

John Benjamins Publishing Company



This is a contribution from *Rhapsodie. A prosodic and syntactic treebank for spoken French*.
Edited by Anne Lacheret-Dujour, Sylvain Kahane and Paola Pietrandrea.
© 2019. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible to members (students and staff) only of the author's/s' institute, it is not permitted to post this PDF on the open internet.

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

Please contact rights@benjamins.nl or consult our website: www.benjamins.com

Tables of Contents, abstracts and guidelines are available at www.benjamins.com

CHAPTER 5

The annotation of list structures

Sylvain Kahane, Paola Pietrandrea and Kim Gerdes

Modyco, Paris Nanterre University & CNRS, France / CNRS, SLT, Lille

University, France / LPP, Sorbonne Nouvelle University & CNRS, Paris, France

This chapter presents phenomena we call “piles” or “lists”, which are characterized by the fact that a list of elements piles up in the same syntactic position. We therefore group the analysis of coordination together with the analysis of other phenomena such as reformulation, disfluency, partial answer, or negotiation. The elements of a pile are linked to one another by a relation that is both syntagmatic (they follow one another) and paradigmatic (they fill the same syntactic slot with respect to their common governor). The syntactic analysis of the other elements – junctors, paradigmatic adverbs, and list completers – is discussed. We also propose a typology of the different cases of pile structure and introduce the seven subcases of paradigmatic links taken into account in the annotation.

1. Introduction

This chapter is devoted to the analysis of various phenomena that we have considered and annotated as special cases of a more general syntactic configuration we call lists. List phenomena include, among others, coordination (1), disfluency (2) and reformulation (3):¹

- (1) *des jeunes { cagoulés | ^et armés } ont { pillé des magasins | brûlé { des pneus
^et des voitures } } //* [Rhap-M2006, Rhapsodie]
'youngsters { hooded | ^and armed } have { looted shops | burned { tires
^and cars } } //
- (2) *alors < { { j'â | j'avais } beaucoup | j'avais beaucoup } trop peur de m'installer
(comme ça) seule { d~ | dans } la brousse //* [Rhap-D2004, Lacheret]
'yet < { { I wã | I was } much | I was much } too frightened to settle (like
that) alone { i~ | in } the bush //

1. Lists are in bold font. Our annotation of lists using the tags “{”, “|”, “}”, and “^” will be presented in Section 2. Other tags correspond to the macro-syntactic annotation and are presented in Chapter 6.

- (3) *tu arrives place aux Herbes avec { une | une } sorte { de halle | "quoi" { de | de | de } structure métallique } //* [Rhap-M0001, Avanzi]
 ‘you reach Herb Square with { a | a } kind { of hall | "you know" { of | of | of } metallic structure} //’

Coordinations, reformulations, and disfluencies are generally analyzed as completely separate phenomena, or, concerning disfluencies, they are frequently not analyzed at all and classified as performance errors, and thus excluded from syntax in spite of their internal “grammar” and some constraints on their realization. The many existing studies on coordination, on the other hand, very rarely make a link with disfluency or reformulation (see however Levelt 1983, cited by Blanche-Benveniste et al. 1990).

We will start our study of list phenomena by arguing why we think that these various phenomena should be brought together. First, the juxtaposition of two phrases belonging to the same category is a priori ambiguous. For instance, in (4), *de la production musicale* ‘of musical production’ could be a reformulation of *de la poïésis* ‘of poïesis’ or a coordinated element.

- (4) *^et si vous faites de la musique < "eh bien" vous avez l'expérience { de la poïésis | { de la | de la } production musicale } //* [Rhap-M2002, Rhapsodie]
 ‘^and if you practice music < "well" you have experience { of poïesis | { of | of } musical production } //’

In the same way, the juxtaposition of *La Seine* ‘The Seine’ and *la Bastille* ‘the Bastille’ in (5) could be either a reformulation or a coordination.

- (5) *on traverse { la Seine | "euh" la Bastille } //* [Rhap-D0001, CFPP2000]
 ‘you cross { the Seine | "um" the Bastille } //’

It was only by closely inspecting the meanings of these two utterances in context and by listening to them (and consequently by capturing their prosody) that we were able to decide that (4) is a reformulation and (5) an additive coordination (‘you cross the Seine and then the Bastille’).

From a strictly syntactic point of view – and this is our second important argument – (4) and (5) are identical structures: in both cases two elements² occupy the same structural position, that is, they have the same syntactic function and the same governor. We call these syntactically identical configurations *lists*.³

2. We use the term *element* to refer to a contiguous segment of text, which commonly corresponds to a connected part of the dependency tree (called a *catena* by Osborne et al. 2012, and a *fragment* by Gerdes & Kahane 2011). Often an element is also a phrase, that is, a saturated subtree.

3. The difference between reformulations and coordinations can be compared to the difference between the interrogative and the declarative sentences *You like it.* vs. *You like it?*. In spite of the

To sum up, the distinction between coordination and reformulation (but also the distinction between reformulation and disfluency) is not a matter of syntax, but it is a matter of semantic interpretation.⁴ We decided therefore to provide a two-step annotation: we first identified on purely syntactic grounds all the elements fulfilling the same syntactic function and having the same governor and annotated them as general list phenomena, without making any distinction; in a second step, we used semantic arguments to functionally tag lists as coordinations, reformulations, or disfluencies (Section 3).

In order to better grasp what a list is from a formal point of view, it is necessary to examine in more detail the historical and theoretical underpinnings of our annotation, which is based on the seminal work by the French linguists Lucien Tesnière and Claire Blanche-Benveniste. Tesnière (1959) was probably the first linguist to characterize coordination as a device orthogonal to dependency and he called this device *junction* (Fr. *jonction*):

Junction consists of the joining of two nodes of the same sort. [...] Two nodes cannot be joined unless they fulfill the condition that they are of the same nature and as a consequence, they appear on the same structural level. The junction node has to be horizontal.

(Tesnière 1959: 323–326)

In the representations proposed by Tesnière, dependencies are vertically aligned, whereas junction phenomena are represented by horizontal lines. Figure 1 illustrates a Tesnière-like representation of example (1).

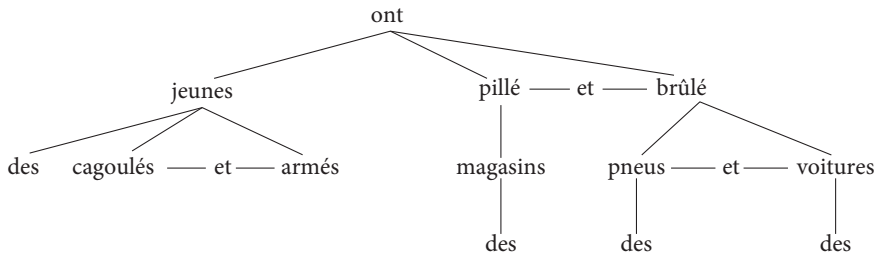


Figure 1. Tesnière-like dependency-based representation of (1)

Claire Blanche-Benveniste – who was the first linguist to acknowledge the importance of list constructions in spontaneous spoken productions and the first to suggest lumping them together on the grounds that they realize the same process

important semantic difference between them, the two utterances are generally considered to have the same syntactic structure.

