French Language Studies **18** (2008), 121–145, © Cambridge University Press
doi:10.1017/S0959269507003195

Tool-assisted analysis of interactional corpora: voilà *in the CLAPI database*

GROUPE ICOR (MICHEL BERT, SYLVIE BRUXELLES, CAROLE ETIENNE, LORENZA MONDADA, VÉRONIQUE TRAVERSO)

ICAR Research Lab (CNRS and University of Lyon)

(Received August 2007; revised September 2007)

ABSTRACT

The aim of this paper is to show how databases and computer tools can be used for archiving and browsing corpora of social interactions. The development of specific search engines allows for both qualitative analysis of naturally occurring interactions and for quantitative explorations of larger corpora. The paper is based on the CLAPI Workbench <http://clapi.univ-lyon2.fr>, an interfaced ensemble of analytic tools which operates on a consistent body of corpora to facilitate their description and theoretical reconstruction. The analytical part of the paper focuses on the uses of a discourse particle in French interaction: *voilà*.

1. INTRODUCTION

This paper explores the uses of a database of corpora of naturally occurring interactions – CLAPI (Corpus de LAngue Parlée en Interaction) – for the systematic study of linguistic phenomena within social interaction.[1] It focuses on the specific complexities of interactional data, on the possibilities for archiving them and making them accessible to the scientific community and on the possibilities for browsing them in order to sustain qualitative analysis and/or to enhance quantitative approaches.

The paper is divided into two parts: the first deals with the presentation of the CLAPI workbench, the objectives, principles and solutions which have been adopted for the conception of its architecture. The second part focuses on the analysis of the discourse particle *voilà*: based on a previous multimodal qualitative analysis of this particle by Bruxelles and Traverso (2006), the study exploits CLAPI larger corpora and search engines in order to verify, generalise and finally reorient previous findings. We demonstrate therefore how it is possible to highlight the specific contributions of particular analyses by applying them to larger quantities of data and by employing specific search tools.

[1] The authors wish to thank three anonymous JFLS referees for their corrections and comments. Special thanks to Kris Lund for revising the English text.

https://doi.org/10.1017/S0959269507003195 Published online by Cambridge University Press

2. CLAPI, A WORKBENCH FOR THE STUDY OF FRENCH LANGUAGE IN INTERACTION

2.1. Which corpora for interactional linguistics?

The need for readily available, naturally occurring audio and video recorded interactional data is now generally accepted by various disciplines within linguistics and the social sciences. On the one hand, social interaction is the 'natural habitat' of language – the site where grammar emerges (Hopper 1987) and changes, where children acquire their first language, where non-natives are socialised in a second language, and where social identities are displayed, negotiated and achieved. On the other hand, only naturalistically gathered data can make available for analysis complex details on which the organisation of talk is based – details which are neither imaginable, for instance by introspection, nor reproducible, for example in experimental tasks. On the basis of such considerations, interactional linguistics has explored both traditional grammatical forms and previously neglected verbal and vocal resources, as well as multimodal resources, as they are used for the organisation of talk in interaction: this has opened up new insights on syntax (Auer 2005; Ford, Fox and Thompson 2002), phonetics (Local, Kelly and Wells 1986), prosody (Couper-Kuhlen and Selting 1996), the lexicon (Clift 2001), and multimodality (Goodwin 2000), in ordinary speakers' talk, as well as in that of children (Wootton 1997), or aphasic participants (Goodwin 2003). This has also fuelled new reflections and higher demands on transcription practices and standards (Jefferson 2004), as well as on audio and video recording practices and technologies. The field of conversation analysis has played an historical role in underlining the necessity for naturalistic data, of repeated listening and visioning of these data by the analyst, and of detailed transcriptions for the study of the way in which participants orient themselves towards the relevant resources mobilised for achieving the interactional order.

These insights have emerged very recently, since the 1960s, in the constitution of large collections of data, as a basis for a systematic exploration of the details of talk in interaction – for example, for the study of sequencing in openings (Schegloff 1968, 1979), or for the study of repair (Schegloff, Jefferson and Sacks 1977; Schegloff 1979). Nowadays, the existence of browsable data banks allows us to enlarge these systematic analyses; conversely, past systematic explorations of data collections fuel the development of search tools and request engines which are not reduced to the identification of single forms but try to access aggregates of forms, characterised by their distribution within various sequential positions. Interactional linguistics (Ochs, Schegloff and Thompson 1996; Selting and Couper-Kuhlen 2001) challenges the way in which searches are done in large corpora, by stressing the importance of time and therefore of a series of successive moments within the turn as well as in the sequence (Mondada 2005; Fox 2007). Turns develop incrementally, step by step, in an emergent way, within the temporal unfolding of their production, sensitive to the contingencies of context and adjusting to the coordinated conduct of other participants: this view invites the development of a 'positionally sensitive grammar' (Schegloff 1996), where grammatical forms – as well

https://doi.org/10.1017/S0959269507003195 Published online by Cambridge University Press

as other resources – are described relative to their contribution to the sequential organisation of interaction. This description focuses on actions implemented in systematic practices mobilising specific (linguistic but also multimodal) resources arranged within specific configurations or constructions, positioned in specific sequential environments. This grounds the hypothesis 'that there may, in fact, be grammatical formats for sequentially-specific actions' (Fox 2007: 304).

This conception of grammar in interaction (Ochs, Schegloff and Thompson 1996) articulates two kinds of intertwined demands for larger corpora: on the one hand, this kind of inquiry is grounded in the details observable in naturally occurring social interactions, as they are recorded, transcribed, and stored in corpora constituted by primary data (recordings) and secondary data (their annotations and representations). On the other hand, this kind of inquiry relies on search engines which are 'positionally sensitive', which allow us to search for aggregates of concomitant forms in specific sequential environments.

Both dimensions inspire the development of specific data banks for interactional corpora. These are still in fieri: in most of the national corpora – such as the British National Corpus (BNC) or the Real Academia Española corpus (Corpus de Referencia del Español Actual, CREA) – the spoken part is quantitatively marginal and accessible only through very simple transcripts; data banks focusing on spoken language recorded in social interactions in natural settings (and not for instance through sociolinguistic interviews) are still very few. A notable exception is the Talk Bank (http://talkbank.talkbank.org/), started in 1999 by Brian MacWhinney, and which is devoted to the study of human and animal communication, but which does not offer any equivalent search mechanism capable of browsing the corpora of the data bank.

2.2. *The development of the CLAPI talkbank*

CLAPI (Corpus de LAngue Parlée en Interaction – http://clapi.univ-lyon2.fr) has been developed in order both to make accessible to a larger scientific community corpora of naturally occurring data and to make possible new kinds of (semi)automatic requests on these data.

The project was launched in the 1990s at the GRIC research lab in Lyon (Groupe de Recherche sur les Interactions Communicatives), which then became the ICAR research lab (Interactions Corpus, Acquisition, Représentations). This research unit had a long history of interactional studies based on corpora collected through fieldwork in various social settings. The first database was aimed at storing and conserving the corpora constituted by the members of the centre, associated with a set of metadata describing them. In this early version, each corpus was illustrated through a downloadable excerpt of sound and transcript, but was not associated with the entire corpus (for a presentation of this stage of the project in 2001, see Bruxelles and Traverso 2002).

The second database, CLAPI, was created in 2001 by the Groupe ICOR; it consists not only of a storage space for corpora but of an articulated workbench,

https://doi.org/10.1017/S0959269507003195 Published online by Cambridge University Press

where primary and secondary data are archived, described by metadata, and made browsable thanks to search engines (for a presentation of the project in 2004, see Balthasar and Bert 2005; for an up-to-date presentation see Groupe ICOR, to appear b).

