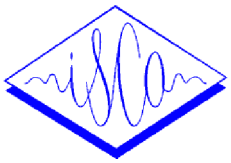

Speech Prosody 2010
Chicago, IL, USA
May 10-14, 2010

Prominence perception and accent detection in French. A corpus-based account

Jean-Philippe Goldman[1,2], *Antoine Auchlin*[1], *Sophie Roekhaut*[3,4], *Anne Catherine Simon*[2], *Mathieu Avanzi*[5,6]

[1]Département de Linguistique, Université de Genève, Suisse,
[2]Institut Langage & Communication, Université catholique de Louvain, Belgique,
[3]CENTAL, Université catholique de Louvain, Belgique
[4]TCTS Lab, Université de Mons - UMONS, Belgique;
[5]Chaire de linguistique française, Université de Neuchâtel, Suisse,
[6]MoDyCo, Université Paris Ouest Nanterre, France.

jeanphilippegoldman@gmail.com, antoine.auchlin@unige.ch, Sophie.Roekhaut@umons.ac.be, anne-catherine.simon@uclouvain.be, mathieu.avanzi@unine.ch

Abstract

The goal of this paper is to shed new light on accentuation in French, more precisely to discuss the role of grammatical constraints and of phonetic factors implicated in the perception of French final and non-final accent. The study is based on the analysis of a 70-minute long corpus, including various speaking styles. The corpus has been annotated manually and automatically for prominence detection and tagged semi-automatically for grammatical categories. We first describe the rate of accentuation for each grammatical category (discussing the notion of "clitic" in French) and then discuss the divergences between manual and automatic prominence detection, in relation with the phonological structure.

Index Terms: prominence detection, French accentuation, clitics.

1. Introduction

Specialists agree on two types of stress in French: an obligatory primary (or final) stress, which falls on the last syllable of a prosodic group (composed of a full word and its most-left adjacent clitics), and an optional secondary (or non-final) stress, which can be on any other syllable of that prosodic group. Classic features such as grammatical category, morpho-syntactic grouping and metrical constraints are known to be influential parameters involved in the realization of these two kinds of stresses in French [1][2][3][4][5][6]. But it has also been demonstrated that external factors such as speaking styles interfered in the realization of final and non-final accents. For example a fast speech rate (like in spontaneous conversations) involves the realization of larger prosodic groups – *i.e.* of less primary accents – than in read-aloud texts, while typical professional style, like news broadcasts or radio interviews, are characterized by a high frequency of non-final accents.

Yet, as far as we know, the question of whether those factors interact and create possible divergences between acoustically measured prominences and auditory perceived accents has never been addressed. Scholars generally assume that acoustic measurements give independent evidence to support the auditory judgments and report high reliability in the establishment of the location of accents [7]. In this paper, we would like to discuss this state of affairs. To this end we propose to pay special attention to those cases where human stress perception does not coincide with automatic acoustic detection in order to bring new evidence for the factual significance of each of the features involved in stress perception (acoustic prominence depending on F0 and duration, grammatical category and word phrasing). The paper is organized as follows. After having presented the corpus (recordings, protocol of annotation) and the tools used to handle it semi-automatically (§2), we give a quick overview of the percentages of primary and secondary stresses according to the words' grammatical category (§3). The following section is devoted to the discussion of the prediction rules for stress assignment, and the effective accentuation of words and adjectives (§4). The last section before the conclusion (§5) proposes a typology and explains blends, viewed here as a subpart of the mismatches between manual and automatic annotation.

2. Material in the database

Our study is based on C-PROM, a multi-level annotated corpus comprising different speaking styles and different regional varieties of spoken French. The corpus is 70 minutes long, and comprises 24 samples from 7 different speaking styles (going from very formal speech – read-aloud texts, political discourses – to less formal conversations, such as map tasks or spontaneous monologues), amounting to 10,477 words (see [8] for more details on the corpus constitution). The entire corpus was annotated with prosodic and grammatical tags, so that information concerning the "stressability" in regard to French accentuation rules could be retrieved for each syllable.

2.1. Prosodic annotations

Sound files were first semi-automatically aligned into phones, syllables and orthographic words within the Easyalign script [9], working under the Praat software [10] (see Table 1). Next, a manual annotation of syllabic prominence was carried out by two transcribers (two of the authors, see [8]). At the same time, specific labels were used to single out those typical syllables found in unprepared speech (interruptions, hesitations, cough, overlap, etc.) and exclude them, so that they would not interfere with the automatic extraction of different acoustic features (including syllable duration, F0 and silent pauses).

