the framework. The following sections describe some possible approaches to the description of the local and the global phenomena in the intonation contours.

1.3.4.1. Describing local phenomena: levels vs. contours

As Cruttenden (1986, p. 45) states, "A basic difference between much British and American intonation analysis in the last forty years has been that, whereas British writers have preferred a 'contour' analysis, American writers have generally preferred a 'levels' analysis".

1.3.4.1.1. The 'contour' approach

The 'contour' analysis presents the local phenomena of the intonation contours as a continuum of movements. The IPO pitch movements, defined from the 'close-copy stylizations', are a good example of contour representations.

The work presented in Garrido (1991), and even the work by Navarro, can be considered as examples of contour analysis of Spanish intonation.

1.3.4.1.2. The 'level' approach

In the 'levels' analysis, the intonation contours are represented as a series of discrete points, each one associated with a level. The description of American English performed by Pierrehumbert (1980) is an example of this type of approach. Most 'contour-based' descriptions, however, also need some kind of level system, as is the case of the IPO approach, that uses three levels - H, M and B - for the definition of the pitch movements.

For Spanish, the descriptions by Quilis (1987, 1993) and Fant (1984) use two systems of transcription of intonation based on the concept of tonal level. Quilis (1993) proposes 3 different tonal levels:

a) /3/ (high): all syllables with 'strong' stress in a utterance.
b) /2/ (medium): corresponds to the weak syllable after the last stressed syllable in a utterance.
c) /1/ (low): after pause or terminal juncture.

This kind of description, considering three levels of syllabic tone, is already present in earlier works, such as Stockwell *et al.* (1956).

Fant's (1984) description, however, considers four tonal levels, defined in the following way:

a) M level (medium): would be the level typical at the boundary between a pretonic and a stressed syllable ("*se localiza en la frontera entre la protónica y la tónica*").
b) B level (*bajo*, or low): would be the level at the end of a declarative utterance.
c) A level (*alto*, or high): the level corresponding to the end of a stressed syllable at medial position.
d) A+ level (*alto extremo*, or extra high): level found at the end of the first stressed syllable of a utterance.

1.3.4.1.3. 'Levels' vs. 'contours'

In general, the 'contour' analyses are related to the stylization procedures, resulting in a description of the **evolution** of the intonation contours. However, 'level' analyses are more related to transcription systems, that usually offer a series of symbols to take the different levels into account.

The question of the adequacy of both types of description to the reality of intonation has been considered in several works. Bolinger (1951), for example, presents the results of a series of experiments to test which of the two approaches is closer to the phonetic and phonological reality of intonation contours. Bolinger concludes that "it has been sufficiently proved that the basic entity of intonation is a pattern, not a pattern in the relatively abstract sense of grammatical recurrences, but in the fundamental, down-to-earth sense of a continuous line that can be traced on a piece of paper" (p. 206), that is, that 'contour' analysis seems a more adequate approach to the study of intonation than the 'levels' one.

Cruttenden (1986) reviews some arguments for and against each approach, to conclude that "the decision to allow an analysis based on contours to predominate in this book remains to some extent an act of faith. But there is at least one factor which tips the balance towards a contour analysis: this is my conviction that there is some basic similarity of meaning in all falling contours as opposed to a basic similarity of meaning in all rising contours" (p. 47).

In fact, both approaches may be complementary. The IPO approach, for example, considers three levels of intonation in the utterance, but it also pays attention to the evolution of the contours from one inflection point to the next. The inflection points, in addition, are not associated to a entire syllable; their location in the syllable may change depending on the type of movement.

1.3.4.2. Describing global phenomena

The description of the global evolution of intonation has been achieved in several ways, using more or less intuitive procedures, as in the case of the Spanish intonation descriptions reviewed in section 1.2., or using more formal systems. In this section a brief review of three of these procedures is presented.

1.3.4.2.1. The stressed-syllable contour approach

A first way of describing the global shape of the intonation contours is as "the course described by the stressed syllables alone" along an intonation contour, as Thorsen (1978, p. 19) suggests. This approach is also present in Bolinger's (1958, 1970) descriptions of American English.

1.3.4.2.2. Navarro's approach

As has been outlined in section 1.2.2.3., Navarro (1944) describes the global shape of the phonic group contours paying attention only to the initial and final parts of the contour. Only variations related to the first and last stressed syllables are considered to be relevant for the description of the global pattern, while the rest of accent movements are 'deleted' of the final representation.

This approach has been used in Garrido (1991) to define patterns for one-phonic-group sentences of different sentence types.

1.3.4.2.3. The topline-bottomline approach

In this approach, the global patterns are described as a set of (more or less) parallel lines that define the maximum and minimum values that the intonation contour can reach at any moment in time. Several variants may be considered in this approach.

