Combining greedy algorithms with expert guided manipulation for the definition of a balanced prosodic Spanish-Catalan radio news corpus

D. Escudero-Mancebo, C. González-Ferreras[1], *J. M. Garrido*[2], *E. Rodero*[3], *L. Aguilar* [4], *A. Bonafonte*[5]

[1] Department of Computer Science, Universidad de Valladolid
[2] Department of Translation and Language Sciences, Universidad Pompeu Fabra
[3] Department of Communication, Universidad Pompeu Fabra
[4] Department of Spanish Philology, Universidad Autonoma de Barcelona
[5] Department of Signal Theory, Universidad Politécnica de Catalunya
descuder@infor.uva.es

Abstract

This article reports the process of building a bilingual (Spanish-Catalan) text corpus balanced in parallel taking into account prosodic features for both languages. We propose an expert guideline for text manipulation that in combination with greedy algorithms significantly improves the quality of the selected corpus. The application of this methodology to a radio news corpus empirically supports the proposed strategy.

1. Introduction

Subcorpus selection is a need in various domains of speech technologies. In text-to-speech, greedy algorithms are used to build the space limited unit-selection data base [1] while in speech recognition the training corpus must be selected to find a representative sample [2]. Although some authors have proposed to randomly choose the subcorpus (see [3]), text selection is broadly accepted as a procedure to ensure the representativeness of the corpus by means of maximising its coverage.

On the other hand, the issue of multilinguism is present in all domains of speech and language technologies and efforts to automatize the collection of data in several languages are made. Parallel corpora (i.e. text paired with its translation into a second language) are very frequently used in statistical machine translation (see conferences LREC) and multilingual speech corpora for specific applications are available for some pairs of languages (Speechdat; Verbmobil; see Linguistic Data Consortium).

This contribution reports the use and the comparison of selection techniques for building a prosodically balanced corpus that intends to be a reference in the prosodic studies. To our knowledge, this is first study in which statistical procedures for text selection are applied to obtain a parallel corpus that is prosodically balanced in two languages (Spanish and Catalan). We focus on the maximum coverage of prosodic properties in the chosen texts, but not in an independent way from the source (Spanish or Catalan), as they are selected simultaneously so the texts can be aligned. This procedure allows us to compare in a detailed way the prosody and intonation of the news reading.

This activity has been done in the framework of the research project Glissando, which main goal is to build a reference prosodic corpus for Spanish and Catalan. It is being developed for a multi-disciplinary user group, and it is going to contain speech from three situational settings, namely, news reading, conversational speech and task-oriented speech. All speech will be orthographically and phonetically transcribed, and a manually verified prosodic annotation will be provided. The selection method presented here has been applied to the selection of the news corpus. Given the large-scale compilation, a text selection procedure, with the most objective criteria, is clearly needed, since the reading of radio news should be limited to a time of thirty minutes for each language.

The major milestone was to select a corpus that contains a balanced sample of prosodic phenomena. A priori, the problem is not very different to the issue of constructing a phonetically balanced subcorpus if we have a reference prosodic unit and the set of prosodic features to characterize it. In this paper we have chosen stress groups as this basic reference prosodic unit and texts have been labelled using it, so that greedy algorithms can devise a prosodically balanced subcorpus. Then selection task has chosen the best set of texts which offer the best coverage for all the predefined stress group types.

A second restriction was to keep the parallelism between the Spanish and Catalan corpora: the two sets of selected texts had to be the same in both languages, to allow future inter-language comparisons, so in this case the problem was to obtain the best set of paired Spanish-Catalan texts that would offer the best coverage of stress group types in both Spanish and Catalan.

The output of the greedy algorithm is expected to be a fully prosodically balanced parallel corpus. Nevertheless, due to the mother corpus limitations this goal is difficult or impossible to be reached. In Spanish texts the relative frequency of *paroxitones* words is tenths more than the frequency of *proparoxitones* words. Similar figures could be obtained in the case of Catalan. In these circumstances, it is normal that the selection algorithm still outputs unbalanced results. In order to improve these results, the subcorpus has been reviewed and corrected. Some actions of this revision were semiautomatic as listed in the *Expert Guideline* that we present below. Next the results of increasing the size of the mother corpus versus the use of greedy algorithms together with expert modifications are compared.