The project has thus evolved in time, taking into consideration various dimensions:

- It evolved from a data bank for the internal purposes of a closed group into a data bank open to the larger scientific community, thus varying the users' profiles (both of data producers and data consumers) implemented through a codification of rights and accesses;
- It evolved from a data bank considered first of all as a storage space into a data bank considered as a collaborative workbench, offering various tools in order to make corpora and phenomena more accessible, thanks to XML tagging;
- It gradually merged older corpora, interesting as regards their legacy and for a diachronic approach to interactional data, and newer corpora, produced using state-of-the-art recording practices and transcribed, aligned, and tagged according to contemporary standards.

These changes show that a data bank is a dynamic object that evolves in time, implying a constant reflexive discussion about its constitution, its principles, and its choices.[2]

2.3. Current state of the CLAPI talk bank

Currently, the data bank comprises 120 hours of audio and video, constituting 296 recordings and 477 transcripts – of which 40 hours are aligned with their transcripts in an XML format (the equivalent of 350,000 words). The collated corpora document various social interactions, from dinner conversations to medical interactions, from legal discussions to service encounters – recorded over the last 25 years: the data bank is dedicated to corpora of social interactions in French, recorded within a diversity of settings in a naturalistic way.

In order to make these data available, the workbench has been developed according to several principles and choices, governing the metadata, the transcript and formats, and the access. At these three levels, the issues are to enhance data accessibility, to offer the possibility of efficiently selecting a subset of data for specific analytical purposes and to provide interoperability through browsable and searchable corpus structures and resource descriptions. In this development, the relevant legal and ethical constraints have been scrupulously taken into account.

2.3.1. Metadata

Metadata represent a key resource that characterises, thanks to a standard set of descriptors, a corpus in sufficient detail and accuracy for users to determine

[2] The background of this discussion, containing tools, resources and references on interactional corpora, is published and updated by the Groupe ICOR in a related site, CORINTE (http://icar.univ-lyon2.fr/projets/corinte).

https://doi.org/10.1017/S0959269507003195 Published online by Cambridge University Press

whether or not it is of relevance to their particular enquiry. Metadata of the CLAPI talk bank are both oriented towards a wide scientific community of users and applications as well as towards the specific features of interactional corpora and needs of interactional research. In order to respond to these rather complex demands, 75 descriptors have been elaborated, taking into account current work on social interaction as well as standards introduced by various international initiatives, such as the Dublin Core Metadata Initiative (which proposes a rather general vocabulary of fifteen properties for use in resource description), the OLAC initiative (which proposes an XML metadata format to interchange language resources within the framework of the Open Archives Initiative), and the ISLE metadata initiative (IMDI, which is oriented to multimedia/multimodal language resources paying special attention to disciplines such as corpus linguistics, anthropology and language and speech engineering).

The descriptors cover a set of generic or specific fields organised as follows:

- General information on the corpus (name, date and place of collection. . .);
- Authors (owner(s), collectors, transcriptors. . .);
- Recordings (name, date and place of recording, length, type: audio/video/computer mediated traces . . .);
- Transcripts (transcript conventions, software, alignment and tagging, information on the quality of the transcript);
- Participants (number, sociolinguistic characteristics of each participant);
- Interactional genre (medical consultation, conversation, etc.);
- Access conditions of the transcript and the recordings;
- Bibliography.

In addition to the standard information that is required in a database, special attention has been devoted to:

1) The different agents who are involved in the creation of a corpus throughout the successive steps of its production (i.e. fieldwork, data recording, digitalisation and anonymisation, transcription, alignment, annotation etc.). It is the 'corpus owner' who decides the conditions under which her/his corpus is made accessible.
2) The scientific documentation of the corpus: previous studies on the corpus are listed (and if possible, they are full-text accessible). When recording a corpus in the database, these descriptors can be filled either by selecting predefined categories or by proposing a descriptive text – which provides for more specific details. When consulting the database, they facilitate the selection of a set of corpora for automatic queries, and they can be used as component of the query itself (see 3.4).

2.3.2. Transcript and formats

As a talk bank specifically devoted to the documentation of naturalistic interactional data, CLAPI is oriented towards the preservation and accessibility of the relevant

https://doi.org/10.1017/S0959269507003195 Published online by Cambridge University Press

detailed features of social interaction. This has consequences for the way in which secondary data (transcription) are made out of the primary data (recordings).

Transcription for interactional research implies that the specificity of spoken production is preserved: a transcript convention has been established to this end (ICOR Convention), inspired by current standards in interactional studies, such as the Jefferson Conventions (Jefferson 2004) and the GAT conventions (Selting et al. 1998). It encompasses a list of relevant phenomena, related to both the verbal production and the multimodal resources (gestures, gazes, body postures, facial expressions, etc.) used by participants in their exchanges, as well as the temporal organisation of these resources within interaction (coordination with other co-participants' conduct).

Excerpt 1. (Mosaic, 1038-1054, 12:09:13-12:09:47)
(see transcription conventions at end of article)

```
1038 M  toi toi t'as peur que le bureau soit trop p'tit
        >>+pointe sur le calque-->
1039 L  oui
1040    (0.6)
1041 M  et:/ ça c'est:+ +c'est large tout ça non+
                  -->+ +geste circulaire       +
1042    (0.3)
1043 L  +oh c'est pitchoune +quand même hein
        +M ramène sa main vers son visage+
1044 M  ((rire)) (.) (non mais +en [même)
                                +M repointe->
1045 C                              [non mais ça peut
        êt' euh:: la bagagerie là\ [typiqu'ment+ +(0.4)+
                                              -->+ +M recule main
```

In this excerpt, the main features of interactional transcription are available:

- each piece of talk is related to a participant (initials);
- the characteristics of talk production are preserved: repetition, false start (1038, 1041 for instance), pronunciation (made visible by the use of an 'adapted orthographic representation', as in 1038 't'as', instead of the standard 'tu as');
- precise temporal indication on turn organisation is given: overlap (indicated by the brackets, in 1044–1045 for instance); gap (1040, 1042);
- multimodal information is provided on a separate line and located in relation to talk (for instance line 1040, + indicates the place where the gesture described in the following line stops or begins).

In addition to the attention devoted to this sort of transcription, the number and complexity of relevant phenomena make it necessary to align the recordings with the transcripts, with one of the currently available pieces of software such as Praat, CLAN, or ELAN.

On the basis of the ICOR transcript convention, tools have been developed for providing an XML tagging of the main interactional phenomena (participants,

https://doi.org/10.1017/S0959269507003195 Published online by Cambridge University Press

tokens, verbal productions, overlaps and pauses) making automatic search possible. The issue of *adapted orthography* has been solved by the development of an automatic orthographic variants recognition tool.

In the current state of the database, each corpus contains both the original transcripts and conventions chosen by the original owner and the ICOR-XML transcript. The result is a preservation of the original features, the use of a standard, and the browsability of a large number of non-standard resources.

2.3.3. Access

CLAPI provides various tools for interactional research, making possible the browsing of the corpora in a selective way.

A first set of tools automatically generates a synoptic view of the main characteristics of each corpus, such as its lexicon and lexical frequencies, the distribution of interactional phenomena (participants, tokens, verbal productions, overlaps and pauses) throughout the entire transcript, and the more frequent correlations between these phenomena.

