

La lutte des classes des unités linguistiques : micro- et macro-syntaxe

The class struggle between linguistic units: micro- and macro-syntax

María Luisa Fernández-Echevarría¹

Abstract: In a synchronic approach to discourse we postulate a dynamic unit that constructs macro-syntactic meaning *as it is uttered*. This conventionalised segment (fr. *période énonciative*), elaborated in order to anticipate the reactions of the hearer, forms part of a linguistic repertory (vocabulary) while at the same time behaving as a space-time module (syllabification). Conceived in this way, this unit is flexible enough to undergo possible deformation and accentual attrition in the evolution of a given language. It is also the basis of a meta-language capable of generating progressively more precise linguistic objects whose categorial membership is a matter of debate. We may observe this conventionalised segment in the analysis of set expressions (*paremies*) that testify to shared meanings and, by default, in automatic deviant sub-titles that reveal the shortcomings of a virtual translator.

Key words: meso-syntax, paremiology, conventionalised segments, phonology, syllabification, virtual translator.

Introduction : séquences médianes, macro-syntaxe et micro-syntaxe²

Pour pouvoir rendre compte de la cohérence et la cohésion textuelles et de l'harmonie entre macro- et micro-syntaxe, nous avons cherché le lieu (trope) où se produit l'interprétation sémantique. Le résultat de notre étude indique qu'elle se produit dans les séquences médianes composées de « mots prosodiques » (Martin³) *signifiants*, c'est-à-dire à l'intérieur des groupes syllabiques dont la longueur peut être enregistrée

¹ Universidad Complutense Madrid/Modyco Paris X ; luisafernandez@ucm.es.

² Je tiens à remercier les deux relecteurs anonymes, ainsi que Liana Pop et Daciana Vlad, pour leurs relectures soignées et précises, qui ont fait possible la parution de cet article dans les meilleures conditions.

³ Voir Martin: http://new.projet-pfc.net/index.php?option=com_docman&view=docum ent&layout=default&alias=285-martin-la-structure-prosodique-cognitive-realisation s-regionales&category_slug=journees-pfc-paris-2011&Itemid=215.

dans la mémoire phonologique (6/7 syllabes en français) et encodée selon des inflexions intono-accentuelles. Cette procédure de sémantisation a un caractère sémiologique car, comme l'indique Coursil (2000 : 73), « la forme d'inscription première d'une pratique dans la langue est le vocabulaire, [...] l'environnement d'un sujet de langage est partout sémiotisé [...] et le langage est présent sous chaque geste du sujet ».

Le caractère phonologique du « mot prosodique » et le caractère sémiologique du vocabulaire font concevoir une unité linguistique qui dispute son statut de mot avec le mot lexical, puisqu'il faut distinguer :

le « lexique » relevant d'une grammaire de valeurs pures, du « vocabulaire » relevant d'une sémiologie de pratique particulière, c'est-à-dire d'un système de signes. Ainsi distingués, on dira que dans la langue, les formes lexicales ont une diachronie, alors que dans une sémiologie, les signes de vocabulaire ont une ethno-histoire. (Coursil 2000 : 73-74)

Dans ce double cadre de signes et formes lexicales nous envisageons trois niveaux :

- le niveau micro-syntaxique d'articulation syllabique (morpho-syntaxe),
- le niveau méso-syntaxique d'articulation phonémique (phonologie/sémiologie),
- le niveau macro-syntaxique de reconstruction textuelle (liens « logiques »).

Ces trois niveaux correspondent aux « trois systèmes de repérages » utilisés par les langues selon Renaud (2005 : VII) : « les repérages déictiques par rapport au moment d'énonciation, les repérages anaphoriques par rapport au point de référence fourni par le contexte et le repérage calendérique ».

Nous avons donc une interaction de paramètres catégoriels pour organiser les icônes signifiants répertoriés dans le système de la langue. Les signifiants se présentent comme des segments syllabiques bornés : les *périodes énonciatives* (PE dans ce qui suit) que l'on va aborder ici par le biais des figements linguistiques (parémies, phraséologies, collocations, locutions, dictons, proverbes...)⁴. Les PE font partie d'un lexique verbalisé où, comme dans les proverbes, « la structure n'est pas indépendante de la syntaxe avec laquelle elle est exprimée » car : « [...] dans la mémoire collective où grand nombre de parémies sont enregistrées, on voit se configurer un réservoir de syntaxe orale et c'est dans ce réservoir que des schémas générateurs de nouvelles formulations se conforment [...] » (Bizarri 2004 : 58).