4. In other words, we consider that all these list phenomena are the product of the same syntactic mechanism, even if they can show prosodic differences.

of listing (Blanche-Benveniste et al. 1979; Blanche-Benveniste et al. 1990) – proposed a similar (albeit inverted) representation of spoken productions in which she aligned horizontally the syntagmatic development of the utterance, and vertically, the paradigmatic lists of elements occupying the same syntactic position (Blanche-Benveniste et al. 1979, but see also Blanche-Benveniste et al. 1990, 1997; Bilger 2000; Pietrandrea 2008; Masini & Pietrandrea 2010; Kahane 2012). Example (6) illustrates the grid analysis of (1).

- (6) *des jeunes cagoulés et armés ont pillé des magasins brûlé des pneus et des voitures*
 youngsters hooded and armed have looted shops burned tires and cars

It is clear from Tesnière's and Blanche-Benveniste's representations that when an element piles up with another element, it has a special relation – which Blanche-Benveniste would call a *paradigmatic relation* – with the first element, and is indirectly governed by the first element's governor.

This triangular relationship cannot be easily rendered in pure government structures. The proof is that any attempt to formalize even the simplest list phenomenon, that is, coordination, in terms of government leads to very complicated modeling. Ivanova et al. (2012), while comparing different dependency schemes from various aspects, note that “the analysis of coordination represents a well-known area of differences” and, even on a simple example like *cotton, soybeans and rice*, “none of the formats agree.”

Syntactic analyses of coordination can generally be divided into two families: symmetrical and asymmetrical analyses (and mixed forms can be placed on a scale between these two families). Symmetrical analyses aim to give equal status to each conjunct, disregarding the linear order. Asymmetrical analyses on the contrary give a special status to one, commonly the first, of the conjuncts, and iteratively place the other conjuncts below the special one. Each of these families has advantages and disadvantages that have been widely discussed in the literature.

We can note that a symmetrical analysis constitutes a higher abstraction from the surface in the sense that paradigmatic phenomena are by definition formed by conjuncts filling a common dependency slot offered by a syntactic governor. However, placing the conjuncts on an equal level poses the problem of choice of the governor among the different participants in the coordination. Some X-bar approaches use co-heads for coordinations in order to avoid the identification of a single head (Jackendoff 1977). Similarly, the Negra-type annotation used in Alpino (van der Beek et al. 2002) and CGN (Dutch Spoken Corpus, Schuurman et al. 2004) falls back on flat headless constituents for the description of coordinations. But dependency annotation assumes that a single word or morpheme can be identified as the head of any substructure. The Prague Dependency treebank

(Hajič 1998) uses the coordinative conjunction (or even punctuation marks) as the head of the coordination in order to obtain a dependency-based symmetrical analysis.

An asymmetrical analysis, also called Mel'čukian analysis of coordination (Mel'čuk 1988, used in CoNLL 2008, Surdeanu et al. 2008), on the contrary, represents the surface configuration better: the coordinating conjunction usually forms a phrase with the following phrase, like *and rice* in the above example, and only an asymmetrical dependency-based analysis contains this segment as a subtree.⁵

Phrase structures, just like dependency trees, therefore excluding multiple governors for the same node, have to choose between a symmetrical and an asymmetrical analysis. In contrast to approaches that tend to encode paradigmatic and syntagmatic relations with the same tools, we follow the two-dimensional analysis of Tesnière and we recognize that there exists a process of listing complementary to government. We do not choose between the symmetrical and asymmetrical analysis of coordination and our analysis subsumes both. It should be incidentally pointed out that our structure, similar to Tesnière's one, is not a tree but a Directed Acyclic Graph (DAG). A DAG differs from a tree by the fact that a vertex can have multiple governors.

All in all, list analysis proposes a unique modeling of all types of paradigmatic phenomena, a modeling that is orthogonal to government. In the following sections, we will present the formal properties of lists, their internal structure, and their encoding in the Rhapsodie treebank. In Section 3 we will highlight the advantages of choosing an annotation in lists. Section 4 is devoted to the categorization of lists into semantically motivated classes, including coordination, disfluency, and reformulation introduced above.

2. Structure of lists

As we have seen above, a list is a syntactic cohesion device that is orthogonal to government. Since we say that an element Y piles up on an element X if it occupies the same syntactic position as X, we claim that Y is in the same *government unit* [GU] as X (Chapter 4). Lists are to be regarded therefore as a particular type of microsyntactic phenomena to be analyzed within the theoretical framework of dependency syntax. In the following sections we will examine the structure of lists in greater detail.

5. It is worth recalling that dependency trees acknowledge some phrases, the maximal projection of words. Indeed the maximal projection of a word is the phrase formed by the words belonging to the subtree rooted by this word.

2.1 Layers

We use the annotation $\{ X \mid Y \}$ to represent the piling up of an element Y on an element X . In the list $\{ X \mid Y \}$, X and Y are the layers of the lists. A list can have two or more layers. When layers are not contiguous we use the following annotation: $\{ X \mid \dots \mid Y \}$, as shown in (7).

- (7) *si je ne craignais pas d'entrer dans le jeu de certains hommes qui abusent de leur condition < je dirais que vous avez donné { quelque chose de plus } à la femme //+ { | des armes de persuasion } //* [Rhap-D2001, Corpus Mertens]
 'if I wasn't afraid to play along with some men who abuse their condition < I would say that you have given { something more } to woman //+ { | arms of persuasion } //'

2.2 Conjuncts

A layer of a list may comprise many elements. The central element of a layer is the *conjunct*. The conjunct is a phrase formally defined by the fact that it can occupy by itself the syntactic slot occupied by the entire list:

- (8) a. *des jeunes { cagoulés | ^et armés }*
 ' { hooded | ^and armed } youngsters'
 b. *des jeunes cagoulés*
 'hooded youngsters'
 c. *des jeunes armés*
 'armed youngsters'

2.3 Junctors

Junctors are the elements that connect the conjuncts of a list (we annotate them by placing a carat $\wedge$ to their left). We adopt here, following Blanche-Benveniste et al. (1990) and Ndiaye (1989), a variant of the term “jonctif” used by Tesnière (1959) (cf. also the term “list marker” used by Gerdes and Kahane (2009)). Junctors coincide more or less with coordinating conjunctions. More precisely, junctors are elements that are part of a list structure, as they form a layer with one of the conjuncts of the list, but they are not part of the conjunct itself as they cannot be maintained if we conserve only one layer of a list:

- (9) a. *des jeunes { cagoulés | ^et armés }*
 ' { hooded | ^and armed } youngsters'
 b. **des jeunes et armés*

2.4 Paradigmatizing adverbs

As shown by Masini and Pietrandrea (2010), some adverbs can join the layer of a conjunct they have scope over, like *peut-être* ‘maybe’ in (10).⁶

- (10) { *la* | *la* } *société est* { *plus stratifiée* | ^*et peut-être plus variée* } { *que* | *qu’* } à la campagne // [Rhap-D1001, ESLO]
 ‘{ the | the } society is { more stratified | ^and maybe more varied } { than | than } in the countryside //’

In the literature, these adverbs are called *paradigmatizing adverbs* (Nölke 1983, 2001) or *focus particles* (König 1991). They are generally placed between the (optional) junctor and the conjunct. The word *paradigmatizing* refers to the semantic fact that these adverbs open a paradigm of choices that could alternatively occupy the element the adverb has scope over. This holds despite the fact that the adverb is commonly analyzed as a dependent of the main verb and it therefore has no syntactic relation with the conjunct it has scope over (here *plus variée* ‘more varied’). The reason for analyzing the adverb as depending on the main verb is that the adverb can appear independently of paradigmatic structures (*la société est peut-être plus variée qu’à la campagne* ‘the society is maybe more varied than in the countryside’) and can be separated from the element it has scope over (*la société a peut-être été plus variée qu’à la campagne* ‘the society has maybe been more varied than in the countryside’).