A second set of tools offers various searching possibilities. The simplest search tools ('*co-occurrences of a token*', '*contextual analysis of a token*') allow for the search of a linguistic form or token in order to make concordances, with the results displayed in an aligned form. More complex search tools allow for the search of various tokens, in various positions within the turn, and taking into consideration their co-occurring within specific interactional structures ('*complex search*', '*requêtes multicritères*'). For example, it is possible to search for two tokens which are separated by three others, the first being at the end of a turn and the second at the beginning of an overlap or just after a pause). Thus, searches are able to deal with linguistic and interactional items, as well as with standard and non-standard original items, thanks to the automatic tagging of the latter. Searches within the transcripts can be combined with searches within the metadata.

2.3.4. Legal and ethical constraints

These possibilities offered by the CLAPI database – accessing metadata, consulting primary as well as secondary data and browsing corpora – are crucially dependent on the legal and ethical constraints which are particularly complex and delicate in the case of naturalistic data.[3] Taking these issues into consideration, in order to respect both the privacy of the recorded participants and provide access to the data for a large community, the CLAPI workbench has implemented a series of solutions. These solutions concern, on the one hand the categorisation of the items to be accessible and, on the other the categorisation of the users.

Corpora items can be accessible in various ways: the whole of a corpus can be consulted (recordings and transcription), or only a part of it can be available

[3] On these questions, which also concern data production and not only data dissemination, see Baude et al. (2006), as well as the section on legal and ethical aspects of the CORINTE website: http://icar.univ-lyon2.fr/projets/corinte.

https://doi.org/10.1017/S0959269507003195 Published online by Cambridge University Press

(i.e. a subpart of the corpus and/or the transcript only, or the transcript cum audio but not video), or the corpus can be accessible by the means of the search mechanisms only (as the result of a query). Various users' profiles have also been implemented, allowing different kinds of accessibility, depending, for restricted corpora, on previous agreements made with the corpus' owner.

In general terms, a good accessibility of the corpora has been made possible, while respecting the lega and ethical demands of the corpus owner(s): a generic user of the talk bank, without special authorisation, can gain access to all the metadata, to the transcript conventions, to the freely accessible corpora, to about 80 exemplary subcorpora (transcripts and signal), and, through search engines, to 70% of the XMLised data set (30h, 265,000 words). The remaining portion of the data bank is accessible by agreement with the corpus owner(s).

3. FROM QUALITATIVE TO QUANTITATIVE ANALYSIS: THE CASE OF VOILÀ

The second part of this article is devoted to the study of the discourse particle *voilà* in the large range of corpora accessible in the CLAPI database.[4] It is based on a previous multimodal interactional study of this particle (Bruxelles and Traverso 2006) that has been carried out on a video recording of a working meeting of architects, the Mosaic corpora (Détienne and Traverso (eds.), to appear). Before presenting these findings, we recall the main outcomes of previous studies of *voilà*.

3.1. Previous studies of voilà

3.1.1. Deictic and Conversation Structuration Marker

The particle *voilà* has a core deictic value, which is still visible in its morphology: *voilà* is composed of the verb *voir* ('to see') in the imperative mood and of the adverb *là* ('here'). *Voilà* was opposed to *voici*, as indicating distance (*vs.* proximity), but this opposition is no longer significant in spoken French. In their 1999 and 2001 papers, Plauché and Bergen explain in detail the diachronic evolution that has led, along with a metaphoric extension, from this core usage of *voilà* as a deictic – in which the particle draws attention to an object in the setting – to its temporal usage ('voilà l'instant que nous attendions', lit. 'here is the moment we were waiting for') and to its usage as a discourse deictic ('voilà deux exemples', lit. 'here are two examples').

V*oilà* has also been studied by the Geneva researchers (Roulet *et al.*, 1985) in their work on *Conversation Structuration Markers* ('marqueurs de la structuration de la conversation', MSC). Following Gülich's pioneer work on *Gliederungssignale* (1970), they define these items as forms (morphemes or syntagms) that are recurrent in discourse production and characterised by a loss of their original lexical meaning. The category of *Conversation Structuration Markers* is based on the role these items

[4] For another example of analysis, see Groupe ICOR (2007) on the grammaticalisation processes affecting the particle *attends*.

https://doi.org/10.1017/S0959269507003195 Published online by Cambridge University Press

play in interaction: they display the way in which participants structure the co-construction of discourse. Roulet *et al.* develop an approach based on a hierarchical model (1985: 99–100) that distinguishes between a 'monological level' (the level of interventions) and a 'dialogical level' (that of exchanges). Within this framework, they identify the following features for *voilà*. At the level of intervention, *voilà* is one of the rare markers that are not constrained in their syntagmatic position (i.e. to the beginning or end of a constituent): *voilà* can precede as well as follow the constituent it marks. As for structure, *voilà* marks the constituent as the main act or intervention in the sequence. In this analysis, *voilà*, as well as *quoi* and *pi(s)*, does not articulate constituents at the exchange level. It mainly marks monological relations and it scarcely appears in dialogical contexts (Roulet *et al.* 1985: 108). For Auchlin (1981: 95, footnote 11), the main value of *voilà* on this monological level is to indicate that the constituent is conclusive. Working within the same framework, Filliettaz (2001) points out that this description does not account for all the usages of *voilà*. Specifically, it does not account for the cases in which *voilà* is used along with a 'non-verbal action' ('action non langagière'), as it is frequently the case in his data (service encounters). He thus proposes that in such usages, *voilà* is not a *Conversation Structuration Marker*, but is used in order to clarify the communicative feature of a non-verbal action (2001: 194–199).

Although we are not working within the same framework, the main outcomes of these studies needs to be kept in mind in order to be verified as part of our quantitative study: *voilà* mainly acts at a monological level (i.e. it is used within a turn and not to articulate turns by different speakers); it has a prominent conclusive value and, in some of its uses, a *co-gestural* value.

3.1.2. A multimodal study of voilà

Bruxelles and Traverso's study (2006) is based on the video recording of a meeting between three architects. Their analysis integrates the gestural dimension of interaction: gaze, gestures to documents, position of the body, etc. They have identified three values for the particle *voilà* in their data: a) a use of *voilà* as a deictic particle (excerpt 2); b) a use of the particle as a *structuration marker*, specifically in transitions between two successive topics or activities (excerpt 3); c) a use of *voilà* as an agreement marker (excerpt 4). This usage is particularly found in highly collaborative sequences, and *voilà* frequently has a conclusive value.

Excerpt 2. [11:34:36]. Deictic

```
155 M  +okay *(.) [je: veux juste revoir un truc
       +M feuillette les plans empilés à D
            *C se met à feuilleter avec elle
156 C              [xxxxxx (ici/)
157     (1.0)
158 M  +euh:: d'accord *okay
       +M tire un des plans de la pile et le regarde
                        *C dégage un peu plus le plan
159    (.)
```


https://doi.org/10.1017/S0959269507003195 Published online by Cambridge University Press

```
-> 160 C  voi[là *c'est là
                *C pointe sur le plan avec geste de D à G -->
   161 M     [non je voulais savoir à quoi
   162    (..)
   163 C  c'est celle-ci* (après)
                      -->*
   164 M  oui oui
   165    (5.0)
```

C's pointing gesture (line 160: C points at the concerned area on the plan and shows it with a repetitive gesture from left to right) is done immediately after he has uttered 'voilà'. This gesture lasts several seconds, and is accompanied by several other deictic verbalisations: '(voilà) c'est là' ('here it is'), 'c'est celle-ci' ('it's this one') (160, 163).