⁴ Voir Sevilla Muñoz & Crida Álvarez (2013) pour des classifications détaillées.

⁵ «La estructura sobre la que se sustenta su formulación, no es independiente de la sintaxis con la que se la expresa [...] en la memoria colectiva, donde se guardan gran cantidad de paremias, se conforma un reservorio que configura la sintaxis oral, y se crean esquemas que luego generan nuevas formulaciones.» (notre traduction).

Ce *réservoir de syntaxe orale* trouve sa place (locus) entre la micro- et la macro-syntaxe. A un niveau intermédiaire (méso-syntaxique), l'énonciateur combine et manipule des structures signifiantes, emmagasinées dans la mémoire collective. Il assure l'interface entre micro- et macro-syntaxe à travers la syllabation, où la contiguïté joue un rôle syntaxique par référence à une réalité « méta/extralinguistique ». Nous cherchons aussi la piste de ces unités signifiantes, modulaires et en constante génération, dans l'analyse de l'activité interprétative du traducteur.

Notre travail est divisé en quatre parties. Dans la 1^{ère} partie nous rappelons notre classification syllabique des expressions figées proverbiales en 6 catégories. Ce premier travail permet de repérer les constantes phonologiques d'ouverture/fermeture prosodique ancrées dans leur constitution, similaires en français et en espagnol, pour les opérateurs que nous proposons. Dans la 2^{ème} partie nous argumentons sur la notion de *télicité* (complétude énonciative) qui met en place un sujet d'énonciation qui organise son discours en fonction de l'entendant. Nous signalons quelques caractéristiques différentielles d'encodage accentuel du français par rapport à l'espagnol pour montrer comment il implique de subtiles transformations dans les unités lexicales. Dans la 3^{ème} partie nous présentons quelques exemples de segments syllabiques a-significatifs figurant dans un sous-titrage automatique sur Internet et les traitons en termes de fréquences morphologiques déviantes par rapport au modèle audio transcrit par nos soins. Dans la 4^{ème} partie nous spécifions quels « types » de séquences signifiantes pourraient servir dans l'étude contrastive des langues et/ou à l'implémentation de systèmes de traduction automatique. Finalement, nous concluons que quatre catégories syntaxiques principales opèrent pour déterminer le choix d'éléments signifiants au niveau de la méso-syntaxe, dont deux nucléaires (syllabiques) et deux par composition (rapports syntagmatiques). Par leur caractère interactif morphologique, ces quatre catégories se constituent en éléments générateurs du discours et peuvent être réduites à deux (signifiants dérivables/non dérivables). Les séquences médianes signifiantes, par leur typologie phono/sémiologique, sont le substrat d'une catégorisation linguistique dynamique présente dans le traitement automatique des textes et pouvant rendre explicite leur contenu pragmatique.

1. La génération des unités linguistiques : le domaine de la méso-syntaxe

Nous définissons la méso-syntaxe comme le domaine où une ouverture/fermeture énonciative détermine des unités signifiantes par rapport à un co(n)texte macro-syntaxique. Ce terrain d'analyse intermédiaire présente des unités lexicales sans lesquelles aucune

catégorisation linguistique n'est possible. Nous choisissons parmi ces catégories les *parémies* qui, ayant été largement étudiées par la phraséologie, ont acquis un statut lexical spécial dans le cadre du discours. Elles sont en effet aisément repérables par tout locuteur expert et sont signifiantes par leur caractère « autonome illocutoire »⁶. Dans ce qui suit, nous en présentons un échantillon d'exemples après avoir réalisé une étude de quelque 600 formes proverbiales⁷ en nous inspirant de l'approche sémiologique de Coursil (2000). Cette étude montre les caractéristiques communes des séquences lexicales médianes en français/espagnol, qui devraient opérer dans toutes les langues, par rapport à 6 critères :

- leurs formes récurrentes assonantes (critère rythmique) ;
- la polarité (opposition forme affirmative *vs* négative) ;
- l'utilisation d'éléments isosyllabiques de reprise (pronominalisation/ déictiques/ embrayeurs) ;
- la dérivation lexicale (morphologie verbale, adjectivale, (pro)nominale) ;
- la modalité allocutive (interrogation, impératives, discours direct/indirect) et, par défaut,
- l'absence de critères de constitution interne (déclaratives pures, simples ou par composition).