The grouping in the same layer of the possible junctor and paradigmaticizing adverbs, and the conjunct (*et + peut-être + plus variée*) and the possibility of using the layer alone (as an answer to a question for example) could give rise to the hypothesis that the layer forms a phrase, and that therefore there should be a syntactic connection between the paradigmaticizing adverb and the rest of the layer. In spite of the strong semantic relationship between the paradigmaticizing adverb and the conjunct induced by the scope relationship, it is difficult to determine the type or the direction of any syntactic relationship that could exist between them. In the Rhapsodie treebank, we thus decided to consider only the adverb’s link with the verb, accompanied by an independent explicit marking of the lists and the layers. As an example, see the dependency from *est* ‘is’ to *peut-être* ‘maybe’ in Figure 2.

6. Paradigmatic adverbs clearly have scope over one particular element of the sentence:

- i. *Peter will maybe give the book to Mary (unless he will only lend it)*
- ii. *Peter will give maybe the book to Mary (or maybe something else)*
- iii. *Peter will give the book maybe to Mary (or maybe to another person)*

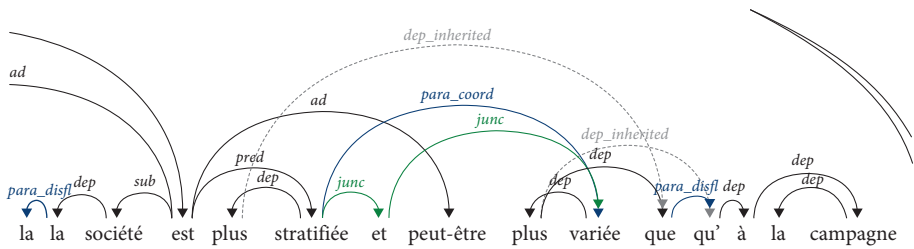


Figure 2. Microsyntactic structure of (10)

2.5. List completers

Lists are sometimes characterized by the presence of what we call, following Jefferson’s (1991) terminology, “list completers”. These elements – also called in the literature “set markers” (Stenström et al. 2002) or “set-marking tags” (Dines 1980; Ward & Birner 1993), or general list completers (Selting 2007) – have the property of closing a list. A typical example of completer is the word *et caetera*:

- (11) *elle voit le garçon pâtissier "euh" qui transporte des plateaux chargé de {
friandises | gâteaux | et caetara } //* [Rhap-M0021, Rhapsodie]
'she sees the young baker "um" who is carrying trays of { sweets | cakes | **et
caetara } //**

These elements cannot occupy the governed position alone: **des plateaux chargés de et caetera*, ‘trays full of et caetera’. Thus these elements, contrary to other conjuncts, do not inherit the dependency, as shown in Figure 3:

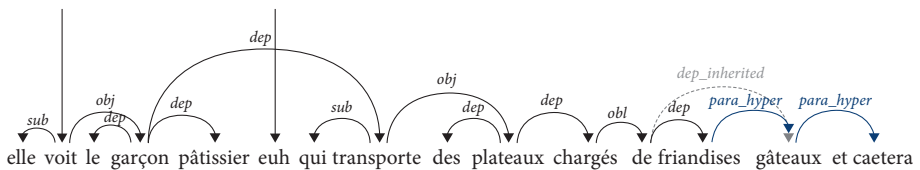


Figure 3. Microsyntactic structure of (11)

It should be mentioned that there also exist a number of completers that are composed of a junctor and a particular type of conjunct, such as *tout ça* ‘all that’ in (12). Following Overstreet (2005), we call these elements “general extenders”.

- (12) *et "euh" "bon" "ben" ça pose des problèmes { de mainten~ | "enfin" de
maintenance "euh" | { { de | de } mise à jour | ^et tout ça } "euh" } //* voilà //
[Rhap-D0005, PFC]
'and "um" "well" "like" that raises problems { of mainten~ | "I mean" of
maintenance "um" | { { of | of } updates | ^and all that } "um" } // you know //

The lexeme *et caetera* acts like an agglutination of a junctor (*et* ‘and’) and an extender (*caetera* ‘others’), which is what *et caetera* used to be in Latin. This explains why completers, like junctors, can only appear within lists.

2.6 Dependency and inherited dependency

A list occupies a governed position in a dependency tree. A tree-based analysis assumes that each governed position receives one and only one dependency from the governor of this position, which imposes an asymmetrical analysis. We make a compromise between a symmetrical and an asymmetrical analysis of lists (see discussion in 2.8 below), considering that one conjunct is really dependent on the governor of the list, while the others inherit this dependency. For instance, in Figure 4, the dependency on the governor of the list (dependency *obj* on *donné*) is attributed to the first conjunct (*quelque chose de plus*), while the second conjunct (*des armes de persuasion*) receives an inherited dependency *obj_inherited*.

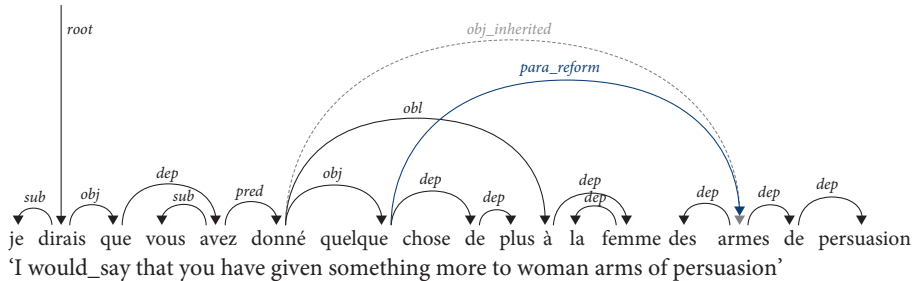


Figure 4. Microsyntactic representation of (7)

It is not always straightforward to decide which conjunct carries the genuine dependency connecting it with the context and which conjunct only inherits this dependency. Except for coordination, we consider that the genuine dependency belongs to the nearest conjunct because this conjunct is generally the one that forms a prosodic constituent with this context, even if the governor of the list follows the list (Figure 5).

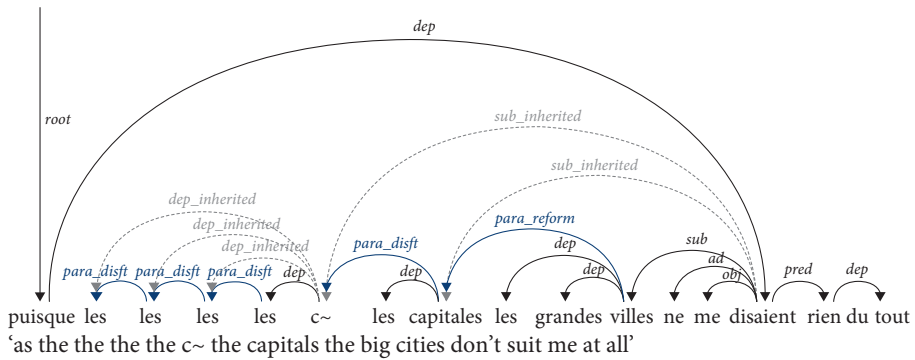


Figure 5. Microsyntactic structure with right to left inheritance

It can be noted that our analysis always tries to limit the use of empty syntactic positions and ellipsis. Thus in (13), we consider that elements of different speech turns form a list and the four IUs form only one GU.