Excerpt 3. [11:27:55]. Transition

```
   1 O  okay//
   2    (0.6)
   3 O  on y va// (1.1)
   4 ?  xxxxx
   5    (0.9)
   6 C  *alors c'est bon (.) alors je vous explique un petit p-*
        *C se penche vers la zone centrale et touche le plan*
-> 7    *(4.0) voilà (1.0)*
        *C replie ses bras*
   8    *alors le projet au niveau du plan masse*
        *C se repenche sur la zone centrale et repose sa main sur
        le plan*
   9    *euh tu l'avais vu le plan masse hein Marie/*
        *tapote le plan masse*
```

The other elements that accompany 'voilà' in this transition are found in nearly all the transitions between activities throughout the entire meeting:[5]

- Other particles 'alors' ('then'), 'bon ('well') and 'c'est bon' ('okay');
- Numerous pauses;
- Gestural markers: the speaker retracts his current gesture and handles the plan that will be used in the next activity.
- Furthermore, in this excerpt, a meta-pragmatic formulation[6] is used (line 3).

[5] See Bruxelles *et al.* (to appear); Traverso (to appear).

[6] Cf. Heritage and Watson (1979): 'We shall speak of conversationalists' practices of saying-in-so-many-words-what-we-are-doing as formulating'; see also Katsiki and Traverso (2004) on the spontaneous denomination of interactional activities in interaction.

https://doi.org/10.1017/S0959269507003195 Published online by Cambridge University Press

Excerpt 4. [11:49:19]. Agreement

```
   481 C  *et puis là on rentre là:/ on sort là:/ (.) et on a
          >>*C était déjà en train de dessiner et continue-->
   482    notre (.) nos salles sous les voûtes\ (0.6) et là
   483    l'accès (2.0) on té- on arrive par là\ et puis là on
   484    va trouver une porte forcé[ment/
   485 M                            [ouais
   486 C  +[par là/ [hein
          +M pointe
-> 487 M   [ouais+ +[voilà\+
              -->+ +M retire sa main+
   488 C  là hein/
-> 489 M  voilà\
   490    (0.8)
   491 L  moi je trouve/* ‡*vraiment dommage que ça ce
                 ------->* *C recule-->>
                          ‡L pointe-->>
```

In this excerpt, two occurrences of 'voilà' are used to answer requests for confirmation marked by the particle 'hein' (translatable as 'you see').

The authors have identified three main features that distinguish the different usages:

- *voilà* isolated or combined. This feature distinguishes cases when *voilà* constitutes the complete turn. It is the case for the deictic marker and for the agreement marker, but not for the structuration marker. A related feature seems to be that isolated *voilà* often has an interjectional feature.
- Monological (within one participant's turn) or dialogical (second pair part) use. Only the structuration marker is clearly used to articulate components within a participant's turn; the two other categories relate turns of different participants.
- Backward or forward orientation. This issue only concerns the agreement marker and the structuration marker, for the deictic marker is related to the local item (be it an object or an idea) that is designated. As to the agreement marker, it mainly has a closing feature (thus a backward orientation), which is displayed by the positive assessments, and the summaries that are often uttered along with it. The issue is less clear for the structuration marker. The authors put forward two alternative views: it could be understood as having an opening feature (thus a forward orientation), for it is followed by topical or actional changes; on the other hand, it can be seen as a constitutive element within a configuration of the form 'marker + pause + marker' that projects a transition. Through such a configuration, participants display that they are achieving a transition and they devote a certain amount of time to this achievement. This permits them to prepare the introduction of the next activity but also gives them the opportunity to go back to the

https://doi.org/10.1017/S0959269507003195 Published online by Cambridge University Press

previous one for an expansion or an addition.[7] This issue of the marker's orientation will be one of those we will try to clarify through the current study.

The authors have identified additional features that can be summarised in the following table:

	1. Deictic marker	2. Structuration marker: transition between topics or activities	3. Agreement marker: conclusive, affiliative
Pause	After a pause	After and before long numerous pauses	Without pause (frequently in latching)
Location within the turn	Alone or in a short turn, "voila" comes after the exclamative particle "ah"	Is not used alone, it appears within a turn or at the end of a turn	Can appear alone. If not, it is located at the beginning of the re active (second pair part), after a request for confirmation, an inserted exchange or in collaborative utterances
Along with	Interjectional feature: rising intonation, lengthening, self-centred talk	Projective feature: falling intonation	Interjectional feature: falling intonation, other-addressed talk
Context	Accompanied by "ah" or by other deictics	Along with other particles ("alors", "bon", etc.) and, in some cases, formulations	Alone or accompanied by affirmative morphemes ("oui")
Repetition	Yes	No	Yes
Overlap	[Not mentioned]	[Not mentioned]	Sometimes. Rather at the beginning of the overlapping turn (receipt token)
Multimodality	After or before pointings; no gaze towards the other participants	Along with gestures to the documents	Gaze towards the other participants, postural movements, gestures to the documents

[7] See Traverso (2005); Mondada and Traverso (2005).

https://doi.org/10.1017/S0959269507003195 Published online by Cambridge University Press

3.2. CLAPI studies of voilà

Before examining how the CLAPI database tools make it possible to check these findings on a larger range of corpora, we can provide general quantitative information about *voilà* on the basis of its frequency in the database.

3.2.1. An overall view of voilà in the CLAPI database

There is no orthographic variant of *voilà* in the database (no shortened version).[8] Seven 'related' forms are found: one misprint ('voila' without the accent on 'a'), and six forms 'voilÀ', in which the capital letter marks that the final vowel is pronounced louder.

There are 871 occurrences of *voilà* in the CLAPI database. A quantitative approach (in addition to a qualitative one) is thus possible and relevant. The particle is one of the most frequent, along with *mais*, *ben*, *alors* (see ICOR, forthcoming a and b).

The distribution of *voilà* in the different corpora accessible in the database is heterogeneous: for a similar length (about 60 minutes), the number of occurrences ranges from 14 (in the corpus 'Français des années 80', which is composed of recordings of sociolinguistic interviews) to 90 (in the corpus 'Vente à domicile', recordings of a door-to-door salesman selling encyclopaedias). The type of activity developed in the interaction seems to affect the use of the particle: for instance, the manipulation of objects seems to be related to a high rank of use, whereas the situation of interview seems to reduce its use. A link also appears with the situation of mediated talk: the use of *voilà* seems to increase in phone-calls. This result is somewhat contradictory to the 'multimodal' use of *voilà*. We can hypothesise that it is linked to the necessity for the participants to display in a very visible manner both the fact that they are on line (receipt tokens) and the way in which they structure their interaction.

3.2.2. Constitution of a subcorpus on the basis of the metadata

In order to carry out this extended study of *voilà*, we have selected in the database a range of corpora that are, to a certain extent, homogeneous with the Mosaic corpus: i.e. XML tagged corpora, recorded in natural settings, in multiparty situations.

The subcorpus is constituted of 23 transcripts. It represents various interaction genres: conversations between friends, working or institutional meetings, medical consultations, service encounters, and classroom interactions. This variety of situations prevents bias that could be linked to an over-representation of a certain type of situation. The number of participants in these situations ranges from two to twelve.

[8] There are no occurrences of the truncated form 'v'là', in spite of its frequent usage in songs and cartoons. This absence can probably be thought of as related to the lack of certain situations in the base, since not all the possible social situations are documented, but above all it underlines the necessity of data to check what the actual usages are.

https://doi.org/10.1017/S0959269507003195 Published online by Cambridge University Press

The subcorpus represents 114,457 tokens, and encompasses 312 occurrences of *voilà*.

3.3. *Analysis of* voilà *with CLAPI database tools*

We aim to show how the CLAPI database tools can be used to confirm or modify the outcomes emanating from the qualitative study. We have used the tools working on the linguistic and interactional context of the particle: co-occurrence of *voilà* with tokens and interactional phenomena that are tagged in the base (i.e. overlap and pauses), and location of the particle within the turn-at-talk.