La 6^{ème} catégorie, non définissable en termes de constitution fréquentielle interne, intègre les structures *tautologiques*⁸, selon Schapira (2012). Nous l'avons répertoriée par défaut et, en tant que PE « pure », elle fait partie de la 3^{ème} partie de notre travail, centrée sur la traductibilité. Les parémies de ce dernier type, malgré leur constitution neutre, sont facilement repérables par leur effet accentuel en position initiale et finale, se conformant assez bien à l'étiquette de « périodes intonatives » dans la terminologie de Lacheret-Dujour & Victorri (2002). Nous les avons analysées en termes d'*extension* pure linéaire en limitant leur *espace* dialogique énonciatif (mésosyntaxe) grâce au repérage de points d'inflexion prosodiques que nous avons symbolisés par le signe # (dièse/hashtag).

2. Prosodie et *périodes énonciatives*

Pour définir les PE nous avons eu recours à leur caractère de *signe linguistique*. Dans la perspective saussurienne, « le signifiant étant de nature auditive, se déroule dans le temps seul et a les caractères qu'il emprunte au temps : a) *il présente une étendue*, et b) *cette étendue*

⁶ Nous empruntons la terminologie de Kahane & Pietrandrea (2009).

⁷ Fernández-Echevarría (2014a&b).

⁸ Ou « assertions superflues » faisant partie de ce qu'Anscombe (2006c, *apud* Schapira 2012 : 105) appelle « le savoir populaire », un savoir primaire et basique, populaire en ce qu'il n'est pas appris à l'école mais gagné par l'expérience du quotidien.

est mesurable dans une seule dimension : c'est une ligne» (Saussure 1972[1915] : 100-103).

Le caractère linéaire du signifiant établit des éléments de comparaison entre les segments énonciatifs eux-mêmes, ou entre ceux appartenant à différentes langues, ou encore entre les modalités (oral, graphique, iconique). Si nous contrastons les PE des langues telles que l'espagnol et le français, nous remarquons que les groupes intonoaccentuel produisent des différences remarquables dans leur constitution à cause de l'origine diverse du vocabulaire et malgré leur similitude étymologique. Nous argumentons alors avec Fabre (2015) que « c'est la dissymétrie plutôt que l'analogie qu'il est intéressant d'analyser » car « la dissymétrie constitue une forme de création lexicale ». Commençons alors par l'analyse d'une parémie basée sur le mot espagnol *aceituna* [olive] d'origine arabe et de son doublet latin *oliva*. Nous constatons que, comme prévoit Fabre en proposant ce même doublet, la création lexicale semble avoir une base rythmique produite par la syllabation dissymétrique d'un même signifié.

2.1. L'exemple de l'accent lexical espagnol

- (1) Olivo y aceituno todo es uno [C'est bonnet blanc et blanc bonnet]
- (2) Lima lima lima [Au long aller la lime mange le fer]
- (3) Llama llama llama [Le feu appelle le feu]

Le rapprochement en (1) des mots *olivo* [olivier] et *aceitun*⁹ (mot inexistant) ne peut s'expliquer que par l'encodage de la forme syllabique *uno* [un] comme forme lexicale ; la récurrence purement formelle du signifiant *uno* a été forcée dans *aceit/uno*. C'est donc **dans les limites d'un énoncé clos que se produit l'ajustement lexical, faisant émerger une sémantique interne d'éléments micro-syntaxiques** qui changent leur catégorie grammaticale pour une autre, d'ordre sémiologique

Comme le montrent aussi les exemples (2) et (3), les rapports syntaxiques des mots dans l'énoncé mettent en place une «période intonative» **signifiante** car autrement les proverbes seraient ininterprétables. Il s'agit donc d'une clause intonative, codifiée sous forme de PE, qui règle des paramètres morfo-phonologiques :

- énonciatifs (linéarité saussurienne), ancrés dans la syllabation et mesurables en phonétique acoustique⁹;
- de lexicalisation ou voca(bu)lisation ethno-historique (pragmatique par rapport au moment d'énonciation dans l'approche de Coursil 2000);

⁹ Durée (millisecondes), fréquence fondamentale (Hz) et amplitude (Db).

