Spontaneous speech corpus data validates prosodic constraints

Philippe Martin

UMR 7110, LLF, UFRL, Université Paris Diderot, ODG, Place Paul Ricœur, 75013 Paris, France
philippe.martin@linguist.univ-paris-diderot.fr

Abstract

In the Autosegmental-Metrical model, the prosodic structure is defined as a hierarchy of Accent Phrases (AP). Groups of AP form intermediate prosodic phrases ip, which in turn are grouped into Intonation Phrases IP, and finally sequences of IP form the sentence intonation unit. In this hierarchy several constraints affect the prosodic structure, such as the AP 7 syllables rule, the stress clash conditions, eurhythmicity and syntactic clash.

These constraints have been established essentially from read sentences data. They lead to an experimental justification in the observed synchronization of AP's syllabic chunking by Delta brain waves.

This paper investigates the validity of the prosodic structure constraints on spontaneous speech data in French, as well as the adequacy of the Delta waves characteristics to synchronize AP data.

Index Terms: prosodic structure, accent phrase, spontaneous speech, Delta waves, eurhythmy.

1. Introduction

In the classical Autosegmental-metrical approach, the prosodic structure is defined as a hierarchical organization of minimal prosodic units, the Accent Phrases (AP, aka prosodic words, rhythmic groups, etc.), a sequence of syllables which contain one lexical stress (or a metrically strong syllable). Groups of AP's form intermediate prosodic phrases ip, groups of ip form Intonation Phrases IP, and groups of IP form the complete sentence intonation. In this hierarchy, the whole sentence intonation can eventually be reduced to one single IP, which may contain a single ip, which itself may include a single AP.

Furthermore, the prosodic structure is constrained by a set of rules [8]:

a) AP 7 syllables rule;
b) Stress clash;
c) Eurhythmy;
d) Syntactic clash.

The maximum number of 7 syllables per AP was already mentioned by Meigret [12]. Stress clash pertains to avoidance of two consecutive stressed vowels in sentence realizations [12]. Eurhythmy ([3], [8], [14]) determines the tendency to either balance the number of syllables in successive IP's (or possibly ip's), or compensate the duration of enunciation of successive IP's containing an unbalance number of syllables. Finally, syntactic clash defines AP's allowed alignments with sequences of grammatical categories, which for instance cannot group a Verb followed by a determinant (something like *which for the* or *followed by a* in a single AP). In these examples, in the syntactic structure, syntactic units (words) are dominated immediately by nodes that group (i.e. are fathers) units which do not belong to the same AP.

Another characteristic pertains to the composition of AP, assumed to contain a single lexical word (Verb, Noun, Adjective or Adverb) possibly accompanied by grammatical words (Pronoun, Conjunction…) [2]. The validity of the rule will be evaluated as well.

2. Testing the hypotheses

The prosodic structure constraints originate mostly from observations pertaining to read sentences built by linguists. The goal of this paper is to evaluate the validity of these constraints for spontaneous speech, and also test a hypothetical cognitive explanation for each of the constraints.

Briefly stated, the cognitive hypothesis assumes that Delta brain waves are synchronized by stressed syllables (ending accent phrases in French) much as syllabic perception is synchronized by Theta waves [5],]6]. This synchronization would operate even if stressed syllables are not in final position.

Delta waves frequency varies from 1 Hz to 4 Hz, i.e. their periods vary from 250 ms to 1000 ms. This suggests that Delta waves are responsible for the conversion of sequences of syllables stored in short-time memory into a higher level linguistic unit, corresponding to AP's [6], [12]. This process timing is limited by the extreme values of Delta periods, whereas minimal period of Theta waves, which synchronize the perception of syllables, is about 100 ms (10 Hz).

The Delta wave hypothesis would be validated for constraints a) and b) if the observed AP longest and shortest duration would not exceed the Delta wave period values, whereas eurhythmicity would be explained if a compression effect would affect AP syllable duration, the shortest AP containing longest syllables, and the longest AP the shortest syllables.

