TRAINING INTONATIONAL PHRASING RULES AUTOMATICALLY FOR ENGLISH AND SPANISH TEXT-TO-SPEECH

Julia Hirschberg

AT&T Bell Laboratories

Pilar Prieto

Universitat Autónoma de Barcelona

Author of record:

Julia Hirschberg

2C-409

AT&T Bell Laboratories

600 Mountain Avenue

Murray Hill NJ

USA, 07974–0636

julia@research.att.com

December 21, 1995

Special Symbols: @, ||

Number of Pages: 24

Number of Tables: 3

Number of Figures: 1

Keywords: intonation, prosodic phrasing

Abstract

We describe a procedure for acquiring intonational phrasing rules for text-to-speech synthesis automatically, from annotated text, and some evaluation of this procedure for English and Spanish. The procedure employs decision trees generated automatically, using Classification and Regression Tree techniques, from text corpora which have been hand-labeled by native speakers with likely locations of intonational boundaries, in conjunction with information available about the text via simple text analysis techniques.

Rules generated by this method have been implemented in the English version of the Bell Laboratories Text-to-Speech System and have been developed for the Mexican Spanish version of that system. These rules currently achieve better than 95% accuracy for English and better than 94% for Spanish.

1. Intonational Phrasing

Assigning appropriate phrasing in text-to-speech systems is important both for naturalness and for intelligibility, particularly in longer sentences and longer texts (Silverman et al., 1993). This paper describes the automatic acquisition of methods of assigning such boundaries in real-time unrestricted text-to-speech.

Intuitively, intonational phrases divide utterances into meaningful 'chunks' of information (Bolinger, 1989). Variation in phrasing can change the meaning hearers assign to utterances of a given sentence. For example, the interpretation of a sentence like '*Bill doesn't drink because he's unhappy*' will vary, depending upon whether it is uttered as one phrase or two. Uttered as a single phrase, this sentence is commonly interpreted as conveying that Bill **does** drink — but the cause of his drinking is **not** his unhappiness. Uttered as two phrases, with an intonational boundary between '*drink*' and '*because*', it is more likely to convey that Bill does **not** drink — and that the reason for his abstinence is his unhappiness.[1]

To characterize this phenomenon phonologically, we adopt Pierrehumbert's intonational description for English (Pierrehumbert, 1980; Beckman and Pierrehumbert, 1986). In this theory, there are two levels of phrasing. An INTERMEDIATE PHRASE consists of one or more PITCH ACCENTS (local minima or maxima in the FUNDAMENTAL FREQUENCY (F0) contour which may occur in six distinct forms) plus a PHRASE ACCENT (a simple high or low tone which controls the pitch from the last pitch accent of one intermediate phrase to the beginning of the next intermediate phrase or the end of the utterance). INTONATIONAL PHRASES consist of one or more intermediate phrases plus a final BOUNDARY TONE, which may also be high or low, and which occurs at the end of the phrase. Thus, an intonational phrase boundary necessarily coincides with an intermediate phrase boundary, but not vice versa. We employ Pierrehumbert's system also for the analysis of our Spanish corpus, with modifications to the inventory of pitch accent types that do not affect the description of phrasing levels.

While we assume phrase boundaries to be perceptual categories, these have been found to be associated with certain physical characteristics of the speech signal. In addition to the tonal features described above, phrases may be identified by one of more of the following features: pauses (which may be filled or not), changes in amplitude, and lengthening of the final syllable in the phrase sometimes accompanied by glottalization of that syllable and perhaps preceding syllables. In general, major phrase boundaries tend to be associated with longer pauses, greater tonal changes, and more final lengthening than minor boundaries. These generalizations appear to hold for both English and Spanish. In the Bell Laboratories Text-to-Speech system (TTS), intonational boundaries are realized by the manipulation of all of these features. However, currently, only one level of phrase boundary, the intonational phrase, is modelled in TTS, so we have only tried to predict this single level in the work described below.

2. Phrasing Prediction for Text-to-Speech

Most text-to-speech systems that handle unrestricted text rely upon simple phrasing algorithms based upon orthographic indicators, keyword or part-of-speech spotting, and simple timing information to assign phrase boundaries (O'Shaughnessy, 1989; Larreur, Emerard, and Marty, 1989; Schnabel and Roth, 1990). More sophisticated rule-based systems have so far been implemented primarily for message-to-speech systems, where syntactic and semantic information is available during the generation process (Young and Fallside, 1979; Danlos, LaPorte, and Emerard, 1986). However, some general proposals have been made which assume the availability of more sophisticated syntactic and semantic information to use in boundary prediction (Altenberg, 1987; Bachenko and Fitzpatrick, 1990; Monaghan, 1991; Quené and Kager, 1992; Bruce et al., 1993), although no current proposal integrating such information into the phrase assignment process has been shown to work well, even from hand-corrected labeled input. Even if accurate syntactic and semantic information could be obtained automatically and in real time for text-to-speech, such hand-crafted rules systems are notoriously difficult to build and to maintain.

