

mais tu veux couper là du coup faut dire pourquoi quoi

Propositions complémentaires pour une segmentation syntaxique du français parlé

Fanni Delferrière* et Nicolas Mazziotta**

* Université de Liège, Belgique

** Université de Liège, UR Traverses, Belgique

fannidelferriere@outlook.fr ; nicolas.mazziotta@uliege.be

Cette contribution¹ se veut le prolongement de l'article intitulé « tu veux couper là faut dire pourquoi. Propositions pour une segmentation syntaxique du français parlé » (Benzitoun et al., 2010), publié dans les actes du CMLF de 2010. Notre texte, dont la visée est méthodologique, a pour objet la question du découpage des transcriptions de productions langagières complètes en unités pouvant être soumises à l'analyse syntaxique. Les enregistrements de productions inanalysées constituent l'unique donnée de l'analyse de la langue parlée, mais il est ordinairement nécessaire de les transcrire pour pouvoir les manipuler et les étudier. Si la transcription ne va jamais de soi, comme le montrent les travaux pionniers (p.ex. Blanche-Benveniste & Jeanjean, 1987), nous considérerons néanmoins ci-après que cette édition constitue notre matériau de travail.

Encore aujourd'hui, le découpage des transcriptions en unités maximales de traitement constitue un enjeu majeur de la constitution de corpus arborés. Avant même l'heure du numérique (Kahane & Mazziotta, 2022), les répertoires de textes analysés en syntaxe sont des répertoires de « phrases ». Que l'on travaille sur des corpus arborés en langue ancienne (Mazziotta, 2009, p. 77-86) ou en langue moderne (Benzitoun et al., 2010), il n'est pas exagéré de dire que ce premier découpage est crucial et conditionne tout le reste de l'analyse syntaxique². La méthode de segmentation en énoncés est ainsi une préoccupation majeure, qui est discutée régulièrement dans la littérature scientifique, comme en témoigne la bibliographie du présent travail.

D'un point de vue terminologique, on peut dire que le texte est découpé en « phrases » ou en « énoncés » sans véritablement définir ces termes³. Les linguistes ne s'entendent à ce propos que tant qu'ils ne rentrent pas dans les détails techniques⁴. Sources de malentendus, ces termes compliquent la discussion, si bien qu'il paraît plus sain de les abandonner dans le cadre d'analyses approfondies (Berrendonner, 2002; on peut mentionner comme exemples marquants: Blanche-Benveniste et al., 1984/1987; Degand & Simon, 2009; Lacheret-Dujour et al., 2019). Nous écarterons donc *phrase* et *énoncé* au profit d'*unité rectionnelle* et d'*unité illocutoire*, qui seront définis ci-dessous (→1.1).

La question de la segmentation rencontre celle de l'*autonomie* (Anscombre et al., 2016), c'est-à-dire celle de l'identification des critères qui permettent de délimiter des unités que l'on considère comme complètes. L'autonomie peut être conçue comme intonative, morphosyntaxique, sémantique ou discursive, les conceptions ne se superposant pas nécessairement. De là naissent certains dialogues de sourds entre chercheurs. Il est donc nécessaire de définir précisément ce qui justifie l'emploi d'un tel qualificatif et, partant, de se positionner d'un point de vue théorique. En l'occurrence, nous adopterons une perspective *strictement syntaxique*, qui exclut notamment tout recours aux marques intonatives. Notre approche est modulaire⁵, au sens où nous séparons conceptuellement les aspects de la langue (syntaxe,

phonématique, prosodie, etc.) pour nous concentrer sur un seul et permettre d'évaluer ultérieurement des congruences entre le module syntaxique et les autres – notamment la prosodie (Degand & Simon, 2009).

De notre point de vue, pour opérer ce découpage, il faut porter une attention particulière aux unités que la tradition nomme « marqueurs discursifs » (désormais « MD ») et les considérer comme des marques grammaticales d'organisation syntaxique. Nous proposerons ici des principes distributionnels fondés sur des jugements d'acceptabilité que nous projetons de tester à grande échelle sur corpus dans des études à venir. Nous commencerons par rappeler les principes de segmentation que les spécialistes du français parlé (en particulier, les acteurs du projet *Rhapsodie*) ont progressivement constitués, ce qui nous permettra également d'en mettre en évidence les limites (→1). Dans un second temps, nous montrerons comment l'analyse distributionnelle des MD permet selon nous de dépasser en partie ces dernières (→2). Nous nous servirons pour ce faire des premiers matériaux exploitables du projet interuniversitaire *Les Vaux* (désormais « LV », matériaux relevés en 2021-2022 ; Glikman & Fauth, 2022). La section →3 synthétise notre proposition.

1 Principes de segmentation acquis

Cette section résume les travaux qui ont été effectués à propos de la segmentation des textes en français parlé en unités maximales de traitement syntaxique dans la continuité des travaux de Claire Blanche-Benveniste, sur lesquels notre cadre théorique est fondé. Sous →1.1, nous rappelons que la distinction entre deux niveaux de syntaxe (micro- et macrosyntaxe) a mené à distinguer deux types d'unités maximales : l'unité rectionnelle et l'unité illocutoire. Sous →1.2, nous présentons brièvement une partie des difficultés que cette distinction ne permet pas de lever aisément.

1.1 Unité rectionnelle et unité illocutoire

Dans le sillage des travaux de Claire Blanche-Benveniste (1990) et d'Alain Berrendonner (2002), les linguistes travaillant sur l'oral ont déjà fourni un effort conséquent pour établir des procédures de découpage en unités de traitement syntaxique. Les continuateurs de l'école d'Aix (Blanche-Benveniste) exploitent ainsi une procédure ascendante (*bottom-up*) pour justifier un regroupement de deux constructions en une seule unité plus large dans des cas comme :

- (1) il suffit que ma sœur passe un disque ou quelqu'un ben moi je danse (Corpaix apud Benzitoun & Sabio, 2010, p. 4)
- (2) je me levais le matin j'étais avec des clients (CRFP apud Benzitoun & Sabio, 2010, p. 5)

La démarche de l'école d'Aix en matière de segmentation du texte est présentée de manière cumulative dans de nombreux travaux à partir de *Pronom et syntaxe* (Blanche-Benveniste et al., 1984/1987). À la base du raisonnement se trouve ce qui sera appelé plus tard l'*unité rectionnelle* (désormais « UR ») par les membres du projet *Rhapsodie* (Benzitoun et al., 2010; Lacheret-Dujour et al., 2019). Blanche-Benveniste et al. (1984/1987, p. 36) posent que le *verbe* doit être considéré comme l'unité première en syntaxe. Envisagé de manière dépendancielle, le verbe agrège autour de lui les éléments qu'il régit et constitue la racine d'une UR – le même raisonnement a été appliqué par la suite à des UR dont la racine n'est pas un verbe. La perspective, ascendante, mène à distinguer ces rapports microsyntaxiques des rapports macrosyntaxiques, qui ne sont pas assimilables à des relations rectionnelles. On peut ainsi

découper (2) en deux UR, comme indiqué par la marque < | > dans (3), au motif que les verbes *levais* et *étais* sont chacun accompagnés de leurs propres éléments régis qu'ils ne partagent pas.