- (13) \$L1 ^et il donne { à Gaga } //+
\$L2 { | à { Gago } } >+ effectivement //+
\$L1 { | Gago } | "pardon" //+
\$L2 { | Gago } { qui est contré | qui est contré } // [Rhap-D2003, Rhapsodie]
'\$L1 ^and he passes { to Gaga } //+
\$L2 { | to { Gago } } >+ indeed //+
\$L1 { | Gago } | "sorry" //+
\$L2 { | Gago } { who is tackled | who is tackled } //

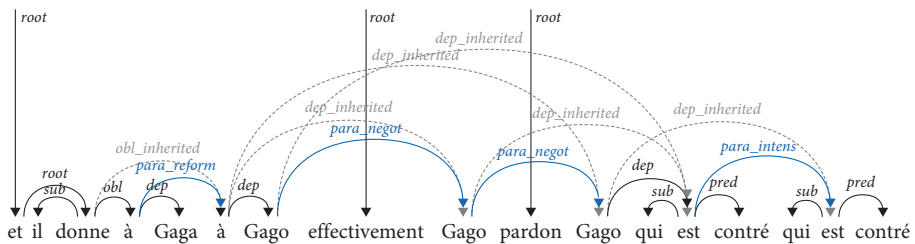


Figure 6. Microsyntactic structure of (13)

2.7 Paradigmatic links

As theorized by Tesnière and Blanche-Benveniste, the conjuncts of a list are not only linked to their governor by a government relation, but they are also linked to each other by a relation that is both syntagmatic (the conjuncts follow one another) and paradigmatic (the conjuncts fill the same syntactic slot with respect to their common governor). We represent this relation through a particular type

of microsyntactic link, called *paradigmatic link*. This link is tagged according to the type of relation holding between the conjuncts (we distinguish between coordination paradigmatic links, reformulation paradigmatic links, disfluency paradigmatic links etc. – see Section 4 for further details). Figure 4 above, for instance, shows a paradigmatic reformulation link between *quelque chose* ‘something’ and *armes* ‘arms’.

The introduction of paradigmatic links as part of the syntactic structure is one of the major contributions of our annotation scheme to the representation of microsyntactic relations in spoken discourse. We have shown in the introduction that existing dependency annotation schemes differ widely as far as the analysis of paradigmatic phenomena is concerned, thus reflecting important underlying syntactic choices, which often remain implicit. We will show in Section 3 that the introduction of paradigmatic links allows for a more comprehensive analysis of paradigmatic phenomena, which has a number of advantages especially for the modeling of well-known difficult cases of coordination and listing.

2.8 Junction links

Simple junctors such as *et* ‘and’, *ou* ‘or’, *mais* ‘but’, etc. are placed between two conjuncts. Following the asymmetrical analysis of coordination (Mel’čuk 1988), we claim that the junctor forms a phrase with the following conjunct and that the phrase adjoins to the preceding conjunct. As the junctor controls the distribution of the layer that it opens (specifically indicating that this phrase is a layer), we therefore treat it as the head of this layer. All in all we thus have the following dependency relations for a list such as { *Mary* | ^and *Peter* }; *Mary* → and → *Peter*. These dependencies are orthogonal to government relations. We tag these dependencies as a particular function *junc* (following Tesnière, who calls this type of relation *jonction* ‘junction’).

Junction links are always overarched by a longer relation. Generally, this relation is a paradigmatic link. It is only in the case of double junctors and junctions without lists (see below) that junction links are overarched by a government relation. The fact that the second layer receives two dependencies from the first conjunct (one junction link to the junctor and one paradigmatic link to the second conjunct) can be related to the well-known fact that the junctor and the conjunct share head properties (see, for instance, Tseng (2002) for the modeling of junctors as weak heads).

A close look at our analysis of example (1) in Figure 7 reveals that the latter is in fact similar to Tesnière’s representation (Figure 1) except for the fact that:

- a. there are paradigmatic links overarching junction links;

- b. junction links are oriented;
- c. we distinguish between “genuine” and inherited government links.

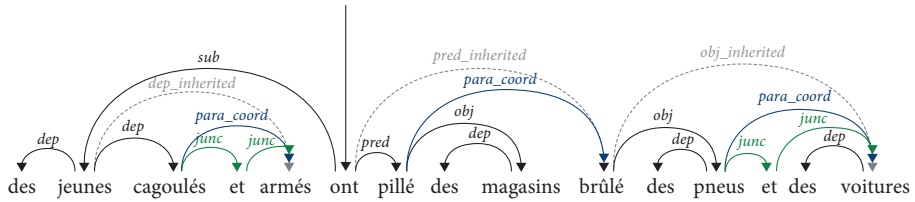


Figure 7. Microsyntactic structure of (1)

3. Resolution of complex cases of coordination

We modelled lists as structures made up of conjuncts, and, optionally, junctors, paradigmatising adverbs and completers. The internal cohesion of these structures is guaranteed by links orthogonal to government proper: paradigmatic links and junction links. The cohesion of lists with the context is guaranteed by the fact that each conjunct is governed by either a direct or an inherited dependency on a governor external to the list. This modeling allowed us to elegantly account for a number of structures which are usually considered as the stumbling block of coordination modeling: scope ambiguities, gapping, correlative structures, embedded coordinations, and junction links outside list structures. Let us examine these particular cases.

3.1 Scope ambiguity

Having posited inherited dependencies in our model, we can easily represent the disambiguation of cases like (14).

- (14) a. *old* { *men* | ^and *women* }
 b. { *old men* | ^and *women* }

In both cases, *old* is a dependent of *men*, but it can optionally be inherited by *women*. This encoding, which still follows the asymmetrical analysis of coordination, subsumes the symmetrical analysis and allows us to compute the desired phrase structure. Whenever a word is governed by an inherited dependency, we know that it is a shared dependent, which is located outside of the list. Each conjunct is the projection of the word linked by the paradigmatic links with the exclusion of shared dependents and the list is the projection of the first conjunct without the shared dependents. We thus obtain the phrases (15a) and (15b).

- (15) a. old ((men) and (women))
 b. ((old men) and (women))

Note also that generally no satisfactory phrase structure representation exists for cases of shared dependents. For example in sentence (16), *flou* ‘vague’ modifies both occurrences of *concept* and none of the layers is a phrase.

- (16) *^mais { c’est un concept | c’est ce qu’on appelle un concept } flou //*
 [Rhap-M2002, Rhapsodie]
^but { it’s a concept | that’s what we call a concept } vague
‘but {it’s a vague | that’s what we call a vague} concept’

The analysis of this example, given in Figure 8, is unproblematic in our annotation scheme.⁷

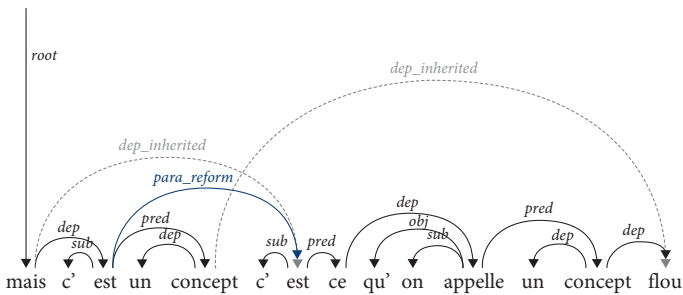


Figure 8. Microsyntactic structure of (16)

3.2 Gapping and non-constituent coordination

Gapping is traditionally analyzed as a coordination of clauses where the second clause is elliptic (see, for instance, Jackendoff 1971). Following Gerdes & Kahane (2009), we consider that both gapping and non-constituent coordination are cases of parallel lists, where a single coordination involves two lists (or more) simultaneously, without any ellipsis.⁸ In (17a), the only example of gapping in the Rhapsodie corpus, *le maire de Pointe-à-Pitre Jacques Bangou* lists on *le LKP*, while simultane-

7. Conversely, this annotation scheme does not exclude cases where the “path” between the head of a conjunct and a shared dependent does not respect some additional constraints as in **(Peter plays on and knows the guy who owns) this piano*. These constraints have to be stated independently. The question of to what extent formal frameworks should intrinsically model language is an interesting subject in itself but cannot be discussed at length here.