2.2. Grammatical annotations

About 30 grammatical categories were annotated automatically and checked manually. Table 1 gives an overview of the number of tokens by category; the smallest categories (like acronyms, discourse particles, etc.) have been excluded.

Table 1. Number of tokens by grammatical categories in C-PROM

Categories	Subcategories	Tokens
NOUN	nouns (1965) and proper names(339)	2304
DET	determiners (definite (800), indef. (412), interrogative (4), multiple words (13), prepositional (262))	1491
PRON	pronouns (includ. 12 different classes)	925
VERB	verbs (701 finite verbs, 304 participles, 313 infinitives)	1318
ADV	adverbs of manner (601), degree (143), negation (116), comparison (43) and interrogation (17)	920
PREP	prepositions	959
ADJ	adjectives	616
CONJ	coordination (371) and subordination (136) conjunctions	507
AUX	verbal auxiliaries (220) and predicative use of "être" (261)	481
NUM	numerals	89

2.3. Summary and description of the database

Scripts were developed in order to retrieve information from the annotation files. Each syllable was described according to the following parameters, all relevant for studying accentuation in French:

- Prominent or not prominent syllable;
- Position of the syllable within the word (final, initial…)
- Position of the word within the chunk (a chunk minimally has a HEAD which is most often a noun or a verb; dependent elements, like adjectives, determiners, conjunctions, etc., have a PRE or POST position depending on their location *vis-à-vis* the HEAD)
- Acoustic description for each syllable (like duration, F0 mean, etc. and measures within the syllabic context).

A database of syllables has been containing quantitative data about the degree of accentuation of certain words or syllables in certain positions or grammatical categories.

3. Accentuation by grammatical category

For each category containing a minimal amount of 150 tokens in our database, we describe the percentage of final accented syllables and non-final accented syllables.

Table 2 shows an interesting and unexpected gradual difference between manual and automatic prominence detection. Grossly speaking, human annotators detect more final prominences than automatic annotation for "lexical" categories (from line 1 "Nouns" to line 7 "Finite Verbs"). The divergence is reversed for "grammatical" categories (from line 8 "Coordinating conjunction" to line 14 "Definite Determiners").

As for non-final accent, results in Table 2 show the same tendency, but the difference does not affect the same categories: human (manual) prominence detection exceeds the automatic one only for Nouns, Proper Names and Adjectives. We further discuss the case of Determiners in Section 5.

Evidence seems to show that human - as compared to automatic acoustic - detection *over-detects* final prominence on lexical categories such as Nouns, Proper Names, etc., and *under-detects* both final prominence on grammatical categories, and non-final prominence on categories such as verbs, either tensed or non-tensed, on adverbs of manner, prepositions, and definite determiners.

Table 2. Percentage of final and non-final accents, according to manual (manu) or automatic (auto) detection, with number of syllables, and words concerned

	N		Final Accents		Non-final Accents	
	syll	w	manu	*auto*	manu	*auto*
Proper Names	777	339	**71.98**	*55.46*	13.93	*13.93*
Nouns	4134	1965	**68.19**	*54.5*	11.66	*12.54*
Adjectives	1432	616	**63.8**	*48.86*	13.85	*12.38*
Infinitives	713	313	**57.19**	*45.37*	10.75	*12.5*
Adv of manner	1108	601	**55.24**	*44.59*	8.28	*12.43*
Past Participle	621	281	**50.18**	*39.86*	6.47	*6.18*
Verbs	1213	701	**37.95**	*31.53*	10.16	*12.11*
Coord. Conj.	375	371	**15.9**	*16.98*	25	*25*
Pred. *'être'*	327	261	**13.41**	*16.86*	9.09	*3.03*
Pers Pronoun	290	290	**8.62**	*10.69*	NA	*NA*
Indefinite Det	464	412	**8.01**	*9.71*	13.46	*23.08*
Relative Pron.	163	156	**7.69**	*8.33*	0	*0*
Preposition	1160	959	**6.05**	*9.49*	14.93	*15.42*
Auxiliary	261	220	**5.91**	*7.27*	2.44	*0*
Determiner	814	800	**4.38**	*6*	0	*0*
Subj Pers Pron	301	301	**2.99**	*8.97*	NA	*NA*
Prep Determ.	262	262	**1.53**	*3.44*	NA	*NA*

Both automatic and human (methodologically controlled) detection are reliable. Even if automatic detection could hypothetically be improved and obtain slightly better agreement scores, our results in Table 2 show that there is more than acoustics involved in human prominence perception. We call this phenomenon "auditory illusion" and we explain that it is linguistically based.