Recently, efforts have been made to acquire phrasing rules for text-to-speech automatically, by training self-organizing procedures on large prosodically labeled corpora (Wang and Hirschberg, 1991a; Wang and Hirschberg, 1991b; Hirschberg, 1991; Wang and Hirschberg, 1992; Ostendorf and Veilleux, 1994). Such methods were used to train a phrasing module for the Bell Laboratories Text-to-Speech system from labeled speech from the DARPA ATIS corpus (Wang and Hirschberg, 1991a; Wang and Hirschberg, 1991b; Hirschberg, 1991; Wang and Hirschberg, 1992); this module predicted intonational phrase boundaries correctly in just over 90% of cases, where data points were defined at the end of every orthographic word.

For this work, Classification and Regression Tree (CART) analysis (Breiman et al., 1984) was used to construct decision trees automatically from sets of continuous and discrete variables. For the phrasing module, these sets included values for all variables which appeared potentially useful in predicting phrasing decisions, and which could be acquired automatically from text analysis in real time.

To produce a decision tree, CART accepts as input a vector of all such independent variable values plus a dependent variable for each data point, and generates a decision tree for the dependent variable. At each node in the generated tree, CART selects the variable which best minimizes prediction error for the remaining unclassified data. In the implementation of CART used in this study (Riley, 1989), all of these decisions are binary, based upon consideration of each possible binary split of values of categorical variables and consideration of different cut-points for values of continuous variables. CART's cross-validated estimates of the generalizability of the trees it produces have proven quite accurate for the current task, when compared with tests on separate data sets; in every case CART predictions for a given prediction tree and that tree's performance on a hand-separated test set fall within a 95% confidence interval.[2]

This procedure performed fairly well, with results reported in Wang and Hirschberg 1992 of a CART cross-validated success rate of 90% correct classification of intonational phrase boundaries for trees grown using only information available automatically and in real time from text analysis. However, the hand-labeling required for the training data (originally, labeling of phrase boundaries and pitch accents was employed, so both of these features had to be identified in the speech corpus) is enormously time-consuming and expensive, requiring well over one person-year to accomplish, for the phrasing procedure described in this section. But automatic labeling of prosodic features does not appear to be reliable enough yet to serve as a substitute, despite some progress made in this area in recent years (Ostendorf et al., 1990; Wightman and Ostendorf, 1994).

2.1. Training Phrasing Procedures on Annotated Text

The current English version of Bell Labs TTS contains a phrasing module which was produced automatically, using procedures similar to those used in (Wang and Hirschberg, 1991a; Wang and Hirschberg, 1991b; Hirschberg, 1991; Wang and Hirschberg, 1992) to train phrasing procedures on hand-labeled speech. However, the prediction tree in this module was itself trained not on prosodically labeled speech but upon a hand-annotated corpus of approximately 87,000 words of text taken from the AP newswire, and labeled for likely prosodic boundaries by one of the authors, a native speaker of standard American English.

The use of such text training data cut the time needed to train a new phrasing module to just two or three days, by eliminating the tedious hand-labeling of prosodic features of recorded speech. While the labeling of the speech corpus used to construct the earlier phrasing module for TTS took a full person year, a better comparison of the amount of time saved by the text-based procedure can be found by comparing more recent labeling of phrase boundaries only in spontaneous and read speech by a similarly trained labeler, using the ToBI system (Pitrelli, Beckman, and Hirschberg, 1994). This labeler averaged 67 words per hour for a 3163-word corpus of recorded speech, compared to 4833 words per hour for the annotator assigning plausible boundaries to text. And note also that the speech labeler's labor does not include the time required to record and process the speech for labeling, which is of course not necessary for the text-based approach.[3] So, the work required to produce a new phrasing module for a TTS system was speeded up by a factor of 70 or better, using the text-based approach. Identifying the prosodic performance of others is considerably more difficult than assigning likely prosody oneself, based on one's own hypothetical performance.

With the text-based method, it is thus possible to retrain the existing TTS phrasing procedure quickly, as deficiencies are uncovered, by the simple addition of exemplars of the (corrected) behavior to the training set. It is also possible to produce phrasing procedures easily for new domains or languages without recording or labeling a large corpus. The Spanish phrasing procedure recently developed is a demonstration of this technique's versatility: a baseline version of this model which performed at about 90% correct was produced in only about a one and one-half person weeks.