A first question is the number of lines that are going to be considered. Some models consider only one line, as is the case of the Pierrehumbert model, that only uses a **baseline** to define the global evolution of the contour. In other cases, two lines, a **top** and a **bottomline**, are used, as in the case of the Lund model (the 'grid'). Finally, the IPO model is an example of use of three lines, a **top**, a **medium** and a **bottom** one. The choice of one or another approach depends on several factors, one of them being the number of 'levels' that are going to be considered in the intonation contour.

A second question is whether these lines can only be descending, or they can be either ascending or descending. The Lund model accepts that top and bottom lines fall or rise depending of the position in the utterance. The IPO model, however, following the declination hypothesis, uses only falling lines.

Finally, the shape of these lines has also been discussed in the literature. The IPO approach assumes that these lines are straight and parallel. More complex shapes, however, have also been suggested. Cooper & Sorensen

(1981), for example, represent the topline as a complex line, with a sharper slope in the initial part of the utterance, and a less sharp one in the second half.

Although this type of representations were initially devised for the description of global patterns along the intonation group, they are also being used to describe higher-level phenomena, such as the 'supra-declination' patterns reported at paragraph level (Garrido *et al.*, 1993, for example).

1.3.4.3. Summary

The representation of local phenomena has been traditionally carried out by means of two types of systems: 'contours' and 'levels'. These two methods are in fact complementary: intonation contours can be first described in terms of contour representations, that can be analysed later in terms of levels. The approach used in this work can be considered as an example of mixed representation: first, a simplified contour is obtained by means of the stylization procedure, as described in chapter 3; second, the stylized contour is interpreted in terms of levels, using the 'transcription' system presented in chapter 4.

Several frameworks have also been proposed for the description of global phenomena. In this case, the topline-bottomline approach has been chosen, because it allows the description of global phenomena at different levels - intonation group, sentence, paragraph - using the same procedure.

1.3.5. Intonation models

The modelling of intonation is a task attempted for several languages - Danish, Dutch, English, French, German, Japanese or Swedish, among others - and with different theoretical approaches. Some of the models that have more extensively described in the literature are the following:

1) O'Shaughnessy & Allen's model (O'Shaughnessy, 1979), for American English.
2) Pierrehumbert *et al.*'s model (Pierrehumbert, 1980; Liberman & Pierrehumbert, 1984; Beckman & Pierrehumbert, 1986; Sosa, 1991), for American English, Japanese and Spanish.
3) The Lund model (Garding, 1979; Garding *et al.*, 1982; Garding, 1983; Garding *et. al.*, 1983; Lindau, 1986; Touati, 1987; Garding, 1989), for Swedish, Chinese, Greek, French and Hausa.
4) Thorsen's model (Thorsen, 1983), for Danish.

5) The Aix model (Hirst, 1983; Hirst & Di Cristo, 1984), for English and French.
6) Fujisaki's model (Fujisaki & Hirose, 1984; Fujisaki *et al.*, 1990; Fujisaki & Mixdorff, 1994; Fujisaki *et al.*, 1994; Fujisaki & Ohno, 1995), for Japanese, British English, Chinese, German and Spanish.
7) The IPO model (De Pijper, 1983; Willems *et al.*, 1988; t' Hart *et al.*, *1990*), for Dutch and British English.
8) Kohler's model (Kohler, 1991) for German.
9) Ladd's model (Ladd, 1983, 1992), for British English.
10) Gussenhoven & Rietveld's model (Gussenhoven & Rietveld, 1992), for Dutch.

As can be observed, only in two cases (Fujisaki's and Pierrehumbert *et al.*'s approaches) there has been an attempt to model Spanish intonation. A general review of the different types of intonation models is not offered here; see Bannert (1991) and Hirst (1991) for a general discussion of the goals and types of intonation models. Only some characteristics of these models are outlined here, and the IPO model, the one that has been followed in this work, is described in more detail.

1.3.5.1. 'Phonological' vs. 'phonetic' models

As discussed earlier (section 1.3.1.), there are phonetic and phonological approaches to the study of intonation. Accordingly, the resulting models of these approaches may be phonological or phonetic.

a) Phonological models take a phonological representation of intonation as starting point for the generation of the intonation contours. They use an inventory of phonological categories, with an associated phonetic representation, as basic units. This is the case of Pierrehumbert *et al.*'s, Ladd's, and Gussenhoven & Rietveld's models.

b) Phonetic models, however, are developed from phonetic data, and try to give a description of F0 movements, linking them to different linguistic functions. Examples of phonetic models are Thorsen's model, Fujisaki's model, O'Shaughnessy & Allen's model, and the IPO model.