- (3) je me levais le matin | j'étais avec des clients

Benzitoun et Sabio (2010) ont montré que ces deux UR devaient à leur tour être regroupées, même en dehors de tout rapport microsyntaxique. Ils ont établi une critériologie syntaxique, sémantique et illocutoire qui évalue systématiquement les propriétés syntaxiques internes et externes des UR pour démontrer s'il existe des contraintes qui justifient leur regroupement⁶. Comme on le voit dans (4), les transformations interrogatives sont acceptables, mais poser une interrogation totale à propos du premier segment uniquement modifie le rapport entre les deux segments. Par ailleurs, c'est le second segment qui est susceptible d'accueillir une modification adverbale circonstancielle (*ibid.* : 20-22) – la visée interrogative est marquée par un point d'interrogation pour faciliter la lecture.

- (4) a. *est-ce* que je me levais le matin j'étais avec des clients ?
b. je me levais le matin *est-ce que* j'étais avec des clients ?
c. je me levais le matin j'étais *forcément/immédiatement/toujours* avec des clients

La première de ces observations met en évidence le fait que l'assemblage des deux UR constitue une unique *unité illocutoire* (désormais « UI »), c'est-à-dire un segment qui réalise un acte de discours selon une certaine visée illocutoire⁷ (Cresti, 1999), qui peut être assertive, interrogative, injonctive ou exclamative. La logique de l'identification des UI implique que toutes les unités segmentales font partie d'une UI.

L'article de *Rhapsodie « tu veux couper là faut dire pourquoi. Propositions pour une segmentation syntaxique du français parlé »* (Benzitoun et al., 2010) a bien décrit les bases de l'articulation entre UR et UI. L'équipe propose ainsi de repérer les limites des UI en effectuant un test d'enchâssement qui consiste à la faire précéder de *je te dis que, je te demande que, je t'ordonne que* (auxquels on peut ajouter *je m'exclame que*, cf. Kahane & Mazziotta, 2015), et, plus tard, un test qui consiste à citer l'UI au discours direct (Pietrandrea & Kahane, 2019, p. 100-101), probablement parce que ce dernier n'altère pas la visée illocutoire du discours cité. Les constructions (5) sont donc toutes deux acceptables.

- (5) a. *je dis que* je me levais le matin j'étais avec des clients
b. *je dis* je me levais le matin j'étais avec des clients

L'analyse des relations qui unissent les UR à l'intérieur des UI mène à distinguer plusieurs *composantes illocutoires* (désormais CI) : le *noyau*, qui porte cette visée illocutoire, et des CI périphériques au noyau appelées *adnoyaux*. Le statut particulier du noyau est démontré par le fait que c'est le noyau lui-même qui subit toutes les altérations illocutoires de l'UI : outre le test d'enchâssement, on peut ajouter que le noyau est macrosyntaxiquement autonome, au sens où il peut réaliser seul l'acte illocutoire qu'il exprime. Le noyau peut être modifié par un adverbe modalisateur (p.ex. *honnêtement*) et peut être remplacé par d'autres noyaux de même contenu locutoire, mais de visée illocutoire différente (Pietrandrea & Kahane, 2019, p. 104) (6). Enfin, si l'on tente de coordonner les deux UR et de les subordonner, le sens est différent du sens obtenu sans coordination, ce qui est un indice de différence syntaxique (Benzitoun et al., 2010, p. 2080) (7) : avec coordination explicite, les deux subordonnées dépendent du verbe *ai dit* au même niveau de hiérarchie syntaxique.

- (6) a. j'étais avec des clients

- b. *honnêtement* j'étais avec des clients
- c. je me levais *est-ce que* j'étais *pas* avec des clients [visée exclamative]
- (7) a. *je lui ai dit que* je me levais *et que* j'étais avec des clients
- b. *je lui ai dit que* je me levais j'étais avec des clients

1.2 Limites

Pour rigoureux et efficaces que soient les critères que nous venons d'examiner, un certain nombre de cas résistent aux procédures mises au jour par les travaux précédents. Singulièrement, il s'agit de cas faisant intervenir ce qu'on peut intuitivement identifier comme des MD (qui seront définis plus précisément sous →2). Ainsi, dans les deux exemples qui suivent⁸, tirés des données orales du projet en cours *Les Vocaux* (Glikman & Fauth, 2022), le statut de *bref* et de *du coup* pose question par rapport à leur rattachement aux unités syntaxiques qui les entourent.

- (8) ça devient trop compliqué ... mais *bref* je voulais savoir si ... hum ... tu trouvais ça ... chouette (LV, 05_04)
- (9) on n'a pas forcément douze mille draps ... *du coup* si vous pouvez ramener ça ... pour votre lit quoi ... euh ... ce serait parfait (LV, 07_08)

Doit-on poser la frontière entre les UI à droite ou à gauche de ces MD ? Ce genre de problème, relevé de longue date concernant toutes sortes de constructions, est généralement levé en ayant recours à la prosodie depuis les travaux du GARS en passant par ceux de l'équipe de *Rhapsodie*⁹ (Bilger et al., 1997, sect. 2 et 3; Lacheret-Dujour et al., 2011). Ci-dessous, conformément à notre visée modulaire, nous nous intéresserons surtout aux critères syntaxiques qui expliquent pourquoi ces constructions sont ambiguës.

Par ailleurs, la méthode d'analyse ne permet pas vraiment de traiter le cas des MD « autonomes », ou « mots-phrases », sans avoir recours à la prosodie (Lacheret-Dujour et al., 2011, sect. 4.3.3). Ces derniers, que Kahane et Mazziotta (2015) nomment *locutifs* (classe qui contient notamment les interjections mais ne comprend pas *oui* et *non*), ont la propriété intrinsèque de servir de noyaux car ils sont pourvus d'une visée illocutoire figée en langue. Ainsi, alors que la visée illocutoire d'une UI fondée sur un verbe fini est libre, *zut* est toujours expressif, *chut* est toujours injonctif, *hein* n'est jamais assertif. Le caractère figé de cette visée se manifeste par l'impossibilité d'utiliser une visée librement choisie ou de retirer la visée par subordination, comme cela a pu être observé concernant les énoncés averbaux de manière générale (voir notamment Lefeuvre, 2016, p. 76). Une partie des tests mentionnés *supra* sont donc inopérants, comme le montrent les exemples agrammaticaux qui suivent :

- (10) a. *je t'ordonne que chut
- c. ?je t'ordonne chut
- b. *est-ce que chut ?

L'approche (tesnièreenne) de Kahane et Mazziotta les mène à affirmer que toute UI est un locutif et que la langue possède des outils capables de transférer les UR en locutifs en les pourvoyant d'une visée illocutoire (p.ex. : *est-ce que* + UR verbale → visée interrogative ; *Comme* + UR verbale → visée exclamative ; etc.)

Les MD posent également problème quand il est question de déterminer où ces éléments se placent dans la hiérarchie syntaxique. Le guide d'annotation d'Orféo, dont l'élaboration repose sur l'expérience de *Rhapsodie*, propose d'ailleurs des règles assez peu satisfaisantes d'un point de vue théorique : « Un élément *dm* [= MD] est rattaché à l'élément qui le précède directement, ou à la racine s'il est en position initiale » (Kahane et al., 2021, sect. 5.2). Dans leurs guides d'annotation les plus récents, des auteurs de la même communauté de recherche affirment encore, malgré l'évolution de leur modèle : « These markers are not clearly linked to the structure of the sentence, except in an expressive way » (guide d'annotation de *SUD*¹⁰).