The paper is organized as follows: First we present the application of greedy algorithms to select a balanced subcorpus, second we present the expert guideline designed to improve results, third we comment on results and conclusions.

Partially founded by the Ministerio de Ciencia e Innovación, Spanish Government Glissando project FFI2008-04982-C003-02

2. Greedy algorithms for text selection

In [4] we formalized the subcorpus selection problem using greedy algorithms as the need to overcome a target vector that indicates the quantity of required stress groups of different types in the final subcorpus. A heuristic rule determines which of the radio news candidates found in a mother corpus is chosen in every step. The selection in every step is done in function of the already selected subcorpus and the target vector. We tested three different heuristic rules found in the state of art [5, 6, 1].

Variability is probably the main characteristics of prosody, with a high number of factors that affects its form and function. Under this condition, it is not easy to avoid a problem that dramatically decreases the success rates of the greedy algorithms: the scarcity of some type of units in the mother corpus (infrequent phonemes in the case of phonetically balanced selection, rare type of stress groups for prosodic corpus). During the process of selection, the greedy algorithms can discard those rare types of units because of their low relative importance with respect to the more frequent types of units. However, as far as prosodic modelling is concerned, the appearance of infrequent situations in the corpus is a must, as they can show relevant intonation shapes and functions. In [4] we tested a set of commonly used strategies that copes with the scarce type unit issue. In the literature we find greedy algorithms with modified heuristics [1] and specialized search strategies [7] to select this class of rare units. Furthermore, we devised a new strategy modifying the value of the target vector as the selected subcorpus grows.

The different greedy heuristics and strategies for the definition of a prosodically balanced corpus were empirically compared using as mother corpus the **ONU Radio Corpus** (3727 radio news items with more than 376.000 stress groups) and the **Cadena SER Corpus** (to be described in section 4). These corpora represent two different scenarios as the **Cadena SER Corpus** is about 37 times smaller. Despite the proficiency of the greedy algorithms, contrasted with several objective metrics, we showed that a simple manipulation of the contents of the smaller corpus allows to significantly improve the quality of the selected subcorpus: Cadena SER ($valUnits = 1645$) Cadena SER Modified ($valUnits = 1833$) and Radio ONU ($valUnits = 1769$) —$valUnits$ measures the quality of the subcorpus as it indicates the number of units in the selected subcorpus that do not exceed the target vector. Exceding units are a problem as ToBI labelling commonly takes 100-200 times real time [8]—.

3. Expert guideline for corpus modification

A first analysis of the texts selected by the algorithms revealed that the target could not be achieved from the mother corpus only by automatic means: after the application of the greedy algorithms the selected corpus is still unbalanced. Several factors can explain it:

1) As it is well-known oxytone stress word pattern is the most frequent in Spanish, followed by the paroxitone one, and very far away in the scale, by the proparoxitone one (see among others, [9]). The situation is similar in the case of Catalan although paroxitone and proparoxiton groups are more balanced. This is the reason why the number of appearances of proparoxitone stress groups is very low compared with the rest of types (see table 2). A very large set of input (and output) text would be necessary to get a significant number of proparoxiton stress groups.

2) It is not possible to predict form the text the final segmentation in phonic groups carried out by readers, because some of the pauses introduced are not induced by punctuation marks but by other factors such as syntactic structure and even individual decisions. For this reason, the automatic estimation of phonic groups for this task, based exclusively in punctuation marks, is necessarily tentative, only a reference to guide the selection. Using this approach, the theoretical phonic groups detected in the mother corpus tend to be rather long. This fact could explain the low number of short phonic groups in the analyzed corpus (that is, GTInicialFinal, or phonic groups containing only one stress group) as indicates table 2.

3) Radio news style has their own convention to mark prosodic boundaries: texts can have long sentences without any punctuation. Besides this, a careful reading of the texts has revealed that the punctuation conventions for Spanish and Catalan are not always respected in the texts written for radio news, since professional radio speakers have their own ones.