1.3.5.2. Linear vs. hierarchical models

As discussed in 1.1.2, intonation may be considered a hierarchical or a purely sequential phenomenon. According to this, intonation models can be divided in two types, as Ladd (1988) suggests:

a) Most intonation models have an underlying hierarchical conception of the intonation phenomena: F0 contours are considered as the result of the superposition of patterns or movements of different scope (sentence, phrase, word, syllable, segment). The models of Fujisaki and Hirose, Thorsen, O'Shaughnessy and Allen, the Lund model, and the IPO model are included in this group.

b) On the other hand, some models consider F0 generation as a sequential process, that is achieved in only one stage from left to right of the utterance. F0 contours are the result of the concatenation of different F0 values assigned to each syllable. Ladd's, and Liberman & Pierrehumbert's approaches fall within this group.

1.3.5.3. A phonetic approach to intonation modelling: the IPO model

As has been stated before, this work follows quite closely the approach to intonation proposed by IPO for Dutch. The IPO model can be considered a good example of phonetic, contour-based and hierarchical approach to the modelling of intonation. The model is mainly presented in 't Hart *et al.* (1990), although 'intermediate' descriptions can be found in 't Hart & Cohen (1973), 't Hart & Collier (1975) or De Pijper (1983), among others.

Using the stylization and standardization procedures described above, a set of intonational patterns has been defined, and some rules predicting the use and combination of these patterns have been established. The patterns are organized in four different levels of description:

a) **Pitch movements**, or "relatively rapid changes in pitch that are typically associated with syllables" (De Pijper, 1983, p. 44).

These pitch movements are the smallest unit of description in the model. A total of 10 different pitch movements has been defined, that can be described by means of 4 different intonational distinctive features: direction of the movement, timing, rate of change, and size. The characterization of the movements in terms of distinctive features is offered in table 1.3.

transcription symbol									
	1	2	3	4	5	A	B	C	D	E
Direction										
rise	x	x	x	x	x					
fall						x	x	x	x	x
Timing										
early	x				x		x			x
late			x			x				
very late		x						x		
Rate of change										
fast	x	x	x		x	x	x	x		x
slow				x					x	
Size										
full	x	x	x	x		x	x	x	x	
half					x					x

Table 1.3. Features defining pitch movements in Dutch (from 't Hart *et al.*, 1990)

b) **Configurations**, or combinations of pitch movements.

The prototypical example of these pitch configurations is the 'hat-shaped' pattern, presented in figure 1.2.
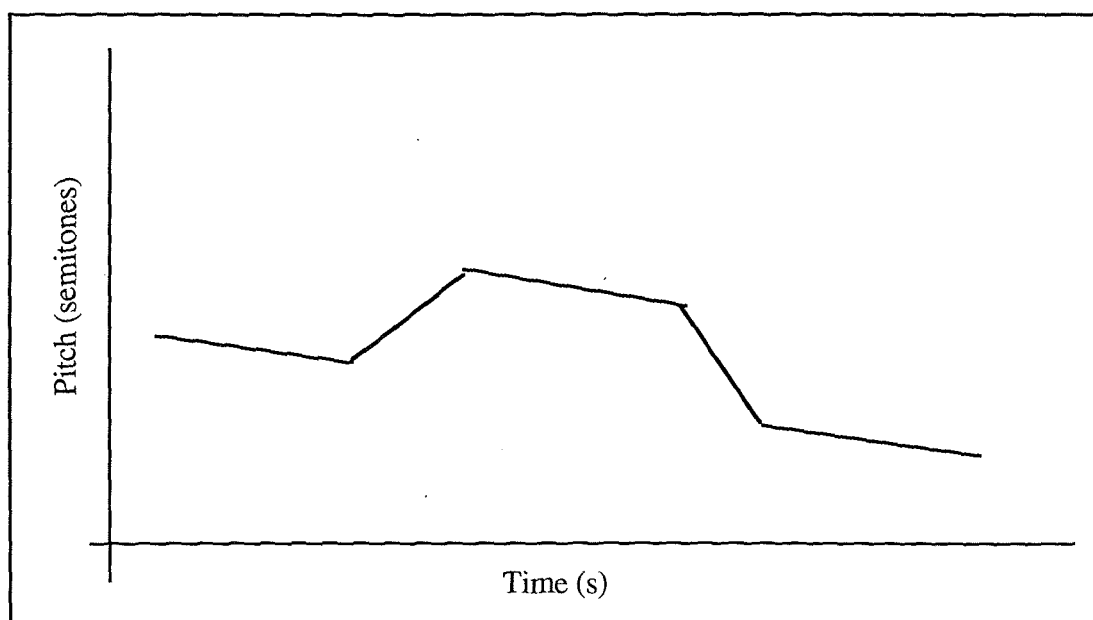
Figure 1.2. Elementary shape of the 'hat-shaped' pattern
(from 't Hart & Cohen, 1973)

Many possible configurations are just variations of this basic pattern.

