Article

Language & Linguistics 語言暨語言學

Language and Linguistics
16(1) 69–92

DOI: 10.1177/1606822X14556605
lin.sagepub.com

SAGE

Processing Units in Conversation: A Comparative Study of French and Mandarin Data*

Laurent Prévot,[1,2] Shu-Chuan Tseng,[3] Klim Peshkov,[1,2] and Alvin Cheng-Hsien Chen[4]

Aix Marseille Université[1]
CNRS[2]
Academia Sinica[3]
National Changhua University of Education[4]

Human spoken language production is directed towards communication delivering comprehensible information to recipients. Speech segmentation into small units efficiently enhances a sensible and interpretable discourse structure. Such processing units in real-life communication may be applied to semantic, syntactic, or prosodic structures. Previous studies have proposed various theories of speech segmentation, mainly based on qualitative analyses. The present study utilizes corpus-based quantitative data to examine how conversational speech in French and Mandarin is structured in terms of three different processing units, and how these units interact with one another. Unit completion location was identified by semantic structure (*discourse unit*), prosodic pattern (*prosodic unit*), and sequences of parts of speech (*chunk*). Quantitative analyses for both languages were carried out by applying comparable processing procedures. This article presents our efforts to establish a dataset for two typologically diverse languages, and to carry out quantitative comparative studies of processing units in face-to-face conversation.

Key words: chunks, discourse, processing units, prosody

1. Introduction

Discussion of processing units of natural spoken language is directly grounded on the theory of discourse structuring (Grosz & Sidner 1986). Semantic processing units may be based on propositional properties and logical relations (Frederiksen 1977; Hobbs 1978), while processing units defined by grammatical structures mainly account for sentence grammar in terms of both surface and deep structures (Ford & Holmes 1978; Myers 1997). In addition to a syntactic account, prosodic presentation, especially intonation contours and pausing in conversation, substantially contributes to discourse structuring (Butterworth 1975; Clark & Wasow 1998; Selting 1996). Although

* This work has been realized thanks to the support of the France-Taiwan ORCHID Program, under grant 100-2911-I-001-504 and the NSC project 100-2410H-001-093 granted to the second author, as well as ANR OTIM BLAN08-2-349062 for initial work on the French data. We would like also to thank our colleagues for help at various stages of the data preparation, in particular Roxane Bertrand, Yi-Fen Liu, Robert Espesser, Stéphane Rauzy, Brigitte Bigi, and Philippe Blache. The third author of this paper would like to thank Provence-Alpes-Côte d'Azur and *Semantia* for funding his PhD.

the main stream of syntax–prosody interface supports the notion that prosodic structure orients to the syntactic pattern (Selkirk 1986), the semantic implications derived from the original syntactic pattern can be overridden by commonly used prosodic means (Ford & Thompson 1996; Selting 1996). A considerable number of discussions on semantic, syntactic, and prosodic processing units have been proposed. Interest in studying the interface of discourse and prosody has accordingly developed in the last decade, as illustrated by the vitality of the events and projects in this domain, for example the Prosody–Discourse Interface conference series. However, despite the considerable number of theoretical proposals and descriptive works, quantitative systematic studies are less widespread due to the cost of creating resources usable for such studies. Indeed, prosodic and discourse analysis are delicate matters requiring lower-level processing, such as alignment with speech signal at syllable level (for prosody), or at least basic syntactic annotation (for discourse). It is also the case that many of these studies deal with read or monologue speech. The extremely spontaneous nature of conversational speech renders the first levels of processing complicated. Previous works (Afantenos et al. 2012; Bertrand et al. 2008; Blache et al. 2009; Chen 2012; Liu & Tseng 2009) gave us the opportunity to produce conversational resources of this kind. We then took advantage of a bilateral project for working on conversational speech in a quantitative fashion, for two typologically diverse languages: French and Taiwan Mandarin. Studies of conversational speech based on large and richly annotated corpora are rare, due to the scarcity of such resources; comparative studies are even more rarely found, because of issues of cross-linguistic comparability in terms of content, genre, and speaking style. We believe this combination of linguistic resources and skills is rather unique, and allows for comparative quantitative experiments on high-level linguistic analyses such as discourse and prosody. Our objective is to understand the commonalities and differences in discourse–prosody interface in these two languages.

This paper is structured as follows. In §2 through §5, we describe how we built a comparable dataset from existing corpora by introducing the characteristics of our datasets: prosodic segmentation, discourse segmentation, and automatic syntactic chunking. In §6, the results of comparing prosodic and discourse units in French and Mandarin are discussed in a comparative and quantitative way, followed by a general discussion with more related work in §7.

2. Building comparable corpora

Our corpora from the two languages were collected in very similar conditions: they both consist of face-to-face conversations recorded in an anechoic chamber. While the recording conditions, settings, and tasks were the same, in the Mandarin data the conversational partners were strangers to each other, while in the French data the partners knew each other well. Despite this difference, we believe the data to be largely comparable across languages, and we point out potential issues at relevant places in the presentation and discussion of results. For the present study, data of a comparable size were extracted from the respective original corpora. While we had to decide which linguistic information and which part of the full corpora to include in our joint dataset, long narrative turns (as in the first two examples) were extracted from the original corpora, as in principle the discourse–prosody interface should be better observed in long speech stretches. However, the consequences of random sampling of two different datasets are unavoidable and future work will work with a bigger and better-controlled sample. As a result, the French dataset is a subset of

three female and three male speakers' data from the 8-hour Corpus for Interactional Data (CID) (Bertrand et al. 2008).[1] The 73-minute speech data are time-aligned with the speech signal at phone level by using forced alignment techniques (Illina et al. 2004), with POS information tagged (Blache et al. 2008). The Mandarin dataset is a subset of seven male and nine female speakers' data from the 42-hour Taiwan Mandarin Conversational Corpus (the TMC Corpus)[2] (Tseng 2013). Like the French dataset, the 205-minute Mandarin speech data are also time-aligned with the speech signal at phone level with automatically tagged POS information (Chen 2012; Liu et al. 2014). Both French and Mandarin datasets were previously annotated with different definitions of prosodic units and discourse units. For the current study, an annotation approach for annotating prosodic phrasing and segmenting the corpus into discourse units had been taken, adjusting the prior annotation to meet cross-linguistically comparable guidelines.[3]

(1) French narrative sequence extracted
c'est vrai que euh # je me suis retrouvée quand même bien souvent prise euh # dans un peu dans un étau # entre finir une these et puis euh # être disponible pour quelqu'un qui supporte difficilement qu'on passe son temps à travailler # alors là quand tu es dans ce genre de situation qu'est-ce que tu veux faire
it is true that uh # I found myself indeed very often taken into uh # in kind of in a vice # between finish a PhD and then uh # be available for someone that has difficulties accepting that we are working all the time # so when you are in this kind of situation what do you want to do

(2) Taiwan Mandarin turn extracted
Nangang guoqu de hua # wo jiu bijiao bu qingchu # xiang wo meitian shangban shi cong # jieyun Yongchun Zhan # da jieyun dao NEIGE # jieyun # Zhongxiaofuxing Zhan # zai zhuan NEIGE Muzhaxian # dao Nanjing Zhan xiache # zai da yicheng gongche Jianguo-beilu kou #
(*If you*) *go from Nangang # I don't know it exactly # For instance I go to work every day from # Yongchun MRT station # take the MRT to THAT* (*hesitation*) *# MRT # Zhongx-iaofuxing station # then transfer THAT* (*hesitation*) *to Muzha line # to Nanjing station get off* (*there*) *# then take a bus to the intersection at Jianguo North Road #*

2.1 Dataset

We kept unchanged the original signal-aligned syllable and word boundaries of both languages and made adjustments to the annotations of prosodic units (PU) and discourse units (DU) to make the definition of the unit boundaries consistent across the two languages. We shall go into the details

[1] For details about the CID Corpus please refer to http://sldr.org/sldr000720.

[2] For details about the TMC Corpus please refer to http://mmc.sinica.edu.tw.

[3] The annotation campaign was financially supported by the French OTIM-ToMA (Blache et al. 2009) and France-Taiwan ORCHID projects.