Finally, the syntactic clash constraint could be validated by the total absence of realizations violating this alignment condition in the corpus. In addition, the validity of the AP composition with lexical words will be questioned, as occasionally some examples show AP containing only grammatical words.

These hypotheses are central in the prosodic incremental storage concatenation process proposed in [11]. In this model, acoustically and phonologically differentiated prosodic events trigger various transfer in the listener memory: a) transfer of syllables into another part of memory synchronized by Theta brain wave, b) transfer of syllabic chunks (correspond to AP's) into another short term memory synchronized by Delta waves, and c) concatenation of these sequences of partial processing into an interpretation module synchronized by differentiated prosodic events (various melodic contours).

3. Data analysis

To test the prosodic structure constraints, French data were selected as the absence of lexical stress in French may a priori

10.21437/SpeechProsody.2014-92

lead to more variations in AP number of syllables, as one single Accent Phrase can contain more than one lexical word, as in *le bilan des ventes* "the stock sales" pronounced rapidly, with two nouns *bilan* and *ventes* belonging to the same AP.

Analyzed data were taken from the C-PROM corpus [16]. C-PROM is a transcribed, aligned and annotated corpus, developed among other applications, to evaluate syllabic prominences in French. It includes 24 recordings belonging to 6 different speech styles of francophone speakers originated from Belgium, France and Switzerland. Only French speakers were retained in this study. Details and transcription formats can be viewed on line.

Primary data were then transferred to WinPitch [17], whose routines allow a direct analysis into an Excel sheet of results, giving automatically in one single mouse click (Fig. 1 and Fig. 2):

a) The number of syllables in each AP;
b) The overall AP duration in ms;
c) The average syllabic duration in a given AP;
d) The AP stressed vowel duration;
e) The API vowel transcription;
f) The AP transcription in API;
g) The AP orthographic transcription;
h) Time references of the events.

	A	B	C	D	E	F	G	H	I	J
1	Nb Syl	Duration	Dur / Syl	Vowel Dur	Vowel	Center	Syllables	Text	Start	End
2										
3										
4	4	1193	298	225	y	1.972	sɛtaptity*	cette aptitude	1.607	2.047
5	4	897	224	39	a	4.372	tẽkõmɛmpa*	incon- même pas	4.294	4.385
6	6	731	121	54	e	5.103	ãvidpɔlemike*	envie de polémiquer	4.974	5.121
7	6	857	142	75	ã	5.96	paskəʃyzokuʁã*	parce que suis au courant	5.858	5.985
8	4	797	199	192	u	6.757	mẽtnãdətu*	maintenant de tout	6.587	6.821
9	7	1058	151	155	ã	7.815	safɛkãɛmvẽsẽkãk*	ça fait quand même vingt cinq ans que	7.622	7.909
10	6	896	149	76	ê	8.711	ʃəkɔnɛasebjê*	je connais assez bien	8.589	8.737
11	4	575	143	104	ê	9.286	sɛtekʁivê*	cet écrivain #	9.19	9.321
12	6	1238	206	75	ã	11.737	tẽkõtɛstabləmã*	incontestablement	11.63	11.762
13	5	846	169	89	ɛ	12.583	ẽtʁɛgʁãpɔɛt*	un très grand poète	12.524	12.738
14	3	719	239	98	u	13.302	avãtu*	avant tout #	13.077	13.335
15	3	467	155	128	ã	15.893	jaləskã*	il|y a le	15.72	15.936
16	4	655	163	84	ɛ	16.548	dalləpɹɔblɛm*	le problème #	16.421	16.683
17	5	794	158	106	u	18.485	dəplyzãplyluʁ*	de plus en plus lourd	18.304	18.61
18	4	700	175	72	ã	19.185	pɥiskəmẽtnã*	puisque maintenant #	19.096	19.209
19	3	897	299	187	õ	20.93	sapasjõ*	sa passion	20.484	20.993
20	3	402	134	69	ɛ	22.282	ʒãdanjɛl*	Jean Daniel	22.153	22.396
21	3	381	127	100	ɛ	22.663	mədizɛ*	me disait	22.536	22.697
22	4	859	214	272	ã	23.522	ynʃɔzmaʁã*	une chose marrante	23.257	23.613
23	2	259	129	35	y	25.001	sɔkyp*	occupe	24.96	25.054

Figure 2. Example of Excel sheet of primary results (*nar-fr speaker*)

The speech styles of the corpus are:

lec-fr: oral reading;
cnf-fr: university conference;
nar-fr: narrative, life story;
pol-fr: political discourse;
jpa-fr: radio news.

