

Propositions et énoncés : unités pour la segmentation des corpus oraux

Gema Sanz Espinar

Universidade Autónoma de Madri (Espanha)

gema.sanz@uam.es

1. Introduction

Une bonne partie de la recherche en Acquisition de Langues Étrangères (ALE) dans l'approche fonctionnaliste et psycholinguistique se base sur l'étude de corpus de productions orales (récits, descriptions, dialogues,...), ce qui amène à des comparaisons entre productions de locuteurs différents, de langues différentes, enfin, de langues étrangères différentes. Cette approche requiert des démarches d'analyses précises que nous voulons expliciter ici.

Le premier pas dans cette démarche est la constitution de corpus oraux, puis leur transcription et découpage en unités linguistiques analysables et comparables. Nous nous arrêterons plus précisément sur les propositions et les énoncés. Il s'agit de notions de nature linguistico-conceptuelle, qui servent à décrire l'architecture de la chaîne parlée qu'est le discours et la segmentation de corpus oraux. Le discours, que l'on peut définir comme la langue mise en action, échange conversationnel et pratique sociale, peut être n'importe quel morceau d'une émission verbale. Mais du point de vue du contenu, on peut isoler des macro-unités discursives que l'on peut délimiter sur des critères de cohérence et de cohésion : les textes, qui peuvent être à leur tour divisés en unités plus petites : les énoncés et les propositions.¹

La segmentation de corpus oraux est une question complexe qui a fait émerger des unités de nature très différente. Levelt (1989, 23) a recensé une série d'unités dans le discours oral qui ont été utilisées par différents auteurs, parmi lesquelles on compte : l'unité informationnelle, la proposition phonémique, la structure propositionnelle, la phrase, la proposition de surface, la proposition profonde, le syntagme, le groupe tonal, l'unité de tour de parole,... La liste de Levelt n'est pas exhaustive et, comme on peut le constater, les niveaux ou les plans auxquels renvoie chacune des unités mentionnées sont multiples : sémantique, phonétique, syntaxique, interactionnelle.

Étant donné que le choix des unités linguistiques dépend en partie du phénomène linguistique que l'on vise à analyser, nous nous centrerons ici sur les énoncés et les propositions en tant qu'unités de base pour l'étude de la référence dans le discours. Ceci implique que l'étude de la référence ne doit pas se borner à ce qui est explicité, mais devrait aussi tenir compte du poids relatif de ce qui est explicité *vs.* ce qui n'est pas explicité dans le discours.

En suivant Levelt (1989), autant l'énoncé que la proposition sont des unités cognitives qui laissent apparaître une certaine conceptualisation opérée «en vue de parler», (*to think for speaking*, SLOBIN, 1991). Levelt (1989) propose un modèle de la production orale en langue maternelle, dans lequel le point de départ de la production est la *conceptualisation*, opération qui génère un *message préverbal* qui deviendra plus tard l'*input* des opérations de *formulation*. Le message préverbal est caractérisé par :

- une *structure propositionnelle* ;
- une *structure thématique* ;
- une *perspective* ;
- des caractéristiques spécifiques aux langues (*language-specific*).

Ainsi, la conceptualisation relève déjà de la construction d'une proposition et d'une entité énonciative. De plus, la génération de messages préverbaux est décrite par Levelt à travers deux opérations : la macro-planification et la micro-planification. Lors de la macro-planification apparaît une intention de communication chez le locuteur, qui sélectionne l'information nécessaire pour verbaliser son intention communicative. Lors de la micro-planification, cette information est hiérarchisée et linéarisée pour être formulée ; la *formulation* étant l'étape suivante dans la production, avant la dernière étape, l'*articulation*.

D'après le modèle de Levelt, au niveau de la conceptualisation, le locuteur fait le choix de la langue à utiliser, ce qui veut dire que l'on trouve, dès le niveau de la conceptualisation (aussi bien dans la macroplanification que dans la microplanification) des caractéristiques spécifiques à la langue parlée.

Lorsque l'on se retrouve face à un discours, ce n'est pas si facile d'identifier les unités conceptuelles derrière les unités de surface². Ainsi, notre propos ici est de relever quelques problèmes de segmentation du discours oral en énoncés et en propositions. Nous ne nous attarderons pas ici sur la notion de *texte*, entité aussi cognitive d'une dimension plus ample, pour laquelle nous renvoyons à Klein & Von Stutterheim (cf. article dans cette revue). Le texte serait, pour ces auteurs, la réponse à ce qu'ils appellent *quaestio*, c'est-à-dire une question virtuelle ou réelle, représentant une macro-intention de communication qui contraint jusqu'à un certain point la *référence*.

Ainsi, lorsque l'on tente de transposer cette modélisation à la production en langue étrangère, il y aura plusieurs questions clés à se poser à titre général pour l'étude de la construction de la référence dans le discours en LE :

- à quel point la conceptualisation et la formulation sont influencées par la langue native ou par la langue cible de l'apprenant (qu'est-ce qui est explicité vs. qu'est-ce qui n'est pas explicité ; quels moyens utilise-t-on pour exprimer la référence) ?

- s'agit-il d'un état des lieux qui varie au fur et à mesure que la langue étrangère est acquise et sous l'influence de quels facteurs (acquisition de la syntaxe, de la morphologie, du lexique, de l'usage pragmatique) ?

- comment se développe la structure propositionnelle et énonciative dans la langue de l'apprenant du point de vue de la conceptualisation et de la formulation ?

2. Le niveau propositionnel

La proposition est définie, du point de vue conceptuel, comme une structure *relationnelle* de type : *procès + environnement actanciel* (LEVELT, 1989 ; FRANÇOIS, 1990). Nous distinguerons la *proposition*, objet conceptuel, de sa corrélation formelle, la *proposition (énoncée) explicitée* (équivalent à *clause*).

Cette structure renvoie à une représentation mentale qui entraîne une sélection et organisation primaire de l'information stockée en mémoire. Cette sélection de l'information destinée à être verbalisée (sens référentiel³) n'empêche pas que d'autres informations soient *véhiculées* dans l'énoncé, non-explicitement (sens à inférer).

Il est tout de même intéressant de souligner que beaucoup d'auteurs défendent, comme Levelt (1989), que l'on ne peut «verbaliser» nos pensées que sous une forme propositionnelle.