Table 1: Data summary

	Word	PU	DU
French	22,359	2,926	2,050
Mandarin	38,704	8,562	6,041

Table 2: The most frequent POS tags in French and Mandarin

Word category	French	Mandarin
Nouns (*N*)	N	N
Pronouns (*Pro*)	P	Nh
Determiners (*Det*)	D	Ne
Verbs (*V*)	V	V
Particles, Discourse markers, Interjections (*Part*)	I	T, I, FW
Adverbs (*Adv*)	R	D
Adjectives (*Adj*)	A	A
Prepositions (*Prep*)	S	P

of annotation criteria in the next section. The end result of our dataset is summarized in Table 1. Given the differences in the POS tagsets used in the French and Mandarin data, we established a list of word categories by collapsing the POS tagsets in the two languages, as shown in Table 2 (Blache et al. 2008; Chen et al. 1996).

3. Annotating prosodic units

Prosodic units were originally annotated in the two languages according to different theories. From the perspective of discourse segmentation, we intended to annotate the prosodic unit completion location, basically following the perception of prosodic phrasing. We acknowledge the fact that language planning should work with a certain kind of structure and hierarchy, which may be expected to result in different types of prosodic phrasing. While we are of the opinion that prosodic phrasing is definitely not purely linear and sequential, it was not our intention to pinpoint the complete prosodic hierarchy presented in our spontaneous conversational speech. But for the current study of discourse segmentation, a design with single-layer prosodic phrasing suffices and makes it easier to achieve reasonable inter-labelers agreement. In praxis, we managed to merge two intermediate levels of the units in the French data to align the prosodic units in it with the definition of prosodic units applied to the Mandarin data. As it is difficult to compare the validity of the defined prosodic units across languages, in order to ensure the comparability of annotation criteria, the authors conducted a cross-language segmentation experiment on a small subset of the data. Each tried to annotate prosodic units in the other language. The annotation results conducted by the non-native labelers confirmed that the main cues used for segmenting the prosodic unit boundaries were in principle uniform, except for those caused by repairs and restarts.

3.1 The Mandarin data

The definition of prosodic units adopted is mainly that of the Intonation Unit found in studies of discourse analysis (Chafe 1994; Tao 1996). The original definition of the Intonation Unit that was developed and applied to earlier Mandarin data relates semantic and pragmatic information to prosodic structures (Tao 1996). But for the current study, the annotation of prosodic units in our Mandarin data was strictly based on prosodic cues, possibly minimizing the interfering effects caused by structuring cues from semantic and syntactic information. The labelers were instructed to annotate locations in the conversation if one of the four prosodic cues was present: (1) pitch reset (a shift upward in overall pitch level), (2) lengthening (changes in duration), (3) alternation of speech rate (changes in rhythm), and (4) occurrences of paralinguistic sounds (disjunction or disruption of utterances such as pauses, inhalation, and laughter). Pitch reset was mentioned by Tao (1996), while temporal and rhythmic cues at prosodic unit completion locations were mentioned by Hirst & Bouzon (2005) and Dankovičová (1997). The annotation of prosodic units in the Mandarin dataset was carried out in an earlier project (Liu & Tseng 2009) and was used for the current study without any changes. In the beginning stages of the annotation of prosodic units, three labelers were trained to annotate prosodic units on a subset of 150 speaker turns until a satisfactory agreement was achieved. More precisely, over 80% of the PU boundaries labeled by all three labelers were consistent (85%, 82%, 84%). Over 90% of the finalized PU boundaries were correctly recognized by all three labelers (93%, 95%, 92%). Disfluencies were not annotated, and the labelers concentrated only on prosodic segmentation. The rest of the dataset (550 speaker turns) was then completed by the three labelers independently.

3.2 The French data

For the French data, the original definition of prosodic units is adopted from prosodic phonology (Nespor & Vogel 1986; Selkirk 1986) that proposed a universal hierarchy of prosodic constituents. At least two levels of phrasing above the word have been admitted in French: the lowest level of phonological phrases (Post 2000) or Accentual Phrases (APs) (Jun & Fougeron 2000) and the highest level of Intonational Phrases (IPs). The accentual phrase is the domain of primary stress. The latter occurs in the final full syllable of a word with longer duration and higher intensity than non-final syllables, and is associated with a melodic movement. Secondary stress, more variable and optional, generally occurs in the initial stressed syllable of the first lexical word. It is associated with a rising movement. The Intonational Phrase contains one or more accentual phrases. It is marked by a major final rise or fall (intonation contour) or a stronger final lengthening, and can be followed by a pause (Hirst & Di Cristo 1984; Jun & Fougeron 2000). More recently, a few studies have attempted to show the existence of an intermediate level of phrasing (intermediate phrase, IP) that would occur with stronger prosodic cues than the ones associated with AP, and weaker than those associated with IP (Michelas & D'Imperio 2010). For the French dataset, once primary and secondary stresses have been identified, the main acoustic cues are (1) specific melodic contour, (2) final lengthening, (3) pitch reset. Disfluencies were annotated separately and silent pauses have not been systematically associated with a boundary (Portes et al. 2011).

For the current study, the extracted French dataset was annotated by naïve labelers in terms of a ToBI-style annotation (0 = no break; 1 = AP break; 2 = ip break; 3 = IP break) in Praat (Boersma & Weenink 2001). To align the prosodic annotation in French with that in Mandarin, any breaks of level 2 or 3 were considered as unit completion boundaries for prosodic units, as were pauses over 400 ms. We computed a kappa-score for our data set by taking each word as a decision point and counting the number of matching and non-matching boundaries across annotators. This method of calculation yielded a kappa score of 0.71 for our dataset, which is a good score for naïve coders on a prosodic phrasing task.[4]

4. Annotating discourse units

Discourse segmentation has been specifically addressed both for written and spoken data including monologues and dialogues, but with differences in focus. Generally speaking, it involves at least two levels of units: utterance- or clause-like units versus the paragraph of topic-like units. The latter is the subject of a vast natural language processing literature both for written data (Hearst 1994) and spoken data (Passonneau & Litman 1997). We focus here on the former, which has been of interest for semanticists and discourse analysts, serving as their basic unit for analyses, and sometimes called the elementary discourse unit (Polanyi & Scha 1984). Relational approaches to discourse have used them as their basic elements for building discourse structure, such as sentential units (Hobbs 1978). Discourse segments viewed in this way are often split between basic or elementary units and complex units (any kind of discourse composition of the basic ones). The type of discourse unit we wanted to annotate is that defined by Polanyi et al. (2004) as a unit that 'communicates information about not more than one "event", "event-type" or "state of affairs in a "possible world" of some type'. They are, therefore, the semantic counterparts of sentences in discourse. However, interactional and dialogic features require their definition to take into account the conversational notion of turn, as is explained below.

As regards the annotation of discourse units, this was not as complicated as it was for prosodic units. The task of annotation in both datasets involved naïve annotators who segmented the whole corpus. This annotation was performed without listening to the signal, but with timing information. In the praxis, it was performed with Praat (Boersma & Weenink 2001), without including the signal window, with only the time-aligned word tiers. The segmentation was performed by adopting a set

[4] Cohen's kappa (Artstein & Poesio 2008; Carletta 1996; Cohen 1960) is a method of measuring inter-coder agreement. It corrects the raw agreement by an estimation of the agreement by chance. The issue here is that it is a segmentation task. We therefore have to decide which are the decision points. We are using words as decision points rather than a fixed sample (as is done in some annotation tools) because the French guidelines use words as the base units indicating where to put the boundaries. Agreement on no-boundary (0-0) is therefore an agreement for this decision task and there is no satisfactory way to evaluate a kappa score if these agreements are left out. Other measures need to be introduced (Fournier & Inkpen 2012; Pevzner & Hearst 2002) if one wants to measure a different aspect of segmentation agreement. For deeper evaluation of the annotation of the whole CID corpus, please refer to Peshkov et al. (2012), and for more on segmentation evaluation metrics applied to this kind of data see Peshkov & Prévot 2014.

of discourse segmentation guidelines, inspired by Muller et al. (2012), Carlson & Marcu (2001), and Chen (2012). We combined the semantic criterion of Vendler's style eventualities identification (Vendler 1957) and Xue's proposition identification (Xue 2008), a discourse criterion (the presence of discourse markers) and a pragmatic criterion (recognition of specific speech acts) to carry out the segmentation. Stede (2012) lists a set of cases that are particularly difficult to handle when crafting such segmentation guidelines: various kinds of ellipsis, relative clauses, complement clauses including direct reported speech, and prepositional phrases that can cause problems when they do not include a verb, but have a discourse role. We adopted Stede's definition as a basic but solid semantic view of Elementary Discourse Unit (Stede 2012:89): 'A span of text, usually a clause, but in general ranging from minimally a (nominalization) NP to maximally a sentence. It denotes a single event or type of events, serving as a complete, distinct unit of information that the subsequent discourse may connect to. An EDU may be structurally embedded in another.'