The iti-fr itinerary style of the corpus was not retained as recordings were considered too short, whereas the lec-fr recording are kept for comparison with the non-read styles.

Since French has no lexical stress, only boundary tones (in AM terminology) are observed. Their effective realization is sometimes difficult to establish as linked to the effective stressed characteristics of given syllables [10], but the error rate can be estimated at less than 5 %.

The original labelling of stressed syllable into two degrees of stress, noted p and P, were carefully revised. As few occurrences were found questionable, informal perception tests were conducted for possible corrections and adjustments.

An example of AP stress syllable revision is given below.

In cnf-fr, a segment was originally transcribed as:
de deux cent soixante-cinq phrases, dədøsãswasãtsœ̃fʁaz*
with seven syllables pronounced in 1491 ms. But listening more carefully two AP were actually realized :
De deux cent dədøsã* and *soixante-cing phrases* swasãtsœ̃fʁaz (« of two hundred sixty five sentences") with two accent phrases, of respectively 3 and 4 syllables.

The AP's duration are taken from the right boundary of the syllable vowel to the next stressed vowel right boundary. When the stressed syllable is preceded by a pause, the end of the pause is retained as the starting time reference to measure the duration of the current AP ended by the next stressed syllable.

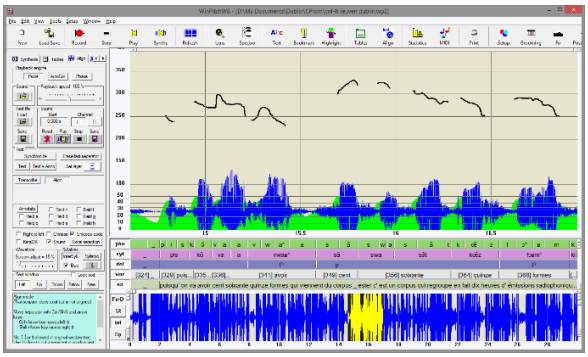


Figure 1. Example of WinPitch display [17]. The second transcription tier displays the syllables in API, with perceived prominence indicated by a star (*cnf-fr speaker*)

4. Results

Table 1 gives the following results pertaining to the hypotheses to be tested:

a) Longest AP duration;
b) Shortest AP duration;

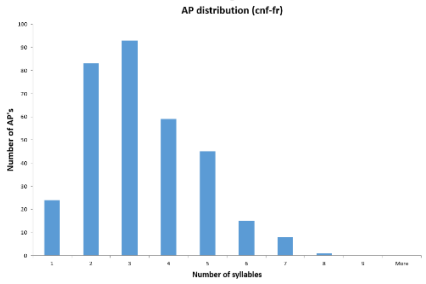
c) Number of stress clash violation.

These values were computed for each recording styles individually, in order to evaluate the possible influence of speech styles on the results.

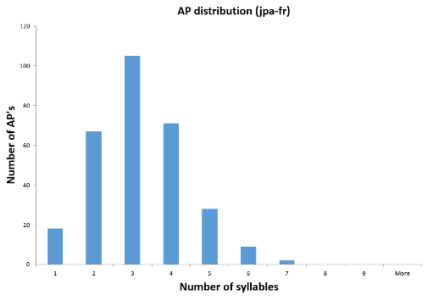
Table 1. Longest and shortest AP duration (in ms) and AP duration vs. number of syntactic clash violation for five speech styles.