The model assumes that pitch movements are superimposed, during the generation process, on a grid of three parallel declination lines, which define the global evolution of the pitch contours. These three declination lines define three different intonational 'levels', H, M and L, that determine the relative height associated to each of the 'target-points' of the pitch movement. An example of these lines with a set of pitch movements superimposed to them building a 'hat-shaped' configuration is given in figure 1.3.
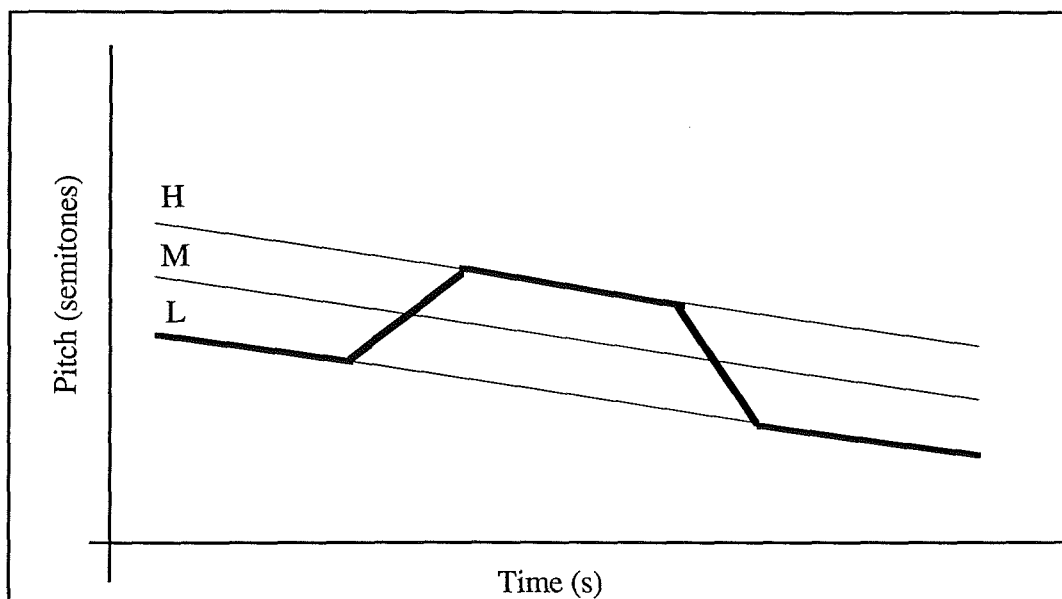
Figure 1.3. An example of 'hat-shaped' pattern superimposed to the grid of declination lines

There are three types of configurations: 'Prefix', 'Root' and 'Suffix' ('t Hart *et al.*, 1990, p. 79):

1) ROOT configurations: "they are obligatory and non-recursive; a contour must contain one such configuration, and not more than one".
2) PREFIX configurations: "they are optional and some of them are recursive; they always precede another Prefix or a Root".
3) SUFFIX configurations: "they are optional and non-recursive; they always follow a Root".

c) **Pitch contours**, or "strings of standardized pitch movements" ('t Hart *et al.*, 1990, p. 56).

Contours are then concatenations of configurations, and they are at least made up of a Root, and optionally one (or several) Prefix, and one Suffix. The allowed combinations of configurations that constitute possible contours in a given language are defined by a grammar.

d) **Intonation patterns**, or mental models that the speaker/listener holds in his mind.

It is important to note, however, that these categories are not linked to intonational functions, but only to mental categories established by the listener of a given language:

> "Neither is it our intention to lay bare the subdivision of speech melody into a limited number of intonation groups on the basis of any kind of linguistic categorization, such as statement vs. question, or of paralinguistic, for instance attitudinal, features. Rather, our experimentation is directed towards getting insight into the listener's internal representation of the intonational system of his native language. If in the experimental outcome the listeners give evidence of sufficient agreement on the categorization of melodic shapes, we shall have to incorporate it in our theoretical account. But in the absence of such evidence, we do not feel entitled to introduce a categorization on a priori grounds" ('t Hart *et al.*, 1990, p. 67).

The IPO approach gives then a complete description and modelling of the intonation of a language, but avoids the connection of the resulting patterns with the linguistic and paralinguistic uses of intonation.

1.3.5.4. Summary

Many intonation models have been proposed for different languages. Apart from other differences and similarities, these models differ in the approach to intonation underlying them (phonetic or phonological) and the conception of intonation they assume (hierarchical or linear). The model presented here should be included in the group of the phonetic models, and in the one of hierarchical models.

The model proposed here follows quite closely the IPO approach: it defines a series of movements that form the basic local patterns, that are combined to form the intonation contour of an intonation group. The global shape of the contour is also defined by top and bottom lines. There are, however, some differences between the two approaches: the main one is that the present approach tries to link intonational patterns to specific linguistic functions.

1.4. General summary: definition of the goal

The discussion of the topics presented in the previous sections allows a more precise definition of the goals and characteristics of this work.