Practically speaking, a discourse unit (DU) consists of a main predicate, and all its related complements and adjuncts as illustrated in (3) and (4). Mandarin spontaneous speech presents an additional challenge in the task of DU annotation because of its lack of a tense-marking verbal system. Our segmentation proceeds on the basis of the semantic bonding between the predicates identified (Givón 1993). Additional cues such as discourse connectives articulating discourse units were also used. Finally, mainly because of the interactive dialogic phenomena, for example question–answer pairs, we added a few pragmatic criteria for allowing short utterances to be acceptable discourse units (Fernandez et al. 2007), for instance *yeah*, and sentence fragments, for example '*where?*'. In other words, our discourse segmentation is first guided by semantics (using syntactic information to obtain the semantic units, and is therefore akin to the rather semantic vision of syntactic completion in Ford & Thompson 1996), but is then refined by discourse and pragmatic considerations. As for discourse, any discourse connective or marker is used as a cue to segment. As for more pragmatic and conversational points, our approach follows the pragmatic completion of Ford & Thompson (1996). Our segmentation guidelines (Prévot 2014) are stated in more speech-act based vocabulary, but with the same intention of identifying *conversational actions*.

We originally allowed for discourse embedded structures, but they were rarely used by the annotators and yielded low inter-rater agreement. We decided therefore to work for the time being with a flat segmentation.

(3) Discourse units in French
[on y va avec des copains]du [on avait pris le ferry en Normandie]du [puisque j'avais un frère qui était en Normandie]du [on traverse]du [on avait passé une nuit épouvantable sur le ferry]du
[*we are going there with friends*]*du* [*we took the ferry in Normandy*]*du* [*since I had a brother who was in Normandy*]*du* [*we cross*]*du* [*we spent a terrible night on the ferry*]*du*

(4) Discourse units in Mandarin
[qishi ta jiang de na ge ren yinwei ta you qu kai guo hui]du [ta hai you jiang]du [keneng shi ye bu zhidao wei she me]du
[*in fact the one he mentioned had the meeting*]du [*he said in addition*]du [*probably* (*he*) *did not know why*, *either*]du

Manual discourse segmentation with our guidelines has proven to be reliable, with kappa scores ranging between 0.74 and 0.85 for the French data and reaching 0.86 for the Taiwan Mandarin data. In addition we distinguished between regular discourse units and abandoned discourse units as illustrated in examples (5) and (6). The abandoned ones are units that are so incomplete that it is impossible to attribute a discourse contribution to them. They are distinguished from false starts, which are included in the DU they contributed to, by the fact that the material they introduced cannot be said to be taken up in the following discourse unit. As a result, abandoned units represent 11% and 6.5% of the annotated discourse units in French and Mandarin, respectively. This is in line with the spontaneous style of conversations in our data.

(5) Abandoned discourse units in French
[et euh mh donc t(u) avais si tu veux le sam+ le]adu [pour savoir qui jouait tu (v)ois]du
[*and uh mmh so you had if you want the sat+ the*]adu [*in order to know who was playing you see*]du

(6) Abandoned discourse units in Mandarin
[danshi muqian]adu [yinwei shengzhiyu]adu [wo you ting renjia jiang]du [man kuazhang] du
[*but for the moment*]adu [*because even though*]adu [*I heard some people say*]du [*it is incredible*]du

5. Automatically derived chunks

Chunks can be seen as an intermediate level of syntactic processing (Abney 1991). They are the basic structures built from the part of speech tags, but do not deal with long dependencies or rich constituency. Chunks are basically units centered on a syntactic head, that is, a content word. As noted by Abney, chunks can be related to sentences (Gee & Grosjean 1983) that have a more intonational nature. An idea defended in these early works is that, from a cognitive viewpoint, chunks are indeed language processing units. The rise of experimental linguistics has renewed interest in this hypothesis, and attempts are being made to make it more precise (Blache 2013) and to relate it to other empirical evidence such as eye-tracking (Blache & Rauzy 2012). With this idea in mind, we shall investigate our prosodic and discourse units in terms of chunk size and constituency. If chunks are a kind of universal processing unit independent of languages, the number of words in chunks should be similar across languages. But PU and DU that represent language-dependent and data-specific properties may not have similar distribution in terms of words. More precisely, we expect a significant variation of PU/DU size across languages in terms of number of words, but not in terms of chunk size.

5.1 Creating chunks

Of the above definition of chunking we have retained the importance of the head. We therefore designed simple rules (see Appendix for a complete list of rules) using POS-tag patterns for

Table 3: Chunk categories

VC	Verbal chunk
NC	Nominal chunk
AdvC	Adverbial chunk
PC	Prepositional chunk
IC	Interactional chunk
DisfError	Disfluencies or tagging errors
AdjC	Adjectival chunk

creating the chunks listed in Table 3. These rules were derived by examining the most frequent patterns at an earlier stage. We then carried out three different steps, applying the following rules:

1. Suggest a chunking rule (e.g. *Pro Pro V -> VC*; *Det N N -> NC*) for the most frequent POS pattern.
2. Suggest a set of rules aggregating tags and chunks into coherent chunks (e.g. *Prep NC*; *-> PC*; *VC Part -> VC*). This is done iteratively until the number of sequences is stabilized.
3. Simplify the sequences by merging certain categories (*Det*, *Pro*) (or sequences of them) into some existing chunks (*e.g* [*Det*|*Pro*] + *VC -> VC*) and simplifying some chunk sequences (*IC IC -> IC*).

The first two rules are strongly language-dependent, while the third is common to both languages. Complete lists of chunking rules for French and Taiwan Mandarin can be found in the Appendix. It was not possible to use pre-trained existing chunkers, as we were concerned with spontaneous spoken constructions. As far as we know, existing chunkers are trained on written data, which does not equip them for our purposes. Moreover, in the rule-based design, the rules are accessible to linguists, and this allows us to compare them directly across languages rather than comparing chunking quality. Indeed, we are not interested in the chunks from an applicative perspective (such as named entity recognition), but as a good approximation of semantic processing units. In the longer term, it could, however, be interesting to improve and evaluate pre-trained chunking procedures, but this will require a large amount of manual work, which we cannot afford for the time being.

6. Interface of discourse and prosody

Utilizing the annotated prosodic units, discourse units, and automatically derived chunks, we were able to conduct quantitative analyses on how prosodic phrasing and discourse segmentation interact with each other in conversational speech.

Table 4: Comparative size of the units produced

	Duration (sec)	Syllables	Words	PU	Chunks
French PU	0.94	5.6	4.9	–	1.76
Mandarin PU	1.44	6.4	4.4	–	2.05
French DU	1.81	10.7	9.3	1.9	3.36
Mandarin DU	2.17	9.6	6.6	1.5	2.27

6.1 Size of units

Table 4 summarizes the results of the three processing units in terms of size and duration. Compared with the Mandarin data, the French PUs are on average shorter in duration, containing more words, but less syllables, which suggests that in conversational speech French may use more monosyllabic reactive words. Another reason for this result, which is not language dependent, is the level of familiarity of the speakers engaged in conversation. We extracted long speaker turns for both languages, but the French data may contain more disfluencies. These disfluencies result in short prosodic units, but create large discourse units. This tendency is also shown in Table 4. The French DUs contain more words and syllables, but are much shorter than the Mandarin DUs. Repeated items in disfluencies are usually reduced, and thus shorter in conversation (Clark & Wasow 1998; Tseng 1999). Despite the discrepancy in duration, the length of prosodic units in terms of words is quite similar: 4.9 in French and 4.4 in Mandarin. Also, the chunk length in terms of words is very similar in French and Mandarin, 2.78 and 2.53 respectively. By contrast, DU differs quite significantly in the two languages. These results suggest that in conversational speech, prosodic units conform more to the automatically derived chunks in both languages. In this first step of the work we did not extract inter-individual variability; however, the next step will be to scale up the study by taking a bigger selection of the dataset for sampling, while preserving an equal proportion of each speaker's speech.