Duration	Lec	Cnf	Nar	Pol	Jpa
AP Max	1260	1134	1058	975	1258
AP Min	435	277	354	438	241
Synclash	0	0	0	0	0

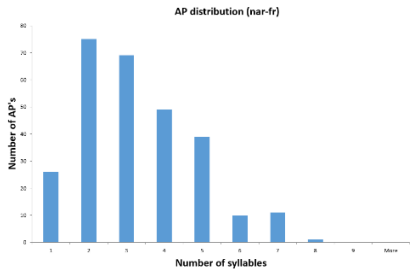
Histograms giving the distribution of the number of syllables per accent phrase are given Fig. 3. They are very similar for the five styles retained.



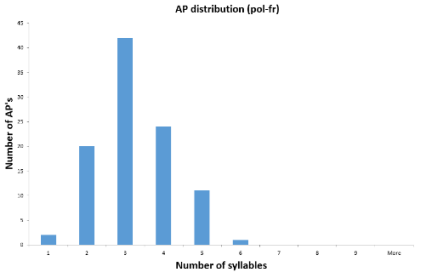
Distribution of cnf-fr number of syllables in AP's



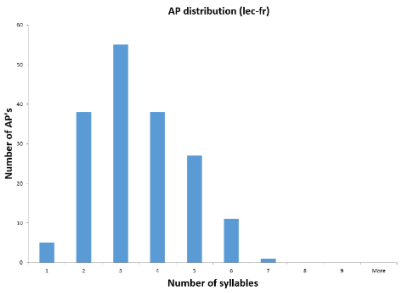
Distribution of jpa-fr number of syllables in AP's



Distribution of nar-fr number of syllables in AP's



Distribution of cpol-fr number of syllables in AP's



Distribution of lec-fr number of syllables in AP's

Fig. 3. Distribution of the number of syllables for the C-PROM styles retained

The only noticeable difference pertains to pol-fr (political speech), which uses a more restrained distribution of AP number of syllables similar to lec-fr, i.e. 3-4 vs 1-7 or 1-8 for the other styles. This suggests that pol-fr style was, at least partially, read speech.

Fig. 4 and Fig. 5 give the corresponding distributions of shortest and longest AP duration in function of the five styles considered.

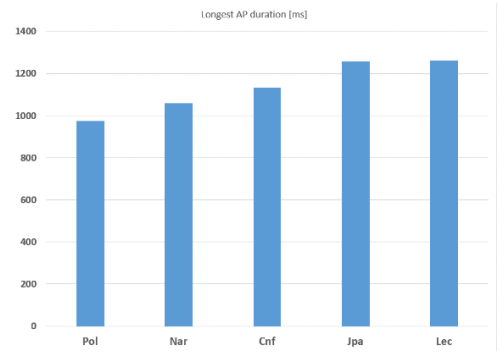


Fig. 4. Longest AP duration in ms

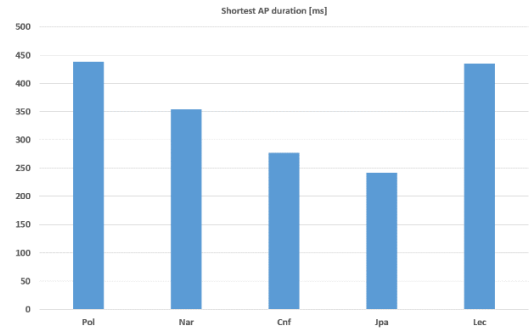


Fig. 5. Shortest AP duration in ms

These distributions show that the spoken news style uses the shortest AP's, whereas the reading and political recordings favor longer minimal AP duration.

The regression line of Fig. 6 demonstrate the compression of AP's duration in function of their number of syllables. When this number increases, the average duration of syllables is reduced allowing a single AP to contain up to about 7 syllables.

The regression lines of the other styles are not shown, as being similar to the one presented (cnf-fr).

This table shows clearly that the hypothesis pertaining to the AP content is invalidated. Not only can an AP contain more than one open class word in French, but spontaneous speech data include a relatively large number of occurrences of AP's with only grammatical words.