8. In Gerdes and Kahane (2015), we extend the present analysis by introducing an additional link between the constituents of the “elliptic” conjunct in order to give an account of the fact it works as a syntactic unit.

ously *un cri d'alarme* piles on *un appel au calme*. In (17b), the only example of non-constituent coordination, *un an* piles on *un an* and *à Lisbonne* on *à Porto*.

- (17) a. { *le LKP lance un appel au calme* | ^et { *le maire de Pointe-à-Pitre* | *Jacques Bangou* } *un cri d'alarme* } // [Rhap-M2006, Rhapsodie]
 ' { the LKP issues an appeal for calm | ^and { the mayor of Pointe-à-Pitre | Jacques Bangou } an alarm } //' [Rhap-D1003, Rhapsodie]
- b. *j'ai été* { *un an à Lisbonne* | ^et *un an à Porto* } { { *dans les* | *dans des* } *collèges* | ^et { *dans des* | *dans des* } *primaires* { *portu~ "euh"* | *"enfin" françaises* } } //
 'I spent { *one year in Lisbon* | ^and *one year in Porto* } { { in the | in } middle school | ^and { in | in } { *Portu~ "um"* | "I mean" French } primary schools } //

We model the fact that there is a single coordination and the junctor *et* marks both paradigmatic links simultaneously by linking it to the conjuncts of both lists (Figure 9).

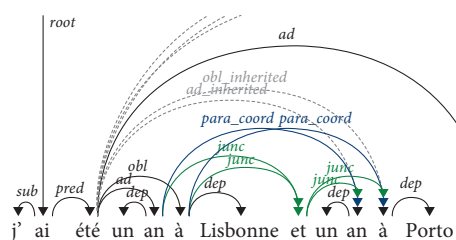


Figure 9. Microsyntactic structure of (17)b

3.3 Embedded coordinations

Consider a classical example like (18).

- (18) *We are looking for someone who speaks French and German or Italian.*

Two interpretations are possible as shown in (19a) and (19b).

- (19) a. { French | ^and { German | ^or Italian } }
 b. { { French | ^and German } | ^or Italian }

In our analysis, in both cases we have the third layer (*or Italian*) attached to the second layer (*and German*): French → and → German → or → Italian.⁹ But in case

9. Mel'čuk (1988) proposes, in case (b), to attach *or Italian* to the head of the group *French and German*, that is to *French*. We disagree with this analysis because *or Italian* is a shared dependent of both *French* and *German*, and as usual a shared dependent must be attached to the nearest

(a), *Italian* inherits a *junc_inherited* dependency from *and* because it is coordinated with the dependent *German* of *and*, while in case (b), *or Italian* is a shared dependent and *or* inherits a *junc_inherited* dependency from *French*, which is coordinated with *German*.

We encountered an example of type (a) in our corpus (see (20)), wherein the third conjunct inherits a link of type *junc_inherited* (see Figure 10).

- (20) *tout chez lui est fait { de compassion | ^et non { de violence | ^ou de châtiement
 } } //* [Rhap-M2003, Rhapsodie]
 ‘everything about him is made { of compassion | ^and not { of violence | ^or
 of punishment } } //’

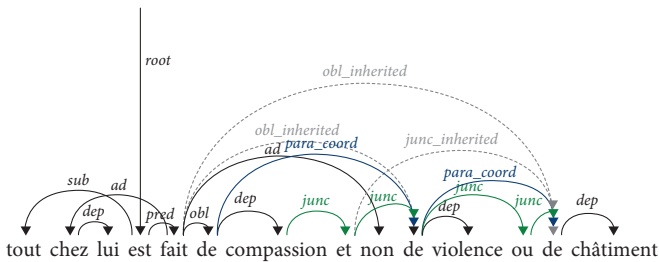


Figure 10. Microsyntactic structure of (20)

3.4 Correlative structures

When the first conjunct of a list is introduced by a correlative junctor such as *soit ... soit* ‘either...or’ as in (21), we represent this junctor as dependent on the governor of the list. See the junction link from *marqué* ‘marked’ to the first occurrence of *soit* ‘either’ in Figure 11.

- (21) *est-ce que vous avez des enseignants { { dont | dont } vous vous souvenez
 particulièrement | qui vous ont marqué "euh" { ^soit au moment de l'école |
 ^soit au moment du lycée } } //* [Rhap-D0001, CFPP2000]
 ‘do you have teachers { { that | that } you remember in particular | who have
 left a mark on you "um" { ^either at school | ^or in college } //’

conjunct it modifies, that is *German*. In any case, the tree Mel’čuk obtains has *French* with two dependents: *German* ← *and* ← *French* → *or* → *Italian*. This tree is semantically very ambiguous and corresponds also to (*French or Italian*) and *German*, which is not at all equivalent to the (b) interpretation of our example.

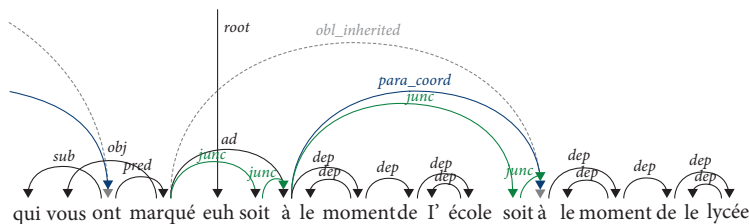


Figure 11. Microsyntactic structure of (21)

This analysis allows for a parallel representation of the two correlative junctors: as shown in Figure 11 they are both taken as governors of a conjunct. Again the *junc* links are overarched by a dependency, but this time it is not a paradigmatic link but a government link. Nevertheless, the principle is similar: the junctor acts as a marker of the link that it doubles.

3.5 Junction without list

In a construction such as *he speaks French and well*, the coordination does not include two conjuncts that fill the same syntactic slot and we do not stipulate the existence of a paradigmatic link. We consider that what we have here is a coordination between *illocutionary units* [IUs].¹⁰ In fact, the speaker makes two assertions ('he speaks French' and 'he speaks French well') in one GU consisting of two IUs: *he speaks French* //+ ^*and well* //. We model these coordinations without the use of ellipsis, only by distinguishing GUs and IUs.

This type of construction appears to be fairly frequent and we have encountered quite a few in our corpus, see examples in (22).

- (22) a. *il frappe de loin //+ ^mais dans les nuages //* [Rhap-D2003, Rhapsodie]
 ‘he kicks from far back //+ ^ but into the sky //’
 b. *on veut bien parler avec vous //+ ^mais { ã | après } le déménagement //*
 [Rhap-D0006, CFPP2000]
 ‘we are willing to talk with you //+ but { ã | after } the removal //’
 c. *normalement < c’est du bois de hêtre dessous //+ ^et { qui est p~ | qui est
 laqué noir } //* [Rhap-D0009, PFC]
 ‘usually < it’s beech wood under it //+ ^and { that is p~ | that is
 lacquered black //’

10. Illocutionary Units (IUs) are sequences encoding one and only one illocutionary act. A definition of IUs is provided in Chapter 6 and other examples of GU beyond IU are studied (see the analysis of epexegesis).