We hypothesize that this is a case of binding. Binding, as explained by [11], corresponds to a first-level conceptual blend in [12] general framework. Binding is defined as the process by which perception compresses information from distinct input spaces into a single, emergent space, i.e. a kind of "improved" or "increased" perception. For example, very distinct neural subsystems operating in parallel ways are implicated when one sees a red ball rolling. Perception, then, is the process involved in our mind's compressing those inputs into a unified perception of a rolling red ball.

Human prominence detection binds information from – at least – two distinct input spaces: (i) the linguistic input space (lexical, grammatical, as well as semantic information), and (ii) the distinct acoustic input subspaces, namely duration properties, F0 proper and relative properties, and F1 to Fn formant characteristics (phonological and non-phonological information). For "full word" categories, this convergence of information would lead to an "end-of-the-word" prominence illusion. Other dimensions of linguistic structuring are implicated in similar blending cases (see Section 5).

4. Accentuation and non-clitic categories

4.1. Nouns and Verbs

Nouns and verbs are classically described as bearing a final, primary accent in French, except when they are followed by a monosyllabic complement (e.g. *prends-le*, with the accent on the "le" pronoun, see the well-known accentual report rule [1][4][7]).

We retrieved all instances of Nouns, Finite Verbs and Infinitives occupying a "head" position within a chunk. Table 3 displays the frequency of accentuation for each category, by distinguishing between monosyllables and polysyllables.

Table 3. Percentage of final, initial and medial prominences (accents) on mono- and polysyllables, as detected manually and automatically.

		Noun	Finite Verb	Infinitive
Final Accent monosyl.	manu	73.4	37.2	63.4
	auto	*62.1*	*35.5*	*56.1*
Final Accent polysyl.	manu	66.2	46.3	57.7
	auto	*51.3*	*36.4*	*43.4*
Initial accent	manu	12.9	12.7	14.5
	auto	*11.9*	*14.8*	*14.5*
Medial accent	manu	1.4	0.09	1.5
	auto	*7.1*	*5.8*	*34.4*

The tendency for human annotators to detect more final accents at final word boundaries has been described in Section 3 as an "end-of-word illusion". The "beginning-of-word" illusion does not seem to be supported by the data (no difference between automatic and manual detection of initial accents), but there is a strong effect preventing humans from hearing an accent on the medial (neither initial nor final) syllable of a word (71 prominences detected automatically against only 15 detected perceptually, among which 7 in common). The same effect applies to all categories, with even more strength on Infinitives (out of 10 medial syllables automatically detected as prominent, only 1 was detected manually).

Nouns are nevertheless characterized by a high rate of final accentuation (51.3% for polysyllables and 62.1% for monosyllables, in auto. annotation), confirming their non clitic nature. The higher score for final accented monosyllables is due to the fact that all prominences are considered final there.

Finite verb accentuation on final syllable amounts to about 35% (although the figure is slightly higher in human perception: 37-46%). One explanation for this fairly low rate is that Verbs hardly ever occupy the last position in the verbal clause, which is frequently the case for infinitives (with a 43.4 to 56.1% of final accentuation).

Even if grammatical category is an important clue for predicting the realization of final accentuation, the number of syllables in the word as well as its position in the clause appear to be of great importance too.

4.2. Adjectives

Adjectives constitute an interesting category: since they form a lexical category, they theoretically bear a final accent. In practice this accentual schema can be modified when the adjective belongs to the same Phonological Phrase (PP) as the Noun it complements [7], in a rephrasing process (at a post-lexical level), such as illustrated in the following examples:

- $[\text{le doux}]_{PP}\ [\text{nom}]_{PP} \rightarrow [\text{le doux nom}]_{PP}$
- $[\text{un long}]_{PP}\ [\text{poème}]_{PP} \rightarrow [\text{un long poème}]_{PP}$

Table 4 Percentage and occurrences of final and initial accent, by syllabic position in Adjectives

	Position (chunk)	Prominence on syllable	Percentages and tokens
Monosyllables	PRE	--	35% (28/80)
Polysyllables	PRE	initial	32.9% (25/76)
	HEAD + POST	final	84.9% (51+237)/(57+282)

Our hypothesis is that initial syllables of polysyllable adjectives as well as monosyllable adjectives PREceding the noun within a clause will have a secondary, non-final accent, instead of a primary, final accent. Non-final accent, traditionally found on the initial syllable in a polysyllable word, creates a kind of "hammock pattern" (*arc accentuel*).