Various means to access these results are usable in the database: either automatic search of co-occurrences ('co-occurrences of a token', 'contextual analysis of a token'), or 'complex search' ('requêtes multicritères') in which the researcher can choose the number and the type of criteria he/she wants to use.

3.3.1. *Co-occurrences of tokens*

The automatic search of co-occurrences shows that the left- and right-hand co-occurrent tokens of *voilà* are mainly other discourse particles. What emerged from the analysis of a single corpus (Bruxelles and Traverso 2006) is thus not linked to a specific corpus or to an interaction genre, but holds for a variety of corpora. The automatic search of co-occurrences indicates that, in more than half of the cases (56%), *voilà* appears in a series of particles.

A second result concerns the particles that are used together with *voilà*. The most frequent is *ben* (which could be translated in many of its usages into 'well'): 16 occurrences in the first left-hand position ('ben voilà'); 20 occurrences in the second left-hand position ('ben x voilà'). This frequent co-occurrence had not been identified in the qualitative study, although it stands for the three usages that have been described: *ah ben voilà* is a typical *deictic* marker; *ben voilà* is used for introducing reported speech (and marks a transition); at the end of a turn, sometimes following 'et' ('and'), 'ben voilà' has a conclusive value (for instance, it is uttered by a dentist at the end of a dental treatment that has lasted for about two minutes – corpus Dentiste).[9] A further examination of this series *ben voilà* would be of interest, specifically when compared with the reverse series *voilà ben*, which only appears four times in the data.

The next most frequent token in the left-hand context of *voilà* is *puis/pis* ('puis/pis voilà', 'then voilà'): 16 occurrences. Interestingly, it also seems to function with a conclusive value, although *puis* ('then') semantically denotes that something will be added (an item in a list, the next episode in a story, etc.).

In the right-hand context, the most frequent co-occurrent token is not a particle, but the pronoun *c'* (19 occurrences, out of which 18 are 'c'est', 'it is'). The second most frequent token is a second occurrence of *voilà* ('voilà voilà'): 12 occurrences (out of which seven take place at the beginning of the turn, and two consist of a

[9] This usage is similar to the *co-gestural voilà* described by Filliettaz (2001).

https://doi.org/10.1017/S0959269507003195 Published online by Cambridge University Press

series of three 'voilà'). The next most frequent token in the right-hand context of *voilà* is the affirmative morpheme *oui*, and its variant *ouais* (9 'oui'; 7 'ouais'). Out of these 16 occurrences, 9 are composed of the two tokens alone 'voilà oui/ouais'. Two additional observations can be made concerning this result. First, it confirms the use of *voilà* as an agreement marker (Bruxelles and Traverso's results, table 1, column 3). Significantly, *non* ('no') rarely occurs along with *voilà* (2 occurrences), and there is no occurrence of *si* (with its negative meaning) in this position. Second, the reverse position of the two tokens ('oui' in the left-hand context of 'voilà', 'oui voilà') only occurs four times in the data. Moreover, this last usage does not correspond to a collocation, but to the successive production of the two tokens separated by a pause or uttered with opposing prosodic formats. This clarification is made possible by the immediate access to the signal that the database permits. This access constitutes an important asset of the database, for it facilitates the checking of the transcript, thus avoiding errors due to an inaccurate transcription.

An important outcome of the automatic search of co-occurrences of *voilà* is to have shown the importance of the clusters of particles. Furthermore, it suggests that an order of particles within the clusters can be significant. This idea of a sequential order of particles and interactional phenomena has already been verified in previous qualitative studies, specifically in transitions (Mondada and Traverso 2005; Bruxelles, Greco and Mondada forthcoming). The possibility of automatic search with the CLAPI database tools is a valuable source of help for this type of investigation.

3.3.2. Location of voilà within the turn

With the aid of the tool 'contextual analysis of a token', we have access to the different positions of a token within the turn. *Voilà* presents a very specific functioning in this respect: it is used as frequently at the beginning of the turn as at the end (in 50% of its occurrences, it is the first token in the turn; in 39%, it is the last). By comparison, the location of the other particles within the turn is clearly different: *ben* is in 40% of its occurrences at the beginning of the turn and in 3% at the end; *donc* is in 25% of its occurrences at the beginning of the turn and in 5% at the end; *mais* is in 25% of its occurrences at the beginning of the turn and in 6% at the end. This specificity of *voilà* is to be understood in relation to the turn length:[10] *voilà* is as frequent at the beginning as at the end of the turn, because it frequently occurs in short turns, i.e. fewer than four tokens. This usage represents 38% of its occurrences (four times more often than the other particles).

This result is consistent with Bruxelles and Traverso's (2006) findings on the use of *voilà* as a deictic or an agreement marker, for which they have identified that the marker constitutes the complete turn or is used in short turns.

[10] Requests on turn length are also available with the tool 'contextual analysis of a token'. The use of the 'complex search' tool, on the other hand, permits us to work in this aspect in a more detailed way, by searching a precise position within the turn and/or a precise turn length.

https://doi.org/10.1017/S0959269507003195 Published online by Cambridge University Press

3.3.3. Co-occurrence of voilà with other interactional phenomena

The two interactional phenomena that are tagged in the CLAPI database (in addition to the *turns*, *tokens* and *participants*) are pauses and overlaps.

- Pauses: *voilà* is used twice as frequently after a pause (12%) than before one (6%). In the latter case (and at a distance of three tokens before the pause), it is noteworthy that the most frequent usages are not *voilà* as deictic, but those of *voilà* in transitions. This also confirms what numerous previous studies on transitions had shown: on achieving transitions, participants make use of long and numerous pauses. This emphasises the temporal unfolding of interactional activities.
- Overlap: *voilà* frequently occurs in overlap, and specifically in the overlapping segment, 15%, (versus 2% in the overlapped segment). This specificity is related to the use of *voilà* as a deictic or an agreement marker (cf. ICOR, forthcoming b).

On the basis of this overall view of *voilà* in the database, we turn now to a detailed examination of the particle in what has emerged as an intriguing and recurrent context: *voilà* preceded by an uncompleted syntactic construction. This configuration has been discovered by observing the possible materials occurring before and after *voilà* and by taking into account their retrospective or prospective orientations, as well as their sequential functions (as closing a particular unit). However, these features have not been identified directly by an automatic search, but thanks rather to the possibility of accessing the transcript and the signal of the stretches of interaction that have been found as hits in response to an automatic query. This configuration had not been noticed by Bruxelles and Traverso (2006). It is nevertheless interesting in comparison with their results for it may challenge their distinction between the two categories of structuration (which has been related to transitions) and agreement marker (to which a closing value has been attributed), as well as between prospective and retrospective orientations of particular markers (such as *et puis*). More specifically, it raises the issue of the possibility and the relevance of sorting out closing and opening values for *voilà* as a structuration marker, since it can co-occur with resources that have either a forward orientation or a backward orientation.

3.4. Back to qualitative analysis: Voilà as an interactional hinge

Automatic searches have confirmed but also revealed different characteristic positions of *voilà*, as well as various other linguistic materials immediately preceding or following it. These findings invite us to return to the corpus for a qualitative in-depth description of some of these environments, in order both to refine the analysis and to give hints for further developments of more sophisticated search tools.

The study of co-occurrences of *voilà*, coupled with tokens, pauses and overlaps by using the 'complex search' tool has enabled us to isolate a recurrent environment

https://doi.org/10.1017/S0959269507003195 Published online by Cambridge University Press

that seems interesting for highlighting some of the jobs achieved by *voilà* as an interactional resource. This environment is constituted of a syntactically uncompleted Gestalt followed by *voilà* in turn-final position.