1.4.1. Object of study: intonation contours

It was established at the beginning that the goal of this work is the modelling of Spanish intonation. It can now be specified that this research will be focusing on the modelling of F0 contours, called here 'intonation contours'. The proposed intonation model is then a description of the shape and organization of the intonation contours in Spanish. Accordingly, intonation in its classical sense is only partially analysed in this study.

1.4.2. Methodology

The review of the literature has also led to the conclusion that the descriptions of Spanish intonation contours available at this moment do not provide the necessary information for the development of a model of the type attempted here.

a) the available studies are not coherent from the point of view of the descriptive framework. The definition of a model needs the use of data obtained using the same descriptive framework.

b) the available descriptions are in many cases not accurate enough to be used in the development of a model oriented to technological applications. Formal descriptions of the intonation patterns are necessary to develop a model with these characteristics.

c) the available descriptions do not always provide a detailed analysis of the relationship between intonation patterns and intonation functions. More research is needed in order to obtain this information. Moreover, information relating intonation phenomena and some non-linguistic factors is also necessary.

For these reasons, the experimental analysis of a sample of speech has been attempted in this work. In the following sections, the characteristics of this experimental analysis are defined in more detail.

1.4.2.1. A phonetic approach to the description and modelling of intonation

It was also discussed earlier that the development of a model of intonation may be attempted from a phonetic or a phonological perspective. In this work, a phonetic approach has been chosen.

The phonetic approach defined here is quite similar to the IPO approach. It involves two main steps, one of stylization and a second one of standardization.

1.4.2.2. A level-by-level analysis of the Spanish intonation contours

The review attempted in the previous sections has shown that there is no agreement among the different descriptions in the definition of intonation units in Spanish. In addition, is is not clear that some of the proposed units coincide with specific intonational patterns. For this reason, these levels are only considered as starting points for the analysis, taking into account the fact they are 'traditional' units in the description of Spanish intonation. The analysis of their status as intonation units in Spanish is also one of the goals of this work.

Four different levels of analysis have been considered:

1) Tonic group
2) Intonation group
3) Sentence
4) Paragraph

The first three levels are present in previous descriptions of Spanish intonation. The tonic group has been defined here as Navarro's or Fant's tonic group, and as López's 'stress group', that is, a group of words containing a stressed word and all its preceding non-stressed words. As far as intonation group is concerned, it is important to note that it has been identified here, for the shake of simplicity, with the phonic groups, or portions of speech between two pauses. Sentences have been defined here following an ortographic criterium, as the portion of utterance corresponding to the text beginning with a blank space, a full stop (.), a question mark (¿) or an exclamation mark (¡), and ending with a full stop, a question mark (?) or an exclamation mark (!).

The paragraph domain has also been considered for this study, taking into account the results of some recent studies, as Garrido *et al.* (1993), that have shown that it may be also a unit of intonation analysis in Spanish. The paragraph has been defined here as a portion of text, including one or more sentences, organized in such a way that its first sentence always starts a new line of the text, and the end of the last sentence also determines the end of the line.

1.4.2.3. A function-oriented analysis of Spanish intonation

It has been established earlier that an intonation model must reflect the structure of F0 contours and describe the patterns that are used to build

them. But it also has to establish a relationship between these patterns and the variables that determine their use.

It is out of the scope of this work to make an exhaustive analysis of all the variables that may influence the shape of F0 contours. A selection of variables has been made, considering their importance in the communication process, and in the building of contours. Linguistic factors seem the most important ones, specially in a model oriented to TTS conversion. The linguistic variables considered in this work are the following:

a) Lexical stress
b) Sentence type
c) Syntactic structure

Among physiological and phonetic variables, only utterance length has been considered. Micromelodic variations have been excluded. All the emotional, sociolinguistic and stylistic variables have also been excluded from this analysis, and they have been controlled in the definition of the corpus as is stated later in chapter 2.

1.4.3. Procedure

The experimental part has been divided in different tasks, which are presented schematically in figure 1.4:

```
┌─────────────────────────────────────────────────────────────┐
│                   ┌──────────────────────────┐               │
│                   │     Corpus definition     │               │
│                   └──────────────────────────┘               │
│                              │                                │
│                              ▼                                │
│                   ┌──────────────────────────┐               │
│                   │       F0 estimation       │               │
│                   └──────────────────────────┘               │
│                              │   Intonation contours          │
│                              ▼                                │
│                   ┌──────────────────────────┐               │
│                   │        Stylization        │               │
│                   └──────────────────────────┘               │
│                              │   'Perceptually equivalent contours'│
│                              ▼                                │
│                   ┌──────────────────────────┐               │
│                   │      'Post-processing'     │               │
│                   └──────────────────────────┘               │
│                              │                                │
│                              ▼                                │
│         ┌──────────────────────────────────────────┐         │
│         │       Intonation analysis                 │         │
│         │       (standardization)                   │         │
│         │                                           │         │
│         │       • descriptive framework             │         │
│         │       • definition of patterns            │         │
│         │       • analysis patterns-functions       │         │
│         │   ┌───────────────────────────────────┐   │         │
│         │   │      Tonic group level             │   │         │
│         │   └───────────────────────────────────┘   │         │
│         │   ┌───────────────────────────────────┐   │         │
│         │   │    Intonation group level          │   │         │
│         │   └───────────────────────────────────┘   │         │
│         │   ┌───────────────────────────────────┐   │         │
│         │   │       Sentence level               │   │         │
│         │   └───────────────────────────────────┘   │         │
│         │   ┌───────────────────────────────────┐   │         │
│         │   │       Paragraph level              │   │         │
│         │   └───────────────────────────────────┘   │         │
│         └──────────────────────────────────────────┘         │
│                              │   Intonation patterns          │
│                              ▼                                │
│                   ┌──────────────────────────┐               │
│                   │    Intonation Modelling   │               │
│                   └──────────────────────────┘               │
│                              │   Grammar of movements         │
│                              ▼                                │
│                              Intonation rules                 │
└─────────────────────────────────────────────────────────────┘
```

