

MIGUELINA GUIRAO

MARIA A. GARCIA JURADO

Laboratorio de Investigaciones Sensoriales
CONICET, Universidad de Buenos Aires
CC 53 1453 Buenos Aires, Argentina

ABSTRACT

In this paper we report an account of Spanish American phonological units. The sample consisted of 74460 syllables distributed in 43306 words. The frequency of occurrence of phonemes are presented each one labeled according to articulatory features. Dentals have an incidence of almost four times more than velars. Palatals have a low frequency. Voiced phonemes gave higher figures than unvoiced ones. A table is presented with a ranking order of the first 50 syllables. Thirty one of these syllables are also words. Included is the percent incidence of each of the first 50 syllables in initial and final word position. Type CV is equally distributed in both positions and CVC tend to terminate words. An observation is made on the most frequent articulatory combinations encountered at both extremes of words.

INTRODUCTION

In this work we intent to address the problem of segmentation of morphemic units in continuous speech on the basis of statistical data. Other aims were to provide useful information for spectrogram reading and for cross language comparison. Our system consists of five vowels and seventeen consonants. All but four phonemes /ʒ/ (spelled *y* and *ll*), /tʃ/ (*ch*), /x/ (*j*) and /ɲ/ (*ñ*) can be represent with orthographic symbols. In spoken Spanish the syllabic structure predominate over the morphemic one. It is also known that CV constitutes more than 50% of all syllabic types. The expansion of this pair produces CV+C, CV+V, CV+V+C which add over to 90% of all syllabic types /1/. It is also true that in Spanish there is a tendency for syllables between words to be fused. This and other phonological changes take place according to principles that could be determined. Taking this facts into account we centered our attention into the syllabic segments that are encountered at both margins of words. Before entering into this problem we examined a previous inventory of phonemes and classified the sounds adopting broad conventional articulatory terms /2/. A more detailed articulatory with allophonic variants is

presented elsewhere /3/. The corpus covered 163861 phonemes and 74460 syllables distributed in 43306 words. The text was extracted from five modern plays written in conversational style, by contemporary argentine authors.

DISTRIBUTION OF VOWELS AND CONSONANTS

In Figure 1 we present the five vowels ordered according to tongue height.

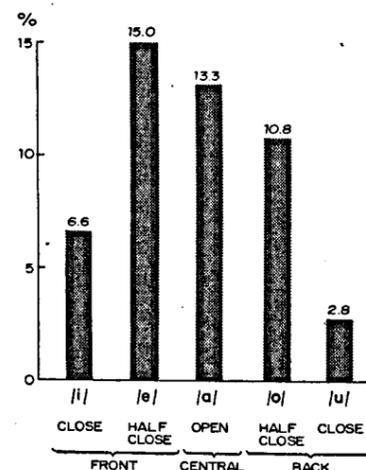


Figure 1. Distribution of Spanish vowels by articulatory configurations.

Taking as a referençe frontal half closed /e/, which is the most used sound, scoring 15%, central /a/ and back closed /o/ follow with relatively small differences. Closed vowels are less used. Frontal /i/ by a factor of two and half and back /u/ by a factor of five. In Figure 2 consonants are distributed according with place, voicing and manner of articulation. Dentals are produced approximately four time more than labials, six times more than velars and twenty six times more than palatals. Voiced phonemes are more frequent than unvoiced ones. As for manner of articulation nasal /n/ is more than twice more recurrent than labial /m/ while

palatal /ɲ/ is infrequently produced. Within the group of liquids vibrant /r/ is more recurrent than lateral /l/ but /rr/ is much less pronounced. The more recurrent stops are dentals /t/ and /d/, velar /k/, followed by labials /b/ and /p/. Velar /g/ is less used.