It was decided then that some kind of manual revision and correction the results would be necessary. To balance as much as possible the number of represented stress groups and (theoretical) phonic groups in the selected texts, without losing the naturalness of the contents of the original corpus, two strategies were applied:

- Modification of the text to include a greater number of proparoxytone words, while preserving as much as possible the naturalness of the contents. The procedure was straightforward: first, a list of proparoxitones Spanish names (6 entries), surnames (10 female and 6 male entries) and names of cities (9 entries) was built; then the proparoxitones proper nouns in the corpus were identified by using FreeLing http://garraf.epsevg.upc.es/freeling (58, 17, 48 and 18 entries respectively) and systematically replaced by the names of the list. The resulting texts were carefully reviewed to avoid repetitions in a given text, and strange results affecting its naturalness. Some extra names were also added or substituted manually where possible. Spanish proper names were used in the modification of both Spanish and Catalan texts: it allowed to improve the comparability of the texts, without loosing naturalness (Spanish names are very common in Catalan texts, and Catalan readers can pronounce them easily).
- Slight modification of the punctuation of the texts, including some extra marks (specially in appositions and long restrictive relative clauses), but trying to keep the balance between control and naturalness of the text (too many punctuation marks in the texts would make them unrealistic, specially considering that they are supposed to be representative of radio news). Some punctuation marks were added in some texts in order to obtain shorter phonic groups (by means of periods, semicolon an colon) and more SGs in positions other than central (by means of the use of commas, whenever the grammatical sense allows it).
- Parallel correction of the punctuation in the Spanish and Catalan versions. The reviewer must take into account that the inclusion of a change in one of the corpus leads to alterations in the other one that must be weighted before proceeding with the modification.

4. Experimental results

4.1. The corpus

The $Glissando$ project requires the recording of 30 minutes of read speech to model the characteristic prosodic features of

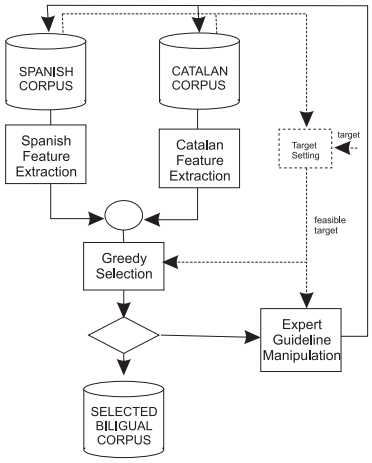

Figure 1: Scheme of the iterative strategy to combine greedy algorithms with the expert guided manipulation

Percentil	SPCAT Ori	SPCATMod
10,00%	0	1
15,00%	1	2
20,00%	1	3
30,00%	4	5
40,00%	6	8
50,00%	10	12
70,00%	23	25
80,00%	38	37
90,00%	68	69
99,00%	156	140

Table 1: Percentils of the quantity of units of the different type stress groups considered. $SPCATOrig$ refers to the selected Spanish-Catalan corpus without modifications and $SPCATMod$ with modifications

different professional radio speakers. The goal is to select a subset of text radio news whose total duration is about half an hour optimizing the prosodic units coverage. Our mother corpus was gently supplied by Cadena SER Radio Station. This corpus is limited in size but we are interested in using it because it was reviewed by an expert according to a set of style conventions [10]. The original corpus is written in Spanish but we have a clone version in Catalan (manually translated from the Spanish version by a bilingual Catalan-Spanish translator). To estimate the normal speed of a professional radio speaker we have used the reading of a corpus from the Cadena SER by a professional speaker.

Candidate news are prosodically analyzed taking into account the type of stress groups. The stress group (SG), defined as a set of words with only one lexical stress, is a unit used in traditional Spanish prosodic studies [11], and it has been revealed to be a good prosodic unit of reference to model prosody (see for example [12, 13]) both for Spanish and Catalan [14]. Other unit that serves us to count the prosodic coverage of texts is the phonic group, that refers to the stretch of speech within two pauses. It is worthwhile noting that the results obtained with the textual analysis will not probably coincide with the final reading of the speakers, since it is well known that speakers attend to more clues than punctuation marks to prosodically organise sentences.