To further validate our hypothesis on the length distribution of PU, DU, and chunks, Figures 1a and 1b show that PU sizes in terms of chunks are quite stable across languages and suggest a kind of universal spoken language planning that will need to be validated on different corpora, and scrutinized, for example, with laboratory experiments. The DU size distribution is less similar even though the annotation process was done more consistently than for PU. Again this may have to do with the level of familiarity of the speakers engaged in these conversations. The French speakers exhibited an especially colloquial style. The speech of many speakers was rich in disfluencies that specifically split into smaller prosodic units, but result in larger discourse units. Our earlier work on disfluencies (Tseng 1999) and detection (Peshkov et al. 2012) mentioned and supported these consequences. Another issue related to chunking is that more careful crafting of joint rules is necessary to deal with the conversational style of speech more appropriately than it is currently. However, we do not have a reasonable size of annotated chunks with this kind of data for training a supervised machine-learning approach. For unsupervised methods our dataset is significant but probably not sufficient.

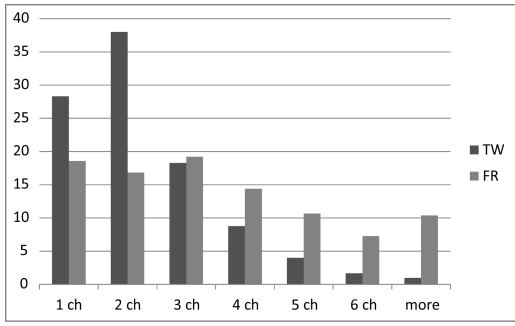

Figure 1a: DU size in chunks

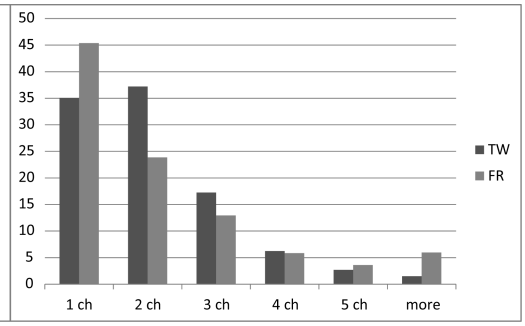

Figure 1b: PU size in chunks

6.2 Association of prosodic and discourse units

We examined different types of association between prosodic and discourse units by means of the boundary alignment proposed by Chen (2012). Our classification system distinguishes complex (com) from simple (sim) discourse units by the presence of a prosodic boundary within the discourse unit. More precisely *simple* discourse units do not include any internal prosodic boundaries, while *complex* discourse units do. As regards the alignment of discourse and prosodic units, four matching cases are distinguished: at the left or right boundary, at both boundaries, or at neither. Such a classification resulted in eight types, as illustrated in Figure 2a.

In the French data, perhaps because of the comparatively smaller prosodic units, discourse units regularly host more prosodic units. It is striking to see in Figure 2b that more than half of the time and for both language, discourse units are providing the starting and ending boundaries for the prosodic units. Overall, we see in Figure 2b that once atomic and composite DUs (simple and complex) are collapsed, their alignment types are highly similar. This result suggests that grammatical structure and prosodic organization significantly match in terms of boundary markedness in conversation.

Examples (3), (4), and (5) illustrate different PU/DU association types in French and in Mandarin. Figures 3a and 3b show the most frequent cases of a perfect match between PU and DU, 40% for Mandarin and 30% for French. We also observed that the syntactic structure of this type of association as modeled by our chunking rules is simple and perfectly standard.

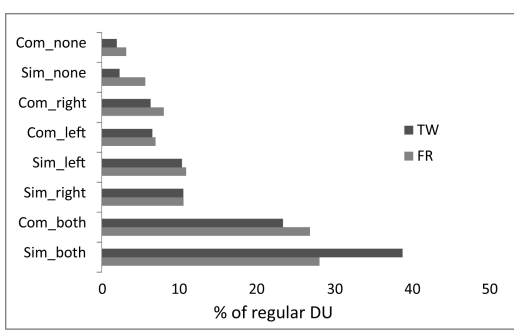

Figure 2a: PU/DU association types

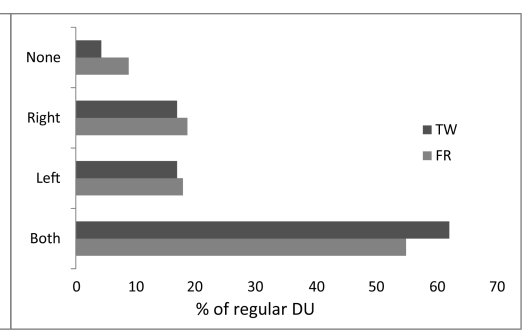

Figure 2b: Simplified association types

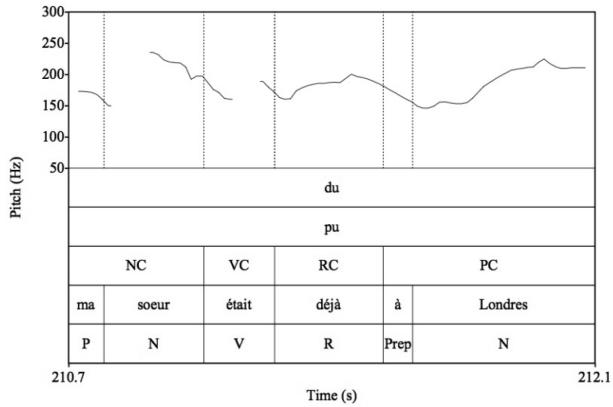

Figure 3a: My sister was already in London (sim-both, French example)

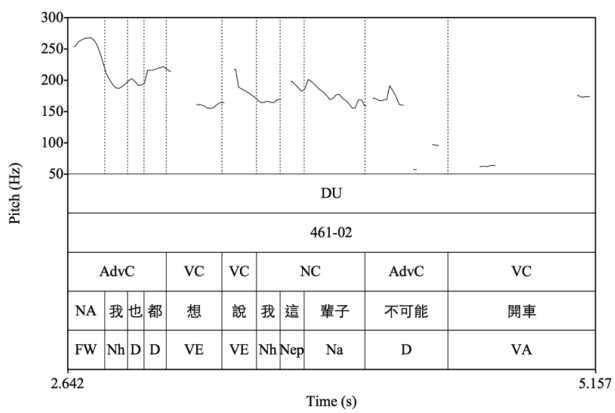

Figure 3b: I thought I could not possibly drive a car in my life (sim-both, Mandarin example)

Figures 4a and 4b illustrate complex discourse units with prosodically aligned boundaries. They seem to correspond to a grouping by syntax according to the terminology proposed by Degand & Simon (2009). However, this pattern often corresponds to different steps in the construction of the full message, with or without disfluences. In our study, disfluencies were not annotated as such at this level of analysis because our focus was on reconstructing the full propositional meaning of the discourse units. In Figure 4a, the French example shows verbal chunks that indeed correspond to the structure of repairs containing a reparans, a pause, and a reparandum (Shriberg 1994). It is therefore more discourse coherence than careful syntax planning that glues these prosodic units together.

Finally, Figures 5a and 5b illustrate both the left and the right alignment.[5] A large number of these cases for the French data correspond to several discourse units grouped under a prosodic unit.

[5] In fact, most discourse boundaries that do not match a prosodic boundary result in both a left-match and a right-match.