To produce this phrasing procedure, or a phrasing procedure for a new application, domain, or language, we proceed as follows: On-line text from an appropriate domain in the language desired is first annotated by a native speaker of that language with plausible intonational boundaries, by identifying locations in the text where the annotator believes boundaries sound 'natural'. We are currently using newswire text from the English and Spanish AP for general TTS training purposes, but other text could be used for particular applications. The unannotated version of the text is itself analyzed to extract values for features at each potential boundary site (defined as each position between two orthographic words $< w_i, w_j >$ in the input) which have been shown or appear likely to correlate with phrase boundary location — and which can be extracted automatically and in real time. For English, these features include:

- a part-of-speech window of four around the site, $< w_{i-1}, w_i, w_j, w_{j+1} >$;
- whether w_i and w_j bear a pitch accent or not;
- the total number of words in the utterance;
- the distance in words from the beginning and end of the utterance to $< w_i, w_j >$;
- the distance in syllables and in stressed syllables of $< w_i, w_j >$ from the beginning of the utterance;
- the total number of syllables in the utterance;
- whether the last syllable in w_i is phonologically strong or weak;
- the distance in words from the previous internal punctuation to w_i;
- the identity of any punctuation occurring at the boundary site;
- whether $< w_i, w_j >$ occurs within or adjacent to an NP;
- if $< w_i, w_j >$ occurs within an NP, the size of that NP in words, and the distance of $< w_i, w_j >$ from the start of the NP

For Spanish, the feature set currently includes only the following features: part-of-speech window of two places to the left and right of $< w_i, w_j >$ (*pos1*$\{a,b\}$...*pos4*$\{a,b\}$), whether or not w_i and w_j are accented (*la* and *ra*), the total number of words in the sentence (*tw*), the distance of the potential site from the beginning (*sw*) and end (*ew*) of the sentence in words and from the beginning of the sentence in syllables (*syls*), the identity of any punctuation occurring at $< w_i, w_j >$ (*punc*), the distance of $< w_i, w_j >$ from the last preceding punctuation mark in the sentence (*dpunc*), and the likelihood of vowel elision occurring across $< w_i, w_j >$ (*wb*). Part-of-speech and accent information was obtained from the front-end preprocessor of AMIGO, the Castilian Spanish text-to-speech system produced by AT&T and Telefónica de España (Rodríguez et al., 1993). Vowel elision likelihood was inferred from AMIGO's dictionary entries. These features are listed in Table 1.

Vectors of independent feature values plus the dependent (observed) value — the annotator's decision as to the plausibility of an intonational boundary between w_i and w_j — are then input to an implementation of CART (Riley, 1989). Note that the features described above represent only a subset of the features originally proposed to the automatic procedure; features which are not useful in prediction at earlier stages of development are simply ignored by CART, and can be omitted from the final tree so that those feature values will not have to be obtained via text analysis in TTS. New features can be proposed to CART as readily as the requisite information can be obtained from the text. Features tested for English which proved not to improve performance over those noted above, include: mutual information scores for words close to $< w_i, w_j >$ (testing the hypothesis that word pairs exhibiting high mutual information scores might tend to occur in the same phrase) and structural syntactic information about constituents bordering on $< w_i, w_j >$ and immediately dominating that site (testing the hypotheses that prosodic phrase boundaries do occur at constituent boundaries and do **not** occur within constituents).

For the phrasing module currently implemented in English TTS, a new matrix of feature vectors can be generated for new text simply by running TTS in training mode, so that it prints the inferred values for independent variables and the annotated values for the dependent variable. The resulting tree is then compiled automatically into *c* code, which can be used for simple prediction in a stand-alone procedure, or which can be substituted for an existing decision tree module in the TTS phrasing module.

3. Analysis and Evaluation of the Spanish Results

Decision trees produced for English using annotated text for training perform somewhat better than trees trained on prosodically labeled speech, probably due to the increased size of the training set. The best result for English annotated text is a cross-validated score of 95.4% correct predictions on an 89,103 word training corpus, compared to around 90% cross-validated accuracy for the best trees trained on a 12,714 word corpus of labeled speech. For Spanish, the best cross-validated success rate is 94.2% correct predictions of phrase boundaries for a 19,473 word (text) corpus. (In all cases, percentages were based on CART cross-validation estimates.) Here, we will discuss results for the Spanish case only. A decision tree for Spanish which is one of those averaged to estimate a 94.2% success rate is presented in Figure 1.[4] Table 1 provides a key to the features referenced in the tree, and Table 2 provides a list of the parts-of-speech labels employed here.

Figure 1 goes here.

Table 1 goes here.