- d. [*il va sans doute faire la même chose qu’ avant //*] >+ *pronostique*
Francis Brochet //+ ^mais autrement // [Rhap-D2013, Rhapsodie]
 ‘[he will no doubt do the same thing as before //] >+ predicts Francis
 Brochet //+ ^but differently //’
- e. *^ou alors < il y a { une petite rue (+ ^mais dont je ne sais pas le nom) |*
une petite rue { en & | qui tourne un peu } } // [Rhap-M0011, Avanzi]
 ‘^or else < there is { a little street (+ ^but whose name I don’t know) | a
 little street { in & | that winds slightly } }’

In such cases, the two junction links are overarched by a government dependency (see Figure 12).

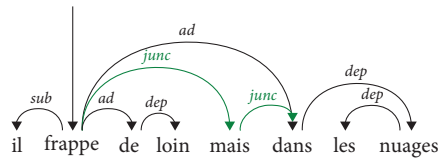


Figure 12. Microsyntactic structure of (22)a

4. Types of lists

In our annotation we did not merely indicate the existence of a syntactic relation of listing between conjuncts, but we also semantically classified these relations. The typology of lists used for our annotation task is a slightly simplified version of the fine-grained typologies proposed by Bonvino et al. (2009) and Kahane and Pietrandrea (2012). Other studies that propose typologies of list phenomena include Blanche-Benveniste et al. (1990), Blanche-Benveniste (1995), Bilger (2000), and Guénot (2006). From a purely theoretical point of view, our classification distinguishes two main types of list phenomena:

- *coordinations* or *de re* lists play the role of constructing an expression with its own denotation; in other words, coordinations are constructions (i.e., linguistic signs) that make a real semantic contribution to the utterance;
- *de dicto* lists play the function of establishing relations between formulations, whether the conjuncts do not have a real denotation (as in disfluencies), or denote the same referent in different ways (reformulations). *De dicto* lists do not provide as such a real semantic contribution and are not linguistic signs, that is, conventional pairs of form and meaning, but simply regular patterns of performance.

For our tagging of lists we distinguished three subtypes of coordinations (relational coordinations, hypernymic coordinations, intensifications) and three subtypes of *de dicto* lists (disfluencies, reformulations, double formulations). We also annotated a last type of list, which we called negotiations.

4.1 Relational coordination (para_coord)

Relational coordinations (or coordinations *tout court*) are coordinations in which every conjunct denotes a different referent and the entire list has a denotation that is a function of the denotation of the conjuncts. We annotate as para_coord two main types of coordinations:

- additive coordinations, that is, lists whose denotation corresponds to the union of the denotation of conjuncts; see (1), (5), (10), (20), (21), and (23).¹¹

(23) *je travaille à la préfecture de Paris qui { n'est pas connue | ^mais néanmoins existe } "euh" //* [Rhap-D0001, CFPP2000]
 'I'm working at the prefecture of Paris that { isn't well known | ^but yet exists } "um" //

- alternative coordinations, that is, lists whose elements are potentially substitutable (Mauri 2008: 47) as in (24).

(24) *allez // avec Messi { qui va chercher le corner | ^et qui va trouver { le corner | ^ou la touche | ^ou la sortie de but } } //* [Rhap-D2003, Rhapsodie]
 'go // with Messi { who is going for the corner | ^and who will make the { the corner | ^or the touch | ^or the back line } } //

4.2 Hypernymic coordination (para_hyper)

Following Bonvino et al. (2009), we acknowledge the existence of many types of non-relational coordinations. These coordinations do not establish a denotation for each conjunct, rather they use the piling of many conjuncts to create a denotation as in (25).

11. Many linguists consider adversative coordinations as a third case of coordination (along with alternative and additive) (see for example Mauri 2008, among others). We consider that the adversative value of junctors such as *mais* 'but' is superposed and simply specializes their additive value.

- (25) \$L1 *mais je trouverai pas de livre sur { Ségolène Royal | ^ou Nicolas Sarkozy }*
 //
 \$L2 "ah" non // ça < { non | non } // ça < *les livres politiques* < "enfin" non //
 [Rhap-D2002, Rhapsodie]
 '\$L1 but I won't find a book on { Ségolène Royal | ^or Nicolas Sarkozy } //
 \$L2 "ah" no // that < { no | no } // that < *political books* < "I mean" no //

In Example (25), the list *Ségolène Royal ou Nicolas Sarkozy* does not denote an alternative between these two politicians, rather it denotes any politician, as shown by the addressee's reference to generic *les livres politiques* 'political books'. Most non-relational coordinations are hypernymic coordinations, that is, coordinations whose meaning denotes a hypernym of the conjuncts, without corresponding to a logical combination of them. See (11) and (12) where hypernyms such as 'sweet things' and 'software management' are constructed. Hypernymic coordination can also coordinate verbal conjuncts as in (26).

- (26) *parce que il a dit [{ elles corromp~ | elles corrompront } tous mes petits "euh" officiers de district //] "euh" { sans | sans } { me connaître | ^ni rien du tout }*
 //
 [Rhap-D2004, Rhapsodie]
 'because he said [{ they cor~ | they corrupt } all my little "um" district officers //] "um" { without | without } { *knowing me* | ^or anything at all }
 //

4.3 Intensification (para_intens)

Intensive coordination has a general function of intensification of the meaning of the repeated conjunct. This function of intensification is specified according to the category of the head of the conjunct. In (27), for example, *des exercices des exercices des exercices* can be paraphrased by 'a lot of exercise'.

- (27) { *le | la | le | le | la* } *grosse recette de Sarah "tu vois" < c'était de de faire { des exercices | des exercices | des exercices } par exemple "tu vois" pour un point de grammaire //*
 (Valibel)
 '{ the | the | the | the | the } big recipe of Sarah "you see" < was to to do { *exercises | exercises | exercises* } for example "you see" for a grammar point //

The reiteration of verbal conjuncts can intensify the duration or the frequency of an action and hence mark the continuative or iterative aspect of the action (see (28)).

- (28) *on pouvait pas s'empêcher à la fin de { Mort | ^et transfiguration } de faire
 { résonner | résonner | ^et résonner | ^et encore } ces accords qui nous
 enchantaient //* [Rhap-D2012, Rhapsodie]
 'we could not help at the end of { Death | ^and Transfiguration } letting
 the chords that enchanted us { resonate | resonate | ^and resonate | ^and
 again } //

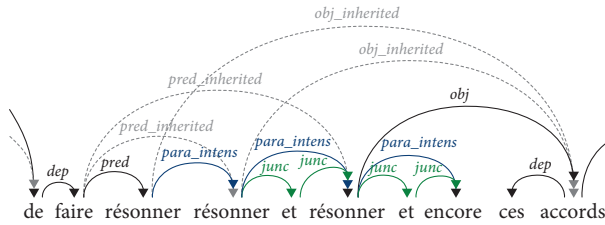


Figure 13. Microsyntactic structure of (28)

4.4 Disfluency (para_disfl)

Disfluencies are characterized by the fact that the speaker stalls at a given syntactic position in order to improve her formulation. Generally speaking, this insistence concerns the first part of a constituent and in particular its grammatical elements (determiners, prepositions, auxiliaries), which are repeated or reformulated while the speaker seeks a lexical element adequate to her discourse. Sometimes disfluencies entail the repetition of quite long segments, as in (2) or in the example of Figure 5. In disfluencies, the elements piled up do not have a true denotation. We annotate a list as *para_disfl* only if there is no lexical change, with the exception though of grammatical words, as in (29).