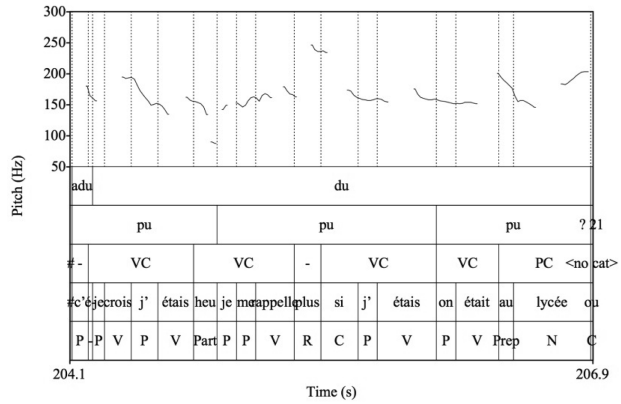

Figure 4a: It I think I was I do not remember anymore whether I was we were in high school (com-both, French example)

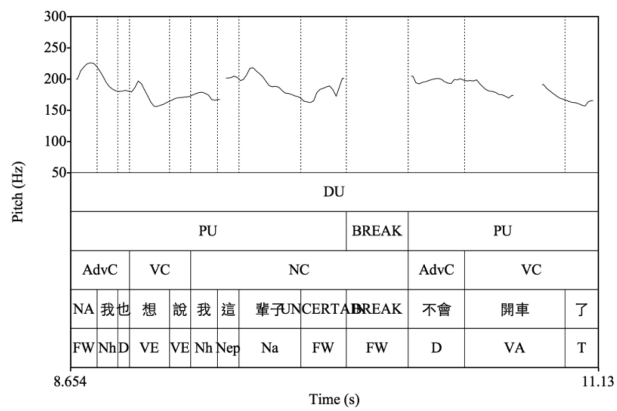

Figure 4b: I thought I in my life will never drive a car any more (com-both, Mandarin example)

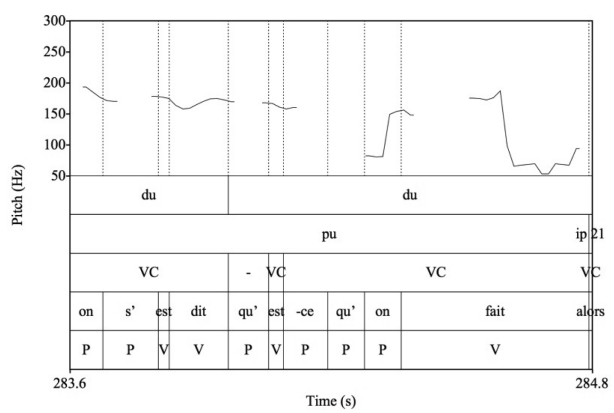

Figure 5a: We asked each other what do we do (sim-left and sim-right)

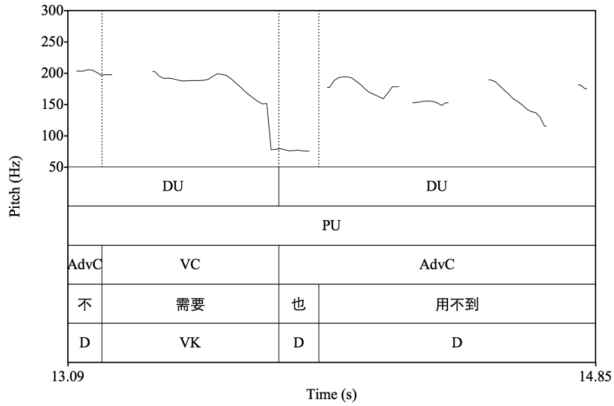

Figure 5b: No need also no necessity (sim-left and sim-right)

In this example, it is reported speech that is split according to our discourse segmentation guidelines, but that has been presented here as a single prosodic unit.

6.3 Syntactic categories at boundaries

Making use of the collapsed POS tags in Table 2, we compared the distribution in terms of word category at the PU-DU matching initial and final boundaries, as shown in Figures 6a and 6b. Interestingly, French units tend to begin more often with connectives and pronouns. In Taiwan Mandarin, the percentage of pronouns is lower and that of adverbs is higher. This may be due to the fact that in conversation sentences are often zero-subject or have the focus moved to sentence-initial positions. For final matching boundaries, Taiwan Mandarin often ends with sentence-final particles, which is expected in conversation. Moreover, French ends more often with nouns than verbs, the reverse being the case in Taiwan Mandarin. Our preliminary studies of the word categories only provide information for the boundary. More work on sentence structure is required to conduct in-depth studies on language production.

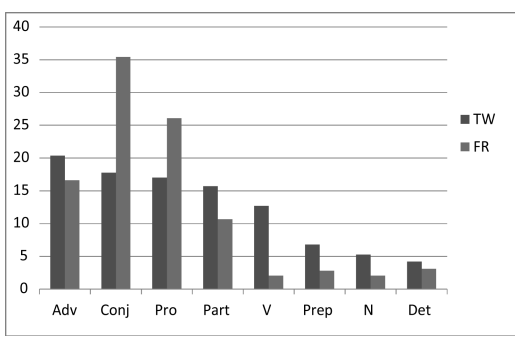

Figure 6a: POS at initial-match boundary

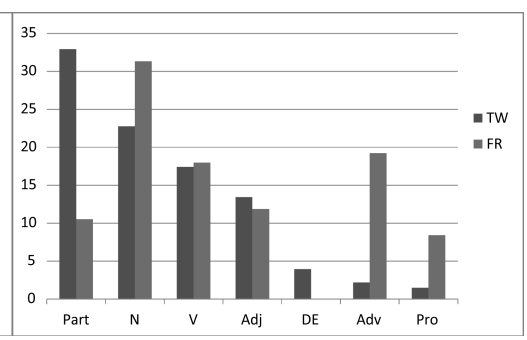

Figure 6b: POS at final-match boundary

7. Discussion

This article sheds light on the study of the discourse unit and on dialogue act automatic segmentation, as well as on descriptive studies of the prosody–discourse interface. Passonneau & Litman (1997) initiated a new research trend by both carrying out an annotation campaign, including inter-annotator agreement evaluation, and setting up a system for discourse segmentation of spoken data. This system combined prosodic (mostly pause duration) and discourse connective information for approaching the suggested human reference. However, their objective is paragraph or topic segmentation rather than elementary unit segmentation.[6] Edlund et al. (2005) took low-level prosodic information to help identify utterance unit segmentation. This is also in line with Pierrehumbert & Hirschberg's (1990) idea of the crucial importance of boundary tones for discourse segmentation. Gross and colleagues (1992) used a boundary tone, a pause in speech longer than a single beat, a resetting of the pitch level, and the start of a new intonational phrase for segmenting discourse. Their attention to task-oriented dialogues led them to include the disfluencies that also introduced a break in the discourse flow.

A general consensus accepted in most of the NLP literature is that low-level prosodic information provides crucial information for discourse segmentation. It is often used directly as meaningful units (Traum & Heeman 1997). Indeed a pilot study of our French data (Peshkov et al. 2012) has also shown that even lower-level prosodic analysis such as IPUs (Inter-Pausal Units for a pause of 200 ms) constituted a baseline for discourse segmentation that was difficult to beat with more sophisticated methods, including finer-grained prosodic analysis or syntactic information.[7] It is also remarkable that the language processing units of phi-sentences (Gee & Grosjean 1983) also have an intonational nature. Finally, also in conversational analysis, Ford and Thompson have shown that intonation plays a crucial role for defining the Turn-Constructional Unit (Sacks et al. 1974).

Many descriptive prosodists, syntacticians, and discourse analysts are, however, much more cautious with regard to the relation of phrasing units to meaning, syntax, and discourse. More precisely, we looked at how prosodic units and discourse units are distributed onto each other. In spirit our work is closely related to that of Degand & Simon (2009), Lacheret et al. (2010), Gerdes et al. (2012), and Beliao et al. (2013). However, our focus here is on insights we might gain from a comparative study. Our dataset has a more conversational nature than the datasets studied in their work. As for the data, Gerdes et al. (2012) wanted to have an interesting spectrum of discourse genres and speaking styles while we focused on conversations, both to make comparative studies possible, and to ensure that enough coherent instances from the perspective of statistical studies were obtained. Also, while Lacheret et al. (2010) took a purely intuitive approach, we sought greater balance by combining explicit criteria from different language domains. Finally, our annotation experiments are largely produced either by automatic tools (trained on experts' data) or by naïve coders. This is a major difference from the studies listed above, which are based on experts' annotations, since it allows us to scale up data size more easily. This contrasts with Degand & Simon's related work, which is based on manual expert analysis, and therefore relates to a relatively small dataset.