In Figure 1, nodes are labeled with the majority decision at that node, which is '*no*' for 'no boundary predicted at this site' and '*yes*' for 'boundary predicted at this site'; for example, the root node is labeled '*no*' because 16,289 sites of the 19,473 in the corpus are correctly predicted **not** to contain phrase boundaries. The fraction of sites correctly identified by this label appears below each node (16,289/19,473 for the root node). The arcs linking nodes are labeled with the variable on which the next split is made and with the values for that variable (recall that only binary splits are allowed in CART) chosen to minimize prediction error. So, the first split in the tree in Figure 1 is on the variable *dpunc*, or 'distance from prior punctuation (if any)', with the value for the (continuous) variable *dpunc* which has been chosen by CART to minimize prediction error being 'greater or less than .5 words distance'. That is, boundary sites less than .5 words (i.e., next to) orthographic punctuation tend overwhelmingly (1900/1936) to be correctly predicted as containing phrase boundaries, while those further from punctuation (greater than .5 words) tend to be correctly predicted as **not** containing phrase boundaries (16253/17537). The left daughter of the root is thus labeled '*yes*' and the right daughter is labeled '*no*'.

As we have just seen, distance from punctuation is the single best predictor of phrase boundaries in our corpus, with boundaries occurring reliably at such punctuation. In addition, later splits in the tree on this same variable indicate that phrase boundaries do **not** tend to occur in close proximity to earlier punctuation; that is, since boundaries **do** tend to occur at punctuation, potential sites that are close to such boundaries tend **not** to contain boundaries themselves. Note also that the overwhelming majority of subsequent decisions in the tree, after the initial split on *dpunc*, are based upon part-of-speech, although considered at different positions within the window of four items around the potential boundary (in the figure, *pos1a,b* denotes the first position in that window — two words to the left of the potential boundary site, *pos2a,b* indicates the second position — immediately to the left of the potential boundary site, and so on). Whether or not the word to the left of the site is predicted to be accented (*la*) or not is also a useful variable, successfully predicting the absence of a phrase boundary after deaccented words in 7753 of 7758 cases. Finally, distance of the site from the beginning of the sentence (*sw*) occurs at the bottom of the tree, with sites occurring later in longer sentences more likely to contain phrase boundaries than sites earlier or in shorter sentences.

When we use our prediction tree to predict the full training corpus, in order to understand the main sources of error in phrasing predictions in the Spanish case, we actually mispredict only 1072 data points, which is somewhat less than the cross-validated error estimate.[5] We first note that a number of the predictions which were erroneous in terms of the training data nonetheless seemed felicitous, while others indeed represent phrasing choices that seemed infelicitous. For example, the predictions of the automatic procedure actually sound more felicitous in some cases than the original hand annotations, as in '*Asimismo, || once hindúes fueron mientras heridos a tiros por agentes policiales @ mientras sus correligionarios || lanzaban bombas incendiarias de fabricación casera || y trapos encendidos @ || contra las improvisadas viviendas de los musulmanes, || agregó.*'[6] However, other automatic predictions are quite infelicitous, as in '*Abogados defensores han expresado su confianza || que muchos de los detenidos podían ser absueltos por el consejo @ || pero también su preocupación por cómo se ha realizado todo el juicio.*'

So, we classified the CART 'errors' into felicitous and infelicitous mispredictions. Of the 1072 mispredictions, 939 turned out to be in fact felicitous, 32 resulted from human error in the original text annotation, and only 101 mispredictions actually produced infelicitous phrasings. Thus, the initial 5.8% error rate in the Spanish case could probably be further reduced if we count only the decisions producing infelicitous readings.

We then scrutinized these 101 infelicitous incorrect predictions to identify common sources of error in our automatic procedure. Table 3 summarizes the sources of infelicitous misprediction and whether they led to the prediction of an infelicitous phrasing boundary (Boundary) or failed to predict a felicitous boundary (NoBoundary).

Table 3 goes here.

Given the importance of part-of-speech information in the tree shown in Figure 1, we wondered whether error in part-of-speech tagging had led to some of the misprediction. However, this did **not** in fact appear to be the case. Overall, we found only two cases in which it was clear that part-of-speech classification error had led to incorrect phrasing prediction. The majority of infelicitous prediction errors derived from the inability of the feature set we used to capture higher-level constituency information, including questions of PP and relative clause attachment, types of coordination, apposition, and 'marked' cases of constituency ordering.

Ambiguous attachment in Spanish, as in English, can be disambiguated through intonational phrasing in speech (Avesani, Hirschberg, and Prieto, 1995) and by punctuation in text. The absence of disambiguating punctuation in our texts, when the attachment itself was semantically if not syntactically unambiguous, led to over half the infelicitous mispredictions in our corpus. For example, in '*Irak || está desplazando baterías de misiles antiaéreos || al sur y al norte de las zonas de no vuelos impuestas por los aliados occidentales @ || en un esfuerzo por confundir || y provocar a sus enemigos,|| dijeron funcionarios norteamericanos.*', the PP '*en un esfuerzo por confundir || y provocar a sus enemigos*' is attached at the S level, and thus is more felicitously set off as a separate prosodic phrase; if it were attached to the NP it would be more likely not to be so set off.