- (29) *est-ce que tu connais { des | un } endroit où je pourrais les laisser pour le
 rempaillage //* [Rhap-D0005, PFC]
 'do you know { some | a } place where I could leave them to have the rush
 seats repaired //

4.5 Reformulation (para_reform)

Reformulations are lists composed of a first denotative formulation which is substituted by one or more denotative formulations of the same referent, as in (3), (4), and (30).

- (30) *^et j'avais { une circonscription | { un | un } rayon d'action } d'à peu près
 "euh" cent kilomètres tout autour de cet endroit //* [Rhap-D2004, Rhapsodie]
*^and I had { an area | { a | a } range } of about "um" one hundred kilometers
 all around that place //*

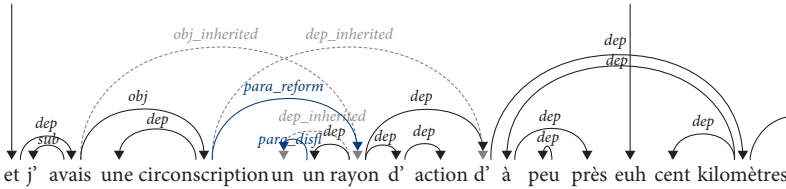


Figure 14. Microsyntactic structure of (30)

The reformulation occurs in general within one and the same IU. However, interruptions, even long interruptions, may occur between two layers of a reformulation. In (31), for example three IUs interrupt the denotative reformulation.

- (31) *dans le vingtième <+ il faudrait { qu'il y ait & | qu'on sépare & | "enfin" qu'il y
 ait des cours de français pour les petits enfants { qui parlent pas français } //+
 (c'est pas compliqué quand même //*
c'est pas très difficile d'apprendre le français à des petits enfants de cet âge-là //
{ ça | ça | ça } se fait assez facilement //)
{| dont les mamans ne parlent pas français } // [Rhap-D0002, Rhapsodie]
*'in Paris' twentieth district <+ it would be necessary { that there be & | that
 they separate & | "I mean" that they organize classes for small children { who
 don't speak French } //+
 (it's really not that difficult //*
it's not very difficult to teach French to kids of that age //
{ that | that | that } can be done easily //)
{| whose mothers don't speak French } //

4.6 Double formulation (para_dform)

The *de dicto* lists we have so far examined (disfluency and denotative reformulations) can be regarded as alternative lists in that the denotation of the last layer replaces the preceding denotations. However there is also a *de dicto* list construction that can be seen as an additive list. This construction consists not in substituting one denotation for another, but in cumulating several denotations for the same referent, see (32) and (33).

- (32) *pour { Philippe Lemaire | (+ l'avocat des parties civiles //) } <+ { c'est d~ | ce sont des } procédés terroristes //* [Rhap-M2006, Rhapsodie]
 'for { Philippe Lemaire | (+ the victims' lawyer //) } <+ { that's | these are } terrorist acts //
- (33) *il y a eu en mille neuf cent dix huit sur l'ensemble de la planète on dit { quarante millions de décès | } // + { ^c'est-à-dire une mortalité effroyable } //* [Rhap-D2008, Rhapsodie]
 'there were in the year 1918 on the whole planet they say { forty million deaths | } // + { ^that ^is an appalling mortality } //

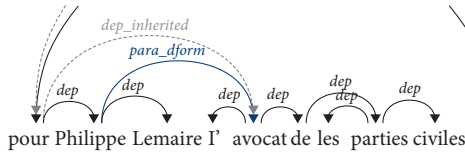


Figure 15. Microsyntactic structure of (32)

The second layer functions as an autonomous IU: it responds in fact to the nuclearity criteria that are defined in Chapter 6. We call this construction *double formulation*. It is quite frequent in spoken as well as in written language and can be regarded as an extension of the notion of non-restrictive apposition (Quirk et al. 1985).

We annotate as double formulation discontinuous lists that, following Quirk et al.'s (1985) terminology, can be regarded as *inclusive* double formulations: the denotation of the second conjunct is not identical to the denotation of the first one, but is included within it. Examples of inclusive double formulations are specifications (the first conjunct is constituted by an indefinite pronoun, such as *quelque chose* 'something' in (7), or a general noun like *endroit* 'place' in (34)), and exemplifications (the second conjunct is constituted by a list made up of co-hyponyms of the first conjunct, as in (35)).

- (34) *et j'ai trouvé { cet endroit | (+ Olkaloo //) } où ils avaient besoin d'un médecin //* [Rhap-D2004, Lacheret]
 'and I found { that place | (+ Olkaloo //) } where they needed a doctor //
- (35) *et j'avais absolument envie d'aller dans { la corne de l'Afrique | } //+ { | { la | l' } Ethiopia | la Somalie | { l'Orā | l'Arabie | } tout ça } } //* [Rhap-D2004, Lacheret]
 'and I definitely wanted to go to { the Horn of Africa | } //+ { | { Ethiopia | Somalia "um" | { Orā | Arabia | } all that } } //

We also include *wh*-question-answer pairs in double formulations. We saw in Chapter 4 that the extension of microsyntactic relations over speech turns is not

only entirely justifiable from a theoretical point of view, but also more simple and efficacious for the practical needs of annotation. *Wh*-question-answer pairs are a particular type of extension of a microsyntactic relation over the borders of a speech turn that we treat as a particular case of double formulation. These structures have all the properties of an inclusive double formulation: the *wh*-pronoun and the head of the second layer occupy the same syntactic position; they denote the same referent; the denotation of the second conjunct is included within the denotation of the first conjunct; the second layer forms a separate IU as in (36).

- (36) *"euh" { pourquoi } j'ai fait du journalisme //+ { parce que "euh" ça se passait tout de suite après la guerre } //* [Rhap-D2001, Mertens]
"um" { why } I got into journalism //+ { because "um" it happened right after the war } //

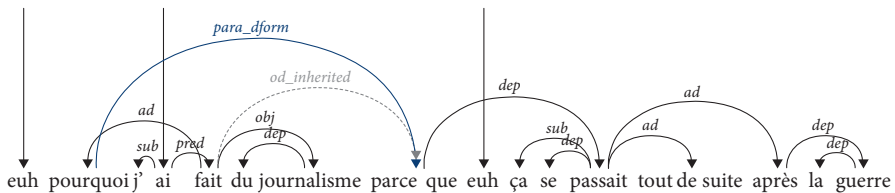


Figure 16. Microsyntactic structure of (36)

4.7 Negotiating formulation (para_negot)

Speakers can negotiate the construction of discourse as well as all the formulations introduced in discourse. As was shown by some pioneering interactional studies (Jefferson 1991; Selting 2007; Traverso 2012), lists are a privileged structure for this kind of discursive operations, which occur mostly, albeit not exclusively, in dialogues. We have identified four operations of negotiation occurring in the list structures of our corpus: confirmation, confirmation request, refutation and correction.

Example (37) illustrates a confirmation request (repetition of *les quais* ‘the platforms’ with an interrogative prosody) and a confirmation (with reformulation in *les quais du tram* ‘the tramway platforms’).¹² Example (38) is another example of confirmation (repetition of *quarante-huit ans* with an assertive prosody).