- d'instanciation du sujet énonciateur ou interprète qui produit les *ajustements des notions et des représentations*¹⁰ objectivés par un temps «calendérique», dans la terminologie de Renaud (2005).

Ce cadre interprétatif triple suggère une conscience lexicale qui joue dans l'expression créative et laisse des traces dans le réservoir de langue collective. C'est ce que montrent les proverbes ci-dessus, intraduisibles sinon par des techniques contrastives de fréquences syllabiques.

2.2. L'accent démarcatif français

Dans les expressions proverbiales qui suivent, les signifiants s'expriment dans des récurrences (coup/choc), l'allocution (geste/mouvement) et la valeur pure de la représentation iconique (espace/étendue)¹¹ :

2.2.1. Rapports rythmiques par assonances récurrentes (accent final)

(4) Bonjour lunettes, / adieu filettes

2.2.2. Rapports polaires : opposition forme affirmative vs négative renforcée par l'assonance (accent initial/final)

(5) Pluie du matin / n'arrête pas le pèlerin

(6) On conduit sa nature, / on ne la change pas

(7) Il y a des larmes pour le bonheur, / il (n²-/j//) en/n/a pas pour les grands malheurs.

La reprise par *en* dans (7) renforce la négation par sa place prosodique dans l'énoncé, renforcée par la liaison /n/. Nous comptabilisons donc le locus de la liaison comme un élément catégoriel (un « type » morphologique), comme nous l'indiquons dans l'exemple (8) ci-dessous et dans l'analyse comparative des sous-titres (exemples 23-52).

2.2.3. Rapports de reprise par des éléments isosyllabiques

(8) Quand/t/ on parle du loup, [#] /on/n/ en voit la queue

(9) Mal prie / qui s'oublie

¹⁰ Dans la perspective d'une « linguistique de l'énonciation » (Culioli 1999).

¹¹ Nous mettons à notre avantage, en les adaptant, les valeurs de la « sémiologie des pratiques » de Coursil (2000).

Les éléments isosyllabiques sont des patrons de fréquence interprétés comme signifiants par leur catégorie syntaxique externe (en grammaire traditionnelle ou dans les approches morphologiques fréquentielles) : *on, en, qui, tel*, etc. (pronoms/déictiques/embrayeurs) sont en effet largement utilisés pour renforcer le rythme.

2.2.4. Rapports à base lexicale par opposition/similitude de morphèmes

(10) Quand les gros maigrissent, /les maigres meurent

(11) Qui se ressemble/s'assemble

2.2.5. Rapports à modalité allocutive (interrogatives, impératives, discours direct et indirect)¹²

(12) Qui ose dire: /# « Monsieur l'éléphant a pété ? »

(13) Si ta gueule est de travers, #/n'accuse pas le miroir

Comme le montre le symbole #, aucun élément formel ne semble unir les constituants de (12) et de (13), car on ne peut identifier aucune récurrence ; les éléments *qui* et *si* sont à l'initiale car les rapports qu'ils établissent sont cataphoriques. Il s'agit d'un dialogisme signifiant par composition qui aide à définir la dernière catégorie par défaut.

2.2.6. Rapports endophoriques (ethno-histoire/mémoire phonologique collective)

(14) Il n'est pour voir / que l'œil du maître

(15) Au long aller / la lime mange le fer

(16) À l'œuvre /on connaît l'artisan

Le critère 2.2.6. est fréquent et productif. Il semble bien ancré dans l'ouverture/fermeture énonciative et responsable de figements linguistiques très vivants dans la langue.