D'autre part, indépendamment des questions qui impliquent le discours cité, que nous n'aborderons pas ici, il arrive que plusieurs UI soient hiérarchisées : en d'autres termes, une unité illocutoire principale peut en contenir une autre, pourvue de sa propre force illocutoire. Par exemple, dans (11), *voilà* est un locutif, une UI contenue dans une UI plus large, sans constituer du discours direct.

- (11) c'est quelqu'un qui qui squatte ... *voilà* ... qui qui est là (LV, 06_07_02)

Kahane et Pietrandrea ont montré que ces UI enchâssées portent une force illocutoire et peuvent constituer un tour de parole autonome, mais ne peuvent se substituer à l'ensemble du noyau ni être modifiées comme pourrait l'être un noyau (2009; et, en partic., 2019, p. 112-114). Or, comme les locutifs sont porteurs d'une visée illocutoire figée, ils peuvent être inclus dans une UI en tant que *noyaux associés*¹¹. Ainsi, *voilà* est intégré à (11), mais il pourrait constituer un tour de parole (12). En outre, *voilà* a une visée nécessairement assertive qui ne serait pas modifiée en cas d'altération de la visée illocutoire de l'UI principale : dans (13), la visée interrogative obtenue à l'aide de *est-ce que* n'affecte pas *voilà*.

- (12) L1 : c'est quelqu'un qui qui squatte

L2 : *voilà*

- (13) est-ce que c'est quelqu'un qui qui squatte ... *voilà* ... qui qui est là ?

Toutefois, d'une part il est extrêmement difficile de déterminer si un locutif au début ou à la suite d'une autre UI (8) doit être considéré comme une UI indépendante ou s'il constitue une unité illocutoire associée, d'autre part ces unités pourraient avoir une distribution contrainte (début, milieu, fin d'UI), ce qui reste à étudier.

2 Distribution des marqueurs discursifs et segmentation

La littérature sur les MD est abondante et hétérogène, tant d'un point de vue général que concernant les lexies particulières. Il ne peut être question ici d'en faire la revue. Notre définition de base des MD est celle de Dostie et Pusch (2007, p. 4-5), que nous reformulons : un MD est un emploi morphologiquement figé d'un signe linguistique, qui n'intervient pas dans le contenu propositionnel, dont l'absence n'entraîne pas d'agrammaticalité et qui ne marque pas de relation microsyntaxique. Du fait de la polycatégorialité des unités qu'elle recouvre, l'étiquette *MD* gagne à être considérée comme la qualification d'un type d'emploi plutôt que le nom d'une classe lexicale comme *substantif* ou *adverbe* (Hansen, 1998). Dans la suite de notre exposé, *MD* signifiera donc plus précisément « unité employée comme un marqueur discursif ».

Les MD posent des difficultés à la méthode de découpage syntaxique présentée sous →1, mais il a été observé de manière récurrente dans divers cadres théoriques que certains d'entre eux sont régulièrement employés pour marquer la structure de la conversation (Auchlin, 1981), ou plus spécifiquement comme

« marqueurs de balisage » à la fin d'une étape dans l'énonciation (v. p.ex. Dostie, 2004, p. 48). On a remarqué leur affinité avec les positions à l'initiale ou à la finale des structures prédicatives (Lefeuve, 2011, 2020 ; Lefeuve et al. 2021) ou des prises de parole, parfois en fonction de valeurs (inter)subjectives spécifiques (Degand, 2014). Ces études exploitent largement les dimensions sémantiques et conversationnelles des MD. Nous pensons quant à nous qu'une analyse fine exclusivement fondée sur l'examen de la distribution de ces unités problématiques permet de résoudre la plupart des problèmes de délimitation des UI. Le rôle des MD est généralement considéré comme lié au niveau macrosyntaxique et nous proposons de les classer en fonction de leur position par rapport aux unités spécifiques à ce niveau. Cette méthode permettra de distinguer les MD qui apparaissent aux limites des UI de ceux dont le rattachement à une UI demeure problématique. Ces derniers sont ceux qui remettent en question l'existence d'UI clairement délimitées d'un point de vue strictement syntaxique. Notre analyse distributionnelle reposera exclusivement sur l'identification de la position des MD dans la séquence linéaire du flux de la parole et nous ne proposerons pas ici d'envisager le contenu pragmatique associé à telle ou telle position (voir Kragh, 2021 pour une approche de ce type).

Nous proposons de différencier les MD tels que *en fait, quoi* (14), *en fait, donc, voilà* (15) ou *du coup* (16) à l'aide de critères positionnels.

- (14) moi euh le temps breton euh ... toute la durée des vacances ... euh *en fait* euh ça me dit bien *quoi* quand il fait pas trop chaud ... éviter la canicule ... ça m'arrange (LV, 01_01)
- (15) parce qu'*en fait* euh ... c'est la semaine de ... euh ... c'est la semaine de prérentrée euh de mon cursus intégré ... *donc voilà* ... j'ai cours à partir de la semaine prochaine (LV, 01_01)
- (16) j'ai décidé de lui faire une longue balade *du coup* on est allé euh quelque part (LV, 04_extrait)

Nous n'aborderons ici que les cas qui concernent le découpage en UI et nous ne nous attarderons pas sur des cas comme *en fait* dans (17) qui est clairement interne à une UI.

- (17) je sais pas *en fait* encore si je suis prête à retourner dans un groupe ou pas ... (LV, 07-10)

Dans le but d'élaborer une typologie qui pourra ultérieurement être employée sur corpus pour segmenter le texte, nous commencerons par présenter les MD que l'on trouve aux limites des UI, que nous nommons *borneurs*. Ces derniers sont de deux types : *introduceurs* et *terminatifs* (→2.1). Nous verrons ensuite que certains MD ont un rattachement ambigu car ils ne peuvent être considérés comme étant inclus de manière exclusive dans l'UI qui les précède ou dans l'UI qui les suit (→2.2). Nous examinerons ensuite les noyaux associés (→2.3), avant de synthétiser nos propositions (→3).

2.1 Définition et identification des borneurs introduceurs et terminatifs

Pour notre classement syntaxique, nous nous appuyons sur les travaux du projet *Rhapsodie* (Pietrandrea & Kahane, 2019) qui distinguent les *introduceurs* (*openers*) des autres constituants de l'unité illocutoire, en particulier des pré-noyaux. En effet, ces derniers se situent par définition avant le noyau (*moi* dans (18)) alors que les introduceurs sont directement placés en début d'UI (comme *mais* dans (18) ou d'autres éléments comme *et*) (*Ibid.* : 111-112). Benzitoun et al. les définissent déjà comme des « conjonctions qui occupent la première position d'une UI et qui dans leur fonctionnement s'apparentent davantage à des marqueurs de pile [≈ articulateurs entre éléments de même rang dépendanciel] qu'à des pré-noyaux » (2010, p. 2084). Les introduceurs répondent à la définition de base des MD, puisqu'ils n'entrent ni dans le contenu propositionnel de l'énoncé, ni dans le noyau qui porte la force illocutoire : si

l'on transforme (18) en UI interrogative (19), il devient évident que *mais* ne fait pas partie du noyau. Toutefois, sa position est stable : lorsque l'on applique les tests évoqués sous →1.1, *mais* se place toujours en début d'UI. Dans les exemples qui suivent, les introducteurs sur lesquels nous nous focalisons sont précédés de la marque « ^ » (cf. Pietrandrea & Kahane, 2019, p. 111).