Figure 1.4. Schematic representation of the different steps of the experimental analysis

Each one of these parts is described in the following chapters:

a) Definition and collection of the speech material (chapter 2)
b) Stylization and post-processing (chapter 3)
c) Level-by-level standardization process:

    1) Tonic group level (chapter 4)
    2) Intonation group level (chapter 5)
    3) Sentence level (chapter 6)
    4) Paragraph level (chapter 7)

d) Modelling (chapter 8)

Chapter 2

SPEECH CORPUS

This chapter describes the process of definition, collection and classification of the speech corpus used for the experimental part of this work. First, the characteristics of the speech corpus are described. The details concerning the collection of the material are given in section 2.2. Finally, the process of classification of the collected material is presented in section 2.3.

2.1. Defining the characteristics of the speech corpus

As defined in chapter 1, this work describes Spanish intonation at four different levels: tonic group, intonation group, sentence and paragraph. The first task was then to define the characteristics of the speech corpus necessary to carry out this four-level description. The definition of the corpus will involve then the setting of the variables that are not going to be considered in the analysis (and that will be controlled in the corpus), and the determination of the variables that will be the focus of the analysis (and that consequently will be systematically modified in the corpus). Llisterri (1991, p. 79 and ff.), following partially Murillo (1981), establishes four set of variables that can be considered in the definition of any phonetic corpus:

a) Elocution variables: those affecting the language, the speaker, the recording order and the speaking rate, among others.
b) Phonetic variables: those related to phonetic factors (location of the target items within the utterance, position in the syllable, lexical stress...)
c) Linguistic variables: those having to do with the linguistic characteristics of the corpus (utterance structure, linguistic function of the elements...)

d) Stylistic variations: those considering differences in speaking style within the corpus.

In the next sections, the different variables that can be taken into account for the definition of the material are reviewed following this classification.

2.1.1. Stylistic variation: the use of text readings

As has been discussed in chapter 1, speaking style is one of the factors which can determine the shape of intonation contours. Accordingly, the style of the speech corpus to be used for the definition of an intonational model has to be controlled in some way in order to obtain a coherent description.

A first problem in the selection of style(s) to be considered lies in the definition of styles itself. Since the classical classification of Labov (1972), who arranged different styles "along a single dimension, measured by the amount of attention paid to speech" (Labov, 1972, p. 208), the research on speaking styles definition has pursued, - see for example the contributions of Gregory & Carroll (1978), Eskénazi (1993), Aguilar, Llisterri & Machuca (1994), or Aguilar & Machuca (1996) -, but for the moment a complete inventory of speaking styles, and of the features defining them, is far from being achieved.

As Eskénazi (1993, p. 502) states, "style reflects the action of the environment upon the individual and the individual upon the environment". In other words, style reflects the influence on the speaker's utterances of the **communicative situation** in which he/she is speaking.

The process of recording a corpus for experimental analysis can be viewed as a special type of communicative situation, which defines a specific type of speaking style. In the definition of a corpus for intonation modelling, the definition of the style to be analysed is then closely related to the selection of the adequate communicative situation in which the speech will be obtained. Following Aguilar, Llisterri & Machuca (1994) and Aguilar & Machuca (1996), the communicative situations related to the collection of material for speech research can be defined according to several factors:

a) the nature of the corpus itself
b) the speaker and listener's attitude
c) the context and the channel

In the following sections, these three sets of factors are considered in more detail.

2.1.1.1. The nature of the corpus

According to the nature of the corpus to be recorded, the obtained speech can be classified into 'scripted', partially 'unscripted' and 'unscripted'. 'Scripted' speech involves some kind of preplanning of the speech material, while the term 'unscripted' refers to non-preplanned speech. However, as Aguilar, Llisterri & Machuca (1994) point out, this preplanning can be related to both the speaker and the researcher. Two different types of preplanning can be then found in the obtained speech, namely, **speaker's preplanning** and **researcher's preplanning**. These two factors constitute two different dimensions which can define different types of speech. These types of speech are considered in the next sections.