---

[6] In fact the candidate boundaries were boundaries of intonational phrases which were obtained directly from the original transcription.

[7] Basically the noise introduced by other information sources tends to damage the reasonable IPU-baseline.

Their work, however, is particularly relevant for our purposes. Degand & Simon (2005, 2009) studied *basic discourse unit*s resulting from the combination of prosodic units and syntactic units. They observed that, despite the high frequency of discourse units grounded on congruent prosodic and syntactic units (called BDU-C), they can also be grouped by syntax (BDU-S) or by prosody (BDU-I). Simon & Degand (2011) stated clearly in their study that following Selting (2000), neither syntactic nor prosodic completeness is sufficient to determine the boundaries of a basic discourse unit. They also show that types of BDU exhibit different distribution across discourse genres and suggest that these types correspond to different discourse strategies. According to Degand & Simon (2009) and Simon & Degand (2011), BDU-C represent a simple and neutral way to present information; BDU-I are used to create an informational macro-unit and BDU-S relate to a more emphatic style, or careful discourse planning.

Beliao et al. (2013) have a very similar approach in terms of synchronized boundaries that are both illocutionary (discourse) units (IU) boundaries and intonational period (IP) boundaries. However, the way these units are defined differs considerably from ours, since the ratio IU / IP is reversed compared to our DU / PU. Their period is indeed defined to be much larger than our PU. To identify a period boundary all the following conditions must be present: a pause of at least 300 ms, significant F0 movement, and pitch reset. Portes & Bertrand (2011) argued for a phonological structuration of spontaneous speech into several phrasing levels like those of read speech. They suggested that this structuration is only affected by disfluencies and interactional processes such as turn-taking and backchannels.

To sum up the comparison of our results with these two traditions of research, we can say that we support the view that using prosodic units as discourse units (as often suggested in Natural Language Processing frameworks) is indeed a very robust and interesting approach, since matching boundaries are the dominant case. However, our results on PU-DU distribution also support the view that a more subtle correlation between prosodic and discourse units exists as is emphasized by descriptive linguistic studies. This observation is not as paradoxical as it seems. Depending on their objective, researchers focus on very different aspects of discourse prosody interfaces. NLP researchers look first for both a wide data coverage and the maximization of F-measures-related evaluation. This leads them to pay relatively little attention to less dominant cases of mismatch between prosodic and discourse units. While, as pointed out in the linguistic literature, there is a variety of cases, quantitatively each of these cases is limited. Consequently, processing them all is likely to damage overall performance. Descriptive linguists, by contrast, are specifically attracted by all the particular phenomena at the interfaces. The first steps presented in this paper constitute an intermediate way that allows a more quantitative approach to prosody–discourse interface analysis.

The way disfluencies (included in the discourse units) were handled in our study prevents a direct comparison with the works described above. However, it is striking that the different types of association are quite similar across languages, as can be seen in Figures 2a and 2b. Our data of POS distribution at boundaries reflect the different syntactic organization of typologically diverse languages such as Mandarin and French. For example, the importance of final particles in Mandarin, the importance of the initial topic position, as well as the frequent zero-anaphora in spontaneous speech are crucial ingredients of the discrepancies between French and Mandarin POS figures (Figures 6a and 6b).

8. Conclusion and future work

This work has shown that in order to create perfectly comparable corpora, a joint design right from the beginning is required. However, this is a rare scenario and most comparative datasets of richly annotated corpora seek to re-use at least part of previous monolingual studies. Here we have tried to make use of very similar resources to produce comparable corpora. We believe that, although this dataset could still be improved and would benefit from an even more similar starting point, we have a unique resource for performing quantitative comparative studies of the kind initiated here. Equipped with this dataset, we are in a position to conduct a series of deeper comparative studies. The next step, that of data management, will be to scale up the study by taking a bigger selection of the dataset to be sampled, while preserving an equal proportion of each speaker's speech.

The chunking systems used in this paper demonstrate our first attempt in this direction. Although the results for the chunk size are not completely conclusive for our hypothesis, we do get a better knowledge of the structures presented in the units we are investigating and we would like to push further our exploration in this direction. We think in particular that, since both languages are subject–verb–object (SVO), detailed structuring of PU and DU in terms of chunk categories (Nominal chunks, Verbal chunks, etc.) should be more stable than their POS distribution. We are currently looking at the distribution of the mono-, bi-, and tri-chunk PU and DU sequences in order to get more precise results in language comparison, without going into a full syntactic analysis, which is out of reach for this kind of data.

Appendix: Chunking rules

The chunking processing is performed from left to right on pairs of POS-tags. Every time the pair matches with one of the rules below, the corresponding tokens are grouped into a chunk. An existing chunk can be further extended if the POS of its left boundary and the next POS match with the rule as well.

French

#	Rule	Left POS	Right
1	Cs .	Subordinative conjunction	Any
2	Cc .	Coordinative conjunction	Any
3	I .	Interjection	Any
4	D .[^C I]	Determiner	Any except conjunction or interjection
5	S .[^C I]	Preposition	Any except conjunction or interjection
6	Pp .[^C I]	Personal pronoun	Any except conjunction or interjection
7	Pr .[^C I]	Relative pronoun	Any except conjunction or interjection
8	Px .[^C I]	Reflexive pronoun	Any except conjunction or interjection
9	Va .[^C I]	Auxiliary verb *avoir*	Any except conjunction or interjection

#	Rule	Left POS	Right
10	Ve .[^C I]	Auxiliary verb *être*	Any except conjunction or interjection
11	Rpd .[^C I]	Adverb of negation (e.g. *que*)	Any except conjunction or interjection
12	Rpn .[^C I]	Adverb of negation (e.g. *ne*)	Any except conjunction or interjection
13	A N	Adjective	Noun
14	Nc Np	Common noun	Proper noun
15	Np Np	Proper noun	Proper noun
16	Nk Nc	Cardinal noun	Common noun
17	Pd V	Demonstrative pronoun	Verb
18	Pd Rpn	Demonstrative pronoun	Adverb of negation (ex. *ne*)
19	Pd Cs	Demonstrative pronoun	Conjunction. of subordination
20	Pd Pr	Demonstrative pronoun	Relative pronoun
21	Rgc P	Comparative adverb	Pronoun
22	Rgp A	Adverb	Adjective
23	Rgp Vmps	Adverb	Past participle
24	V V	Verb	Verb
25	V Rgd	Verb	Adverb of negation (ex. *pas*)
26	V Rgn	Verb	Adverb of negation (ex. *plus*)
27	Pd P[^t]	Demonstrative pronoun	Any pronoun except interrogative
28	Rgn A	Adverb of negation (e.g. *ne*)	Adjective

Taiwan Mandarin

#	Rule	Left POS	Right POS
1	Ne .[^Cbb]	Determinative	Any except correlative conjunction
2	I .[^Cbb]	Interjection	Any except correlative conjunction
3	P .[^Cbb]	Preposition	Any except correlative conjunction
4	Nh .[^Cbb]	Pronoun	Any except correlative conjunction
5	Nf .[^Cbb]	Measure noun	Any except correlative conjunction
6	FW .[^Cbb]	Foreign word	Any except correlative conjunction
7	SHI .[^Cbb]	Copula *shi4*	Any except correlative conjunction
8	V_2 .[^Cbb]	Verb *you3*	Any except correlative conjunction
9	Cbb .[^Cbb]	Correlative conjunction	Any except correlative conjunction
10	Caa .[^Cbb]	Coordinate conjunction	Any except correlative conjunction
11	. T	Any	Particle
12	. Di	Any	Aspectual adverb

#	Rule	Left POS	Right POS
13	. DE	Any	Structural particle *de5*, *zhi1*, *de2*, *di4*
14	. Ng	Any	Postposition noun
15	. FW	Any	Foreign word
16	. Cba	Any	Conjunction *de5hua4*
17	. Cab	Any	Conjunction *deng3deng3*
18	V V	Verb	Verb
19	V SHI	Verb	Copula *shi4*
20	V Nf	Verb	Measure noun
21	V Ncd	Verb	Localizer
22	VH11 VH11	Stative intransitive verb	Stative intransitive verb
23	Na Ncd	Common noun	Localizer
24	Nc Ncd	Place noun	Localizer
25	A A	Non-predicative adjective	Non-predicative adjective
26	D D	Adverb	Adverb
27	D V	Adverb	Verb
28	DE Na	Structural particle *de5*, *zhi1*, *de2*, *di4*	Common noun

Idioms

Idioms are always glued together and separated from other chunks, for instance *shi bu shi*, *you mei you*, *hao bu hao*, and *dui bu dui*.