Results obtained in previous studies [13] for Spanish led us to use the following prosodic features to characterize the stress groups in both corpora: position of the phonic group in the sentence (Initial, Central, Final and Initial-Final), SG's position in the phonic group (Initial, Central, Final and Initial-Final), stressed syllable position within the SG (Oxiton, Paroxiton, Proparoxiton) and the number of syllables that the SG contains (one, two, three or more than three syllables). Of course other features and/or more values could be used but we were selective to avoid the combinatorial explosion and the drastic reduction of the number of samples per feature combination. The total number of possible feature combinations is 144 (some of the combinations are impossible); in other words, 144 different types of stress groups should be found in the Catalan and Spanish corpora.

Table 2 depicts the contents of the SER corpus in terms of type of units. Data reveals the main limitation of this corpus: 12% of the classes has no samples and 40% of the type of units has less than 18 units. The corpus is clearly unbalanced.

4.2. Procedure

Figure 1 shows the iterative operative. $CatalanCorpus$ is a translated mirror of the $SpanishCorpus$ that includes 100 text radio news tokens. The $FeatureExtraction$ task was done automatically with the text analysis module of the Ogmios Text-to-speech system [15] to obtain a prosodic labelling that takes into account the stress groups and their prosodic features. The Catalan and Spanish features are combined into a unique vector that feeds the $GreedySelection$ task.

The $GreedySelection$ task outputs the best set of candidates (radio news tokens). This half an hour subcorpus is manipulated by an expert according to the guideline. The manipulated subcorpus is integrated in the initial corpus and the $GreedySelection$ task is applied again. This is done because different combinations of news can appear as the radio news tokens are modified. This iterative process is stopped when all the tokens in the selected subcorpus have been already manipulated by the expert.

The $TargetSetting$ task (dotted lines) needs the $targetvector$ as an input. With the information in the corpus, it sets a $FeasibleTarget$ vector ($\bar{T}^f$) as some of the required classes of stress groups could be scarce in the corpora. The vector $\bar{T}^f$ governs the operative of the $GreedySelection$ and $ExpertManipulation$ tasks.

4.3. Results

Table 3 reports the figures of the selected subcorpus. Note that a balanced corpus is not a flat corpus where all the classes have the same cardinality. This, if possible, would give an artificial aspect to the corpus and one of the goals was to keep the speech style of professional news reading. The term "balanced" refers to the decreases of distances between the larger classes of stress groups with respect to the more scarce ones (for example the ratio Paroxitones vs Proparoxitones is the Spanish corpus is 336/6819 (table 2) and in the selected corpus is 277/2231 (table 3. This fact is reflected also in the percentils of the table 1: the result of applying the expert manipulation is the reduction of the most populated classes to increase the cardinality of the more scarce. Total number of differences between the subcorpus $SPCATOrig$ and $SPCATMod$ is 1523, which gives an idea of the extensive application of the Expert Guideline.

Note that table 2 and table 3, both have empty or almost empty $SGInitialFinal$ rows. We did not enter units of this type as they are very rare in radio news style.

Spanish

PhGInitial	SGInitial	Oxitone: 15 30 21 22	Paroxitone: 79 50 102	Proparoxitone: 0 12
	SGCentral	Oxitone: 119 104 79 108	Paroxitone: 236 316 528	Proparoxitone: 11 39
	SGFinal	Oxitone: 17 20 39 24	Paroxitone: 41 75 101	Proparoxitone: 3 10
	SGInitialFinal	Oxitone: 2 7 5 5	Paroxitone: 4 16 24	Proparoxitone: 0 0
PhGCentral	SGInitial	Oxitone: 18 45 21 26	Paroxitone: 85 64 121	Proparoxitone: 5 9
	SGCentral	Oxitone: 105 90 74 91	Paroxitone: 178 306 368	Proparoxitone: 9 26
	SGFinal	Oxitone: 9 20 27 17	Paroxitone: 64 89 144	Proparoxitone: 6 17
	SGInitialFinal	Oxitone: 5 3 7 4	Paroxitone: 16 13 28	Proparoxitone: 0 1
PhGFinal	SGInitial	Oxitone: 20 29 18 29	Paroxitone: 95 71 116	Proparoxitone: 1 7
	SGCentral	Oxitone: 145 125 117 145	Paroxitone: 307 383 589	Proparoxitone: 11 47
	SGFinal	Oxitone: 10 22 23 22	Paroxitone: 49 91 142	Proparoxitone: 10 17
	SGInitialFinal	Oxitone: 0 0 0 0	Paroxitone: 0 2 1	Proparoxitone: 1 0
PhGInitialFinal	SGInitial	Oxitone: 19 18 7 14	Paroxitone: 38 46 80	Proparoxitone: 1 5
	SGCentral	Oxitone: 171 174 140 169	Paroxitone: 377 506 724	Proparoxitone: 12 63
	SGFinal	Oxitone: 7 9 20 25	Paroxitone: 25 51 78	Proparoxitone: 3 10
	SGInitialFinal	Oxitone: 0 0 0 0	Paroxitone: 0 0 0	Proparoxitone: 0 0