Our lack of real constituency information also led to mispredictions about phrasing decisions between conjuncts and disjuncts in some cases. In Spanish, as in English, we found, for example, that sentence conjunction is more likely to be accompanied by an intonational phrase boundary before the conjunction or disjunction than is coordination involving NP conjuncts or disjuncts. For example, the S conjunction in '*El consejo supremo, || integrado por oficiales de la más alta graduación @ || y algunos de ellos en puesto de mando de las Fuerzas Armadas, || deberán conocer la decisión del consejo de guerra || y oír si se les permitiera a los abogados de los procesados.*' is more felicitous when separated by a phrase boundary, but the procedure does **not** predict such a boundary. However, it **does** predict a boundary between the conjuncts in '*tal y como*' in '*Además de la nula visibilidad || desde el lugar donde se encuentra la oficina meteorológica || (tal @ y como se ilustra en el gráfico) || , la garita instalada en la terraza || incumple las condiciones.*'.

Our error analysis also uncovered a number of cases in which the ordering of words or phrases appeared 'marked'; it appeared that our phrasing annotation decisions, which were **not** captured by our decision tree, were associated with this markedness in word order. For example, in '*La pregunta la formuló, || el pasado día 13 || un lector de Madrid || que había leído || que el Ayuntamiento había acusado en el juzgado de guardia @ || a un policía municipal || por malos tratos a varios ciudadanos.*', the phrase '*a un policía municipal*' would be more likely to appear directly after the verb; in its actual position, a common construction in newspaper text, it sounds more felicitous when set off from the rest of the sentence as a separate prosodic phrase. Similarly, in '*Harkin dijo en una conferencia de prensa || que espera que el presidente electo Bill Clinton || enviará al congreso || antes del plazo del 21 de mayo @ || la necesaria notificación del inicio de las negociaciones con Chile.*', the direct object, '*la necesaria notificación*', appears in a relatively marked position, and sounds more felicitous when separated from the sentence by a phrase boundary.

Finally, our procedure was often unable to distinguish NP from S coordination, simply from our short part-of-speech window. In one case, an appositive phrase were not handled properly by our decision tree, since the phrase was not set off orthographically in the text.

Our procedure also failed to capture a general tendency we subsequently observed in our corpus to vary boundary placement based upon the length of constituents to the right or left of the potential boundary site. In most cases, longer constituents tended to be set apart from preceding or succeeding material, where the same shorter constituent did not. In '*La presencia española presente en la reunión de Roma @ || no se ha comprometido a aportar cantidad alguna.*', the long subject NP sounds more felicitous when set off from the verb by a phrase boundary, where shorter subjects do not.

4. Conclusion

We have described a procedure for training intonational phrasing decisions for unrestricted text-to-speech on annotated text, using CART techniques to generate phrasing decision trees automatically. This procedure has been used to build a phrasing module which is incorporated in the English Bell Labs TTS system and has also been used to construct a stand-alone procedure for the Mexican Spanish version of this system, for eventual inclusion in the Mexican TTS system. The advantages of this procedure are several: It makes updating an existing phrasing procedure simple and rapid: One need only provide a new or additional set of annotated text. Phrasing errors observed in the use of the TTS system can often be corrected simply by providing correctly annotated exemplars of the observed error, although, in some cases correcting one error by introducing new training material does introduce new errors in phenomena previously handled correctly. Phrasing for new domains can also be modeled easily, simply by having a native speaker annotate text for the new domain. And phrasing rules can be acquired for new languages easily, limited mainly by the tools for text analysis available for the new language.

Notes

[1]Accent may also vary between the two readings, but, for English, this variation is not as consistent as phrasing (Avesani, Hirschberg, and Prieto, 1995).

[2]CART cross-validation estimates are derived in (roughly) the following way: CART separates input training data into training and test sets (90% and 10% of the input data in the implementation used here), grows a subtree on the training data and tests on the test data, repeats this process a number of times (five, in the implementation used here), and computes an average result for the subtrees.

[3]While some additional work was needed to select and pre-process the text for annotation, this was all done via some simple shell scripts, which are easily re-used.

[4]See the procedure described on page 5 for a discussion of how such trees are generated.

[5]Recall that the cross-validated error is averaged over the predictions of multiple trees of the same length, so any single tree may perform better or worse.

[6]In this and subsequent examples, "||" indicates the presence of a phrase boundary at a potential boundary site in the annotated training corpus; "@" indicates a misprediction, according to our hand-annotated data; so, "@ ||" indicates that our procedure failed to predict a boundary which appears in the annotated corpus, while "@" alone indicates that our decision procedure predicted a boundary which does **not** appear in the annotation. English translations of these examples appear at the end of the paper.