References

Abney, Steven P. 1991. Parsing by chunks. *Principle-based Parsing: Computation and Psycholinguistics*, ed. by Robert C. Berwick, Steven P. Abney & Carol Tenny, 257–278. Dordrecht: Kluwer.

Afantenos, Stergos, Nicholas Asher, Farah Benamara, Myriam Bras, Cécile Fabre, Mai Ho-dac, Anne Le Draoulec, Philippe Muller, Marie-Paule Péry-Woodley, Laurent Prévot, Josette Rebeyrolle, Ludovic Tanguy, Marianne Vergez-Couret, & Laure Vieu. 2012. An empirical resource for discovering cognitive principles of discourse organisation: the ANNODIS corpus. *Proceedings of LREC 2012*, 2727–2734. http://www.lrec-conf.org/proceedings/lrec2012/pdf/ 836_Paper.pdf.

Artstein, Ron, & Massimo Poesio. 2008. Inter-coder agreement for computational linguistics. *Computational Linguistics* 34.4:555–596.

Beliao, Julie, Sylvain Kahane, & Anne Lacheret. 2013. Modéliser l'interface intonosyntaxique. *Proceedings of the Prosody-Discourse Interface Conference 2013* (*IDP-2013*), 21–26. http://wwwling.arts.kuleuven.be/franitalco/idp2013/Proceedings.html.

Bertrand, Roxane, Philippe Blache, Robert Espesser, Gaëlle Ferré, Christine Meunier, Béatrice Priego-Valverde, & Stéphane Rauzy. 2008. Le CID—corpus of interactional data—annotation et exploitation multimodale de parole conversationnelle. *Traitement Automatique des Langues* 49.3:1–30.

Blache, Philippe. 2013. Chunks et activation: un modèle de facilitation du traitement linguistique. *Proceedings of TALaRE 2013*, 229–242. http://www.aclweb.org/anthology/F13-1017.

Blache, Philippe, Roxane Bertrand, & Gaëlle Ferré. 2009. Creating and exploiting multimodal annotated corpora: the ToMA project. *Multimodal Corpora: From Models of Natural Interaction to Systems and Applications*, ed. by Michael Kipp, Jean-Claude Martin, Patrizia Paggio & Dirk Heylen, 38–53. Berlin & Heidelberg: Springer.

Blache, Philippe, & Stéphane Rauzy. 2008. Influence de la qualité de l'étiquetage sur le chunking: une corrélation dépendant de la taille des chunks. *Proceedings of TALN 2008*, 290–299. Avignon, France.

Blache, Philippe, & Stéphane Rauzy. 2012. Robustness and processing difficulty models. A pilot study for eye-tracking data on the *French Treebank*. *Proceedings of the First Workshop on Eye-tracking and Natural Language Processing*, 21–36. http://aclweb.org/anthology/W12-4903.

Boersma, Paul, & David Weenink. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5.9–10:341–345.

Butterworth, Brian. 1975. Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research* 4.1:75–87.

Carletta, Jean. 1996. Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics* 22.2:249–254.

Carlson, Lynn, & Daniel Marcu. 2001. Discourse tagging reference manual. ISI Technical Report ISI-TR-545.

Chafe, Wallace L. 1994. *Discourse*, *Consciousness*, *and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: University of Chicago Press.

Chen, Alvin Cheng-Hsien. 2012. *Prosodic Phrasing in Mandarin Spontaneous Speech: A Computational-acoustic Perspective*. Taipei: National Taiwan University dissertation.

Chen, Keh-Jiann, Chu-Ren Huang, Li-Ping Chang, & Hui-Li Hsu. 1996. Sinica Corpus: design methodology for balanced corpora. *Proceedings of the Eleventh Pacific Asia Conference on Language*, *Information and Computation*, 167–176. Seoul, Korea.

Clark, Herbert H., & Thomas Wasow. 1998. Repeating words in spontaneous speech. *Cognitive Psychology* 37.3:201–242.

Cohen, Jacob. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20.1:37–46.

Dankovičová, J. 1997. The domain of articulation rate variation in Czech. *Journal of Phonetics* 25.3:287–312.

Degand, Liesbeth, & Anne Catherine Simon. 2005. Minimal Discourse Units: Can we define them, and why should we? *Proceedings of SEM-05*. *Connectors*, *Discourse Framing and Discourse Structure: From Corpus-based and Experimental Analyses to Discourse Theories*, 65–74.

Degand, Liesbeth, & Anne Catherine Simon. 2009. On identifying basic discourse units in speech: theoretical and empirical issues. *Discours: Revue de linguistique*, *psycholinguistique et informatique* 4. http://discours.revues.org/5852.

Edlund, Jens, Mattias Heldner, & Joakim Gustafson. 2005. Utterance segmentation and turn-taking in spoken dialogue systems. *Sprachtechnologie*, *mobile Kommunikation und linguistische Ressourcen*, ed. by Bernhard Fisseni, Hans-Christian Schmitz, Bernhard Schröder & Petra Wagner, 576–587. Frankfurt am Main & New York: Lang.

Fernandez, Raquel, Jonathan Ginzburg, Howard Gregory, & Shalom Lappin. 2007. Shards: fragment resolution in dialogue. *Computing Meaning*, Vol. 3, ed. by Harry C. Bunt & Reinhard Muskens, 125–144. Dordrecht: Springer.

Ford, Cecilia E., & Sandra A. Thompson. 1996. Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns. *Interaction and Grammar*, ed. by Elinor Ochs, Emanuel A. Schegloff & Sandra A. Thompson, 134–184. Cambridge & New York: Cambridge University Press.

Ford, Marilyn, & Virginia M. Holmes. 1978. Planning units and syntax in sentence production. *Cognition* 6.1:35–53.

Fournier, Chris, & Diana Inkpen. 2012. Segmentation similarity and agreement. *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 152–161. Montréal, Canada.

Frederiksen, Carl H. 1977. Semantic processing units in understanding text. *Discourse Production and Comprehension*, ed. by Roy O. Freedle, 57–87. Norwood: Ablex.

Gee, James Paul, & François Grosjean. 1983. Performance structures: a psycholinguistic and linguistic appraisal. *Cognitive Psychology* 15.4:411–458.

Gerdes, Kim, Sylvain Kahane, Anne Lacheret, Arthur Truong, & Paola Pietrandrea. 2012. Intonosyntactic data structures: the rhapsodie treebank of spoken French. *Proceedings of the 6th Linguistic Annotation Workshop*, 85–94. Jeju Island, Korea.

Givón, Talmy. 1993. *English Grammar: A Function-based Introduction*, Vol. 2. Amsterdam & Philadelphia: John Benjamins.

Gross, Derek, James F. Allen, & David R. Traum. 1992. The Trains 91 Dialogues. TRAINS Technical Note 92-1. Rochester: Department of Computer Science, University of Rochester.

Grosz, Barbara J., & Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics* 12.3:175–204.

Hearst, Marti A. 1994. Multi-paragraph segmentation of expository text. *Proceedings of the 32nd Annual Meeting on Association for Computational Linguistics*, 9–16. Las Cruces, New Mexico, USA.