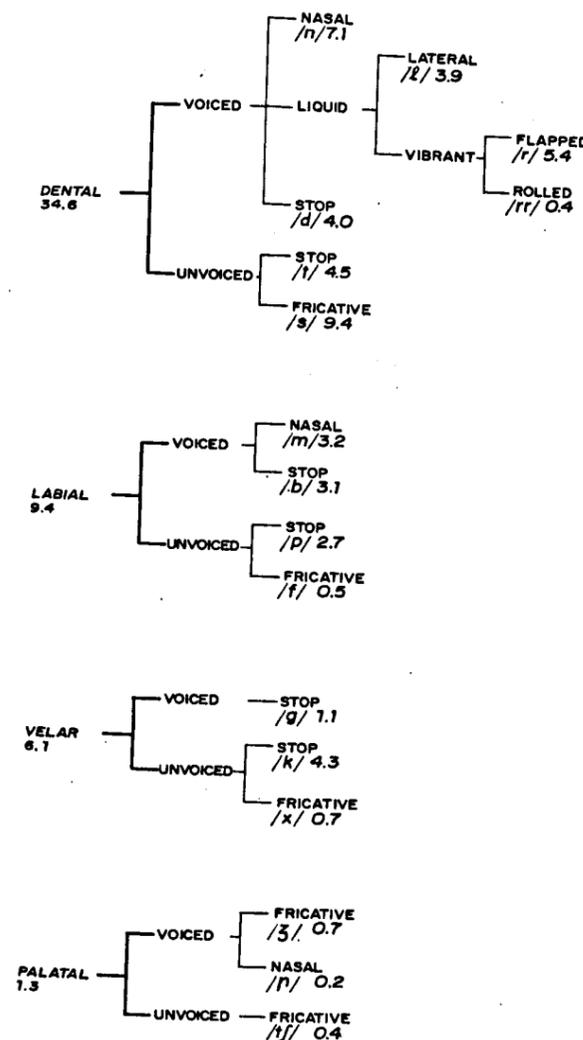


Figure 2. Distribution of Spanish consonants by articulatory configurations.

In general we observe that a large number of sounds are produced at the front of the mouth. Two frontal vowels plus dentals and labials summate over three thirds of all phonemes occurrences. Voicing is another apparent feature vowels and voiced consonants added together make more than three fourths of the total occurrences. A comparison with the Peninsular Spanish pronunciation gave minor deviations which are correlated with some differences in the phonological system /2/.

FREQUENTLY REPEATED SYLLABLES

It is noted that the high incidence of certain phonemes result from a group of frequently repeated syllables.

Table 1. The first fifty syllables (with asterisk are also words)

Syllables	Percent	Syllables	Percent
/a/*	5.529	/e1/*	1.050
/ke/*	3.920	/ro/	1.043
/no/*	3.826	/pe/	1.019
/de/*	2.484	/en/*	0.992
/se/*	2.460	/so/	0.957
/es/*	2.375	/le/*	0.870
/i/*	2.080	/por/*	0.864
/te/*	2.042	/di/*	0.835
/si/*	1.775	/mo/	0.819
/do/*	1.676	/u/*	0.768
/ta/	1.649	/ma/	0.742
/to/	1.598	/be/*	0.687
/la/*	1.518	/mi/*	0.679
/me/*	1.489	/ne/	0.640
/ra/	1.449	/bi/*	0.620
/sa/	1.407	/po/	0.608
/ko/	1.321	/ʒa/*	0.596
/na/	1.304	/ʒo/*	0.593
/pa/	1.286	/kon/*	0.589
/e/*	1.249	/go/	0.585
/ba/*	1.200	/un/*	0.549
/lo/*	1.193	/mos/	0.531
/da/*	1.136	/mas/*	0.529
/o/*	1.075	/des/	0.486
/ka/	1.073	/ga/	0.474

In Table 1 we present the first 50 items which represent 66.3% of the total sample but only 7% of the different syllables. Thirty seven are formed by a CV pair, five by CVC, five by V and four by VC. The most frequently pronounced consonants are dentals /d t s n l/ which summate 27%, labials /m p b/ 10% and velar /k/ 7%. These consonantal components are mainly combined with the three called strong vowels /e a o/. Thirty one of the units listed in Table 1 are monosyllabic words. These units appear with a relative much higher incidence 48.7% than bisyllabic 34.7%, trisyllabic 13.0% or tetrasyllabic words 3.1%. However when looking at the number of different words which in the total sample are 3993, we found that those formed by one syllable are only 3.3% while those of two are 34.8%, of three 41.2% and of four 17.1%.