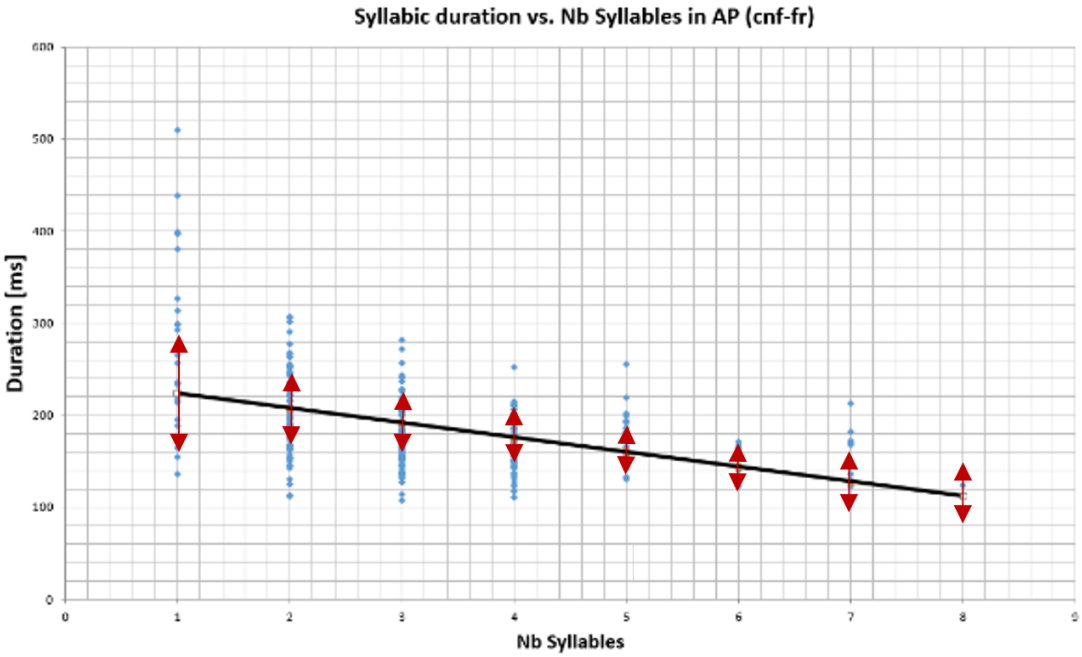


Figure 6. Syllabic duration in function of the number of syllables in AP regression line (*nar-fr*). Standard deviation is indicated by double arrows.

These regression lines correspond to the results given in [12] and [1], the longest the AP duration, the shortest are the syllables they contain.

5. Interpretation of results

1.1 Maximum and minimum values of AP's duration

Table 1 and the following figures suggest that Delta brain waves synchronization hypothesis is validated, as the minimum and maximum values do correspond convincingly to the maximal and minimal values of AP duration values.

1.2 Eurhythmy

Fig. 6 shows that eurhythmy is obtained by compression of the syllabic duration for long AP's.

1.3 AP's grammatical words

Table 2. Number of AP's containing only closed class words (Conjunctions, determinants, etc.) over the total number of AP's in each recording.

	Lec	Cnf	Nar	Pol	Jpa
Cases	**4/175**	**18/223**	**2/290**	**0/100**	**11/300**
	2%	**8%**	**0.6%**	**0%**	**3.6%**

6. Conclusion

The analyzed data on various styles of spontaneous speech data validate the proposed explanation for prosodic structure constrains, namely:

a) The maximum number of syllables in a given AP is indeed of the order of 7 to 8, but the actual limit is given by the largest possible Delta wave period, about 1200 ms. Examples of AP containing up to 12 syllables in are found for example in [7], but even in this fast speech rate case their duration is below the Delta brain wave limit of 1200 ms;

b) Successive stressed syllables are found ("stress clash", corresponding in French to one single syllable AP following the first AP) but there is a minimal amount of time between two consecutive stressed syllables (actually between two consecutive stressed vowels). This observation confirms the hypothesis about Delta brain waves synchronizing the perception of AP, in the case at a maximum frequency, i.e. a minimal period of about 250 ms.