Consequently, we think that the acoustic correlates of prominence of those 2 types (lines 1 and 2 in Table 4) will diverge from the prominence on final syllables of adjectives that are HEAD of a clause, or after the noun (POST)(line 3).

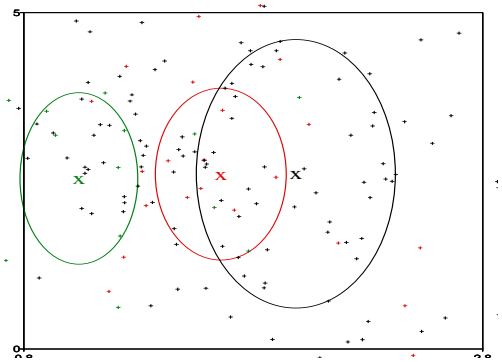

Figure 1. Relative syllable duration vs. rel. F0 of prominent syllables of adjectives as in Table 4. Ellipses are 1 std. dev. away from mean.

As expected, Figure 1 shows the 3 types of syllables mentioned in Table 4 distinguished by relative duration rather than by relative F0. All the differences are very significant at $p<0.01$, except between HEAD+POST final syllables and PRE monosyllabic ($p<0.02$). The latter category may comprise initial and final accents despite the PRE position in a clause.

5. Perceptually blended accents

From a phonological perspective, lexical categories tend to have a final accent, may lead to "end-of-word" (accentual group) illusion (see Section 3). We now present results regarding a subset of grammatical categories, traditionally considered as clitics, and thus unstressed. Here again, the two kinds of disagreement between automatic and human prominence detection (looking for human under- and over-detection) may be accounted for using a blending/binding framework. We distinguish three different cases, into which most if not all mismatches fall.

- **Clitic negative illusion** concerns determiners and other grammatical categories [13] showing a general tendency for human under-detection. Indefinite determiners (like *un, une, des*, etc.) and multiple-word determiners (like *davantage de, plus de, plein de*, etc.) demonstrate an even stronger illusion (Table 5).

Out of 800 determiners, 53 were coded differently by the machine and by the human annotators. Of these 53, 39 are detected as prominent by the automatic procedure only (and not by the human). They illustrate the case of negative clitic illusion: although they are acoustically salient, humans do not match the acoustic prominence to the realization of an accent.

Table 5. Rate of accentuation on determiners, comparing manual and automatic detection of accent.

	Accented syllables	
	manu	auto
Definite determiner (n=800)	4.38	6
Indefinite determiner (n=412)	8.01	13.46
Multiple-words determiner (n=13)	35.7	21.4

As far as multiple-word determiners are concerned, manual and automatic prominence detection highly diverge. Out of 27 syllables, only 2 were detected as accented both by manual and automatic annotation. This can be explained by another perceptive illusion.

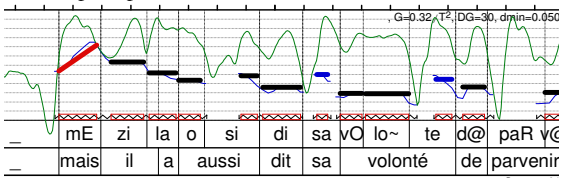

Figure 2. Prosogram of example *mais il a aussi dit sa volonté de parvenir* ("but he also proclaimed his wish to come to"). Prominence on determiner *sa* (his) has been detected by machine only.

- **Positive semantic quantity illusion** concerns multiple word determiners expressing "lots of" (*plein de*) (Figure 3). This is a marginal case in our data (only 5 tokens of *plein de*) but it uncovers what seems at work in the prominence/accentuation articulation.

Four out of five tokens of *plein de* were detected as prominent by human annotators and none of them by the automatic detection. Considering the acoustic parameters of those occurrences, we reach the conclusion that they do not stand out against their local context. Only the semantic strength of the word "lots of" contributes to their perception as prominent (see Figure 3).

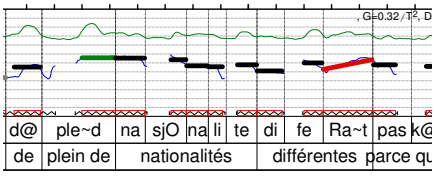

Figure 3. Prosogram of example *de plein de nationalités différentes* (lots of different nationalities), with the word *plein* being perceived as accented by the human annotators.