Pour terminer cette analyse de proverbes, il faut indiquer que plusieurs critères sont applicables dans une même PE pour rendre les parémies plus expressives. On trouve même des énonciations longues faites de suites de parémies ; en espagnol cette procédure a été largement étudiée dans l'œuvre majeure de Cervantes. Ainsi les

¹² Les modalités impératives, interrogatives ou le discours direct impliquent l'instance allocutaire et sont reconnaissables par la ponctuation à l'écrit, par l'intonation à l'oral. Dans le style indirect libre elles se dégagent de même par la rupture syntaxique dans le discours.

signifiants s'organisent selon les rapports à la fois syntagmatiques et paradigmatiques des séquences médianes et répondant à des critères sémiologiques dans l'acte de verbalisation. Dans cette « sémiologie des pratiques langagières » intervient une sémiotique d'ordre plus général¹³ qui organise la méso-syntaxe, à l'instar des phraséologies, selon les *outils* récurrents d'ordre :

- syllabiques, interprétés comme un *coup* et/ou un *choc* (2.2.1. /2.2.2.) ;
- dialogiques, impliquant un *geste* et/ou un *mouvement* vers l'observateur/interprète (2.2.3./ 2.2.4.) ;
- endophoriques¹⁴, re-présentant un *espace* et/ou une *étendue* dans le système de la langue (2.2.5./2.2.6.).

3. Unités linguistiques au-delà du mot lexical

Nous avons vu comment les unités proverbiales signifient par leur complétude énonciative dans un temps discursif qui répond à trois paramètres. Cette *complétude* (« téléicité », pour emprunter la terminologie de Renaud 2005) implique une structure prosodique repérable par des critères autres que le signifié, car le « mot (lexical) » isolé ne signifie pas :

[...] employé seul, il est sans valeur, et c'est par groupes de mots que nos pensées s'expriment. Nous recevons en effet du fonds commun un certain nombre de groupements pour lequel le sujet parlant n'a pas à fournir l'effort d'assembler les éléments. Il les trouve déjà formés par un long et lent travail qui les a réunis pour donner à l'ensemble une signification unique. (Asharf & Miannay : 1995¹⁵)

Si nous faisons intervenir la grammaire contrastive, nous voyons dans les exemples suivants comment un trait aspectuel interprété dans un sens large (télique) peut être interprété en termes d'ouverture/ fermeture énonciative :

- (17) Appeler un chat, un chat [Al pan pan y al vino vino] (Le pain (c'est du) pain et le vin (du) vin)
- (18) Quand on parle du loup, on en voit la queue [Hablando del rey de Roma, por la puerta asoma] (Parler du roi de Rome et le voir paraître à la porte)

Si la complétude énonciative se fait en (17) par une récurrence sur un complément d'objet (structure SVO--O) en français, il n'en est pas

¹³ Nous pensons aux langues des signes et à d'autres systèmes de communication.

¹⁴ Cata/métaphoriques.

¹⁵ Les auteurs citent F. Brunot, *La pensée et la langue*, 3ème édition, Masson, 1965, p. 3 et 4.

de même en espagnol, où un énoncé averbal est suffisant pour produire le même signifié : le rapprochement entre les deux proverbes se fait par effet de récurrence syllabique.

En (18) le proverbe français et son correspondant espagnol contiennent le verbe *parler* ; mais le français recourt aux deux occurrences de *on* pour renforcer le poids signifiant (récurrence fréquentielle de type 2.2.3), alors que l'espagnol renforce par l'assonance (type 2.2.1.)

La catégorisation grammaticale doit rendre compte de ces faits de langue en assumant et en répertoriant comme formes lexicales les figements linguistiques divers, générés d'ailleurs aussi par modification (dé-figements) dans un but expressif. Leur génération est analysable dans la reconnaissance du domaine où le sujet énonciateur est *pris* dans sa verbalisation et cela même quand il observe et construit l'objet linguistique. Le lexicologue ne peut être que juge et partie dans sa recherche scientifique, qui est, comme l'affirme Dufaye (2009), « une activité de réélaboration permanente du métalinguistique ».

Le terrain de la méso-syntaxe est donc le « trope » où se produit la *lutte de classes des unités linguistiques*, toujours renouvelée et aujourd'hui de plus en plus vivante par l'influence des nouvelles technologies.

