Aix MapTask: A new French resource for prosodic and discourse studies

Ellen Gurman Bard[1], *Corine Astésano*[2,3], *Mariapaola D'Imperio*[3], *Alice Turk*[1], *Noël Nguyen*[3], *Laurent Prévot*[3], *Brigitte Bigi*[3]

[1]The University of Edinburgh, English and Language Studies & Human Communication Research Center, Edinburgh, United Kingdom
[2]Toulouse Université and CNRS, Octogone, Toulouse, France
[3]LPL, CNRS, Aix-Marseille Université, Aix-en-Provence, France

ellen@ling.ed.ac.uk, corine.astesano@univ-tlse2.fr, mariapaola.dimperio@lpl-aix.fr, turk@ling.ed.ac.uk, noel.nguyen@lpl-aix.fr, laurent.prevot@lpl-aix.fr, brigitte.bigi@lpl-aix.fr

Abstract

This paper introduces the Aix MapTask corpus. This corpus was modelled after the original HCRC Maptask. Lexical material selection has been carefully crafted for speech and prosodic analysis [1]. We present the design of the lexical material, the protocol and basic quantitative facts about the existing corpus. We also describe an additional face-to-face condition now being collected. Finally, we explain how the material has been transcribed and processed.

Index Terms: corpus, maptask, French

1. Introduction

Leading on from pioneering work on communicative skills [2], the Map Task protocol had been designed in Edinburgh with the HCRC Map Task corpus [3]. The usefulness of the data produced with this protocol has led many teams to create their own Map Task corpora on various languages including Italian (different varieties), Japanese or Occitan. However, until now no Map Task Corpus was available for French.

Map Task corpora are interesting in particular because they can be simultaneously well controlled (in terms of lexical material, difficulty of the task, participant pairings....), while allowing genuine spontaneous speech exhibiting all the phenomena of speech production (pauses, disfluencies, etc.). The lack of Map Task for the French language is therefore, at a general level, a missing element for approaching speech and discourse in French and comparing certain phenomena across languages. Moreover, some of the authors wanted to investigate the findings from previous work [1] on spontaneous speech and the Map Task protocol was the perfect one for achieving this goal.

Thanks to EU Marie-Curie funding, the corpus was recorded and transcribed in 2002.With additional funding from ANR projects PhonIACog[1] and CoFee[2] [4] it has been developed for further use. We have gathered data and metadata and archived it in the Ortolang speech and language repository[3]. We are now working on a new set of recordings in a face-to-face condition.

[1]http://aune.lpl.univ-aix.fr/∼phoniacog/
[2]http://cofee.hypotheses.org/
[3]The corpus itself is archived at http://sldr.org/sldr000732

The present paper sets out the experimental design of the corpus (Section 2), explains how it has been processed (Section 3) and provides some quantitative information (Section 4) before introducing ongoing work and planned research (Section 5).

2. Lexical Material and Design

2.1. Lexical Material

The critical lexical material used for the Aix Map Task is a subset of the material used in [1]. That corpus consists of syntactically ambiguous sentences which prosodic cues (namely boundaries, Final Accent –FA– and Initial Accent –IA–) help to disambiguate. Syntactic ambiguity is created by manipulating adjective scope as in 'les gants et les bas lisses', where the adjective (A) 'lisses' either qualifies:

1. the second noun 'bas' (N2) only: [les gants][et les bas lisses], with an intermediate phrase (ip) boundary (B2) between N1 and N2, and a word (w) boundary (B5) between N2 and A (hereafter Case 1 or C1);
2. or the two nouns 'gants et bas' (N1 and N2): ([les gants et les bas][lisses], with an ip boundary (B5) between N2 and A, and an accentual phrase (ap) boundary (B2) between N1 and N2 (hereafter Case 2 or C2).

The manipulation of adjective scope thus yields 4 sites of interest (C1-B2 ; C1-B5 ; C2-B2 ; C2-B5) for observing indications of prosodic boundaries via FA and IA (see Figure 1).

The prosodic structure is also manipulated with regard to constituents' length, nouns and adjectives ranging from one to four syllables, in all possible combinations (eg. 'les gants et les bas lisses' vs. 'les bonimenteurs et les baratineurs fabulateurs').

Results from [1] showed that IA was a consistent marker of structure. More than its 'classic' rhythmic role as a marker of long stretches of speech, IA was shown to preferentially be used as a marker of constituency over FA, especially at the minor-phrase (ap) level, thus clarifying its role and putting it at the centre of the prosodic description of French.