There is more than a single «language of thought»; however, if a thought is to be expressed in natural language, the mediating code must be propositional. (LEVELT, 1989, 71)

Les travaux de Kintsch & Van Dijk (1977) sur la compréhension du langage, ainsi que les travaux sur la représentation de la connaissance dans le domaine de l'Intelligence Artificielle, montrent que la conceptualisation liée à la compréhension se fait aussi par des opérations de propositionnalisation. Par ailleurs, on souligne le rôle important, pour une bonne compréhension, de la restitution des chaînes causales entre les propositions, ainsi que le rôle des connaissances préalables sur le thème du discours ou sur la construction d'un texte (genres discursifs), comme l'évoquent Lambert & Voutsinas (cf. article dans cette revue Letras).

En suivant Klein & Von Stutterheim (1989, 1991 et article dans cette revue), la référence peut être étudiée à travers certains domaines référentiels de base souvent présents dans toute proposition : la temporalité, la spatialité, la modalité, les procès, et les entités. On pourrait y ajouter le domaine de la causalité (TRABASSO et NICKELS, 1992 ; TRABASSO et al : 1992), la qualification ou la manière (TALMY, 1985, 1988), voire le macro-domaine de la force, où Talmy (1988) inclut la causalité, la conséquence, la modalité.

La proposition semble le cadre minimal pour l'étude de la référence dans une approche discursivo-textuelle basée sur l'analyse des corpus. Elle nous permet d'analyser autant le niveau local (déploiement de la référence dans une proposition explicitée, type et moyens de cohésion) que le niveau global (hiérarchisation de l'information). Sur le plan formel, c'est aussi le contexte minimal pour regarder de près l'interaction des différents éléments linguistiques (sujet-verbe-compléments à l'intérieur de la proposition explicitée, mais aussi les rapports entre propositions).

Nous illustrerons l'avantage de cette approche à partir de la comparaison de deux études comparatives sur la conceptualisation et la lexicalisation⁴ des procès de mouvement dans différentes langues.

D'une part, Talmy (1985) compare les unités lexicales utilisées pour référer à des événements de mouvement dans différentes langues, notamment l'anglais *vs.* les langues romanes. Ainsi, le fameux exemple de *La bouteille est sortie de la cave en flottant / La botella salió de la cueva flotando vs. The bottle floated out of the cave* montre les tendances suivantes au niveau de la conceptualisation :

a) En anglais, on fait référence aux procès de mouvement par un schème récurrent de confluence des traits sémantiques⁵ [procès + manière + mouvement], la direction du mouvement étant exprimée par une particule ou *satellite*. Cette information est exprimée à travers un schème de lexicalisation récurrent⁶ [V+prep]. Par contre, dans les langues romanes, la confluence de traits sémantiques dans le verbe est différente [procès + direction du mouvement], de telle sorte que le trait [manière] doit être exprimée par un moyen adverbial séparé (« en flottant »).

b) L'analyse de la référence en contexte de Slobin (1996) fait encore un pas en avant en élargissant le cadre d'analyse au-delà du cadre de la proposition. Dans son étude, il prend deux types de données : des romans en anglais et leurs traductions en espagnol, puis des romans en espagnol et leurs traductions en anglais. D'autre part, il se base aussi sur un corpus translinguistique de récits oraux suscités à partir du même support (un livre sur images). Grâce à cette méthodologie, il peut non seulement retrouver les caractéristiques contrastives dont parlait Talmy au niveau de la conceptualisation et de la formulation, mais il montre aussi que certaines informations sont souvent omises en espagnol, mais explicitées en anglais, notamment la manière du mouvement, grâce au lexique verbal de l'anglais, tandis qu'en espagnol, on fait moins référence à la manière et on ne fait pas si souvent mention du *fond* (moins d'éléments locatifs explicités). Bref, au moment de comparer l'expression de la référence dans des langues différentes sur le plan de la conceptualisation et de la formulation, il vaut mieux analyser des données textuelles et non pas se borner à une comparaison hors contexte, puisque tout en disposant d'équivalents sémantiques, les langues peuvent en faire usage différemment.

Enfin, cet exemple renvoie à certains domaines référentiels, mais il sera possible de cerner la relation entre implicite et explicite pour chaque domaine référentiel (les entités, la spatialité, la causalité, la temporalité, les procès⁷, la modalité).

Pour terminer, nous rappellerons la définition de proposition de Klein (1995), qui dépasse le niveau purement lexical (sens lexical des mots), et qui renvoie à un contenu lexical structuré. Le niveau propositionnel comprend l'information relative aux anaphores, ellipses et éléments déictiques, c'est-à-dire, des éléments référentiels.

Ainsi, la proposition serait une conceptualisation *en vue de parler* qui s'insère dans une structure référentielle plus large (les anaphores et les ellipses relèvent d'une gestion, au moins locale, du déploiement de la référence dans un texte).

3. Le niveau énonciatif

La prise en compte parallèle de l'énoncé et de la proposition, nous oblige à préciser les rapports qu'ils entretiennent.

L'énoncé est le lieu où sont actualisés les contenus propositionnels, où l'on pourra attribuer le temps et le lieu de l'énonciation à l'énoncé, ainsi que reconstruire des inférences. Par ailleurs, le repérage des limites d'un énoncé multipropositionnel implique la prise en considération de l'interrelation entre les propositions (coordination, subordination, juxtaposition).

La segmentation en énoncés est loin d'être évidente. En effet, on entend par *énonciatif* autant des questions relatives à la structure informationnelle, c'est-à-dire le statut de topique et de focus de l'information (KLEIN&VON STUTTERHEIM, dans cette revue), qu'à la thématisation ou à la modalité. Maingueneau (1981) inclut les propriétés suivantes qui sont effectives dans tout énoncé⁸ :

- renvoyer à l'acte même d'énonciation (« franchement »)
- thématiser (« elle cueille des fleurs, Charlotte »)
- référer (« le tableau de Jean »)
- action, acte de langage ou force illocutoire (« promesse, question,... »)
- exprimer une modalité (« il se peut que »)
- argumenter (« donc, certes, puisque »)
- montrer l'intention de communication de façon non explicite par des sous-entendus.