c) Cases where eurhythmy is obtained at the expense of congruence of the prosodic structure with syntax are rare so the eurhythmic compensation is done by compressing the syllabic duration in AP with many vowels. This was already observed empirically in [4], [8], [15] and more recently in [1]. One of the reason why balancing of the number of syllables is not frequent in spontaneous data may pertain to the fact that such balancing requires preplanning essentially possible for read speech (cf. the

read phrasing [*Marie adore*] [*les chocolats*] vs. the spontaneous [*Marie*] [*adore les chocolats*]). It seems that speakers realize eurhythmic phrasing when the syntactic constraint is weak or absent, i.e. for enumeration, short read sentences, etc.

d) No cases of syntactic clash were observed;

e) However, occurrences of AP containing no lexical words and only grammatical word are not infrequent.

The next step in this research would concern other Romance languages with lexical stress, and later tone languages such as Mandarin with no lexical stress.

Romance languages other than French may show the coexistence of a lexical stress and a tone boundary sometimes combines on the same AP final syllable (in Italian for example). These two prosodic events may play the same role (or complement each other) in the storage concatenation process proposed by [11].

7. References

[1] Avanzi, Mathieu, Lucie Rousier-Vercruyssen et al. (2013) C-PROM-Task. A New Annotated Dataset for the Study of French Speech Prosody, Proceedings TRASP 2013, Aix-en-Provence, 27-30.

[2] Beckman, Mary E. & Janet B. Pierrehumbert (1986) Intonational structure in Japanese and English, Phonology Yearbook 3, 255-309.

[3] Dell, François (1984) L'accentuation dans les phrases en français, in Dell F., Hirst D. & Vergnaud J.R. (éds), *Formes sonores du langage*, Hermann, Paris, 65-122.

[4] Fónagy, Ivan & Magdics, Klara (1960) Speed of utterances in phrases of different lengths, Language and Speech, 3, 179-192.

[5] Friederici, Angela & Wartenburger, Isabell (2010) Language and brain, *Cognitive Science*, (10) 150-159.

[6] Ghitza1, Oded, Giraud, Anne-Lise and Poeppel, David (2013) Neuronal oscillations and speech perception: critical-band temporal envelopes are the essence, Frontiers in Human Neuroscience, www.frontiersin.org, January2013, Volume 6, Article340.

[7] Lehka Irina. & David Le Gac (2004) Etude d'un marqueur prosodique de l'accent de banlieue, *Actes des XXIIIème Journées d'Etudes sur la Parole*, avril 2004, Fèz, Maroc

[8] Malécot, André, Johnston, R., & Kizzlar, P. A. (1972) Syllabic rate and utterance length in French. Phonetica, 26, 235-251.

[9] Martin, Philippe (1986) Structures prosodiques et structures rythmiques, Actes des 13ème JEP, Aix-en-Provence, l986.

[10] Martin, Philippe (2005) La transcription des proéminences accentuelles : mission impossible ? *Revue PFC*, septembre 2005.

[11] Martin, Philippe (2009) *Intonation du français*, Armand Colin, Paris, 256 p.

[12] Martin, Philippe (2013) Contraintes phonologiques de l'intonation de la phrase réinterprétées à la lumière des recherches récentes en neurophysiologie, *La Linguistique*, 2013/1.

[13] Meigret, Louis (1550) *Le treté de grammere francoeze*, Réimpression chez Slatkine, Genève, 1972.

[14] Pasdeloup, Valérie (2004) Le rythme n'est pas élastique : étude préliminaire de l'influence du débit de parole sur la structuration temporelle, Actes des JEP 2004, Fès (Maroc) - 19-22 avril 2004.

[15] Wioland, François (1984) Organisation temporelle des structures rythmiques du français parlé, *Bulletin de Linguistique de Lausanne*, 6, 293-322.

[16] C-PROM (2010) Corpus libre de parole multigenre, https://sites.google.com/site/corpusprom/

[17] WinPitch, www.winpitch.com