2.1.1.1.1. Speaker's preplanning: 'spontaneous' vs. 'read' speech

As Blaauw (1995) has pointed out, the **level of preparedness** is the main difference between 'read' and 'spontaneous' speech. Read speech can be largely preplanned before production, while spontaneous speech is not usually prepared. She also suggests (p. 8) that "Perhaps spontaneous and read speech can be defined as the clusters of speech styles at the extreme ends of a continuum, which is crucially defined by the extent to which the speech has been prepared before".

As discussed in chapter 1, preplanning is a crucial factor in the organization of the intonational structure of a utterance. The selection of one type of speech or another is then important in a corpus of intonation.

For this study, **text reading** has been the selected type of speech to be analysed. The main reason for this choice has been that the description of Spanish intonation presented here aims at developing a model that could be used in TTS systems. The main task of TTS systems is to read aloud written text, so it seems more adequate to describe intonation of text readings. Also, the use of text instead of oral realizations ensured a better control of variables like syntactic structure.

Within the reading style, several types can also be established. A second decision that had to be taken was the type of texts to be used. The traditional descriptions of Spanish intonation are mainly based on literary readings, as already discussed. However, a more neutral type of text would be preferable for a study like the one presented here. **News readings** was then the selected type, taking into account two main reasons: first, news readings are less affected by expressive variations, and expressive information was one of the factors to be controlled in the collected material;

second, news readings is one of the most common applications of TTS systems.

2.1.1.1.2. Researcher's preplanning: 'controlled' vs. 'uncontrolled' speech

The contents of a corpus can also be more or less prepared in advance by the researcher. Different degrees of researcher's preplanning can be also defined within read text: from the (almost) complete preparedness of an 'ad hoc' corpus of sentences, where almost any word has been chosen by the experimenter, to less preparedness in the case of 'real' text chosen from books or newspapers. The use of a corpus preplanned by the researcher offers the advantage that the different variables can be controlled in a easier way, but it has the disadvantage that this method can lead to artificial results. In the case of 'uncontrolled' speech, however, it is necessary to collect a large amount of material to obtain samples of the target phenomena. The main advantage of this type of material is that it allows to obtain data about speech in real communicative situations. These considerations apply also for the definition of a corpus oriented to the development of an intonation model.

The analysis of 'uncontrolled' speech instead of 'controlled' material has been preferred here. Several factors have been taken into consideration for this decision: first, the description of 'uncontrolled' speech seems much more adequate than 'controlled' material for the development of a model specially oriented to TTS applications; second, the characteristics of the description, including large-domain units as paragraphs, seem to favour the use of this type of speech.

However, the use of completely raw and uncontrolled material does not allow any systematic analysis of intonational phenomena. For this reason, the collected material has been classified and filtered, in order to obtain a set of representative samples of utterances to be analysed. After this classification process, four different sub-corpora were defined, one for each of the established analysis levels: tonic groups, intonation groups, sentences and paragraphs. Each of these sub-corpora is made up of a set of representative units extracted from the spontaneous material recordings. These sub-corpora were used for the analyses presented in the following chapters.

By this method, then, some type of **half-controlled** material has been obtained, combining the characteristics of both types of material: the naturalness of the 'uncontrolled' corpora, and the control of the variables that the 'ad hoc' material offers, necessary to obtain some reliable conclusions.

2.1.1.2. The speaker and the listener

Aguilar, Llisterri & Machuca (1994) consider several factors involving the speaker and the listener in the definition of communicative situations:

a) the presence or absence of the listener in the recording situation
b) the identification of the listener with the researcher in some special collecting procedures (such as interviews, for example)
c) the use of professional or non-professional speakers
d) the degree of familiarity of the speaker with the listener or researcher.

Some of these factors are discussed in the following sections.

2.1.1.2.1. The listener

Aguilar, Llisterri & Machuca (1994) consider three possibilities in a recording situation regarding the presence or absence of the listener

a) There is no specific listener during the recording. This is the case of read-aloud texts in a recording room, for example.

b) There is one listener present, which can even participate as a speaker in the communicative situation. This is the typical case of an interview, or a more or less casual conversation.

c) There are several listeners. They can be physically present, and they can participate in some occasions, as in the case of a lecture, or they can be remote, as in the case of radio and TV broadcasts.

The presence of a large audience determines the use of some resources by the speaker to maintain the attention of the listener. These resources are usually related to prosody. Then, the presence or absence of listeners seems to be an important factor affecting the intonation of the resulting speech:

a) if listeners are present, the speaker has to try to make his speech intelligible. In the case of text readings, this means that the speaker has to make an effort to read in a correct way, placing pauses in the right position and with a 'comprehensive' intonation.

b) if there are listeners, the speaker has also to maintain their attention, and intonation variation is one of the resources which can be used to achieve this goal. Lowering and rising intonation is a way of avoiding a monotonous speech. This is the case of lecturers, for example, or the radio and TV speakers.