SYLLABLES AT BOTH MARGINS OF WORDS

To count syllabic sounds located either at the onset or at the offset of words, we took again a sample formed by the first 50 syllables. Most of these items resulted from bisyllabic words. The data are displayed in Table 2 and 3. The units in Table 2 represent 74% of the total corpus and 1.7% of the different syllables initiating words. The range of scores is in the order of 1 to 20. To facilitate the inspection of our data we have divided the table in two sets of 25 syllables. The first from 9.8 to 1.0% and the second from 0.97 to 0.49%. In the first set we see that vowel alone or in VC pair are the most frequently produced sounds initiating syllables. Open vowel /a/ 9.8% is first. Vowel /e/ alone and combined with a fricative in /es/ and a nasal in /en/ summate 10.4%. Back closed vowel /u/ alone and /us/ make 3.8%. With lower scores isolated vowels /o/ and /i/ appeared also in the first set. The rest of the list is completed by three labials /p b m/, four dentals /t d s n/ and one velar /k/.

Table 2. Syllables in initial word position

CV: 35.838		CVV: 4.229
/pa/ 3.697	/rre/ 0.720	/bue/ 1.307
/ko/ 3.070	/rra/ 0.559	/tie/ 1.051
/pe/ 2.878	/mo/ 0.501	/pue/ 0.716
/to/ 2.143	/ti/ 0.492	/kie/ 0.635
/di/ 2.072	/ʒe/ 0.492	/bie/ 0.591
/ka/ 1.987		
/de/ 1.969	V: 18.566	CVC: 3.997
/sa/ 1.964		
/se/ 1.575	/a/ 9.802	/des/ 0.971
/po/ 1.315	/e/ 3.334	/kon/ 0.725
/na/ 1.293	/u/ 2.542	/tam/ 0.649
/ma/ 1.172	/o/ 1.602	/ten/ 0.604
/me/ 1.114	/i/ 1.284	/por/ 0.595
/te/ 1.033		/pen/ 0.452
/ba/ 1.002		
/mi/ 0.917	VC: 8.965	CCV: 1.597
/bi/ 0.890	/es/ 4.735	/kla/ 0.819
/ke/ 0.877	/en/ 1.378	/tra/ 0.778
/mu/ 0.872	/us/ 1.307	
/be/ 0.774	/al/ 0.868	CVVC: 0.693
/no/ 0.747	/an/ 0.675	/kuan/0.693

Among the labials stop /p/ articulated in pair with strong vowels constitute the favourite combination. Follows a voiced stop in /bue ba/ and a nasal in /ma me/. In the category of dentals stops /t d/ in /to te ti/, /di de/ and fricative /s/ in /sa se/ ranked first. Velar /k/ appeared relatively often

in /ko ka/. Subtotals are vowels 26.9, labials 12.2 dentals 12.1 and velars 5%. Most of the consonantal sounds listed in the first set appeared again in the second but in pair with vowels of lower incidence or in CVC or CCV context. The distribution of all items is quite even. This time initial vowels-with exception of /a/ and /an/-were not registered. Fricative /s/ was also missing but another dental in /rra rre/ and a palatal in /ʒe/ entered in the list. Subtotals are: vowel 1.5, labials 5.8, dentals 5.6 and velars 3.7%. Syllables in Table 3 represent 70% of the total sample and 1.8% of those terminating words. In this list scores are ranging from 5.3 to 0.49%. This proportion, close over 1 to 10, is practically half of the range observed in Table 2. In this table we observed a clear predominance of dentals. This category covers about two thirds of the total percentage listed in the table. The first eleven sounds in CV pair are dentals. Also, in CVC all but two, belong to this category. Stops /t/ and /d/ fricative /s/ nasal /n/ and vibrant /r/ combined mainly