L'énoncé serait le dernier palier pour la compréhension du discours, comme le dit Klein (1995) qui distingue les niveaux sémantiques suivants :

Niveau lexical (contenus lexicaux): sens lexical, des unités lexicales

Niveau propositionnel (contenus propositionnels): sens lexical + information contextuelle structurée (déixis, anaphores, ellipses)

Niveau de l'énoncé : information inférable

Ceci nous permet d'opposer les propositions et les énoncés de la façon suivante:

Énoncés : sens = référence + inférences

Propositions : référence explicite (lexicale, morphologique, anaphorique, déictique) et référence non explicite (ellipses, anaphore zéro)

D'autre part, les énoncés contiennent une ou plusieurs propositions⁹. Chaque proposition explicitée a une structure, disons *canonique*, de type *sujet + verbe + (compléments)*. Mais, on dit bien *canonique*, puisque pour certains énoncés, ce n'est pas si évident

d'y retrouver une forme canonique en surface : « Paul ! Oh !, Direction, la prison ! » ou « Un peu de vin, s'il vous plaît ! ».

Si pour être verbalisée ou compréhensible, toute séquence énoncée reconnaissable comme du langage doit contenir au moins une proposition, c'est donc au niveau conceptuel que l'on retrouve les propositions (procès + environnement actantiel). Les unités syntaxiques classiques telles que le sujet ou le verbe se retrouvent au niveau formel. En fait, le même problème pourrait s'appliquer aux formules figées, métaphores, expressions figurées ou interjections. C'est-à-dire, qu'il peut y avoir des formulations que l'on ne peut pas facilement relier à des énoncés informatifs classiques (rappelons que l'on parle aussi d'énoncés performatifs ou allocutifs où la relation entre le sens et la référence est moins directe).

Enfin, la notion de *proposition cognitive* n'est pas incompatible avec les deux ensembles d'énoncés : informatifs, non-informatifs. Même si certains énoncés ne semblent pas référer à des entités, des procès ou des propriétés, moyennant des parties du discours tels que les noms, les verbes ou les adjectifs, il semble tout autant bizarre de considérer qu'un énoncé puisse ne pas faire du tout référence à quelque chose. Ainsi, lorsque l'on dit « Oh ! » ou « Mince ! », on applique une certaine subjectivité (surprise, contrariété) à une situation précise, autrement, on demanderait au locuteur pourquoi il énonce cela.

Enfin, on constate bien l'existence d'énoncés incomplets, vagues ou ambigus dans la communication quotidienne et ils sont aussi fréquents dans les productions en langue étrangère surtout aux premiers stades. La non-compréhension d'un énoncé ou d'une partie d'énoncé nous oblige normalement à le laisser de côté dans nos analyses. D'où l'intérêt, en général, de trouver des matériaux d'appui pour la compréhension (version filmée des productions, confrontation du locuteur avec sa production, production parallèle en langue maternelle et en langue étrangère de textes à partir du même support).

4. La segmentation du discours en propositions

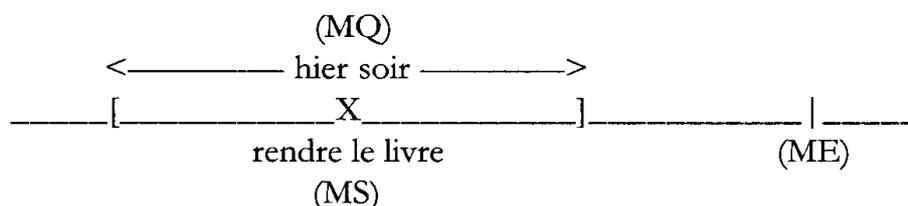
Au moment de segmenter le discours en propositions, les problèmes relèvent surtout du fait que la corrélation entre la proposition et son expression formelle n'est pas toujours basée sur une correspondance entité-nom, procès-verbe...

4.1 Typologie des propositions

Parmi les domaines référentiels, les procès constituent la base des propositions. Il peut y avoir une référence floue à des entités (« Il pleut », « Ça sent mauvais »), mais dans toute proposition on réfère à des procès et à leur ancrage temporel. Ainsi, Klein (1995) propose une classification de base des contenus lexicaux de la proposition du point de vue de la temporalité : contenus ou procès à 2 états, à 1 état ou à 0 état. Ces procès se distinguent par leurs propriétés temporelles, plus précisément, par le type de contraste qui s'établit entre l'intervalle temporel du procès (Moment de la Situation, MS) et le Moment en Question (MQ ou *Topic Time*). Le MQ se définit comme l'intervalle de temps par rapport auquel une assertion (ou bien d'autres

types de valeurs illocutoires) est faite. Il ne faut pas confondre le MQ avec le Moment de l'Énonciation (ME, *Utterance Time*), par rapport auquel se situe le MQ.

La figure suivante illustre le jeu d'intervalles qui sont à l'oeuvre dans un énoncé comme : "Hier soir, j'ai rendu le livre à la bibliothèque".



Voyons de plus près les caractéristiques de chaque type de procès :

- **Propositions à 0 état** : Il n'existe pas de contraste entre le MQ et n'importe quel autre moment envisageable sur l'axe du temps. La mise en relation de deux éléments (l'un nominal, l'autre qualitatif) se fait dans un MQ précis (bien que celui-ci ne puisse jamais offrir un contraste qualitatif avec d'autres intervalles temporels envisageables) : "La grenouille est verte" ; "Il boit". (au sens il est alcoolique).

- **Propositions à 1 état (états, processus, activités, certains événements)** : Il existe un contraste entre le MQ de l'assertion et un autre moment de l'axe du temps, mais ce contraste n'est pas véhiculé par le contenu lexical de la proposition. Le contraste temporel est extérieur à celle-ci. On dira donc que le procès, dans ces propositions, n'a pas de borne temporelle intrinsèque.

Ce type de contenu propositionnel correspond aux procès de type : état ("je suis fâché"), processus ("je grandis"), activités ("il court") et certains événements ("il pleut"). "Hier dans la fête, ils ont bu, ils ont dansé et ils ont fait beaucoup de bruit".

- **Propositions à 2 états (accomplissements, achèvements, actions, événements)** : Il existe un contraste temporel interne, borne interne ou passage d'un état initial à un état final distinct. Dans "J'ai bu mon café", le MQ de l'énoncé contient de l'information sur deux états qualitativement distincts : *un café à boire / pas de café à boire*. Ce type de contenu propositionnel correspond à plusieurs types de procès : achèvement ("il a bu son café"), action ("il lui a donné un coup"), certains événements ("il y a un accident").

4.2 Problèmes de découpage

En premier lieu, nous traiterons les problèmes généraux d'élision et de reformulation, qui peuvent apparaître dans n'importe quelle langue.