As a follow-up to this first study on controlled speech, we wanted to test whether IA's role as structure marker would apply in more spontaneous speech. We ask whether IA will still

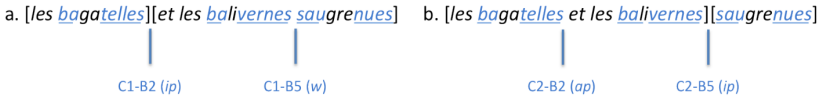

Figure 1: *The 4 prosodic sites of interest. Underscored syllables are where FA and IA potentially can occur to mark prosodic structure.*

be elicited as a structure marker in dialogue, for the same controlled target words and phrases as were used in the previous study. A subset of the corpus noun phrases was thus chosen to be represented within a Map Task design for semi-guided speech. Our goal is to compare IA occurrence in guided dialogues with our previous results on read speech. The target words and phrases chosen to appear on the maps are described below.

2.2. Material Design

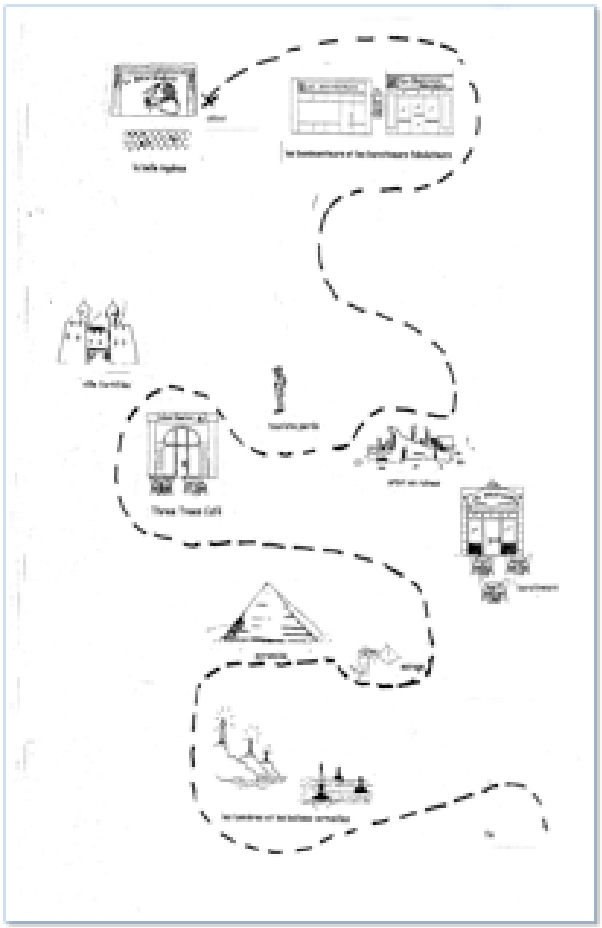

Figure 2: *Instruction Giver's map includes a route.*

To elicit spontaneous speech forms for comparison with the read speech examples used in our earlier work, we collected and transcribed a corpus of task-oriented dialogues, following the general method used in the HCRC Map Task Corpus [3]. In this task two players collaborate to reproduce on the map before one of them the route drawn on one of the player's maps (Figures 2 and 3). Neither can see the other's map. They know that the maps describe the same features but that some details may differ. In fact, the maps differ in alternate route-critical landmarks, so that discussion of the mismatches is common. To hold prior experience constant, the maps are of imaginary places and the players proceed through a series of different maps in a way that balances their experience in Instruction Giver and Instruction Follower roles. Whatever their assigned roles, players are allowed to say anything necessary to accomplish their communicative goals. However, since the two speakers cannot see each other, gestures would be ineffective.

Figure 3: *Instruction Follower's map does not include a route.*

The maps were designed around labelled cartoon landmarks, the names of which gave us the freedom to elicit nominals of the desired structure. So as not to make the names suspiciously alike, we included materials for several experiments by colleagues – on /r/ placement and on final high vowels. The critical landmark names, however, were chosen from the conjoint noun phrases of our read materials as described above. Thus there was one set of lighthouses at the top of a cliff with tall

buoys tilting on a choppy sea (*les lumières et les balises vertigineuses*: broad adjective scope) and another sketch with lighthouses at sea level and tall buoys (narrow adjective scope). The length of the second noun and of the following adjective were varied to see whether IA was encouraged by longer phonological words (N) or by longer phrases (N + Adj). In half the cases, the two players' versions of a landmark matched in adjective scope, but elsewhere they did not. To make the two Ns necessary to the naming process, single exemplars (a lone lighthouse) were also found on maps.

Sixteen pairs of speakers performed 8 map tasks each. Digital channel per-speaker stereo recordings were made in studio conditions and transcribed by native speakers of French.

The distribution of target words (and therefore landmarks) across the maps is set out in Table 3. In this table, each cell corresponds to a map. The condition $IG = IF$ means that the corresponding landmark has the same scope on follower and giver's maps ($IG \neq IF$ if not). More precisely, the conditions are Broad Broad, Narrow Narrow, Broad Narrow and Narrow Broad. Finally the four colors correspond to the four dyads of participants (each dyad had 8 maps to communicate and participants switched role after 4 maps).