(18) ^*mais* moi j'arrête de boire hein (LV, 06_04)

(19) ^*mais* est-ce que moi j'arrête de boire hein ?

Souvent qualifié de « connecteur », terme que nous évitons en raison de sa connotation argumentative, le lexème *mais* a été identifié de longue date comme un marqueur de début de tour de parole (ou de « réplique » dans le théâtre ; Bruxelles et al., 1976/1980). La littérature cherche surtout à déterminer la valeur sémantique. Cette position initiale nous permet de le considérer comme un *introduceur prototypique*.

À partir de la brève étude de Andrews (1989), nous proposons également de qualifier certains MD de *terminatifs*, c'est-à-dire la classe des MD qui se trouvent en fin d'UI et indiquent à l'interlocuteur qu'elle est terminée (Andrews, 1989, p. 194). Il paraît raisonnable d'augmenter la typologie des CI macrosyntaxiques en ajoutant le pendant final des introducteurs : les terminatifs concluent les UI comme les introducteurs les ouvrent. Dans les exemples qui suivent, les terminatifs dont il est question dans l'exposé sont suivis de la marque « \$ »¹². Il a été montré que le lexème *quoi* est majoritairement employé sans visée illocutoire interrogative en position de fin de rhème (Lefevre et al., 2011), ce qui le rend compatible avec un rôle de terminatif. Nous considérons qu'il s'agit d'un *terminatif prototypique*.

Sans recourir à cette terminologie et dans une optique convoquant l'analyse pragmatique, Lefevre a montré que ^*bon* et *quoi*\$ étaient utilisés de manière privilégiée pour délimiter des unités syntaxiques autonomes (2011). Le critère distributionnel prévaut pour définir ces deux paradigmes de MD que nous qualifions de *borneurs*. Cela nécessite de déterminer à quelle UI le MD se rattache. À partir de là, l'enjeu est d'objectiver ce classement sur la base de tests qui sont fondés sur le comportement de *quoi* (20) et de *mais* (21)¹³.

(20) moi euh le temps breton euh ... toute la durée des vacances ... euh en fait euh ça me dit bien *quoi* quand il fait pas trop chaud ... éviter la canicule ... ça m'arrange (LV, 01_01)

(21) je vous l'envverrai par la poste *mais* ça tu me rediras (LV, 02_02)

Pour résoudre la question, nous proposons de dégager les tests permettant de déterminer à quelle UI se rattachent les MD au départ du postulat suivant : *quoi* se rattache à « ça me dit bien » et non à « quand il fait pas trop chaud » (20) et *mais* lie certes « je vous l'envverrai par la poste » et « ça tu me rediras » (21), mais se rattache davantage au second. Le MD *mais* est un introducteur prototypique (→2.1.1), alors que *quoi* est un terminatif prototypique (→2.1.2)¹⁴. Ces borneurs indiquent les limites des UI et peuvent être exploités pour élaborer une procédure de segmentation du discours (→2.1.3).

Comme nous allons le voir, notre postulat intuitif est validé par des critères d'identification des deux types de borneurs, qui se déterminent mutuellement de manière interdépendante. Il n'est en effet pas possible de justifier qu'un terme est introducteur sans recourir à l'emploi d'un terminatif et inversement. La cohérence empirique des observations justifie la démarche. Dernière précaution : ces tests ne prennent pas en considération la possibilité universelle d'avorter une UI en cours d'énonciation.

2.1.1 Introduceurs

Pour identifier les introduceurs, nous proposons les tests suivants :

1. On peut ajouter un terminatif (*quoi*) devant le MD, marquant ainsi la fin de l'UI qui précède.
(22) je vous l'enverrai par la poste *quoi* \$ // *^mais* ça tu me rediras
2. On ne peut ajouter un terminatif (*quoi*) derrière le MD sans affecter ce terminatif d'une visée illocutoire propre qu'il n'a pas en tant qu'introduceur.
(23) *je vous l'enverrai par la poste // *^mais quoi* \$ // ça tu me rediras
(24) je vous l'enverrai par la poste // *^mais quoi* \$? // ça tu me rediras
3. On ne peut pas utiliser le MD en finale absolue d'UI.
(25) *je vous l'enverrai par la poste *^mais*
4. On ne peut utiliser le MD comme un tour de parole sans lui associer une visée illocutoire propre qu'il n'a pas en tant qu'introduceur.
(26) L1 : je vous l'enverrai par la poste
L2 : **^mais*

Ces introduceurs peuvent être accompagnés d'autres MD selon des principes combinatoires plus ou moins figés (Dargnat & Jayez, 2021; Dostie, 2013).

- (27) *^mais en fait* je ... je préfère (LV, 01_02)
- (28) *^et euh et du coup euh ben* déjà ils ont recom- ... ils ont refait des pâtes ... pour moi genre mon repas (LV, 03_05)

Ils sont en outre compatibles avec d'autres morphèmes appartenant au continuum conjonction-MD, tel que *parce que*, dont le niveau d'intégration syntaxique variable a contribué à illustrer la notion d'*insubordination* en français (v. notamment Debaisieux, 1994, 2013; Deulofeu et al., 2019). (29) prouve ainsi la non-intégration microsyntaxique de l'UR débutant par *mais parce que*.

- (29) j'envoie tout valser et tout *^mais* parce qu'en fait j'ai tellement du mal à exprimer en fait euh s- s- tu vois mes angoisses (LV, 106_17)

2.1.2 Terminatifs

Pour identifier les terminatifs, nous proposons les tests suivants :

1. On ne peut ajouter un introduceur (*mais*) devant le MD sans lui associer une visée illocutoire propre qu'il n'a pas en tant que terminatif.
(30) *moi euh le temps breton toute la durée des vacances en fait ça me dit bien // *^mais quoi* \$ // quand il fait pas trop chaud éviter la canicule ça m'arrange
2. On peut ajouter un introduceur (*mais*) derrière le MD.
(31) moi euh le temps breton toute la durée des vacances en fait ça me dit bien *quoi* \$ // *^mais* quand il fait pas trop chaud éviter la canicule ça m'arrange

3. On ne peut pas utiliser le MD à l'initiale absolue d'une UI.

(32) **quoi\$* moi euh le temps breton toute la durée des vacances en fait ça me dit bien quand il fait pas trop chaud éviter la canicule ça m'arrange (LV, 01_01)

4. On ne peut utiliser le MD comme un tour de parole sans lui associer une vidée illocutoire propre qu'il n'a pas en tant que terminatif.