En ce qui concerne l'élision d'un verbe, on considérera logiquement qu'il y a une proposition sous-jacente¹⁰ :

ELECÉL2/ 10 : y qué pasa después? eh parece que Chaplin hace cárcel / va a la cárcel // pero la chica no

FLMCHA30. donc accident de voiture // Charlot <<qui s'était précipité sur la jeune fille>> la jeune fille et l'agent de police sont éjectés du camion

FLMJEN20. il en fait continue [sic. à faire ce genre de choses] // en volant un cigare

// en donnant des cartes aux enfants et cetera qui passent
FLMMAT32. donc direction la prison
FLMMAT36. lui assez galant // il lui cède sa place

Par ailleurs, un cas plus délicat apparaît lorsqu'il y a des des énumérations de propriétés avec omission des copules : *avoir/haber/tener* ; *être/ser/estar*

FLMLOI55. donc c'est hors de la ville aussi // et en fait c'est une baraque toute vieille // toute fatiguée // toute délabrée

Les textes descriptifs se prêtent souvent à ce genre d'ellipses.

Enfin, il ne faut pas confondre la proposition (ou le procès) avec la prédication (BAUDET, 1990 ; FRANÇOIS, 1990, 1991 ; BRUM DE PAULA, 2005) : *La grenouille verte vs. La grenouille est verte*. D'après Langacker (1991), au niveau cognitif, un procès (dans les termes de Langacker, un «verbe cognitif») est une opération de séquentialisation d'un état de faits. Or, dans le premier cas, on ne retrouve pas d'assertion liée à la prédication, ni de MQ lié à un ME, tandis que, dans le deuxième cas, il s'agit vraiment de l'assertion d'un MQ qui sert à ancrer la propriété *être verte* à l'axe temporel.

Par ailleurs, les reformulations obligent souvent à prendre des décisions méthodologiques. Bien qu'un même procès puisse être formulé de plusieurs façons, corrigé, précisé, etc. parfois les chercheurs se penchent soit sur la première formulation (pour être la plus spontanée), soit sur la dernière (celle *approuvée* par le locuteur après révision), soit ils gardent les deux pour l'analyse.

ELEEMM14 : il l'embarque

15. il le fait dans un fourgon

On trouvera que parfois la double formulation d'un procès ne relève pas toujours d'une correction ni d'une substitution proprement dites, mais de deux façons d'aborder le même événement, complémentaires, souvent sur deux plans qui contrastent de deux façons possibles (résumé vs détaillé, général ou explicatif vs. particulier ou concret). Parfois, ces doubles formulations se retrouvent lorsqu'il y a des retours en arrière dans le récit.

Il fait plusieurs choses pour retourner en prison, il vole un cigare il mange et il ne paye pas.

ELMCJES 33: y intenta // escaparse // atacando al policía // que estaba guardando la entrada (et elle essaie de s'échapper en attaquant le policier qui gardait l'entrée)

On ne confondra pas ce cas avec les répétitions d'action.

FLMCHA35. évidemment il rentre // il se prend la poutre sur la tête // il casse la table // la chaise // 'fin tout...

FLMCHA36. et donc ils arrivent à vivre // à installer la maison plus ou moins

Enfin, lors de l'analyse du corpus on trouvera des problèmes liés aux matériaux linguistiques propres des langues romanes : des verbes fléchis qui ne correspondent pas toujours à une proposition (« C'est Chaplin qui a volé le pain ») ; des propositions sans verbe (« Direction, la prison ») ou avec plus d'un verbe (« Il va partir »), des procès nominalisés (« Il est accusé du vol du pain ») et des compléments adverbiaux à statut propositionnel (« Sur son chemin, il rencontre la fille »).

Voici quelques exemples tirés du corpus :

ELMCJUA38: entonces ven una escena pues // de un hombre que sale de su casa / / y una parejita de recién casados probablemente // y la mujer muy feliz // y el hombre al trabajo. (trad. litt.: alors ils voient une scène donc d'un homme qui sort de sa maison et un couple de jeunes mariés probablement et la femme très heureuse et l'homme au boulot)

ELMCJES33: y intenta escaparse // atacando al policía // que estaba guardando la entrada (trad. litt.: et elle essaie de s'échapper en attaquant le policier qui gardait l'entrée)

Au niveau formel, on n'aura donc que des *pistes* (éventuellement des fausses pistes) pour la segmentation en propositions ; lorsqu'il y aura plusieurs langues à segmenter, les pistes formelles pourront aussi varier.

Les problèmes pour la segmentation en propositions relevés pour le français et l'espagnol dont nous ferons le commentaire sont les suivants :

- la présence ou non des verbes fléchis ou non-fléchis, où l'on peut distinguer plusieurs cas :

- a) les périphrases temporo-aspectuelles
- b) les procès de déplacement
- c) les cas liés à la dynamique des forces
- d) les structures de focalisation.

- les procès nominalisés et pronominalisés

4.2.1 Les verbes fléchis vs. non-fléchis

Levelt (1989) explique qu'une segmentation en «propositions de surface fléchies» (*finite surface clause partitioning*) amène à considérer l'exemple 1.a comme une seule unité formelle :

1.a) « Attila believe the world to be flat » (Attila croit que le monde est plat)

1.b) « Attila believe // the world to be flat » (Attila croit //que le monde est plat)

tandis que la segmentation en «propositions de base» (*basic clause partitioning*), correspondant à 1.b amène à considérer deux unités ; on pourrait donc parler d'une proposition principale et d'une proposition infinitive, les verbes, conjugués ou non, servant de critère pour la segmentation.

Pour aborder cette question, nous évoquerons ici la notion de *finiteness* (KLEIN,

1998 et KLEIN, à paraître), dont la traduction correspondrait à la notion *d'actualisation* dans le discours. Quand on parle de *finite verbs*, on fait normalement référence aux verbes fléchis. Or, étant donné que toutes les langues n'ont pas toutes de verbes fléchis, Klein redéfinit la notion en termes fonctionnels : il s'agit donc de l'actualisation de la référence dans le discours, de l'énonciation, normalement liée à une assertion (dans les énoncés informatifs), voire à la force illocutoire en général. C'est une opération d'ancrage discursif qui serait présente aussi dans les langues sans flexion, comme le chinois, où l'on trouve des particules déclaratives qui donneraient aux phrases le formatage nécessaire pour le passage de la langue au discours (KLEIN, 1998).