3. Transcribing and Processing the data

When a speech corpus is transcribed into a written text, the transcriber is immediately confronted with the following question: How to reflect the reality of oral speech in a corpus? Conventions are thus designed to provide a set of rules for writing speech corpora. These conventions establish which phenomena have to be annotated and also how to annotate them.

The corpus was transcribed in standard French orthography, using Transcriber [5]. The transcription includes short pauses, truncated words and hesitations.

SPPAS is a tool to produce automatic segmentations from a recorded speech sound and its transcription [6]. The resulting segmentations are represented in a set of TextGrid files, the native file format of the Praat software [7].

SPPAS tools and resources are currently available under the GNU Public License, at the URL:

http://www.lpl-aix.fr/∼bigi/sppas/

SPPAS generates separate TextGrid files for utterance, word, syllable, and phoneme segmentations. (i) utterance segmentation, (ii) word segmentation, (iii) syllable segmentation and (iv) phoneme segmentation. An example of SPPAS output is represented in Figure 4.

4. Quantitative aspects

The combination of the four dyads explaining each other 8 maps (alternating roles) provides 32 dialogues of an average duration of 6 minutes 52 seconds. The corpus includes about 50 000 tokens but with a vocabulary size of only 1500 different forms.

As can be seen in table 1, which lists the corpus frequencies the 30 most frequent words with their number of occurrences, other than function words, feedback items (*ouais, mh, d'accord, voilà, oui, non*) are well represented. Words related to space are also extremely frequent (*gauche / left , droite / right, vers / towards, vas / go*[4]*, sur / on , dessous / under*).

[4]Since the corpus has not been lemmatized yet, this table may not give other spatial verbs their true rank.

2389	tu	703	à	347	du
1146	la	643	les	337	vas
1077	ouais	507	un	332	ai
1032	euh	494	gauche	331	voilà
1017	de	446	d'accord	325	sur
927	et	438	donc	321	c'est
815	le	420	droite	317	oui
812	mh	414	as	314	j'
800	je	397	là	310	non
704	en	375	vers	293	dessous

Table 1: *Most frequent forms with number of occurrences*

word	occurrences	mean duration (sec)
balises	187	0.3538
lumières	120	0.3675
bonimenteurs	105	0.6777
baratineurs	82	0.5809
vertigineuses	81	0.6823
fameux	45	0.4388
vermeilles	41	0.4440
fabulateurs	23	0.7513

Table 2: *Target words with number of occurrences and average duration*

Table 2 displays the number of occurrences of target words and their average durations. Due to their presence on all the maps, the nouns are used more often than the adjectives. Noun frequencies are however difficult to interpret at this early stage. The spread of frequencies for the adjectives is bigger with *vertigineuses* occurring about four times more than *baratineurs*. This is however difficult to explain without analyzing the dialogues more deeply . For example, though all the design-critical landmarks were also route-critical, some landmarks may in the end have been less useful or easier to use than others, yielding less discussion and resulting in a lower frequency of occurrence.

Finally the target phrases occurrences were:

- les balises [euh] vermeilles 25
- les balises [pause] vertigineuses 42
- les baratineurs fameux 15
- les baratineurs [pause] fabulateurs 13

5. On-going work and perspectives

We are currently working in two directions on this corpus: (i) shallow natural language processing for extracting more linguistic information of the corpus, (ii) constitution of an additional condition with participants seeing one another as they work. Natural language processing includes POS tagging with probabilistic tagger trained on written and spoken data [8], lemmatizing and chunking. The additional condition is basically a replication of the existing corpus but allowing participants to see each other (but of course not each other's maps). The recordings are once more performed in an anechoic room with a high-quality headset and 3 cameras (1 for each participant and general one) following a technical setting already used in the lab [9].

At this stage, the main studies planned include the annotation of feedback items (which are extremely frequent as it can