Translations of Examples

p. 11 *Asimismo, once hindúes fueron mientras heridos a tiros por agentes policiales mientras sus correligionarios lanzaban bombas incendiarias de fabricación casera y trapos encendidos contra las improvisadas viviendas de los musulmanes, agregó.*

Likewise, eleven Hindus were shot and wounded by policemen while their sympathisers were throwing homemade incendiary bombs and burning rags against improvised Muslim homes, he added.

p. 11 *Abogados defensores han expresado su confianza que muchos de los detenidos podían ser absueltos por el consejo pero también su preocupación por cómo se ha realizado todo el juicio.*

Defense lawyers have expressed their confidence that many of the arrested could be released by the council, but they have also expressed their concern about the way in which the trial has been conducted.

p. 12 *Irak está desplazando baterías de misiles antiaéreos al sur y al norte de las zonas de no vuelos impuestas por los aliados occidentales en un esfuerzo por confundir y provocar a sus enemigos, dijeron funcionarios norteamericanos.*

Iraq is moving units of anti-aircraft missiles north and south of the no-fly zones imposed by the Western allies, in an effort to confuse and provoke their enemies.

p. 13 *El consejo supremo, integrado por oficiales de la más alta graduación y algunos de ellos en puesto de mando de las Fuerzas Armadas, deberán conocer la decisión del consejo de guerra y oír si se les permitiera a los abogados de los procesados.*

The supreme council, formed by high-ranking officers (some of whom are high commanders in the Armed Forces), ought to know the court-martial's decision and listen, if they are allowed to, to the defendant's lawyers.

p. 13 *Además de la nula visibilidad desde el lugar donde se encuentra la oficina meteorológica (tal y como se ilustra en el gráfico), la garita instalada en la terraza incumple las condiciones.*

The cabin installed in the terrace does not meet the requirements, not to mention the very poor visibility from the metereological observatory (as shown in the graph).

p. 13 *La pregunta la formuló, el pasado día 13 un lector de Madrid que había leído que el Ayuntamiento había acusado en el juzgado de guardia a un policía municipal por malos tratos a varios ciudadanos.*

The question was asked on the 13th of this month by a Madrid reader. The reader had read that the City had pressed charges against a city policeman for mistreating several local residents.

p. 13 *Harkin dijo en una conferencia de prensa que espera que el presidente electo Bill Clinton enviará al congreso antes del plazo del 21 de mayo la necesaria notificación del inicio de las negociaciones con Chile.*

Harkin said in a press conference that he hopes that president-elect Bill Clinton will send the necessary notification of the start of the negotiations with Chile to Congress before May 21.

p. 14 *La presencia española presente en la reunión de Roma no se ha comprometido a aportar cantidad alguna.*

The Spanish delegation present at the Rome meeting did not make a commitment to contribute any money whatsoever.

References

B. Altenberg (1987), *Prosodic Patterns in Spoken English: Studies in the Correlation between Prosody and Grammar for Text-to-Speech Conversion*, Lund Studies in English, Vol. 76, (Lund University Press, Lund).

C. Avesani, J. Hirschberg, and P. Prieto (1995), "The intonational disambiguation of potentially ambiguous utterances in English, Italian, and Spanish", *Proc. XIIIth Internat. Cong. Phonetic Sciences*, Vol. 1, pp. 174–177.

J. Bachenko and E. Fitzpatrick (1990), "A computational grammar of discourse-neutral prosodic phrasing in English", *Computational Linguistics*, Vol. 16, No. 3, pp. 155–170.

M. Beckman and J. Pierrehumbert (1986), "Intonational structure in Japanese and English", *Phonology Yearbook*, Vol. 3, pp. 15–70.

D. Bolinger (1989), *Intonation and Its Uses: Melody in Grammar and Discourse*, (Edward Arnold, London).

L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone (1984), *Classification and Regression Trees*, (Wadsworth & Brooks, Pacific Grove CA).

G. Bruce, B. Granström, K. Gustafson, and D. House (1993), "Prosodic modelling of phrasing in Swedish", *Proc. ESCA Workshop on Prosody*, ed. by D. House and P. Touati, Lund University Dept. of Linguistics Working Papers, Vol. 41, pp. 180–183.

L. Danlos, E. LaPorte, and F. Emerard (1986), "Synthesis of spoken messages from semantic representations", COLING-86, *Proc. 11th Internat. Conf. Computational Linguistics*, pp. 599–604.

J. Hirschberg (1991), "Using text analysis to predict intonational boundaries", In *Proc. 1991 EuroSpeech.*

D. Larreur, F. Emerard, and F. Marty (1989), "Linguistics and prosodic processing for a text-to-speech synthesis system", *Proc. 1989 EuroSpeech*, Vol. 1, pp. 510–513.