(33) L1 : moi le temps breton toute la durée des vacances en fait ça me dit bien

L2 : **quoi\$*

2.1.3 Exploitation

En prenant appui sur l'identification des borneurs, il est possible de segmenter l'extrait (34) en plaçant une limite à chaque endroit compatible simultanément avec l'enchaînement d'un terminatif et d'un introducteur – ces derniers étant le plus souvent déjà présents.

(34) ... tu vas voir // je t' ai envoyé un mail ... // ^et du coup il y a mon CV et ma lettre de motiv' // ^mais bon ça faut juste ... jeter un coup d' oeil // ^mais a priori euh il y a pas de faute ... (LV, 02_08)

On remarque que certaines lexies pouvant être employées comme terminatifs (35) (Lefeuve et al., 2011) peuvent apparaître dans d'autres positions. Cette compatibilité n'invalide cependant pas notre démarche car ces positions refusent que ce MD soit suivi d'un introducteur (36). Ce blocage démontre que la lexie en question n'est pas employée comme borneur d'UI.

(35) mais a priori *quoi* euh il y a pas de faute

(36) **mais a priori quoi // mais* euh il y a pas de faute

Il arrive toutefois que certains MD rendent plus difficile l'identification des limites des UI car ils peuvent à la fois être précédés d'un terminatif ou suivis d'un introducteur. C'est ce qui va nous occuper dans les sous-sections qui suivent.

2.2 Marqueurs discursifs au rattachement ambigu

Certains MD sont utilisés tant en position initiale absolue qu'en position finale absolue d'UI. Ils ne peuvent être considérés ni exclusivement comme des introducteurs, ni exclusivement comme des terminatifs. Ils posent un problème de segmentation spécifique, que nous allons présenter sous →2.2.1, avant d'examiner plus en profondeur le comportement de ces MD, que nous qualifierons selon les cas de *pivots* ou de *locutifs* (→2.2.2).

2.2.1 Description du problème

Les cas comme ceux de *bref* (37) ou de *du coup* (38) mettent en question l'efficacité des tests basés sur les borneurs que nous venons d'appliquer. *Bref* et *du coup* sont tous deux précédés et suivis de segments qui pourraient être considérés comme des UI. En nous référant uniquement aux tests d'insertion de borneurs, ces unités ont un comportement ambigu – indépendamment du fait qu'on peut intuitivement les rattacher davantage à gauche ou à droite à la lecture de l'exemple.

(37) la tentation est trop forte *bref* j'espère que ça va s'arranger (LV, 01_03)

(38) j''ai décidé de lui faire une longue balade *du coup* on est allé euh quelque part [= (16)]

Dans ces deux exemples, les MD soulignés pourraient tout aussi bien faire partie de l'UI qui les précède que de celle qui les suit. Que faire des unités qui occupent une position macrosyntaxique confuse, qui rend difficile l'identification de l'UI à laquelle ils se rattachent ?

Une partie des tests peuvent être appliqués : *mais* peut être ajouté après le MD, ce qui correspond plutôt au comportement des terminatifs, comme on le voit dans les manipulations (39) et (40).

(39) j'ai décidé de lui faire une longue balade *du coup* // ^*mais* on est allé euh quelque part

(40) la tentation est trop forte *bref* // ^*mais* j'espère que ça va s'arranger

Cependant, un terminatif comme *quoi* peut être ajouté avant le MD, ce qui correspond plutôt au comportement des introducteurs.

(41) j'ai décidé de lui faire une longue balade *quoi* \$ // *du coup* on est allé euh quelque part

(42) la tentation est trop forte *quoi* \$ // *bref* j'espère que ça va s'arranger

En revanche, alors que, de manière générale, les borneurs ne sont pas mobiles (les terminatifs se trouvent en début d'UI, les introducteurs à la fin), ces MD peuvent être déplacés :

(43) a. j'ai décidé de lui faire une longue balade on est allé euh quelque part *du coup*

b. *du coup* j'ai décidé de lui faire une longue balade on est allé euh quelque part

(44) a. *bref* la tentation est trop forte j'espère que ça va s'arranger

b. la tentation est trop forte j'espère que ça va s'arranger *bref*

Rappelons que toutes les unités segmentales font nécessairement partie d'une UI selon notre cadre théorique (→1.1). Or, selon notre méthodologie, il n'est pas possible de déterminer si ces MD appartiennent clairement à une UI ou à l'autre : ils seront considérés comme inclus dans l'UI qui suit ou dans celle qui précède en fonction de l'élément rajouté (un terminatif ou un introducteur). Les limites de l'UI pourront être déplacées en conséquence. Le MD ne peut être considéré comme faisant partie d'une seule des deux UI auxquelles il pourrait appartenir. Il n'y a selon nous aucune solution proprement syntaxique à ce problème. Il est cependant vrai que l'acte d'énonciation considéré dans son ensemble peut fournir les moyens de déterminer si un MD doit être rattaché à l'UI qui le précède ou à l'UI qui le suit. À l'écoute, l'intonation de (16) indique que *du coup* est associé à l'UI qui suit, mais cette observation ne peut venir que dans un second temps, au moment de l'évaluation des congruences entre la syntaxe et les autres modules de description de la langue.

2.2.2 Pivots et locutifs

Une nuance supplémentaire doit être apportée pour différencier les cas similaires à *du coup* de ceux semblables à *bref*. En effet, bien que placés tous deux aux bornes des UI, ils ne se comportent pas de la même façon et n'ont donc pas la même incidence sur le découpage macrosyntaxique.

Ainsi, *du coup* ne peut constituer un tour de parole autonome avec une visée assertive, et, en outre, il ne peut être à la fois précédé d'un terminatif et suivi d'un introducteur ou inversement (45). Par contre, *bref* peut clairement être encadré d'un terminatif à sa gauche et d'un introducteur à sa droite (46).

- (45) a. *j'ai décidé de lui faire une longue balade // ^mais du coup
 b. *j'ai décidé de lui faire une longue balade quoi\$ // du coup // ^mais on est allé euh quelque part
 c. *j'ai décidé de lui faire une longue balade // ^mais du coup quoi\$ // on est allé euh quelque part
- (46) a. la tentation est trop forte // ^mais bref
 b. la tentation est trop forte quoi\$ // bref // ^mais j'espère que ça va s'arranger
 c. la tentation est trop forte // ^mais bref quoi\$ // j'espère que ça va s'arranger

Ces tests nous permettent de différencier les *locutifs*, les lexies employées comme des MD qui possèdent une autonomie en tant qu'UI complètes (→1.2), des lexies employées comme des MD non autonomes, que nous nommerons *pivots*, puisqu'ils se trouvent à la frontière de deux unités macrosyntaxiques et ne comportent aucune visée illocutoire intrinsèque – le terme n'a donc ici ni un sens lexicométrique, ni un sens conversationnel. Nous proposons de distinguer comme suit la notation de ces deux cas de figure :

- (47) j'ai décidé de lui faire une longue balade / du coup / on est allé euh quelque part
 (48) la tentation est trop forte // bref // j'espère que ça va s'arranger

Parmi les pivots, on trouve notamment : *alors, disons, du coup, en fait, genre, on va dire*.