Enfin, les verbes non-fléchis peuvent renvoyer à des propositions ou non. Dans l'exemple de Talmy « La bouteille sale de la cueva flotando », le deuxième verbe fonctionnerait tout simplement comme un complément adverbial de manière, mais pas à un procès différent. On peut tout de même imaginer des exemples formellement semblables qui renverraient à deux propositions différentes : « La bouteille sort de la cave en montrant à tout le monde qu'il y avait un passage à l'extérieur ».

Par la suite, nous traiterons plusieurs phénomènes liés à la présence de verbes fléchis à sens grammatical, et à la présence de verbes non fléchis qui renvoient à une proposition.

a) Les périphrases temporo-aspectuelles

Les périphrases temporo-aspectuelles sont relativement fréquentes en espagnol et en français¹¹. Les valeurs sont multiples et les verbes impliqués constituent un ensemble important. Les périphrases sont des suites de verbe fléchi + verbe non fléchi, où le verbe non fléchi porte le sens lexical et le verbe fléchi est réduit à une valeur grammaticale. Ainsi, on trouve des périphrases terminatives, inchoatives, duratives, progressives...

- Périphrases terminatives : qui focalisent sur la phase finale du procès (*cesser de/finir de/terminar de/acabar de/dejar de ; finir par/acabar por*)

FLMCLG18. il finissent par s'échapper

- Périphrases progressives : « va andando por la calle » (il marche dans la rue)

- Périphrases duratives : « il est en train de manger »

- Périphrases inchoatives : qui focalisent sur la phase initiale du procès.

« sale corriendo » (il part en courant)

FLMVIR46. et donc là Chaplin et la fille se mettent à rêver

ELELAU13. y están a punto de ir a la cárcel // cuando el camión de policía tiene un accidente (ils sont sur le point d'aller en prison mais le camion de police a un accident)

- Périphrases de répétition :

FLMCLG10. mais la femme continue à accuser la jeune fille

- Périphases temporelles (futur proche) : Normalement, le contexte et l'intonation permettent de distinguer les emplois du verbe *aller* comme verbe de mouvement (2 propositions) ou comme verbe périphrastique *aller + infinitif/ir a + infinitivo*.

Il va faire plusieurs choses pour retourner en prison

b) Les verbes de mouvement

On tiendra compte du fait que les verbes de mouvement suivis du participe présent forment une seule proposition où le participe présent indique la manière du mouvement.

ELEANB24. y caminan dándose la mano

FLMCLG4: et donc la jeune fille part en courant avec son pain

Cependant, d'autres structures du type *V de mouvement + (a [espagnol]/ pour/ø [français]) + infinitif*, correspondent à deux procès successifs très proches dans le temps. Il s'agit de procédés de cohésion textuelle qui s'opèrent lorsque les sujets des deux procès coïncident. Le phénomène de la réduction de la morphologie n'est, en fait, pas rare dans les langues romanes :

Sale a despedirle (elle sort lui dire au revoir).

Se sientan a descansar (litt: ils s'asseoient se reposer)

Se va a trabajar (litt: il part travailler)

Se mete en un restaurante a comer (il rentre dans un restaurant manger)

FLMCHA38 : il vient // déjeuner

FLMCHA12. donc Charlot retrouve sa liberté

13. ils vont // attraper la fille

FLMMAT9. donc le boulanger accuse la fille auprès de la police

10. qui va // pour l'arrêter

FLMANB3. et elle court // pour s'échapper

c) La dynamique de la force

Le domaine référentiel de la *force* impliquerait cognitivement la représentation de deux forces opposées chacune liée à une entité : l'agoniste (entité sur laquelle on focalise à laquelle on attribue une tendance ou force), l'antagoniste (entité qui s'y oppose). On pourra focaliser sur un moment de l'interaction, soit le résultat de l'interaction (résistance, soumission à la force). Par ailleurs, la force peut se manifester vers le mouvement, l'action ou vers le repos, la non-action. C'est ainsi que lorsque Talmy (1988) parle de la dynamique de la force (*force dynamics*), il se réfère autant à des verbes modaux et d'intention (pouvoir, devoir, vouloir) qu'au procès causatifs ou résultatifs, y compris la voix passive et les expressions de l'agentivité, l'autoagentivité ou l'instrument, sur le plan soit physique soit psychologique ou social.

Modalité

Les complexes verbaux avec les verbes *pouvoir, devoir, vouloir* sont souvent analysés comme des périphrases modales. Cependant, pour certains verbes liés à la dynamique des forces, on peut repérer deux procès. Dans ces cas, l'agoniste serait encodé comme sujet grammatical.

Vouloir/querer, préférer/preferir, aimer/gustar : les verbes exprimant un désir ou un vœu sont en eux-mêmes des procès ancrés sur l'axe temporel qui introduisent une proposition à la modalité irréaliste comme un de ses arguments.

ELEJEN19b porque prefería (parce qu'il préférait)

19c vivir ahí (vivre là-bas)

FLMDAV29. et eux deux lorsqu'i voient ça // c'est en fait un peu ce qu'ils aimeraient // vivre

Poder/pouvoir : les verbes *poder/pouvoir* peuvent porter sur toute la proposition (possibilité de l'avènement d'un procès, voir exemples ci-dessous), soit référer à la capacité de quelqu'un (« il peut conduire » = il est capable de conduire). Seulement dans ce dernier cas, on pourrait envisager deux propositions.

ELEJEN21. y pues comió (après, il a mangé)

22. y no pudo // pagar (et il n'a pas pu payer)

Devoir/deber, être obligé de/tener que : le verbe *devoir* peut exprimer le doute (périphrase verbale) mais aussi l'obligation (*tener que, être obligé de*). Dans ce dernier cas, il s'agit de deux procès : le devoir attribué à quelqu'un et l'action visée.

FLMCLG23. et Charlie Chaplin sort de prison _donc on pense qu'il a dû aller en prison_

ELEISA20. por eso el agente no puede // hacer nada // y es obligado // llevarlo a la cárcel [et il est obligé de l'emmener en prison]

Structures causatives, agentives, résultatives, factitives, passives

Les procès à 2 états impliquent toujours un trait de *force*, ou de cause, dérivé de l'assertion d'un changement dans l'état de choses. La décomposition de la transformation peut entraîner des structures qui focalisent sur la cause ou la volonté d'agir, le processus ou l'état final.

Structures passives et résultatives (mise en relief du patient/agoniste à travers sa position en début de phrase et son encodage dans le sujet grammatical).