A. Monaghan, (1991), *Intonation in a Text-to-Speech Conversion System*, Ph.D. thesis, University of Edinburgh, Edinburgh.

D. O'Shaughnessy (1989), "Parsing with a small dictionary for applictions such as text to speech", *Computational Linguistics*, Vol. 15, No. 2, pp. 97–108.

M. Ostendorf, P. Price, J. Bear, and C. W. Wightman (1990), “The use of relative duration in syntactic disambiguation”, *DARPA, Proc. Speech and Natural Language Workshop*, pp 26–31.

M. Ostendorf and N. Veilleux (1994), “A hierarchical stochastic model for automatic prediction of prosodic boundary location”, *Computational Linguistics*, Vol. 20, No. 1, pp. 27–54.

J. B. Pierrehumbert (1980), *The Phonology and Phonetics of English Intonation*, Ph.D. thesis, Massachusetts Institute of Technology.

J. Pitrelli, M. Beckman, and J. Hirschberg (1994), “Evaluation of prosodic transcription labeling reliability in the ToBI framework”, *ICSLP-94, Internat. Conf. on Spoken Language Processing*, Vol. 2, pp. 123–126.

H. Quené and R. Kager (1992), “The derivation of prosody for text-to-speech from prosodic sentence structure”, *Computer Speech and Language*, Vol. 6, pp. 77–98.

M. D. Riley (1989), “Some applications of tree-based modelling to speech and language”, *DARPA, Proc. Speech and Natural Language Workshop.*

M. A. Rodríguez, J. G. Escalada, A. Macarrón, and L. Monzón (1993), “AMIGO: Un conversor texto-voz para el españ”, *Boletín de la Sociedad Española para el Procesamiento del Lenguaje Natural*, Vol. 13, pp. 389–400.

B. Schnabel and H. Roth (1990), “Automatic linguistic processing in a German text-to-speech synthesis system”, *Proc. ESCA Workshop on Speech Synthesis*, pp. 121–124.

K. Silverman, A. Kalyanswamy, J. Silverman, S. Basson, and D. Yashchin (1993), “Synthesiser intelligibility in the context of a name-and-address information service”, *Proc. 1993 EuroSpeech*, Vol. 3, pp. 2169–2172.

M. Q. Wang and J. Hirschberg (1991a), “Predicting intonational boundaries automatically from text: The ATIS domain”, *DARPA, Proc. Speech and Natural Language Workshop*, pp. 378–383.

M. Q. Wang and J. Hirschberg (1991b), "Predicting intonational phrasing from text", *ACL, Proc. 29th Annual Meeting.*

M. Q. Wang and J. Hirschberg (1992), "Automatic classification of intonational phrase boundaries", *Computer Speech and Language*, Vol. 6, pp. 175–196.

C. W. Wightman, C. W. and M. Ostendorf (1994), "Automatic labeling of prosodic patterns", *IEEE Trans. Speech and Audio Processing*, Vol. 2, No. 4, pp. 469–481.

S. J. Young and F. Fallside (1979), "Speech synthesis from concept: A method for speech output from information systems", *J. Acoust. Soc. Amer.* , Vol. 66, No. 3, pp. 685–695.

Label	Description
dpunc	distance from previous punctuation in words
ew	distance in words of $< w_i, w_j >$ from end of sentence
la	is w_i accented?
pos1$\{a,b\}$	part-of-speech of w_{i-1}
pos2$\{a,b\}$	part-of-speech of w_i
pos3$\{a,b\}$	part-of-speech of w_j
pos4$\{a,b\}$	part-of-speech of w_{j+1}
punc	identity of any punctuation at $< w_i, w_j >$
ra	is w_j accented
sw	distance of $< w_i, w_j >$ from beginning of sentence, in words
syls	distance of $< w_i, w_j >$ from beginning of sentence in syllables
tw	total words in sentence
wb	would vowel elision occur across $< w_i, w_j >$?

Table 1: Key to Features Used for Spanish

Label	Part-of-Speech
ADEM	demonstrative adjective (e.g. 'esta', **this** (feminine))
ADVO	adverb
APOS	possessive adjective (e.g.' sus', **his** (object is plural))
ARTI	article
CNTR	contraction (e.g. *'del'*, **of the** (masculine))
COOR	coordination particle (e.g. '*y*', **and**)
LADV	complex adverbials (e.g. '*por lo menos*', **at least**)
LPRE	linked preposition (e.g. '*de*' in '*a través de*', **through**)
LSUB	linked subordinate particle (e.g. '*sin embargo*', **however**)
NOMB	noun or adjective
NOMB_^ hline NPRO	second element of complex noun proper noun
NUME	number
PPER	personal pronoun (e.g. '*le*', **him** (indirect object))
PREL	relative pronoun
PREP	preposition
SIG_ORTO	orthographic sign
SUBO	subordinate particle (e.g. '*que*', **that**)
SUBO_^	subcategorized subordinate particle (e.g. '*que*' in '*tengo que ir*', **I have to go**)
VERB	verb
VERB_^	second element of complex verb (e.g. '*mezclado*', **mixed**, in '*esta mezclado*', **is mixed**)