Attribute				
posSTSG	Oxitone=2657	Paroxitone=6819	Proparoxitone=336	
posSGIG	SGInitial=1339	SGCentral=6992	SGFinal=1337	SGInitialFinal=144
posIGSE	PhGInitial=2264	PhGCentral=2111	PhGFinal=2645	PhGInitialFinal=2792
nSyl	1Syl=662	2Syl=2290	3Syl=2750	m3Syl=4110

Catalan

PhGInitial	SGInitial	Oxitone: 29 69 41 54	Paroxitone: 31 46 44	Proparoxitone: 0 11
	SGCentral	Oxitone: 161 214 201 276	Paroxitone: 117 195 227	Proparoxitone: 7 45
	SGFinal	Oxitone: 31 42 52 48	Paroxitone: 22 46 63	Proparoxitone: 3 18
	SGInitialFinal	Oxitone: 2 22 14 15	Paroxitone: 2 7 13	Proparoxitone: 0 1
PhGCentral	SGInitial	Oxitone: 42 100 49 63	Paroxitone: 33 44 70	Proparoxitone: 6 6
	SGCentral	Oxitone: 150 198 163 215	Paroxitone: 124 184 191	Proparoxitone: 7 41
	SGFinal	Oxitone: 29 45 63 59	Paroxitone: 49 75 74	Proparoxitone: 1 18
	SGInitialFinal	Oxitone: 14 15 13 7	Paroxitone: 27 10 18	Proparoxitone: 0 7
PhGFinal	SGInitial	Oxitone: 35 77 49 62	Paroxitone: 45 45 65	Proparoxitone: 2 12
	SGCentral	Oxitone: 216 293 234 366	Paroxitone: 169 239 299	Proparoxitone: 8 59
	SGFinal	Oxitone: 30 54 58 69	Paroxitone: 49 49 56	Proparoxitone: 5 22
	SGInitialFinal	Oxitone: 0 1 1 3	Paroxitone: 0 0 1	Proparoxitone: 1 0
PhGInitialFinal	SGInitial	Oxitone: 22 50 28 35	Paroxitone: 16 35 29	Proparoxitone: 1 7
	SGCentral	Oxitone: 234 322 331 402	Paroxitone: 204 286 348	Proparoxitone: 10 48
	SGFinal	Oxitone: 25 34 30 45	Paroxitone: 15 22 39	Proparoxitone: 0 13
	SGInitialFinal	Oxitone: 0 0 0 0	Paroxitone: 0 0 0	Proparoxitone: 0 0

Attribute				
posSTSG	Oxitone=5602	Paroxitone=3723	Proparoxitone=359	
posSGIG	SGInitial=1353	SGCentral=6784	SGFinal=1353	SGInitialFinal=194
posIGSE	PhGInitial=2169	PhGCentral=2210	PhGFinal=2674	PhGInitialFinal=2631
nSyl	1Syl=1020	2Syl=2439	3Syl=2661	m3Syl=3564

Table 2: Cardinality of the type of units in the **Cadena SER corpus**. First column is the position of the phonic group in the sentence. Second column is the position of the stress group (SG) in the phonic group (PhG). Column 3, 4 and 5 correspond with the different positions of the stress in the word and size of the stress group: Oxitone (1, 2, 3 and more that 3 syllables); Paroxitone (2, 3 and more that 3 syllables); Proparoxitone (3 and more that 3 syllables).