En français : *être + participe ; se voir + infinitif ; se faire + infinitif* (passif) ; *se trouver + participe ; se retrouver + participe* (résultatif).

En espagnol : *ser + participio* (passif), *estar + participio* (résultatif).

FLMANB16. il fait venir un policier // pour faire constater le fait // et se faire arrêter

FLMDEL11. ils se retrouvent dehors // expulsés du camion

Structures factitives (constituent des cas de co-lexicalisation qui renvoie à un seul procès. Souvent la combinaison *faire/hacer* + *infinitif* (Vintransitif) devient une construction transitive).

FLMANB16. il fait venir un policier // pour faire constater le fait // et se faire arrêter

ELEJEN26. y el [policía] le hizo entrar en el camión (et le policier l'a fait monter dans le camion)

Structures agentives : on en trouve quelques-unes en espagnol. Il s'agit des complexes verbaux *coger y V2* ; *agarrar y V2* ; *ir y V2*. Il ne faut pas comprendre ici que le verbe *coger* garde son sens original référent à l'action de *prendre*. Par contre, le sujet dans ces structures doit être [+humain] et on l'interprétera facilement comme l'agent d'une action (exemple *d'autoagentivité*).

FLEMAG33. y mientras viene el coche coge y se fuma un puro de un estanco
[et pendant qu'il attend la voiture, ça lui prend et il fume un cigare dans un tabac]

On peut considérer que les verbes *essayer de/intentar*, *décider de/decidir*, sont aussi des verbes agentifs : ils introduisent un procès ancré dans le temps (un essai, une décision) et un procès visé, dont on ne fait pas l'assertion, bien que l'on puisse parfois la déduire.

et alors la fille décide // de s'échapper

ELEJEN6. que intentó // ponerla en prisión (et on essaie de l'arrêter)

Structures causatives (*laisser/dejar*, *éviter/evitar*, *empêcher/impedir*, *permettre/permitir*, *provoquer/provocar...*) : on y trouvera l'antagoniste comme sujet grammatical. Enfin, il s'agit d'une force en faveur d'une situation qui est réussie (*laisser faire quelque chose*) ou à l'encontre d'une situation virtuelle (*il l'empêche de partir*). C'est le sens qui nous permettra de discerner les cas qui relèvent d'un seul procès (*cause+état résultatif*) ou de deux procès (*prise de décision/action + procès qui suit comme conséquence*). Ainsi, on trouve parfois le même verbe pouvant jouer les deux emplois : « Deja a Chaplin suelto » (il laisse Chaplin relâché/libre) vs « deja a Chaplin // irse » (il laisse Chaplin // s'en aller).

FLMDEL3. mais le policier ne le croit pas et //donc il lui laisse s'en aller // et ils embarquent la jeune fille

FLMISA6. il dit que c'est lui qui a volé le pain que c'est pas elle // ce qui permet à la jeune fille // de fuir // et ce qui permet à Chaplin // de retourner en prison...

FLMCLG7. et Charlie Chaplin dit que c'est lui qui a volé le pain pour éviter à la jeune fille // d'aller en prison

Ces exemples diffèrent sémantiquement sur un point qui nous semble clé : *laisser s'en aller* et *permettre de retourner* impliquent l'action de *permettre* puis l'expression d'une conséquence qui peut *ne pas* arriver (X peut permettre Y de faire quelque chose, mais Y peut ne pas le faire). Par contre, *dejar suelto* est plutôt une périphrase agentive+résultative. Lorsque l'on laisse quelqu'un libre, il est libre.

Il faut tout de même préciser que le statut de réalité assertée de *s'en aller* dans *laisser s'en aller* ou de *permettre de s'en aller* est compliqué à expliquer. Souvent, on est tenté d'interpréter que l'événement a lieu, mais en fait cela n'est pas forcément le cas. *Il l'a laissé partir, mais il n'est pas parti*. En fait, le locuteur est en train de mettre au premier plan le rapport des forces mais c'est à l'interlocuteur de reconstruire la suite, l'événement résultant.

Structures résultatives (qui expriment l'accomplissement de l'action comme quelque chose d'attendu ou de positif: *parvenir à/ réussissent à/ arriver à; conseguir +inf/ lograr+inf / conseguir+ que + VSubjonctif / lograr + que+ VSubjonctif*).

ELECEL9. y logran fugarse

FLMCEL5. donc la jeune fille arrive à s'enfuir

Un sous-groupe particulier de verbes de cognition exprime le fait que quelqu'un évalue une situation pour réussir à atteindre un objectif (*en profiter pour/ aprovechar (para/y), saisir l'occasion pour/ aprovechar la ocasión...*).

FLMJEN9. et puis <<en se relevant >> Charlie Chaplin << qui lui n'avait qu'une idée // se retrouver en prison // pour pouvoir être nourri logé et blanchi >> saisit l'occasion // pour dire // que c'est lui et non pas la fille qui a volé le pain

ELEMRO5. y finalmente ha encontrado la oportunidad // de volver a la cárcel // comiendo algo en el restaurante // y sin pagar la cuenta

d) Structures de focalisation

Les verbes qui sont souvent impliqués dans des structures de focalisation en français et en espagnol sont *ser/être* ou *haber/y avoir*. La structure de focalisation comporte normalement l'un de ces verbes suivi du complément focalisé + relatif + proposition. Cependant, on trouve des variantes. Chez les natifs, il manque parfois le verbe initial et, chez les apprenants, on peut trouver des emplois de *estar* à la place de *ser* en espagnol (ce qui est fautif). Chacune de ces structures (*être vs. avoir*) a des contextes d'emploi particuliers et elles ne sont pas tout à fait équivalents en français et en espagnol).

ELENCL11. pero una mujer <<que conoce la verdad>> dice al policía // que Chaplin no ha robado sino que es la chica quien ha robado

FLMCLG24. et là c'est la jeune fille qui l'attend à la sortie

ELEMAT56. y después bueno otra vez se rompe una silla // no se rompe // sino que es el suelo que se cae bajo el peso de la silla

Les constructions avec *haber/y avoir* peuvent garder le sens locatif (au lieu d'existential) auquel cas on pourrait le considérer comme une proposition à part « là-bas, il y a une fille // qui... » En tout cas, il s'agit toujours d'une opération de focalisation qui introduit normalement en début de phrase des entités non introduites auparavant dans le discours (notamment c'est la fonction de cette structure en espagnol). Dans tous les exemples, le verbe *y avoir* semble beaucoup plus désémantisé en français qu'en espagnol.