Il faut remarquer qu'un certain nombre de lexies utilisables comme MD peuvent constituer des tours de parole autonomes à l'instar des locutifs comme *bref* (37). Ils se voient alors associés à une visée illocutoire spécifique dont il est difficile de déterminer si elle est présente en cas d'emploi non isolé. *Hein ! Bah ! Enfin !* illustrent ce genre de phénomène, qui est l'objet de la sous-section suivante.

2.3 Noyaux associés

Cette distinction entre, d'une part, les MD borneurs et pivots et, d'autre part, les locutifs nous amène à une dernière nuance de classement qui concerne les noyaux associés, qui sont nécessairement des locutifs selon notre définition (→1.2). Les formes propositionnelles comme *tu vois* dans (49), qui est introduit par *enfin* et est intégré à une autre UI, pourraient être considérées comme des UI autonomes (→1.1) dans un autre contexte. Il en va de même pour *bah* (50) ou *hein* dans (51).

- (49) moi j'ai l'impression enfin *tu vois* que ce genre ... d'acte manqué .. en fait c'est pas mal ... (LV, 01_01 ; la conjonction qui suit *tu vois* introduit la conjonctive qui dépend du substantif *impression*)
 (50) euh et puis s'il le faut ... dès que je l' ai *bah* je vous l' enverrai par la poste (LV, 02_02)
 (51) et je viens de voir ça là maintenant ... mais moi j'arrête de boire *hein* (LV, 06_04)

Certaines unités sont complètement bornables, quelle que soit leur position (voir notamment Kragh, 2021). C'est le cas de *tu vois*, qui peut être encadrée de ^mais et quoi\$. Il en va de même pour *bon, bref, écoute, enfin, je t'explique, putain, tu sais, voilà*. La distribution précise de ces MD autonomisables doit être évaluée au cas par cas pour affiner leur typologie distributionnelle. Par exemple, *écoute* n'est pratiquement jamais final (Mazziotta & Glikman, 2023, p. 32). On a coutume de dire que ces unités complètement bornables peuvent constituer des tours de parole et sont donc associées à une force illocutoire (→1.2). Cependant, il nous semble que la visée illocutoire de ces unités n'est jamais clairement identifiable lorsqu'elles sont utilisées de manière associée. On peut noter ces noyaux associés à l'aide de parenthèses, comme dans (52), où nous ajoutons les deux borneurs.

- (52) j'ai l'impression enfin (^mais tu vois quoi\$) que ce genre ... d'acte manqué .. en fait c'est pas mal [= (49)]

Les questions du discours rapporté et des « greffes » (UI intégrant une structure rectionnelle sans mécanisme classique de subordination, v. Béguelin & Corminboeuf, 2020) mériteraient une étude spécifique exploitant les mêmes tests. Il semblerait notamment que plus le degré de figement des greffes est élevé, moins les borneurs peuvent être mobilisés – cf. le figement de *je sais pas quoi* dans (53).

- (53) a. si tu veux euh passer à la maison pour prendre tes toiles ... si tu vas ... au marché de bourg la reine ou *je sais pas quoi* (LV, 86_14)
b. *si tu veux euh passer à la maison pour prendre tes toiles ... si tu vas ... au marché de bourg la reine ou ^*mais je sais pas quoi quoi\$*
- (54) a. il y a des sons ils sont grave bien (LV, 367_03)
b. il y a des sons ^*mais ils sont grave bien quoi\$*

Nous n'approfondirons pas cette question ici. Une application systématique des tests proposés permettrait sans doute de compléter la typologie des greffes.

Bah, ben et *hein* – analysés également comme des locutifs par Kahane et Mazziotta (2015) en raison de leur capacité à constituer un tour de parole – se comportent selon les cas soit comme des noyaux associés, soit comme des borneurs. Ainsi, les locutifs introducteurs (*bah, ben*) ne sont pas complètement bornables : ils ne peuvent pas être suivis d'un terminatif. Inversement, un locutif terminatif comme *hein* ne peut être précédé d'un introducteur. Dans la transcription, on peut rendre compte du caractère de noyau associé de ces borneurs en ajoutant le symbole correspondant à la borne devant ou derrière les parenthèses, comme dans (55).

- (55) ^*mais moi j'arrête de boire* (hein)\$

3 Synthèse et conclusion

Nous avons montré que les MD pouvaient être classés selon leur comportement distributionnel et que ce classement pouvait être exploité pour la segmentation en unités de traitement macrosyntaxique (UI).

Les tests destinés au classement des MD en fonction de leur position macrosyntaxique et pouvant être mobilisés pour segmenter le texte en UI sont résumés dans le Tab.1 (p. suivante). Les catégories de MD sont distinguées en fonction de la possibilité (« + ») ou non (« – ») qu'ont ces MD d'être précédés ou suivis d'un borneur terminatif ou d'un introducteur. Ces tests distributionnels mettent en outre en évidence les propriétés de ces classes de MD en termes d'autonomie strictement syntaxique et de mobilité. La définition des MD en fonction de leur distribution permet ainsi conjointement de les classer sur la base de critères homogènes et de résoudre une partie des cas problématiques que pose la segmentation.

Ces critères seront appliqués au corpus du projet *Les Vocaux* pour évaluer leur efficacité et proposer une annotation qui tiendra compte des ambiguïtés de segmentation. L'approche, qui se veut entièrement syntaxique, met en évidence ses propres limites : certains MD sont notamment compatibles à la fois avec la position initiale et avec la position finale dans une UI. Tout comme on ne pourra jamais proposer d'analyse non ambiguë de *La petite brise la glace* sans tenir compte du contexte, on ne pourra jamais lever les cas ambigus de rattachement des pivots (→2.2.2) ou la distinction entre locutifs employés de

manière autonome et locutifs borneurs (→2.3) sans mobiliser le contexte et l'intonation (à l'oral). Similairement, nous ne proposons pas non plus une solution définitive à la question des « segments flottants » (Béguelin & Corminboeuf, 2020), mais une autre manière de les envisager. Il nous paraît cependant essentiel de procéder en deux temps : pour que la modularisation de l'analyse soit opératoire et permette d'étudier les liens entre prosodie et syntaxe, nous préférons ne pas mêler les deux *a priori*. Cette démarche permet de mieux comprendre le fonctionnement des MD dans la construction du discours, comme marqueurs de continuité/discontinuité (Crible, 2017). Elle permettra, en comparant les distributions de ces MD dans des corpus variés, d'envisager les relations entre leur utilisation et celle d'autres marques pour articuler les UI.

Type de MD	Critères					
	Ajout				Propriété	
	terminatif devant	terminatif derrière	introduceur derrière	introduceur devant	Autonomisable	Mobile
Locutif bornable	+	+	+	+	+	+
Introduceur	+	–	/	/	–	–
Locutif introduceur	+	–	/	/	+	–
Terminatif	/	/	+	–	–	–
Locutif terminatif	/	/	+	–	+	–
Pivot	±	±	±	±	–	+

Tab. 1. – Critères de distinction des types distributionnels de MD (« + » : critère validé ; « – » critère non validé ; « ± » critère validé selon les cas ; « / » critère non pertinent) ; « x » représente le MD évalué.

Références

Corpus

LV = Glikman, Julie, Camille Fauth, Christophe Benzitoun et Nicolas Mazziotta. 2021-2022. Corpus *Les Vocaux*. <<https://www.atilf.fr/ressources/corpus-les-vocaux/>>.