Spanish

PhGInitial	SGInitial	Oxitone: 11 14 7 9	Paroxitone: 31 27 35	Proparoxitone: 6 6
	SGCentral	Oxitone: 48 46 21 35	Paroxitone: 86 101 177	Proparoxitone: 8 20
	SGFinal	Oxitone: 8 4 15 7	Paroxitone: 23 28 48	Proparoxitone: 3 10
	SGInitialFinal	Oxitone: 1 5 4 4	Paroxitone: 4 6 12	Proparoxitone: 2 0
PhGCentral	SGInitial	Oxitone: 19 27 16 19	Paroxitone: 45 41 69	Proparoxitone: 14 10
	SGCentral	Oxitone: 44 30 34 44	Paroxitone: 76 140 159	Proparoxitone: 12 30
	SGFinal	Oxitone: 6 17 19 15	Paroxitone: 35 42 74	Proparoxitone: 21 31
	SGInitialFinal	Oxitone: 4 5 7 6	Paroxitone: 14 9 17	Proparoxitone: 1 3
PhGFinal	SGInitial	Oxitone: 5 12 11 12	Paroxitone: 37 39 46	Proparoxitone: 7 8
	SGCentral	Oxitone: 46 40 32 36	Paroxitone: 81 106 139	Proparoxitone: 2 30
	SGFinal	Oxitone: 4 7 8 6	Paroxitone: 22 47 57	Proparoxitone: 12 14
	SGInitialFinal	Oxitone: 0 0 0 0	Paroxitone: 0 3 2	Proparoxitone: 1 1
PhGInitialFinal	SGInitial	Oxitone: 6 5 2 2	Paroxitone: 10 10 21	Proparoxitone: 0 1
	SGCentral	Oxitone: 23 35 24 32	Paroxitone: 63 89 122	Proparoxitone: 5 12
	SGFinal	Oxitone: 1 3 3 5	Paroxitone: 4 12 22	Proparoxitone: 2 5
	SGInitialFinal	Oxitone: 0 0 0 0	Paroxitone: 0 0 0	Proparoxitone: 0 0

Attribute				
posSTSG	Oxitone=911	Paroxitone=2231	Proparoxitone=277	
posSGIG	SGInitial=640	SGCentral=2028	SGFinal=640	SGInitialFinal=111
posIGSE	PhGInitial=872	PhGCentral=1155	PhGFinal=873	PhGInitialFinal=519
nSyl	1Syl=226	2Syl=781	3Syl=999	m3Syl=1413

Catalan

PhGInitial	SGInitial	Oxitone: 12 27 15 18	Paroxitone: 10 25 14	Proparoxitone: 4 4
	SGCentral	Oxitone: 72 75 68 95	Paroxitone: 44 56 92	Proparoxitone: 5 18
	SGFinal	Oxitone: 15 11 22 20	Paroxitone: 10 14 22	Proparoxitone: 4 11
	SGInitialFinal	Oxitone: 1 15 6 5	Paroxitone: 1 1 6	Proparoxitone: 2 0
PhGCentral	SGInitial	Oxitone: 17 49 29 34	Paroxitone: 14 17 33	Proparoxitone: 14 8
	SGCentral	Oxitone: 53 76 61 80	Paroxitone: 53 77 69	Proparoxitone: 10 35
	SGFinal	Oxitone: 17 23 24 35	Paroxitone: 23 27 30	Proparoxitone: 13 23
	SGInitialFinal	Oxitone: 8 10 5 7	Paroxitone: 5 4 4	Proparoxitone: 1 2
PhGFinal	SGInitial	Oxitone: 13 30 16 25	Paroxitone: 20 17 19	Proparoxitone: 9 6
	SGCentral	Oxitone: 70 84 73 93	Paroxitone: 44 73 62	Proparoxitone: 4 27
	SGFinal	Oxitone: 12 26 13 23	Paroxitone: 22 18 21	Proparoxitone: 8 12
	SGInitialFinal	Oxitone: 0 1 1 2	Paroxitone: 0 2 3	Proparoxitone: 1 1
PhGInitialFinal	SGInitial	Oxitone: 6 17 11 10	Paroxitone: 9 8 9	Proparoxitone: 1 2
	SGCentral	Oxitone: 39 90 69 110	Paroxitone: 53 76 80	Proparoxitone: 7 27
	SGFinal	Oxitone: 7 17 7 14	Paroxitone: 4 4 12	Proparoxitone: 2 6
	SGInitialFinal	Oxitone: 0 0 0 0	Paroxitone: 0 0 0	Proparoxitone: 0 0