12. A negotiation list can be a simple repetition or include a reformulation. As shown in Chapter 7 this leads the annotators to frequent confusions between negotiation and reformulation.

- (37) \$L1 *vous allez longer { les quais } //+*
(vous allez passer devant { u~ | une } banque //
à l' angle <+ il y a une banque pour vous repérer //)
 \$L2 { *les quais* } //+
 \$L1 "euh" { *les quais du tram* } //
 \$L2 "ah" d'accord // [Rhap-D0008, Avanzi]
 ' \$L1 you will go along { the platforms } //+
 (you will pass in front of { ã | a } bank //
 at the corner <+ there is a bank, just to help you find your way //)
 \$L2 { the platforms } //+
 \$L1 "um" { the tram platforms } //
 \$L2 "ah" OK //'
- (38) \$L1 *puisque finalement < ça fait "euh" { quarante-huit ans } que vous êtes au Kenya //*
 \$L2 { *quarante-huit ans* } // { *oui | oui* } // [Rhap-D2004, Lacheret]
 ' \$L1 because finally < it's "um" { forty-eight years } that you've been in Kenya //
 \$L2 { *forty-eight years* } // { *yes | yes* } //'

The exchange in (13) above includes an example of correction (*à Gaga* corrected by *à Gago*) and a confirmation (*Gago pardon* 'Gago sorry'). Repetitions can also be used, with an appropriate prosody, to refute the repeated element. In (39), the refutation is introduced by the particle *enfin*, which appears to be specialized for this type of usage.

- (39) *c'est la crise générale { { des | des } Français | "enfin" des Français | pas simplement des Français "hein" | { des | de } l'humanité | ^et de la lecture } } //*
 [Rhap-D0004, Rhapsodie]
 'it's a general crisis { { of | of } Frenchmen | "well" Frenchmen | not only Frenchmen "you know" | { of the | of } humanity | ^and of literacy } } //'

5. Conclusion

The proposed general annotation scheme of lists subsumes other analyses of coordination, and an annotation in this format can be exported, by simple projections, into any other treebank format. The inverse does not hold, and it commonly requires additional linguistic analyses to obtain the Rhapsodie scheme from other annotations. For disfluencies and reformulations, many treebanks simply skip the first conjunct and do not analyze these phenomena as lists. Disfluencies are generally studied as a specific problem of spontaneous speech and receive a specific

annotation schema in spoken corpora (Shriberg 1994). We believe that it is important to include disfluencies in the higher-level annotation schemes of spoken corpora, for theoretical as well as for practical reasons: Encoding disfluencies reveals basic aspects of utterance planning and constitutes essential training data for future parsers of spontaneous speech. In the Spoken Dutch Corpus, for example, only repairs are taken into account when constructing the syntactic structures. When complete constituents are repeated, they will all be constructed up to that level, but only the last one will be part of the structure assigned to the utterance as a whole (Schuurman et al. 2004). In our annotation, the whole text is analyzed and disfluencies are just taken as a special case of lists. Extending the scope of syntax in a controlled manner into what up to now fell under performance errors is an essential step for bringing theoretical syntax and Natural Language Processing back together as today's parsers tackle the last frontier of spoken language while no syntactic theory describes what is actually observed.¹³

As for coordination, the Rhapsodie scheme contains all the relations from both the symmetrical and the asymmetrical analyses of coordination and the fine-grained classification of the paradigmatic relations subsumes most other categorizations of coordination (see Gerdes & Kahane 2015 for an extension of our analysis of list to non-constituent coordination). Note however, that the resulting analysis is a DAG because it requires multiple governors. While, on the one hand, it is true that it is more onerous to obtain the general scheme, on the other hand we believe that treebank annotations that attempt to anticipate the parts of the paradigmatic analysis to be used later create in fact only partial analyses of their data.

The micro-syntactic relations that make up the Rhapsodie scheme are of three types: government links, paradigmatic links, and junction links. It is useful to conceive a paradigmatic link as orthogonal to government links; it relates the choices of filling a government link. As government relations are already 2-dimensional (a tree structure is 2-dimensional), we need to imagine the paradigmatic links in the 3rd dimension, as in Figure 17.

13. The usual strategy in the automatic parsing of spoken data is to preannotate disfluencies, by eliminating repetition for instance, but Honnibal and Johnson (2014) show that we can obtain better results by performing the disfluency detection jointly with the parsing strictly speaking. We go one step further by including the disfluencies in the syntactic structure.

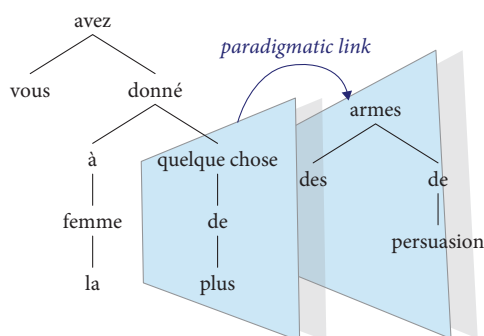


Figure 17. 3-dimensional view of the microsyntactic structure

Since the 1970s, the syntactic analysis of spoken French has used grid analysis to describe paradigmatic phenomena (cf. Section 1). On this base, the Rhapsodie project has advanced in three directions. We have provided:

1. A formalization and modeling of the notion of list.
2. An extension of the notion of list to a wider range of paradigmatic phenomena, such as partial answers and negotiations, which allows us to avoid ellipsis in the syntactic analysis of these phenomena.
3. A typology of list classes. The typology is shown in Figure 18.

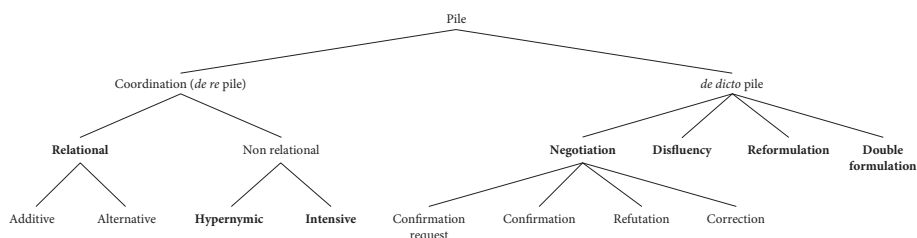


Figure 18. Hierarchy of list types (the types annotated in Rhapsodie are in bold)

Some of the classes are very rarely studied, such as for example non-relational coordinations or some types of double formulation. Only the empirical perspective of our corpus study allowed us to devise this unified view of coordination, reformulation, and other paradigmatic phenomena and to develop this hierarchy of list classes. It should be highlighted that providing a unified account for different list phenomena has important consequences for the modeling of spoken languages. Indeed, while coordination phenomena have received a great deal of attention in language modeling, to our knowledge, no formal model has provided an adequate representation for the totality of list phenomena. In particular, reformulations and disfluencies are not taken into account in the majority of models, due to the fact that they are considered as performance rather than as competence

phenomena. For coordinations we opposed two basic types which we name respectively *relational* and *non-relational* coordinations, and only the former is generally considered in the literature, whereas the latter makes up close to half of all paradigmatic occurrences in certain genres. It is interesting to note that lists carry out operations whose value is shared by grammatical elements: intensification, hypernymic derivation, determination, etc. Indeed, the constant speech flow obliges the speaker to leave numerous traces of the meaning construction and of the choices of denomination (Blanche-Benveniste et al., 1990).