Here is a first set of excerpts showing this phenomenon in a particular configuration:[11]

Excerpt 5. (LyonSaxe 54.30)

```
  1 SOP  donc tu sais euh::
  2 FAB  oui tu sais à quoi tu t'exposes/=
  3 SOP  =+voilà#
          +geste symétrique, palmes ouvertes, verticales-->
    im         #im1
  4       (0.3)+
           -->+
  5 FAB   t'es [pas pris en traît'e]:
->6 SOP        [donc t' acceptes le-] (.) +voilà\+
                                          +hoche la tête+
```



1

Excerpt 6. (Mosaic 11.46.55)

```
   1 C  est-ce que les gen:s/ (.) quand tu vas arriver ici/[si&
   2 M                                                     [mm
   3 C  &tu vois les portes des sanitaires/ ça c'est pa::s=
   4 M  =non mais ça a pas besoin d'avoir
   5    une porte [forcément\ (.) tu peux tu] peux&
-> 6 L              [non mais tu peux avoir: euh voilà/]
   7 M  &avoir [un couloir
   8 C         [un retrait
```

In both cases, *voilà* in final position closes a turn, which is produced by a second speaker overlapping the current speaker. In both cases, this turn is syntactically incomplete: in excerpt 5, line 6, the article ('le') projects a noun which is not uttered; in excerpt 6, line 6, the verb 'avoir' projects an argument which is not uttered. In both cases, the overlapping speakers withdraw in favour of the overlapped speakers. Moreover, both produce a contribution that acts out the same kind of

[11] All the excerpts analysed in this part are extracted from video-recorded data, in order to give access to the multimodal features of the participants' practises.

https://doi.org/10.1017/S0959269507003195 Published online by Cambridge University Press

relevant next action at that sequential moment: in excerpt 5, after a slight pause (4), FAB and SOP self-select almost at the same time, for proposing a kind of conclusion to what precedes; in excerpt 6, L answers to C with a negative utterance (5) which is very similar to M's answer (4).

So, in both cases, two speakers are not only overlapping but also producing a similar kind of turn, responsive to a previous one. The second speaker seems to withdraw before completing the turn s/he is engaged in. S/he thus displays that her/his own turn needs not to be completed, since it is achieved by the first, overlapped turn. In this sense, the syntactical incompleteness of the second turn is a product of the orientation towards the fact that action completion is achieved by the first. In this environment, *voilà* introduces a right boundary of an otherwise syntactically uncompleted turn: it closes the Gestalt.

We can notice that, in excerpt 5, a first collaborative conclusion has already been produced just before (1–2), closed by the one-unit turn 'voilà', co-occurring with a *closing gesture* (image 1).

The environment described in these two excerpts displays one of the interactional jobs achieved by *voilà*, consisting in closing a Gestalt which otherwise could appear as (syntactically) unachieved.

This invites us to search for other cases in which *voilà* occupies the final position of an uncompleted turn:

Excerpt 7. (Mosaic 11.49.22) [cf. excerpt 3)

```
1 C  donc on arrive par là/ et puis là on va trouver une
2    porte forcé [ment par [là/
3 M              [ouais    [je ouais [voilà
4 C                                  [là hein/
```

Excerpt 8. (LS 1.26.26)

```
1 JEB tu prends (juste) du mieux
2     (0.2)
3 FAB hm
4 SOP tu prends le [moins] le [moins] euh [voilà\]
5 JEB              [hm]       [hm]        [et en] plus/il est
6     situé comme ça/ c'à-dire géographi[quement le le premier prix]
```

Excerpt 9. (LS 41.15)

```
1 SOP donc on (.) i's m'avaient dit entre sept et huit cent/
2     (0.7)
3 FAB .h::\
4 SOP et pis on partait sur cinq cents xx/ comme ça/ moi
5     ça m' fait/ euh mille cinq/ euh::::
6     (0.9)
7 SOP °et puis euh voilà/°
8     (1)
```


https://doi.org/10.1017/S0959269507003195 Published online by Cambridge University Press

In these three excerpts, unlike the case in the previous ones (5, 6), we don't have two concurrent turns achieving the same action, but we still have a syntactically uncompleted turn, closed by a final 'voilà'.

In excerpt 7, M begins with 'je' (3) projecting a verb but not producing it; in excerpt 8, SOP engages in the production of a comparative expression with 'le moins' (4) but self-repairs it and does not complete it; in excerpt 9, SOP, after a pause, adds an extension to her previous turn with 'et puis euh' (7) but does not continue. In all the cases, the speaker starts a syntactical structure but stops before completing it: (s)he closes the turn by 'voilà', thus displaying that, although syntactically uncompleted, it has reached its interactional completion.

Thus, *voilà* is a resource by which speakers mark the turn as complete, in spite of its syntactic incompleteness. This usage shows that participants in interaction rely on various kinds of turn completion and that completion is locally defined within the ongoing course of interaction, in a flexible and changeable way.

This feature of *voilà* is also attested in a similar environment, constituted by 'et puis voilà' (as in excerpt 9 above) or 'et voilà', in which the particle immediately closes a construction, even if something more has been projected.

In the following two cases, it is possible to contrast, within the same excerpt, two uses of 'et puis': the first projects *and* implements more to come; the second is immediately followed by 'voilà':

Excerpt 10. (LS 43.00)

```
   1 JEB  [ouais:/ euh je pense que il faut pas l' rater:/]
-> 2 SOP  [et pis o- et pIS/ on IRA VOIR la direction]
   3      marketing/ de ces mêmes marques là/
   4      (0.+5)+
     sop     +..+
-> 5 SOP  +et pis [voiLA/]
          +se tourne vers JEB--->
   6 FAB         [(si) on arrive] à rentrer là de+dans puis
     sop                                       ---+
   7      développer +sur le [fond
   8 SOP                     [mais oui/ (.) voi+là=
     sop             +reg vers JEB-------------+reg bas-->
```

Excerpt 11. (LS 1.20.44)

```
   1 FAB  le le truc de base c'est: les [premiers prix&
   2 JEB                               [ouais]
   3 FAB  &en bas/ l[a marque distributeur] au niveau des yeux/
   4 JEB           [les premiers prix en bas]
   5 SOP  [ouais]
   6 JEB  [ouais]
   7      (0.2)
-> 8 JEB  et puis euh:
-> 9 FAB  et les marques euh:: LE haut d' gamme en haut/
   10 JEB  hm hm
   11 FAB  et les les marques leaders autour de la mar[que euh]&
```


https://doi.org/10.1017/S0959269507003195 Published online by Cambridge University Press

```
   12 SOP                                                    [d'accord]
   13 FAB  &de la èm dé dé\#
      im                    #im1
   14      (0.2)
-> 15 SOP  +et# [+voi#là] #+
           +....+geste horiz de la main balayant dev elle+
      im      #im2 #im3 #im4
```









In excerpt 10, SOP is proposing various initiatives to take within a business strategy: the first one is formulated with 'et pis' + an action description ('on IRA VOIR la direction', 2); the second one, added after a pause (4), is not completed and followed by 'voilà' (5).

Similarly, in excerpt 11, the participants are collaboratively elaborating the description of the way in which objects are spatially distributed in supermarkets: FAB introduces a first explanation (1–3), which is reformulated by JEB, line 4. JEB pursues his description, in line 8 ('et puis euh'), and is collaboratively joined by FAB, who produces the next part of the turn (9 and then 11, 13). In line 15, SOP closes this explanation by 'et voilà', accompanied by a *closing gesture* (images 2, 3 and 4).