Attribute				
posSTSG	Oxitone=1884	Paroxitone=1207	Proparoxitone=267	
posSGIG	SGInitial=572	SGCentral=2120	SGFinal=572	SGInitialFinal=94
posIGSE	PhGInitial=820	PhGCentral=990	PhGFinal=851	PhGInitialFinal=697
nSyl	1Syl=342	2Syl=863	3Syl=924	m3Syl=1229

Table 3: Cardinality of the type of units in the **Selected Corpus**. Same interpretation as table 2

5. Conclusions

We have presented a procedure to select the contents of a corpus for prosodic studies. Its application has permitted to obtain a relatively short set of radio news tokens , using a rather small mother corpus as input, and showing a balance in terms of type of stress groups close to the obtained automatically when using a larger input corpus (the ONU corpus text).

6. References

[1] J. van Santen and A. Buchsbaum, "Methods for optimal text selection," pp. 553–556, 1997.

[2] A. Nagorski, L. Boves, and Y. Steeneken, "Optimal selection of speech data for automatic speech recognition systems," pp. 555–558, 2473-2476 1992.

[3] T. Lambert, N. Braunshweiler, and S. Buchholz, "How (not) to select your voice corpus: Random selection vs. phonologically balanced," pp. 264–269, 2007.

[4] D. Escudero-Mancebo, L. Aguilar, A. Bonafonte, and J. Garrido-Alminana, "On the definition of a prosodically balanced corpus: combining greedy algorithms with expert guided manipulation," *Procesamiento del Lenguaje Natural*, no. 43, pp. 93–102, 2009.

[5] J. Matousek, D. Tihelka, and J. Rompuortl, "Building of a speech corpus optimised for unit selection tts synthsis," 2008.

[6] J. van Santen, "Diagnostic perceptual experiments for text-to-speech system evaluation," pp. 555–558, 1992.

[7] J. Zhang and S. Nakamura, "Least-to-most ordered search for minimum sentence set for collecting speech database," *Proceedings of ASJ*, pp. 145–146, October 2001.

[8] A. K. Syrdal, J. Hirschberg, J. McGory, and M. Beckman, "Automatic ToBI prediction and alignment to speed manual labeling prosody," *Speech Communications*, no. 33, pp. 135–151, 2001.

[9] M. Canellada and J. Madsen, *Pronunciación del español. Lengua hablada y literaria*. Madrid: Catalia, 1987.

[10] E. Rodero-Antón, *Locución radiofónica*, 1st ed. IORTV, 2003.

[11] T. Navarro-Tomás, *Manual de Entonación Española*. Madrid, Guadarrama, 1944.

[12] J. M. Garrido, "Modelling spanish intonation for text-to-speech applications," Ph.D. dissertation, Facultat de Lletres, Universitat de Barcelona, España, 1996.

[13] D. Escudero and V. Cardeñoso, "Applying data mining techniques to corpus based prosodic modeling speech," *Speech Communication*, vol. 49, pp. 213–229, 2007.

[14] D. Escudero, V. Cardeñoso, and A. Bonafonte, "On the comparison of catalan-spanish intonation systems using statistical corpus modeling and objective metrics," in *Proceedings of Prosody 2008*, 2008.

[15] A. Bonafonte, A. Moreno, J. Adell, P. D. Agüero, E. Banos, D. Erro, I. Esquerra, J. Pérez, and T. Polyakova, "The UPC TTS system description for the 2008 blizzard challenge," in *Proc of the Blizzard Challenge*, Brisbane, Australia, Sept. 2008.