Études

- Andrews, B. J. (1989). Terminating devices in spoken french. *International review of applied linguistics in language teaching*, 27(3), 193-216.
- Anscombre, J.-C., Darbord, B., Oddo, A., & García de Lucas, C. (Éds.). (2016). *La phrase autonome: Théorie et manifestations*. P.I.E. Peter Lang.
- Auchlin, A. (1981). *Mais heu, pis bon, ben alors, voilà, quoi!* Marqueurs de structuration de la conversation et complétude. *Cahiers de linguistique française*, 2, 141-159.

- Béguelin, M.-J., & Corminboeuf, G. (2020). Segmentation en unités : Le cas des “greffes” et des “segments flottants”. In M.-J. Béguelin, G. Corminboeuf, & F. Lefeuve, *Types d'unités et procédures de segmentation* (p. 99-129). Lambert-Lucas.
- Benzitoun, C. (2010). Quelle(s) unité(s) syntaxique(s) maximale(s) en français parlé ? Discussions autour de quelques problèmes rencontrés. *Travaux de linguistique*, 60(1), 109-126. Cairn.info. <https://doi.org/10.3917/tl.060.0109>
- Benzitoun, C., Dister, A., Gerdes, K., Kahane, S., Pietrandrea, P., Sabio, F., & Debaisieux, J.-M. (2010). *tu veux couper là faut dire pourquoi*. Propositions pour une segmentation syntaxique du français parlé. 2^e Congrès Mondial de Linguistique Française, 2075-2090. <https://doi.org/10.1051/cmlf/2010201>
- Benzitoun, C., & Sabio, F. (2010). Où finit la phrase ? Où commence le texte ? L'exemple des regroupements de constructions verbales. *Discours*, 7, 3-25. <https://doi.org/10.4000/discours.7966>
- Berrendonner, A. (2002). Les deux syntaxes. *Verbum*, 24(1-2), 23-35.
- Berrendonner, A. (2021). La notion de phrase. In *Encyclopédie grammaticale du français*. <http://encyclogram.fr>
- Bilger, M., Blasco, M., Cappeau, P., Pallaud, B., Sabio, F., & Savelli, M.-J. (1997). Transcription de l'oral et interprétation. Illustration de quelques difficultés. *Recherches sur le français parlé*, 14, 57-86.
- Blanche-Benveniste, C., Bilger, M., Rouget, C., & van den Eynde, K. (1990). *Le français parlé : Études grammaticales*. CNRS Éditions.
- Blanche-Benveniste, C., Deulofeu, J., Stefanini, J., & Van den Eynde, K. (1987). *Pronom et syntaxe : L'approche pronominale et son application au français*. SELAF. (Édition originale 1984)
- Blanche-Benveniste, C., & Jeanjean, C. (1987). *Le français parlé : Transcription et édition*. Didier.
- Bruxelles, S., Ducrot, O., Fouquier, É., Gouazé, J., Dos Reis Nunes, G., & Rémis, A. (1980). Mais occupe-toi d'Amélie. In O. Ducrot (Éd.), *Les mots du discours* (p. 93-130). Minuit. (Édition originale 1976)
- Cresti, E. (1999). Force illocutoire, articulation topic/comment et contour prosodique en italien parlé. *Faits de langues*, 13, 168-181.
- Crible, L. (2017). *Discourse markers and (dis)fluency across registers : A contrastive usage-based study in English and French* [Thèse de doctorat, UCLouvain]. <http://hdl.handle.net/2078.1/182932>
- Dargnat, M., & Jayez, J. (2021). Mais et associés. *Lexique*, 29, 211-228.
- Debaisieux, J.-M. (1994). *Le fonctionnement de parce que en français parlé contemporain : Description linguistique et implications didactiques* [Thèse de doctorat]. Université de Nancy II.
- Debaisieux, J.-M. (Éd.). (2013). Autour de *parce que* et de *puisque*. In *Analyses linguistiques sur corpus : Subordination et insubordination en français* (p. 185-247). Hermès Science, Lavoisier.
- Degand, L. (2014). So very fast then, Discourse markers at left and right periphery in spoken French. In K. Beeching & U. Detges (Éds.), *Discourse functions at the left and right periphery : Crosslinguistic investigations of language use and language change* (p. 151-178). Brill.
- Degand, L., & Simon, A. C. (2009). On identifying basic discourse units in speech : Theoretical and empirical issues. *Discours*, 4. <https://doi.org/10.4000/discours.5852>
- Deulofeu, H.-J., Debaisieux, J.-M., & Martin, P. (2019). Apparent insubordination as discourse patterns in French. In K. Beijering, G. Kaltenböck, & M. Sol Sansiñena (Éds.), *Insubordination Theoretical and empirical issues* (p. 349-383). De Gruyter. <https://hal.archives-ouvertes.fr/hal-03172591>
- Dostie, G. (2004). *Pragmaticalisation et marqueurs discursifs. Analyse sémantique et traitement lexicographique*. De Boeck Supérieur.
- Dostie, G. (2013). Les associations de marqueurs discursifs de la cooccurrence libre à la collocation. *Linguistik online*, 62(5), 15-45.
- Dostie, G., & Pusch, C. D. (2007). Présentation. Les marqueurs discursifs. Sens et variation: *Langue française*, n° 154(2), 3-12. <https://doi.org/10.3917/lf.154.0003>
- Gerdes, K., Guillaume, B., Kahane, S., & Perrier, G. (2021). Starting a new treebank ? Go SUD ! In N. Mazziotta & S. Mille (Éds.), *Proceedings of the Sixth International Conference on Dependency Linguistics (Depling, SyntaxFest 2021)* (p. 35-46). Association for Computational Linguistics. <https://aclanthology.org/2021.depling-1.4>
- Glikman, J., & Fauth, C. (2022). Un nouvel accès à la parole spontanée : Les vocaux. XXXIV^e Journées d'Études Sur La Parole – JEP 2022, 154-162. <https://doi.org/10.21437/JEP.2022-17>
- Grevisse, M., & Goosse, A. (2011). *Le bon usage : Grammaire française* (15^e éd.). De Boeck-Duculot. (Édition originale 1936)
- Hansen, M.-B. M. (1998). *The function of discourse particles : A study with special reference to spoken standard French*. John Benjamins.
- Hjelmslev, L. (1968). *Prolégomènes à une théorie du langage. Suivi de : La structure fondamentale du langage* (U. Canger, A. Wewer, & A. M. Léonard, Trad.). Les Éditions de Minuit. (Édition originale 1966)
- Kahane, S., Deulofeu, H.-J., Gerdes, K., Valli, A., & Nasr, A. (2021). *Guide d'annotation syntaxique Orféo (version Platinum)*. <https://shs.hal.science/halshs-01740488>