Table 2: Part-of-Speech Labels Used for Spanish

Error Type	NoBoundary	Boundary
POS Error		2
Attachment	53	1
Coordination	8	
Constituent Length	23	1
Apposition		1
‘Marked’ Word Order	13	

Table 3: Sources of Error in Spanish Infelicitous Phrasing Prediction

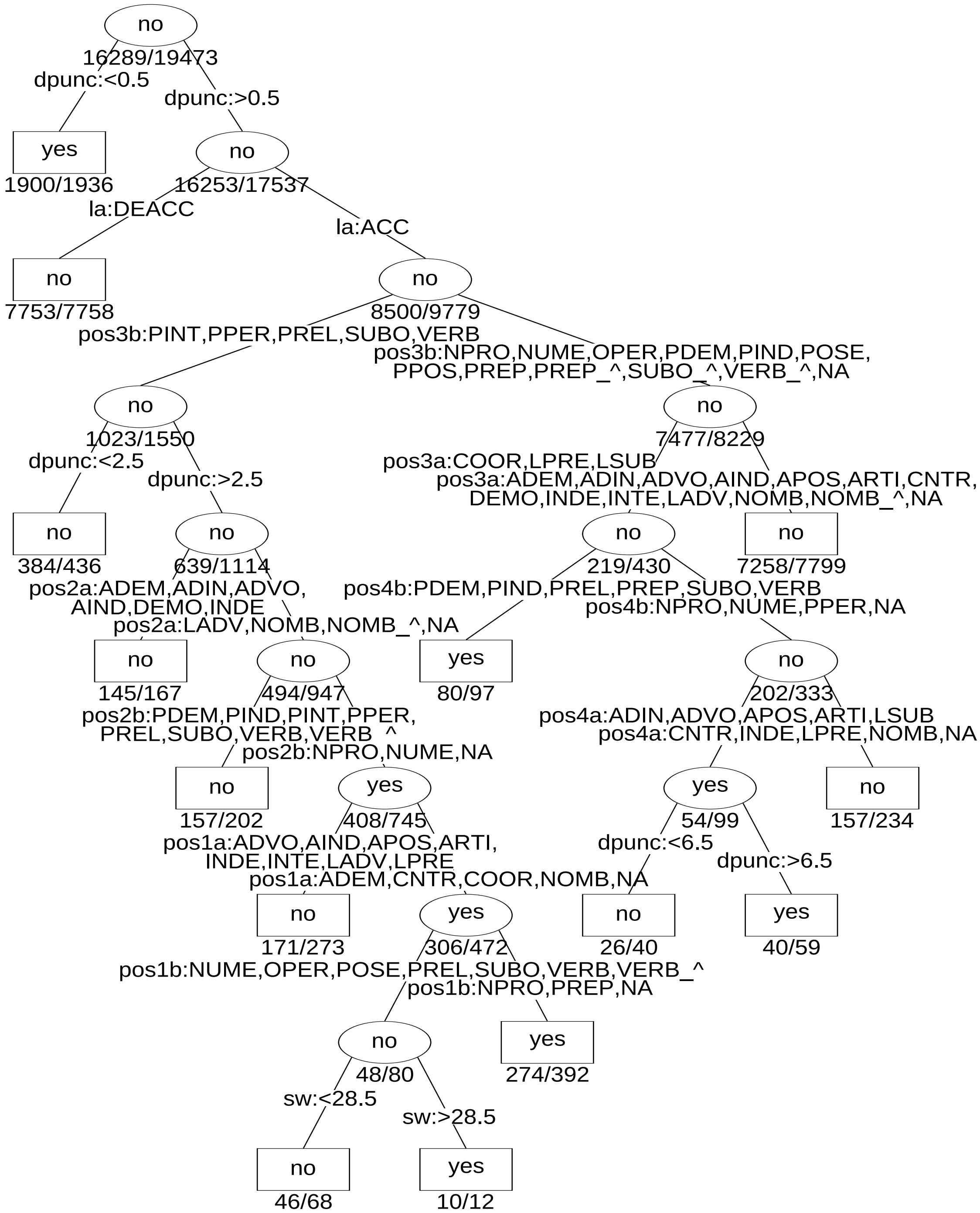

Figure 1: CART Prediction Tree for Spanish Phrasing