Thus, in both cases, we have first a use of *et puis* followed by a syntactical turn completion and a second use of *et (puis)* followed by 'voilà'. In the latter case, 'voilà' closes down a Gestalt that is still open.

These analyses focus on environments where *voilà* closes an uncompleted Gestalt. This accounts for the participants using *voilà* as an interactional resource for doing

https://doi.org/10.1017/S0959269507003195 Published online by Cambridge University Press

the closings in an environment where completion is not achieved at the formal level. Moreover, this draws attention not only to the position of *voilà* within the turn, but also to the resources that co-occur before and after it. These resources can be characterised by:

- their completion or incompletion;
- their stronger or weaker projective potential;
- their possible orientation not only onwards – prospectively – but also backwards – retrospectively.

Voilà deals with these features, which can coexist within a multi-unit turn: a multi-unit turn can be incrementally formatted in such a way that it is oriented either to the previous or to the next turn, this direction being possibly reoriented during the temporal unfolding of the turn. *Voilà* works as a closing device, distinguishing between these Gestalts and providing for their boundaries. This can be seen in the following two excerpts:

Excerpt 12. (LS 1.18.12)

```
   1 JEB  donc tu vois (0.3) ET les <((en riant)) biscuits
   2      apéro> à l'asperge/ (.) pa'ce que le mec le lend'main
   3      matin ah h <((en riant)) tu puais l'asperge> ((rire))
   4 SOP  j' les ai jamais vus moi ceux là
   5 FAB  non i's sont pas restés longtemps [xx]
   6 JEB                                   [BON\]
   7      (0.4)
-> 8 JEB  c'est euh (0.4) VOIlà\ (.) Donc tu as cet enjeu d'
   9      l'innovation ((continues))
```

Excerpt 13. (LS 1.19.45)

```
   1 SOP  et là après tu rentres euh produits euh: euh
   2      classiques entre gui'mets/ après t' as les c' qu'i's
   3      appellent euh pâtisseries euh:: bronson euh: machin là
   4 FAB  ouais pâtisseries élabo[rées]
-> 5 SOP                        [ouais] voilà:/ après t' as
   6      les biscuits d'antan euh les recettes euh grand mère
   7      euh tartinette et compagnie/
```

In excerpt 12, JEB tells an anecdote about an asparagus-flavoured biscuit and then goes on with a more general topic concerning innovative products. Line 8, after the punch line of the story (3), and SOP's (4) and FAB's (5) comments, JEB initiates the closing of this sequence in overlap (6). After a pause (7), he nevertheless initiates a turn projecting more to come ('c'est euh' 8), then immediately closes it by 'voilà', although it is uncompleted (see the previous excerpts for similar cases). After it, 'donc' opens another prospective development.

https://doi.org/10.1017/S0959269507003195 Published online by Cambridge University Press

In excerpt 13, we have another kind of material preceding 'voilà': 'ouais' orients retrospectively in a responsive and aligning way to what precedes; after 'voilà', 'après' is clearly oriented forwards.

Thus, *voilà* can work as a hinge between different kinds of materials, oriented to different directions. Further inquiries will have to explore both sequential positions (before and after *voilà*) and formal resources, which can be implemented within them, having different orientations (prospective or retrospective). Together, these configurations of forms and positions define a *positional grammar* invoked by interactional linguistics.

4. CONCLUSIONS

In this paper, we have given an overall view of how quantitative and qualitative approaches can be used jointly in interactional linguistic research. Specifically, we have underlined how automatic research in a large range of data can support detailed analysis of interactional phenomena.

The first part of the paper was devoted to the CLAPI database, presenting its functionalities and the various opportunities it offers for studies in different linguistic fields concerned with the spoken language. The CLAPI database is both an archive of naturally-occurring interactional data, easily browsable thanks to a rich set of metadata, and a workbench encompassing a series of automatic search tools. Although the base has been designed and developed from an interactional point of view, it is also a resource for other types of approaches to spoken interaction.

The analysis presented in the second part of the paper has shown how a careful consideration of similar sequential environments leads to identification of recurring interactional configurations, made of various interactional resources used by participants. It has illustrated how the constitution of a collection of cases for exploring these sequential environments is made easier and even made possible in larger corpora by search tools able to identify various items (such as *voilà* combined in an ordered way with other items, *et puis* and/or *après*, *donc*, etc.), to specify the sequential position of certain items (*voilà* in a turn-final position, *voilà* within or just after an overlap), to recognise concomitant gestures ('closing gesture' simultaneous with *voilà*) and even to detect uncompleted turn-constructional units. As the study has shown, currently available CLAPI tools do not automatically produce the collection of relevant cases: this has to be done manually. Therefore, search tools represent a more heuristic way of browsing large corpora than a tool that automatically produces final results; moreover, this manual dimension of the search also involves the essential check and refinement of the transcripts, allowing for the discovery of new relevant features (to give only one example, the resolution onset of the overlap has been transcribed during the analysis, since it was not annotated in the original transcripts; all transcripts have been refined for the particular purposes of the analysis). The use of the database involves both automatic and manual analyses: in turn, the latter allow the development of new search tools or functions, which originate in analytical practices and respond to analytical needs.

https://doi.org/10.1017/S0959269507003195 Published online by Cambridge University Press

Address for correspondence:
Lorenza Mondada, Véronique Traverso
ICAR UMR 5191 CNRS
Ecole Normale Supérieure Lettres et Sciences Humaines
15, Parvis René Descartes
BP 7000
69342 LYON cedex 07
France
e-mail: lorenza.mondada@univ-lyon2.fr, veronique_travers@univ-lyon2.fr

Transcription conventions

Data were transcribed according to the ICOR convention (see http://icar.univ-lyon2.fr/projets/ICOR/ICAR_Conventions_ICOR.pdf).

Multimodal details have been transcribed according to the following conventions:

**	delimitate participant's gestures and actions descriptions.
++	(one symbol per participant).
*–>	gesture or action described continues across subsequent lines.
*–>>	gesture or action described continues until and after excerpt's end.
*>>	gesture or action described begins before end of excerpt.
—>*	gesture or action described continues until the same symbol is reached.
....	gesture's preparation.
—	gesture's apex is reached and maintained.
,,,,,	gesture's retraction.
im	the exact point where screen shot (image) has been taken is indicated,
#	with a specific sign showing its position within turn-at-talk.

REFERENCES

Auchlin, A. (1981). Réflexions sur les marqueurs de structuration de la conversation. *Etudes de Linguistique Appliquée*, 44: 88–104.

Auer, P. (2005). Syntax als Prozess. *InLiSt*, 41, URL <http://www.uni-potsdam.de/u/inlist/issues/41/index.htm>

Balthasar, L. and Bert, M. (2005). La plate-forme Corpus de Langues Parlées en Interaction (CLAPI) : Historique, état des lieux, perspectives. *Lidil*, 31: 13–33.

Bange, P. (1987). *La Dame de Caluire : une consultation*. Bern: Peter Lang.

Baude, O. (ed.) (2006). *Corpus oraux. Guide des bonnes pratiques 2006*. Paris-Orléans: CNRS éditions and PUO.

Bruxelles, S., Greco, L. and Mondada, L. (to appear). Pratiques de transition: ressources multimodales pour la structuration de l'activité. In: F. Détienne and V. Traverso (eds.), *Méthodologies d'analyse de situations coopératives de conception: Corpus MOSAIC*. Nancy: PUN.