- Kahane, S., & Mazziotta, N. (2015). Quel classement syntaxique pour les « marqueurs discursifs », « mots-phrases » et autres « inserts »? Prédicatifs et locutifs. *Travaux de linguistique*, 71(2), 7-42. Cairn.info. <https://doi.org/10.3917/tl.071.0007>
- Kahane, S., & Mazziotta, N. (2022). Les corpus arborés avant et après le numérique. *TAL*, 63(3), 63-88.
- Kahane, S., & Pietrandrea, P. (2009). Les parenthétiques comme “Unités Illocutoires Associées”. *Linx*, 61, 49-70.
- Kragh, K. A. J. (2021). Proposition d’une classification des marqueurs discursifs comme membres d’un paradigme. *Langue française*, 209(1), Art. 1. <https://doi.org/10.3917/lf.209.0119>
- Lacheret-Dujour, A., Kahane, S., Avanzi, M., Pietrandrea, P., & Victorri, B. (2011). Oui mais elle est où la coupure, là? Quand syntaxe et prosodie s’entraident ou se complètent: *Langue française*, n°170(2), 61-79. <https://doi.org/10.3917/lf.170.0061>
- Lacheret-Dujour, A., Kahane, S., & Pietrandrea, P. (Éds.). (2019). *Rhapsodie: A prosodic and syntactic treebank for spoken French*. John Benjamins Publishing Company.
- Lefeuve, F. (2011). *Bon et quoi* à l’oral: Marqueurs d’ouverture et de fermeture d’unités syntaxiques à l’oral. *Linx*, 64-65, 223-240. <https://doi.org/10.4000/linx.1422>
- Lefeuve, F. (2016). Les énoncés averbaux autonomes: Approche syntaxique et discursive. In J.-C. Anscombe, B. Darbord, A. Oddo, & C. García de Lucas (Éds.), *La phrase autonome: Théorie et manifestations* (p. 73-87). P.I.E. Peter Lang.
- Lefeuve, F. (2020). Les marqueurs discursifs averbaux résomptifs. In F. Diémoz, G. Dostie, P. Hadermann, & F. Lefeuve (Éds.), *Le français innovant* (p. 225-243). Peter Lang.
- Lefeuve, F., Morel, M.-A., & Teston-Bonnard, S. (2011). Valeurs prototypiques de *quoi* à travers ses usages en français oral. *Neuphilologische Mitteilungen*, 112(1), 37-59.
- Martinet, A., & Martinet, J. (1979). *Grammaire fonctionnelle du français*. Didier.
- Mazziotta, N. (2009). *Ponctuation et syntaxe dans la langue française médiévale: Étude d’un corpus de chartes originales écrites à Liège entre 1236 et 1291*. Walter de Gruyter. <https://doi.org/10.1515/9783110231069>
- Mazziotta, N., & Glikman, J. (2023). Emplois discursifs et pragmatiques des formes du verbe *écouter*: Observations sur les corpus *88milSMS* et *Les Vocaux*. In M. Saiz-Sánchez & S. Gómez-Jordana Ferary (Éds.), *Études de sémantique et pragmatique en synchronie et diachronie. Hommage à Amalia Rodríguez Somolinos*. Presses Universitaires de Savoie Mont Blanc. <https://hdl.handle.net/2268/304614>
- Mel’čuk, I. A. (2009). Dependency in natural language. In A. Polguère & I. A. Mel’čuk (Éds.), *Studies in Language Companion Series* (Vol. 111, p. 1-110). John Benjamins Publishing Company. <https://doi.org/10.1075/slcs.111.03mel>
- Pietrandrea, P., & Kahane, S. (2019). Chapter 6. Macrosyntactic annotation. In A. Lacheret-Dujour, S. Kahane, & P. Pietrandrea (Éds.), *Rhapsodie: A prosodic and syntactic treebank for spoken French* (p. 97-125). John Benjamins Publishing Company.
- Riegel, M., Pellat, J.-C., & Rioul, R. (2018). *Grammaire méthodique du français* (7^e éd.). PUF. (Édition originale 1994)
- Siouffi, G. (Éd.). (2020). *Une histoire de la phrase française: Des Serments de Strasbourg aux écritures numériques*. Actes Sud.
- Wilmet, M. (2010). *Grammaire critique du français* (5^e éd.). De Boeck.

- 1 L'autrice et l'auteur ont participé de manière équivalente à la rédaction de cet article. Nous remercions Christophe Benzitoun, Julie Glikman, Nicolas Gregov, Sylvain Kahane, ainsi que les personnes ayant évalué ce travail pour leurs commentaires et corrections.
- 2 On s'étonnera d'ailleurs de la nonchalance avec laquelle cette question est parfois traitée à coup d'analyse automatique fondée le plus généralement sur des critères graphiques. À titre d'exemple, voir les routines proposées par la suite logicielle *spaCy*, qui segmentent le texte en phrases en fonction des occurrences des signes de ponctuation <<https://spacy.io/api/sentencizer/>> [consulté le 7/9/2023]. De nombreuses grammaires dénoncent l'imperfection de ce type de démarche (Grevisse & Goosse, 1936/2011, p. 511; Martinet & Martinet, 1979, p. 17; Riegel et al., 1994/2018, p. 201-202; Wilmet, 2010, paragr. 537). Qui plus est, le code graphique est inopérant pour les écrits non normés et, bien entendu, pour les données orales, sur lesquelles nous nous focalisons.
- 3 Concernant le problème terminologique, nous renvoyons à la présentation synthétique de Berrendonner (2021, sect. 1.3).
- 4 On peut ainsi écrire *Une histoire de la phrase française* sans que le terme *phrase* ne soit précisément défini (Siouffi, 2020).
- 5 Nous suivons ici l'exemple de la Théorie Sens-Texte, dont l'approche modulaire permet de comprendre les interfaces entre les différentes composantes de la langue (Mel'čuk, 2009).
- 6 Sans avoir recours à la prosodie, pourtant évoquée comme une piste envisageable dans une publication antérieure (Benzitoun, 2010, p. 122).
- 7 Nous utilisons le terme *visée* plutôt que *force* pour souligner le fait que l'acte illocutoire ne se réalise pas nécessairement.
- 8 La transcription orthographique du projet LV rétablit les formes canoniques en cas d'élision (*j'te dis pas* → *je te dis pas*) : la prononciation authentique est enregistrée à un autre niveau. Les « ... » indiquent des pauses. La modalité interrogative est notée par « ? » pour faciliter la lecture.
- 9 La démarche inverse est également possible (Lacheret-Dujour et al., 2011, sect. 4.2).
- 10 SUD = *Surface syntactic universal Dependencies* (Gerdes et al., 2021). Guide d'annotation en ligne: <https://surfacesyntacticud.github.io/guidelines/u/oral_language/discourse/> [consulté le 8/12/2023].
- 11 Les auteurs qualifiaient ces unités d'*UI associées* avant la parution de l'ouvrage de synthèse sur *Rhapsodie*, mais ils s'expliquent du changement de terminologie en indiquant qu'elles se comportent davantage comme des noyaux (une seule composante d'UI) que comme des UI (Pietrandrea & Kahane, 2019).
- 12 Comme « ^ », similairement aux conventions de notation des expressions régulières en informatique.
- 13 Et non de *bon* comme le fait Lefevre (2011), en vertu de la capacité de *bon* à être employé comme tour de parole autonome (voir →2.3).
- 14 Le caractère prototypique de ces borneurs a été plaisamment mis en évidence par Auchlin dans le titre de son article de 1981 : « *Mais heu, pis bon, ben alors, voilà, quoi !* Marqueurs de structuration de la conversation et complétude ».