Hirst, Daniel, & Caroline Bouzon. 2005. The effect of stress and boundaries on segmental duration in a corpus of authentic speech (British English). *Proceedings of INTERSPEECH 2005*, 29–32, Lisbon, Portugal.

Hirst, Daniel, & Albert Di Cristo. 1984. French intonation: a parametric approach. *Die Neueren Sprachen* 83.5:554–569.

Hobbs, Jerry R. 1978. Why is discourse coherent? SRI Technical Note 176 (No. SRI-TN-176). Menlo Park: SRI International.

Illina, Irina, Dominique Fohr, Odile Mella, & Christophe Cerisara. 2004. The automatic news transcription system: ANTS, some real time experiments. Paper presented at the 8th International Conference on Spoken Language Processing—ICSLP'04, October 4–8, 2004. Jeju Island, Korea.

Jun, Sun-Ah, & Cécile Fougeron. 2000. A phonological model of French intonation. *Intonation: Analysis*, *Modeling and Technology*, ed. by Antonis Botinis, 209–242. Dordrecht & Boston: Kluwer.

Lacheret, Anne, Nicolas Obin, & Mathieu Avanzi. 2010. Design and evaluation of shared prosodic annotation for spontaneous French speech: from expert knowledge to non-expert annotation. *Proceedings of the Fourth Linguistic Annotation Workshop*, *ACL 2010*, 265–273. Uppsala, Sweden.

Liu, Yi-Fen, & Shu-Chuan Tseng. 2009. Linguistic patterns detected through a prosodic segmentation in spontaneous Taiwan Mandarin speech. *Linguistic Patterns in Spontaneous Speech*, ed. by Shu-Chuan Tseng, 147–166. Taipei: Institute of Linguistics, Academia Sinica.

Liu, Yi-Fen, Shu-Chuan Tseng, & J.-S. Roger Jang. 2014. Phone boundary annotation in conversational speech. *Proceedings of the 9th International Conference on Language Resources and Evaluation* (*LREC-2014*), 848–853. Reykjavik, Iceland.

Michelas, Amandine, & Mariapaola D'Imperio. 2010. Accentual phrase boundaries and lexical access in French. *Proceedings of Speech Prosody 2010*. Chicago, Illinois, USA.

Muller, Philippe, Marianne Vergez-Couret, Laurent Prévot, Nicholas Asher, Benamara Farah, Myriam Bras, Anne Le Draoulec, & Laure Vieu. 2012. Manuel d'annotation en relations de discours du projet annodis. Technical Report 21, CLLE-ERSS, Toulouse University.

Myers, Antje S. 1997. Conceptual influences on grammatical planning units. *Language and Cognitive Processes* 12.5–6:859–863.

Nespor, Marina, & Irene Vogel. 1986. *Prosodic Phonology*. Dordrecht: Foris.

Passonneau, Rebecca J., & Diane J. Litman. 1997. Discourse segmentation by human and automated means. *Computational Linguistics* 23.1:103–139.

Peshkov, Klim, & Laurent Prévot. 2014. Segmentation evaluation metrics, a comparison grounded on prosodic and discourse units. *Proceedings of the 9th International Conference on Language Resources and Evaluation* (*LREC-2014*), 321–325. Reykjavik, Iceland.

Peshkov, Klim, Laurent Prévot, Roxane Bertrand, Stéphane Rauzy, & Philippe Blache. 2012. Quantitative experiments on prosodic and discourse units in the corpus of interactional data. *Proceedings of SemDial 2012: The 16th Workshop on the Semantics and Pragmatics of Dialogue*, 19–21.

Pevzner, Lev, & Marti A. Hearst. 2002. A critique and improvement of an evaluation metric for text segmentation. *Computational Linguistics* 28.1:19–36.

Pierrehumbert, Janet, & Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. *Intentions in Communication*, ed. by Philip R. Cohen, Jerry L. Morgan & Martha E. Pollack, 271–311. Cambridge: MIT Press.

Polanyi, Livia, Chris Culy, Martin van den Berg, Gian Lorenzo Thione, & David Ahn. 2004. A rule based approach to discourse parsing. *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue at HLT-NAACL 2004*, 108–117.

Polanyi, Livia, & Remko Scha. 1984. A syntactic approach to discourse semantics. *Proceedings of the 10th International Conference on Computational Linguistics*, 413–419.

Portes, Cristel, & Roxane Bertrand. 2011. Permanence et variation des unités prosodiques dans le discours et l'interaction. *Journal of French Language Studies* 21.1:97–110.

Post, Brechtje. 2000. *Tonal and Phrasal Structures in French Intonation*. The Hague: Thesus.

Prévot, Laurent 2014. Manuel de segmentation de parole conversationnelle en unités discursives. Technical report. LPL, Aix Marseille Université.
Sacks, Harvey, Emanuel A. Schegloff, & Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50.4:696–735.
Selkirk, Elisabeth O. 1986. *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge: MIT Press.
Selting, Margret. 1996. On the interplay of syntax and prosody in the constitution of turn-constructional units and turns in conversation. *Pragmatics* 6.3:371–388.
Selting, Margret. 2000. The construction of units in conversational talk. *Language in Society* 29.4: 477–517.
Shriberg, Elizabeth Ellen. 1994. *Preliminaries to a Theory of Speech Disfluencies*. Berkeley: University of California at Berkeley dissertation.
Simon, Anne-Catherine, & Liesbeth Degand. 2011. L'analyse en unités discursives de base: pourquoi et comment? *Langue Française* 170:45–59.
Stede, Manfred. 2012. *Discourse Processing*. San Rafael: Morgan & Claypool Publishers.
Tao, Hongyin. 1996. *Units in Mandarin Conversation: Prosody*, *Discourse*, *and Grammar*. Amsterdam & Philadelphia: John Benjamins.
Traum, David R., & Peter A. Heeman. 1997. Utterance units in spoken dialogue. *Dialogue Processing in Spoken Language Systems*, 125–140. Berlin & New York: Springer.
Tseng, Shu-Chuan. 1999. *Grammar*, *Prosody and Speech Disfluencies in Spoken Dialogues*. Bielefeld: University of Bielefeld dissertation.
Tseng, Shu-Chuan. 2013. Lexical coverage in Taiwan Mandarin conversation. *Computational Linguistics and Chinese Language Processing* 18.1:1–18.
Vendler, Zeno. 1957. Verbs and times. *Philosophical Review* 66.2:143–160.
Xue, Nianwen. 2008. Labeling Chinese predicates with semantic roles. *Computational Linguistics* 34.2:225–255.

[Received 17 February 2014; revised 2 May 2014; accepted 21 June 2014]

Laurent Prévot (corresponding author)
Laboratoire Parole et Langage
Université de Provence
5 Avenue Pasteur
13100 Aix-en-Provence, France
laurent.prevot@lpl-aix.fr

Shu-Chuan Tseng
tsengsc@gate.sinica.edu.tw

Klim Peshkov
klim.peshkov@lpl-aix.fr

Alvin Cheng-Hsien Chen
alvinworks@gmail.com

自然對話的處理單位：法語與漢語的比較研究

Laurent Prévot[1,2]　曾淑娟[3]　Klim Peshkov[1,2]　陳正賢[4]

馬賽大學[1]
法國國家科學研究中心[2]
中央研究院[3]
國立彰化師範大學[4]

人類口語產製主要目的是為了傳達訊息給予聽話者，進而達到溝通目的。因此，長段話語切分為較小的處理單位將更易於建構意義清楚的言談架構。在自然對話中，此類型的處理單位可能參照語意、句法或韻律層次的結構。過去文獻曾提出不同的理論嘗試解釋言談切分的運作，但多半屬於質化分析研究。本研究利用以語料庫為本的量化資料，根據語意訊息結構（言談單位）、韻律表現（音韻單位）和詞類標記序列（詞組）標記個別處理單位之邊界。以此為根基比較台灣華語和法語在這三種處理單位上的結構表現，並檢視這三種處理單位之間的互動關係。本研究以語料庫語言學方法為兩個類型迥異的語言建築一個對等的比較平台，進而針對自然對話的語言表現，作跨語言之量化比較研究。

關鍵詞：言談，韻律，語言處理單位，詞組