Bruxelles, S. and Traverso, V. (2002). Les corpus de langue parlée en interaction au GRIC. In: K. Pusch and W. Raible (eds.), *Romanistiche Korpuslinguistik*. Tübingen: Gunter Narr Verlag, pp. 59–70.

https://doi.org/10.1017/S0959269507003195 Published online by Cambridge University Press

Bruxelles, S. and Traverso, V. (2006). Usages de la particule *voilà* dans une réunion de travail: analyse multimodale. In: M. Drescher and B. Job (eds.), *Les marqueurs discursifs dans les langues romanes: approches théoriques et méthodologiques*. Bern: Lang, pp. 71–92.

Clift, R. (2001). Meaning in interaction. The case of *actually*. *Language*, 77/2: 245–291.

Cosnier J. and Kerbrat-Orecchioni C. (eds.) (1987). *Décrire la conversation*. Lyon: PUL.

Couper-Kuhlen, E. and Selting, M. (eds.) (1996). *Prosody in Conversation*. Cambridge: Cambridge University Press.

Détienne, F. and Traverso, V. (eds.) (to appear). *Méthodologies d'analyse de situations coopératives de conception: Corpus MOSAIC*. Nancy: PUN.

Filliettaz, L. (2001). L'hétérogénéité sémiotique de la gestualité en contexte transactionnel. De la gestualité coverbale à la verbalité cogestuelle. In: C. Cavé, I. Guaïtella and S. Santi. (eds.), *Oralité et gestualité. Interactions et comportements multimodaux dans la communication*. Paris: L'Harmattan, pp. 401–404.

Ford, C., Fox, B., and Thompson, S. (2002). Constituency and the grammar of turn increments. In: C. Ford, B. Fox and S. Thompson. (eds.), *The Language of Turn and Sequence*. Oxford: Oxford University Press, pp. 14–38.

Fox, B. (2007). Principles shaping grammatical practices: an exploration. *Discourse Studies*, 9: 299–318.

Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, 32: 1489–1522.

Goodwin, C. (2003). Conversational frameworks for the accomplishment of meaning in aphasia. In: C. Goodwin. (ed.), *Conversation and Brain Damage*. Oxford: Oxford University Press, pp. 90–116.

Groupe ICOR (2007). Variations interactionnelles et changement catégoriel: l'exemple de 'attends'. *Langues, Cultures, Interaction. Actes du Colloque du Réseau Français de Sociolinguistique, Paris, Oct. 2005*. Paris, L'Harmattan.

Groupe ICOR (to appear a). *Voilà*, approche qualitative et investigations quantitatives. L'apport de la plate-forme outillée de Corpus de LAngue Parlée en Interaction (CLAPI). *3e rencontre Fribourgeoise de la linguistique de corpus*. Tübingen: ScriptOralia, Günter Narr.

Groupe ICOR (to appear b). Exploitation de la plate-forme Corpus de Langue Parlée en Interaction (CLAPI) pour l'étude interactionnelle des particules discursives en chevauchement. In: S. Bruxelles, L. Mondada, A.-C. Simon and V. Traverso. (eds.), *Grands corpus de français parlé: bilan historique et perspectives, Interactions & Langages*, 1.

Heritage, J. and Watson, D. (1979). Formulations as conversational objects. In: G. Psathas. (ed.), *Everyday Language*. New York: Irvington Publishers, pp. 123–163.

Hopper, P. (1987). Emergent grammar. *BLS*, 13: 139–157.

Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In: G. Lerner. (ed.), *Conversation Analysis: Studies from the First Generation*. Amsterdam: Benjamins, pp. 13–31.

Katsiki, S. and Traverso, V. (2004). Les dénominations ordinaires spontanées des activités langagières et la question des équivalences entre les communautés discursives. *Langages*, 154: 47–59.

Local, J., Kelly, J. and Wells, W. (1986). Towards a phonology of conversation: turn-taking in Tyneside English. *Journal of Linguistics*, 22: 411–437.

Mondada, L. (2005). L'analyse de corpus en linguistique interactionnelle: de l'étude de cas singuliers à l'étude de collections. In: A. Condamine (ed), *Sémantique et corpus*. Paris: Hermès, pp. 76–108.

https://doi.org/10.1017/S0959269507003195 Published online by Cambridge University Press

Mondada, L. and Traverso, V. (2005). (Dés)alignements en clôture. *Lidil*, 31: 35–59.
Ochs E., Schegloff E. A. and Thompson S. (eds.) (1996). *Interaction and Grammar*. Cambridge: Cambridge University Press.
Plauché, M. and Bergen, B. (1999). The evolution of binary spatial deictics: French *voilà* and *voici*. In: S. Chang, L. Liaw and J. Ruppenhofer. (eds.), *Proceedings of the 25th Annual Meeting of the Berkeley Linguistics Society*, URL <http://www.icsi.berkeley.edu/~bbergen/MarkednessPaper.PDF>
Plauché, M. and Bergen, B. (2001). *Voilà voilà*: extensions of deictic constructions in French. In: A. Cienk, B. Luka and M. Smith. (eds.), *Conceptual and Discourse Factors in Linguistic Structure*, URL <http://www.icsi.berkeley.edu/NTL/papers/BergenCSDL4.pdf>
Roulet, E., Auchlin, A., Moeschler, J., Rubattel, C. and Schelling, M. (1985). *L'articulation du discours en français contemporain*. Bern, Frankfurt, New York, Paris: Peter Lang.
Schegloff, E. A. (1968). Sequencing in conversational openings. *American Anthropologist*, 70/6: 1075–1095.
Schegloff, E. A. (1979). The relevance of repair to syntax-for-conversation. In: T. Givon (ed.), *Syntax and Semantics,* 12: *Discourse and Syntax*. New York: Academic Press, pp. 261–286.
Schegloff, E. A. (1996). Issues of relevance for discourse analysis: Contingency in action, interaction and co-participant context. In: E. H. Hovy and D. Scott. (eds.), *Computational and Conversational Discourse: Burning Issues. An Interdisciplinary Account*. Heidelberg: Springer Verlag, pp. 3–38.
Schegloff, E. A., Jefferson, G. and Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 53/2: 361–382.
Schegloff, E. A. (1979). Identification and recognition in telephone conversation openings. In: G. Psathas. (ed.), *Everyday Language: Studies in Ethnomethodology*. New York: Irvington Publishers, pp. 23–78.
Selting M. and Couper-Kuhlen E. (eds.) (2001). *Studies in Interactional Linguistics*. Amsterdam: Benjamins.
Selting, M., Auer, P., Barden, B., Bergmann, J., Couper-Kuhlen, E., Günthner, S., Quasthoff, U., Meier, C., Schlobinski, P. and Uhmann, S. (1998). Gesprächanalytische Transkriptionsysteme (GAT). *Linguistische Berichte*, 173: 91–122.
Traverso, V. (2005). Quelques formats intégrant la répétition comme ressource pour le développement thématique dans la conversation ordinaire. *Rivista di Psicolinguistica Applicata*, Special Issue on Conversation Analysis, IV-2/3: 153–166.
Traverso, V. (to appear). Co-élaboration de solutions et rôle du graphico-gestuel: analyse interactionnelle. In: F. Détienne and V. Traverso (eds.), *Méthodologies d'analyse de situations coopératives de conception: Corpus MOSAIC*. Nancy: PUN.
Wootton, A. J. (1997). *Interaction and the Development of Mind*. Cambridge: Cambridge University Press.

Internet websites

CLAPI (Corpus de LAngue Parlée en Interaction), URL <http://clapi.univ-lyon2.fr>
CORINTE (CORpus d'INTEraction): URL <http://icar.univ-lyon2.fr/projets/corinte/>
Talk Bank: URL <http://talkbank.talkbank.org/>

https://doi.org/10.1017/S0959269507003195 Published online by Cambridge University Press