Table 3. Syllables in final word position

CV: 55.064		
/do/ 5.308	/xo/ 0.863	/nas/ 0.559
/ra/ 4.435	/ga/ 0.738	/tas/ 0.550
/ta/ 4.122	/ke/ 0.734	/res/ 0.541
/ro/ 3.450	/ʒa/ 0.671	/nos/ 0.510
/to/ 2.989	/ma/ 0.671	/sɪr/ 0.510
/na/ 2.801	/po/ 0.514	/mas/ 0.492
/te/ 2.788	/xa/ 0.505	
/da/ 2.743	/tʃo/ 0.505	V: 3.083
/no/ 2.667		
/so/ 2.327	CVC: 11.144	/a/ 2.475
/se/ 2.193		/i/ 0.608
/mo/ 2.112	/mos/ 1.696	
/sa/ 2.027	/ted/ 0.944	CVVC: 1.145
/go/ 1.785	/nes/ 0.801	
/de/ 1.383	/ses/ 0.796	/pues/0.598
/ne/ 1.351	/tar/ 0.734	/bien/0.555
/ko/ 1.239	/des/ 0.626	
/ba/ 1.159	/dos/ 0.622	
/be/ 1.025	/ser/ 0.617	CCV: 0.496
/ka/ 0.989	/sas/ 0.577	
/si/ 0.957	/ron/ 0.564	/bre/ 0.496

with the strong vowels scored each an average close to 10%. Labials /m/ in /mo mos ma mas/ and /b/ in /ba be bien bre/ follow with 4.7 and 3% respectively. Among the velars stop /k/ and /g/ ranked in /ko ka ke/ 2.8% and in /go ga/ 2.4%. Fricative/x/ in /xa/ 1.3%. Palatals in /ʒa/ and /ʃo/

scored only 0.6 and 0.5% respectively. Two isolated vowels are found in this list /a/ 2.4% and /i/ 0.6%. In comparing Tables 1, 2 and 3 we observe that eighteen of the items are common to the three lists. Nine are words /a i ke de te se ba be no/. Forty percent of the items in Table 1 are not found either in Table 2 or 3. Among these there are six words /la lo el le o un/ which because of their grammatical function (articles and pronouns) frequently occur as isolated one syllable words. It was also noticed than when forming part of larger segments eleven words /es me e o en por di u mi bi kon/ appeared in initial position and five /si do a mas da/ were located at the end of words.

FINAL REMARKS

The range of frequently repeated syllabic sounds at both margins of words is twice as big for those initiating words. Combinations of CV type are almost equally distributed while CVC is more frequent terminating words. Types V, VC, CVV and CCV are mainly in initial position. As for distribution of articulatory combinations we differentiate four significant categories. Two are prevalent at the onset of words: a) Vowels alone and in VC pair /es/ and /en/; b) Labials /p b m/ plus strong vowels. The third category is formed by dentals combined with strong vowels which are limiting both sides being three times more frequent at the end of words. Velar /k/ is similarly distributed at both extremes. Monosyllabic words that are part of larger lexical units are more frequently occurring in initial than in final position.

REFERENCES

- /1/ M. Guirao, M.A. García Jurado, "Frequency of acoustical patterns in Spanish syllables". *Journal of the Acoustical Society of America* Suppl.1, vol.78:555, 1985.
- /2/ M. Guirao, M.A. García Jurado, "Frequency of occurrence of phonemes in American Spanish" (to be published).
- /3/ M. Guirao, M. Borzone de Manrique, "Fonemas, sílabas y palabras del español de Buenos Aires" *Filología* 16: 136-165, 1972.