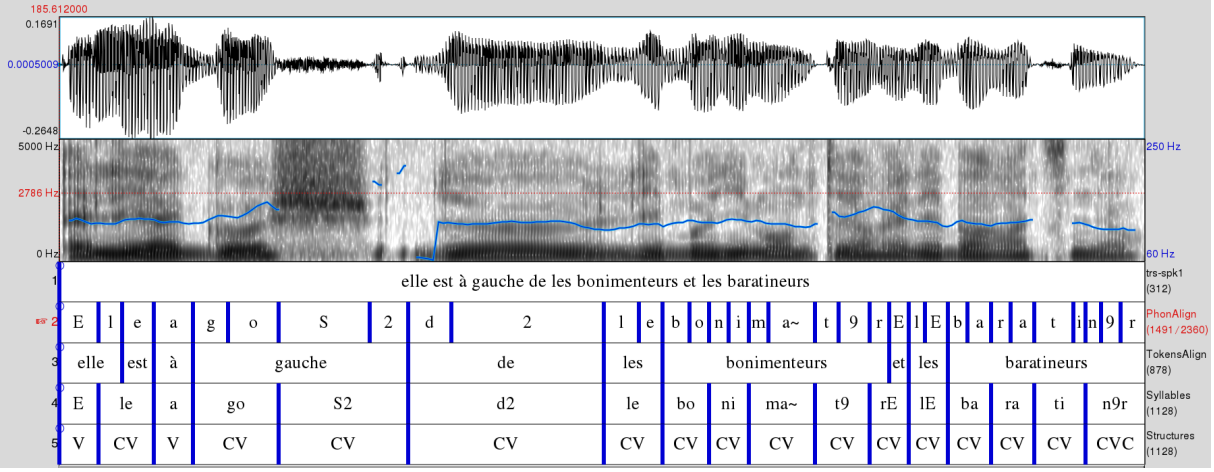

Figure 4: *SPPAS output example.*

be seen in table 1) in the context of a broader study on multimodal feedback, which will focus on the different distributions of feedback between verbal and visual modalities with the absence v presence of a visual channel.

A second study planned consists in comparing compare IA occurrence on guided dialogues with previous results on controlled speech [1]. The number of occurrences (Table 2) of the target words are encouraging in this respect .

6. Acknowledgements

This work was funded by a European Community Marie Curie Fellowship award to Corine Astésano (HPMF-CT-2000-00623). We are grateful for the collaboration, assistance and support of colleagues the Laboratoire Parole et Langage, Bernard Teston, Cheryl Frenck- Mestre, Mariapaola d'Imperio, Robert Espesser, Louis Seimandi, Annelise Coquillon, Ludovic Jankowski and from the University of Edinburgh, Eddie Dubourg and Ziggy Campbell.

7. References

[1] C. Astésano, E. Bard, and A. Turk, "Structural influences on initial accent placement in french," *Language and Speech*, vol. 50, no. 3, pp. 423–446, 2007.

[2] A. Anderson, G. Brown, R. Shillcock, and G. Yule, *Teaching talk: Strategies for production and assessment*. Cambridge University Press New York, 1984.

[3] A. Anderson, M. Bader, E. Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. Thompson, and R. Weinert, "The HCRC Map Task corpus," *Language and speech*, vol. 34, no. 4, pp. 351–366, 1991.

[4] L. Prévot and R. Bertrand, "CoFee-toward a multidimensional analysis of conversational feedback, the case of French language," in *Proceedings of the Workshop on Feedback Behaviors*, Stevenson, 2012, (poster).

[5] C. Barras, E. Geoffrois, Z. Wu, , and M. Liberman, "Transcriber: a free tool for segmenting, labeling and transcribing speech," in *Language Resources and Evaluation Conference*, Granada (Spain), 1998, pp. 1373–1376.

[6] B. Bigi, "SPPAS: a tool for the phonetic segmentation of speech," in *Language Resource and Evaluation Conference*, Istanbul (Turkey), 2012, pp. 1748–1755, ISBN 978–2–9 517 408–7–7.

[7] P. Boersma, "Praat, a system for doing phonetics by computer," *Glot International*, vol. 5, no. 9/10, pp. 341–345, 2001.

[8] S. Rauzy and P. Blache, "Un point sur les outils du LPL pour l'analyse syntaxique du français," in *Actes du workshop ATALA 'Quels analyseurs syntaxiques pour le français ?'*, Paris, France, 2009, pp. 1–6.

[9] J. Saubesty, "L'activité gestuelle en situation d'incompréhension," 2013, Master Thesis, Aix Marseille Université.

			noun 4-syllables + adjective 2-syllables *les bonimenteurs et les baratineurs fameux*				noun 4-syllables + adjective 4-syllables *les bonimenteurs et les baratineurs fabulateurs*			
			IG=IF		IG≠IF		IG=IF		IG≠IF	
			BB	NN	BN	NB	BB	NN	BN	NB
noun 2-syllables + adjective 2-syllables *les lumières et les balises vermeilles*	IG=IF	BB			Ai1	Aiii1			Bi6	Biii6
		NN			Aii1	Aiv1			Bii6	Biv6
	IG≠IF	BN	Av5	Avii5			Bv2	Bvii2		
		NB	Avi5	Aviii5			Bvi2	Bviii2		
noun 2-syllables + adjective 4-syllables *les lumières et les balises vertigineuses*	IG=IF	BB			Ci3	Ciii3			Di8	Diii8
		NN			Cii3	Civ3			Dii8	Div8
	IG≠IF	BN	Cv7	Cvii7			Dv4	Dvii4		
		NB	Cvi7	Cviii7			Dvi4	Dviii4		

Table 3: *Map composition in terms of landmarks (target words)*