The last case of illusion tries to account for the opposite case of "negative clitic illusion": out of 53 disagreements between automatic and manual annotation, 14 concern determiners perceived by humans only as accented.

- **Positive constructional hammock-pattern illusion** concerns human-only prominent determiners seemingly opening a complex semantic construction.

Manually detected prominence acts as the first arch of a bridge over the construction whose second arch is the next prominence at the end of a word. Fig. 4 shows a human-machine disagreement as to this second arch's position (adj. *mineurs*): the human detects it earlier than the machine does – yet they agree on the construction's end.

Such positive initial accent illusion may have several explanations. The first one would come from the human detection procedure (listening three times to a 3-4 second sound segment) [8] – that would explain some *a posteriori* binding effect. A second one could come from intrinsic and relational syllable properties that are too small to be considered by our detection algorithm (such as voicing onset or voice quality) and which could be perceived as an initial boundary.

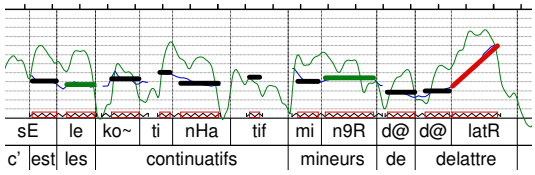

Figure 4. *puisque c'est les continuatifs mineurs de Delattre* (because it's THE minor continuative in Delattre's terminology)

6. Conclusions

This paper explores a large database containing about 12,000 words and more than 18,000 syllables with grammatical and acoustic annotations. Two main conclusions deserve to be recalled.

First, we describe the effective accentuation of a wide range of grammatical categories. The notion of "clitic" is empirically sustained as being a gradual one (grammatical categories can be attributed a "gradient of cliticity" according to their effective tendency to be stressed).

Most interestingly, we systematically compared perceptual and acoustical detection of prominence, in order to demonstrate that perception of prominence, and therefore of accent, is biased by expectations based on grammar or meaning formation. Both *over-* and *under*-perception can be accounted for as cases of *binding* perception to linguistic, lexical, syntactic and semantic knowledge.

7. Acknowledgements

This research was supported by grant no. 0616422 from the Walloon Region, Belgium (EXPRESSIVE). Mathieu Avanzi is grateful for financial support from the Swiss National Science Foundation under grants n° PBNEP1-127788 and n°100012-113726/1).

8. References

[1] Garde, P. L'accent, Paris, PUF, 1968.
[2] Martin, P. "Prosodic and rhythmic structure in French", Linguistics, 5/5, 925-949, 1987.
[3] Delais-Roussarie, E. "Phonological Phrasing and Accentuation in French", *in* M. Nespor & N. Smith Eds, Dam Phonology:HIL Phonology, Holland Academic Graphics, La Haye, 1-38, 1996.
[4] Lacheret, A. et al. La prosodie du français, Paris, CNRS, 1999
[5] Jun, S.A. and Fougeron, C. "Realizations of accentual phrases in French intonation", Probus, 14:147-172, 2002.
[6] Welby, P. "French intonational structure: Evidence from tonal alignment", Journal of Phonetics, 34/3, 343–371, 2006.
[7] Post, B. "French phrasing and accentuation in different speaking styles", Oxford University Working Papers in Linguistics, Philology and Phonetics, 8:69-83, 2003.
[8] Avanzi, M. Goldman, J.-P. & A.C. Simon, "C-PROM. An Annotated Corpus for French Prominence Studies", Prosodic Prominence: Perceptual and Automatic Identification (Speech Prosody 2010 workshop), Chicago, USA, 2010.
[9] Goldman, J.-P. "EasyAlign:a semiautomatic phonetic alignment tool under Praat", http://latlcui.unige.ch/phonetique, 2008
[10] Boersma, P. & D. Weenink, Praat: doing phonetics by computer (Version 5.1). www.praat.org, 2009.
[11] Bache C. 2005, "Constraining conceptual integration theory: Levels of blending and disintegration", Journal of Pragmatics 37, 1615-1635.
[12] Fauconnier G. & M. Turner (2002), The way we think. Conceptual blending and the mind's hidden complexities, New York, Basic Books.
[13] Delais-Roussarie, E., "Prosodie des clitiques en français", in C. Müller et al. Eds, Clitiques et cliticisation: actes du colloque de Bordeaux, 1998. Paris, Honoré Champion, 227-249, 2001.