It was decided that the speech used as corpus for this analysis had to be read 'as if' one or several listener were (physically or remotely) present. Speech produced by TTS systems has to sound as read to an actual audience, and accordingly the speech used to generate intonation models has to meet also this condition. This does not mean, however, that listeners have to be really present during the recording of the speech. The speaker can be encouraged to read 'as if' they were present.

2.1.1.2.2. The speaker

As inferred from the previous section, an important feature of the speaker involved in the recording of an intonation corpus is his/her skill in producing the target material. This skill can be specified in the following characteristics:

a) the ability to read texts in a 'correct' way; in this case, 'correct' has to interpreted as 'understandable', that is, that reflects the syntactic structure and semantic contents of the text.
b) the ability to read the texts in a natural and non monotonous way.

Professional speakers, from radio and TV, seem good candidates to fulfill these conditions. It was decided then to include recordings of this type of speakers in the collected speech corpus. However, their way of reading may be considered too 'marked' by some listeners. For this reason, the use of non-professional speakers was also considered.

Thus, the collected speech corpus contains readings from professional and non-professional speakers. However, preliminary analyses performed with samples extracted from both sets of material showed differences between the reading style of both types of speakers (Garrido, 1993). For this reason, to preserve the homegeneity of the description, the study presented here has been carried using exclusively the material extracted from the 'non-professional' set.

2.1.1.3. Context and situation: 'lab' material vs. 'natural' material

The place at which a corpus is recorded is another important factor that can affect the style of the obtained material. 'Lab' recordings tend to be less natural than those recordings obtained in 'real world' environments. However, recording conditions are usually better in a treated recording room than in a classroom, for example, and the experimental analysis of intonation usually needs recordings free of undesirable background noise. For this reason, it was decided to pay special attention to recording

conditions in the material collected for the corpus. In the case of professional speakers, the recordings were obtained from the source radio station when possible, as is described bellow. In the case of the non-professional recordings, they were made in a sound-treated room with a high-quality material.

2.1.2. Elocution variables: speakers

Within the elocution variables, the characteristics of the speaker are an important source of variation in the intonation contours. Accordingly, special attention has to be paid to the selection of the speakers in the definition of a corpus for the analysis of intonation.

The first question is the number of speakers to be used. Intonation models for TTS systems are usually based in the description of a single speaker, as is in López (1993). In this case, the use of samples of speech coming from different speakers has been preferred, in order to obtain a more general description.

A final aspect concerns the characteristics of the speakers to be selected. These characteristics can be classified in two main groups, according to Llisterri (1991, pp. 107-113):

   a) intrinsic characteristics: age, sex.
   b) extrinsic characteristics: 'ethnicity' or nationality, mother tongue, social class.

In order to obtain an homogeneous sample of speech, all these variables, except sex, have been controlled in the selection of the 'non-professional' set of speakers. In the 'professional' set, the nature of the material (recordings from radio and TV) did not allow the application of this type of control.

2.1.3. Phonetic variables

As was outlined in chapter 1, several phonetic factors can determine the shape of intonation contours. The segmental content of the utterances, for example, introduce small modifications of in the contours, that can be interesting to take into account if the micromelodic variations are going to be analysed. If not, these variations can make difficult the analysis of intonation contours. For this reason, many studies control the segmental content of the utterances. In Garrido (1991), or Garrido et al. (1993), for example, only voiced segments were used to build the target sentences, in order to obtain continuous contours, not interrupted by unvoiced segments. This procedure, however, can give unnatural results in the target sentences,

and it is difficult to apply when analyzing larger domain units, as complete paragraphs. In this case, the use of 'uncontrolled' speech did not allow the control of this type of variable. However, the use of the stylization procedure described in the following chapter allowed to partially avoid the problem of micromelodic influence.

As established in chapter 1, the only phonetic factor considered in this study has been the length of the units. This variable has been taken into account in the collection and classification of the corpus material, and also in the definition of the different sub-corpora, as described in more detail in the corresponding chapters.

2.1.4. Linguistic variables

The linguistic factors taken into account for this study were defined in chapter 1. These linguistic variables are lexical stress, sentence type and syntactic structure. They have been taken into account during the collection of the corpus: for example, the use of both news and commentaries has been considered to obtain examples of utterances with different sentence types: sentences in news readings are mostly declarative, and it was also necessary to obtain examples of interrogative and exclamative sentences. These linguistic variables have been considered as well during the classification process, and also in the preparation of the sub-corpora.

2.1.5. Summary

According to these considerations, a set of real-text readings was defined for this work. A summary of the contents and organization of the material is presented in table 2.1.