FLMMEL17. et puis en fait + qu'est-ce qui ?/ ah oui en fait voilà il y a un moment où ils se retrouvent dans une étendue sur l'herbe

FLMMRO4. mais il y a une dame qui dit à la police // que c'était pas du tout lui // et que c'était la fille d'avant

ELEMAT50. hay una escoba // que <<cuando la sacan>> // se cae la mitad del techo

Il existe aussi d'autres types de structures de focalisation, notamment en français, que l'on associe souvent à la langue orale : *SN + relatif + proposition*

FLMMEL18. et ils voient un couple

18.a qui sort d'une maison

18.b le mari qui part travailler

Il y a enfin des cas d'ellipse associés à la structure de focalisation :

La dame dit // que ce n'est pas Chaplin mais la fille qui a volé le pain

La dame dit // que ce n'est pas Chaplin

La dame dit // que ce n'est pas Chaplin mais la fille

Dans les trois cas on parlera d'une proposition même si la complexité et l'explicitation est différente dans chacune d'elles. Notamment, le procès est omis dans les deux derniers.

4.2.2 Les procès nominalisés et pronominalisés

On peut distinguer les procès nominalisés dans un complément nominal ou adverbial (*être accusé du vol ; sur son chemin*) et les procès pronominalisés (*il a vu ça*).¹²

ELECÉL2/4 pero la gente no se lo creyó (mais les gens n'ont pas cru à cela)

FLMDAV29. et eux deux lorsqu'i voient ça // c'est en fait un peu ce qu'ils aimeraient // vivre

FLMELO10. ensuite ils arrivent à s'échapper de la voiture de police // après un accident

FLMNCL10. et Chaplin <<qui est tombé sous le charme de la jeune fille>> il décide // de la protéger // en s'accusant // du vol

Les compléments adverbiaux ayant des noms déverbaux, voire des infinitifs ou ce que l'on pourrait appeler noms processifs (de par leurs caractéristiques temporelles) renvoient souvent à des propositions indépendantes :

FLMDAV25. ensuite bon au cours de la fuite // on arrive à une maison

FLMJEN38. et eh donc sur leur chemin // ils s'arrêtent pour souffler un peu

FLMMAT24. et au moment de payer // donc il prend le ticket

Dans les cas de nominalisation, il est parfois difficile de trancher si l'on doit identifier une proposition à part. Dans ces 3 cas ci-dessus, les noms se trouvent dans des compléments adverbiaux en début de phrase, en topique. On sait que la nominalisation est un phénomène formel, mais aussi cognitif (LANGACKER, 1991). Il est souvent lié à des réductions de l'environnement actantiel dans la structure de surface, ce qui n'est pas suffisant pour empêcher l'analyse en propositions cognitives. En effet, *Après un accident* garde plus clairement son caractère processif que *Elle l'accuse du vol*.

Par ailleurs, les procès peuvent être enchâssés, subordonnés à des noms ou des adjectifs. Les infinitifs introduits par *pour/para* montrent une certaine ambiguïté. On pourrait penser qu'il s'agit de propositions de but ou bien qu'elles forment une tournure pour exprimer une entité nominale complexe.

ELEDEL13. y luego no sé eh encuestran un sitio // para dormir... (et après ché pas ils trouvent un lieu pour dormir)

FLMMAT28. et en fait il constate // que Chaplin ben il a le ticket de caisse // pour payer // mais il paye pas

Par contre, la structure française *N+à+Infinitif* est plutôt une qualification qu'une action située dans le temps avant ou après le procès principal :

FLMMAT23. et i prend deux plateaux archipleins de choses à manger et à boire

Nous concluons cette partie en soulignant que bien que certains de ces découpages peuvent être discutables, une fois que l'on aura décidé la démarche d'analyse pour les différents cas, il faudra garder la cohérence dans l'ensemble des analyses.

5. La segmentation en énoncés

Parmi les problèmes de segmentation en énoncés, on trouve essentiellement les cas de coordination et de juxtaposition. Quant aux subordonnées, il est assez facile de les relier aux propositions principales, la seule précaution qu'il faut avoir lors de la transcription, c'est de marquer très tôt, soit l'intonation, soit la place occupée par la subordonnée, puisque parfois l'intonation est déterminante pour la situer, comme on peut déduire à partir de ces exemples :

a) « il voulait retourner en prison après le départ de la fille

il rentre dans un restaurant »
b) « il voulait retourner en prison
après le départ de la fille il rentre dans un restaurant »

Quant à la coordination, nous montrerons un exemple à partir de la transcription d'un récit suivi :

(1) « donc il n'est pas content (2) parce qu'il voulait retourner en prison (3) et donc il entre dans un restaurant (4) et il mange tout (5) ce qu'il peut (6) et à la fin s'en va (7) sans payer (8) donc la police l'arrête »

Comme on le voit en 3, 4 et 6, à l'oral sont fréquentes les séquences reliées par *et*, en tant que connecteur strictement cohésif (traditionnellement conjonction de coordination). Si on considérait *et* comme conjonction de coordination, cela nous amènerait à garder de longues séquences de discours. Ainsi, il faudra bien distinguer entre *et* coordonnant (qui relie des entités au même niveau sémantique *blanc et noir, il rit et il pleure*) ou cohésif (qui relie souvent des propositions adjacentes dans les récits).

Par ailleurs, la juxtaposition (lien sans connecteur explicite), peut être à la base d'un rapport de coordination ou de subordination.

« Donc il veut retourner en prison
Il entre dans un restaurant
Il mange tout ce qu'il peut
au moment de payer, il n'a pas d'argent
et il est arrêté »

Dans ces cas, il nous semble qu'il faudra distinguer le plan de *l'énoncé* et le plan textuel. Pour cet exemple, on peut travailler la cohérence textuelle avec la notion d'*épisode* (TRABASSO et NICKELS, 1992 ; TRABASSO et al., 1992), qui est un groupement de propositions basé sur les notions de *but, tentatives (actions, événements orientés en faveur ou à l'encontre du but) et conclusion*. En effet, dans l'exemple, il existe des liens paratactiques qui rendent une suite chronologique de situations précédées de l'expression d'un but, les 3 situations qui suivent sont des tentatives/événements orientés et la dernière, la conséquence des actions. On pourrait donc établir des relations locales entre propositions, mais la fonction du *et* ou de la juxtaposition est liée à l'établissement de rapports hiérarchiques (au niveau textuel) entre des énoncés relativement indépendants du point de vue syntaxique.

En fait, les termes de *coordination, subordination et juxtaposition* sont souvent mal définis et renvoient à une hétérogénéité de cas. D'ailleurs, la notion de dépendance est souvent peu évidente : dépendance syntaxique, sémantique, intonative ? (LÓPEZ GARCÍA, 1999).

Enfin, pour la segmentation en énoncés, on devra faire attention à plusieurs critères : le critère intonatif¹³, le critère énonciatif (rapports interpropositionnels de dépendance, notamment le lien entre les propositions présupposées ou en topique et les propositions posées), le critère syntaxique (phrase complète ou non complète,

équivalence ou non des éléments liés pour identifier la coordination) (LÓPEZ GARCÍA, 1999 ; KIRCHMEYER, 2002 ; NOYAU, 1997).

6. Conclusions

Dans cet article, nous avons voulu établir un dialogue entre une approche théorique cognitive et psycholinguistique de la production langagière et une démarche méthodologique. En effet, nous sommes partie de la définition de deux unités linguistico-cognitives : l'énoncé et la proposition, puis nous avons essayé de rattacher ces définitions à l'analyse concrète de ces unités, en relevant des problèmes et en proposant des critères pour le découpage de la chaîne parlée en propositions et en énoncés.

Nous avons montré des problèmes généraux liés à l'analyse des corpus en général, mais puisque nous nous sommes basée sur un corpus de récits en espagnol et en français LM et LE, nous avons aussi pu montrer toute une série de problèmes propres aux langues française et espagnole en particulier.

Par ailleurs, nous nous sommes intéressée à ces deux unités : la proposition et l'énoncé, en tant qu'unités de base pour l'étude de la référence dans le discours. Comme nous l'avons montré, pour ne pas être partielle, l'étude comparée de la référence devrait se faire dans le cadre de contextes plus larges (c'est-à-dire dans des textes produits dans des conditions de communication).

Le découpage du discours permet enfin des études quantitatives appuyant les études qualitatives : nombre de propositions conceptualisées (en tant que mesure qualitative de la longueur du discours), nombre de propositions par énoncé et liens entre elles, densité sémantique dans une proposition et nombre de propositions pour conceptualiser un certain événement (ce qui correspond aux notions de «condensation», «granularité sémique» et «granularité temporelle» présentes chez NOYAU, 1997 ; NOYAU et PAPROCKA, 2000 ; NOYAU et al., 2005), rôle des propositions lors du déploiement de la référence dans des épisodes (tel que la notion est définie par TRABASSO et al., 1992 ; TRABASSO et NICKELS, 1992), déploiement de la référence dans le texte (notions de «maintien», «introduction», «décalage», «réintroduction» de la référence, introduction en «topique» ou en «focus», chez KLEIN & Von STUTTERHEIM, 1989, 1991, et l'article dans cette revue Letras).

En conclusion, lorsqu'une étude de la référence veut partir de données discursives, on doit tenir compte du niveau de la conceptualisation et de la formulation dans un cadre plus large montrant ce qui est explicité *vs.* ce qui n'est pas explicité dans le discours. Cette voie ne peut que montrer des résultats intéressants en RALE, puisque l'on réussirait à décrire non seulement la compétence linguistique des apprenants de LE mais aussi la compétence textuelle et discursive en LE.

Notes

¹ Nous ne parlons pas de *phrase* à ce niveau, puisque nous voulons pour l'instant nous centrer sur des unités de nature conceptuelle, plutôt que formelle ; la phrase étant le matériau linguistique organisé syntaxiquement avant l'actualisation dans le discours au moyen d'un acte d'énonciation (NEVEU, 2000).

² Nous illustrerons ces critères par des énoncés tirés de 4 sous-corpus d'espagnol et de français langues maternelles (ELM, FLM) et langues étrangères (ELE, FLE) de récits de fiction suscités après le visionnement d'un court-métrage des **Temps Modernes** de Chaplin.

³ Par *référence*, nous n'entendons point ici la référence extralinguistique, mais la représentation mentale des situations sous-jacente au discours (intralinguistique).

⁴ Ici, il faut entendre par *lexicalisation* le choix d'unités lexicales *référentielles* qui rendent l'information relative aux procès, c'est-à-dire essentiellement les verbes et les noms déverbaux.

⁵ *Conceptualization patterns.*

⁶ *Lexicalization patterns.*

⁷ Cf. Pour la relation entre explicite et implicite dans le domaine des procès on peut souligner les notions de granularité sémique et temporelle (NOYAU, 1997 ; NOYAU et PAPROCKA, 2000 ; NOYAU et al., 2005 ; SANZ ESPINAR, 1999).

⁸ Notons que toutes ces propriétés sont placées dans le modèle de Levelt (1989) au niveau de la conceptualisation.

⁹ Comme le dit Maingueneau (1981), parmi d'autres fonctions, les énoncés "réfèrent".

¹⁰ Les trois/quatre premières lettres majuscules renvoient à la langue utilisée (ELM - espagnol langue maternelle dans le groupe de contrôle; ELE - espagnol langue étrangère, ELMC - groupe de contrôle ELM, FLMC - groupe de contrôle FLM); les trois lettres suivantes renvoient au nom du locuteur) ; enfin, le numéro renvoie au segment du récit et la lettre (a, b, c) à une proposition faisant partie d'un énoncé multipropositionnel. Le signe / correspond au découpage en propositions. Le signe // à une pause dans le discours. Le signe <<—>> au découpage d'une proposition qui est insérée dans une autre.

¹¹ On peut dire qu'en espagnol il y a, tout de même, un éventail plus large de périphrases verbales.

¹² Bowerman (1984) montre, à propos de l'acquisition de la temporalité que la conceptualisation d'un procès comme un argument d'un autre est une opération plus coûteuse que la conceptualisation indépendante des deux procès, qu'il y ait reprise anaphorique (au moyen d'un pronom) ou pas : *Le train est parti et ils ont vu ça // Le train est parti et ils l'ont vu.* Ainsi, dans l'acquisition enfantine au moins, la conceptualisation *enchâssée* d'une proposition apparaît plus tard, après la reprise anaphorique.

¹³ Il faudra, soit rendre l'intonation par des marques ou symboles non-ambigus dans le récit transcrit suivi, soit éviter l'ambiguïté au moment de la lecture de la transcription par un découpage en énoncés à partir de l'écoute de l'enregistrement.