3.1. À propos de types et tokens dans l'analyse textuelle : les PE dans les énoncés non parémiques

Une étude discursive *dans une approche fréquentielle* (récurrences) peut être entreprise en termes de *types* (sélection/catégorisation de faits de langue dans un but opératif concret) et de *tokens* (nombre d'occurrences des *types* retenus dans le discours effectif à analyser). Nous avons décrit les valeurs formelles des énoncés et nous nous intéressons maintenant à leur organisation au niveau textuel. Nous allons procéder par la sélection d'un extrait de transcription automatique d'un texte oral, puisque nous avons défini les critères par référence au « réservoir de syntaxe orale » que sont les parémies. La comparaison entre les PE signifiantes du texte transcrit par nos soins et les PE a-signifiantes du sous-titrage déviant tel qu'il se présente sur Internet, nous permet d'étiqueter les positions syllabiques non reconnues par le traducteur/transcripteur virtuel. Une telle procédure interprète, par instances, des segments syllabiques, en comparant le texte oral transcrit manuellement et sa transcription automatique en tenant compte du signifié. Les segments significatifs (PE) sont alors catégorisés comme « types » et les a-signifiants comme « tokens ». Une comparaison entre les occurrences des uns et des autres nous permet d'établir des catégories formelles pour en tirer les

conséquences qui justifieront la pertinence d'un domaine méso-syntaxique dans le cadre de l'analyse textuelle.

3.2. Entre micro- et macro-syntaxe : le niveau de (re)construction du sens

Nous appelons « types » les segments textuels pouvant être décodés par un interprète non virtuel. En effet, une interprétation syntaxique déficiente produit des ambiguïtés et des quiproquos utilisés souvent consciemment pour produire des effets expressifs par le décryptage des normes altérées¹⁶. De la même façon, les troubles du langage peuvent trouver leur explication dans une mauvaise catégorisation des types linguistiques à emmagasiner par la mémoire phonologique. Comme on va le montrer par les déviations du sous-titrage automatique, les *types* catégoriels ne peuvent se définir qu'au niveau de l'énoncé (niveau explicatif) et nécessitent un interprète non virtuel avec une mémoire (ortho-)phonologique de langue(s). Le rôle des bases de données est alors de re-construire une macro-syntaxe pour cribler les *tokens* sur lequel l'interprète virtuel opère. Ce travail préalable commence à donner des résultats dans l'encodage des messages vocaux en téléphonie mobile. Notre proposition est de le faire sur une catégorisation phonologique type (langue + co(n)texte) qui tienne compte des éléments prosodiques de la syllabation standard d'un locuteur standard. Nous cherchons alors des « types » de segments syllabiques par leur nombre de positions phonologiques effectives (locus), dans les différentes occurrences (*tokens*) et les comptabilisons au niveau méso-syntaxique des représentations significantes. Nous présentons un exemple de cette procédure en 4.

4. Repérage de patrons syllabiques en sous-titrage automatique

L'hypothèse du niveau méso-syntaxique dans l'énonciation (cadre des PE ou périodes intonatives significantes) va nous servir à repérer les constantes phonologiques de la verbalisation. En effet, le signal sonore n'est pas discret et c'est le locuteur qui organise son découpage sémiologique par le marquage des silences prosodiques. Ce travail de marquage, fait à la fois par rapport au signifiant et au signifié, suppose une sorte de *entre-les-lignes* (un caché phonologique), interprétable par référence au co(n)texte. Quelques

¹⁶ Voir par exemple Rossi (1979) : *J'en parle/Jean parle ; Je les dessinais / Gelée dessinée ; On s'en dégoute/ On sent des gouttes ; Cette ville est agréable/ Sète-ville est agréable ; L'attentat réussit/La tante a réussi...* ; des exemples de la publicité, des enseignes et des titres de presse, ou encore les pages Internet humoristiques et les tweets.

pistes phonologiques vont nous aider à interpréter ce *caché* phonologique inscrit dans la syllabation. Prenons l'exemple de l'accent lexical espagnol : il suppose un relâchement des consonnes en syllabe ouverte finale qui marque les effets intono-accentuels et le rythme de la langue. Ce n'est pas le cas du français où l'accent démarcatif fait disparaître des consonnes à l'intérieur de l'énoncé ; ces particularités des syllabations impliquent des organisations différentes des séquences médianes et sont indispensables pour l'analyse en termes de fréquences des *tokens* prosodiques et en conséquence des *types* qui vont servir à les étiqueter. Voyons ces différences en espagnol et français.

4.1. Pistes métriques pour le repérage de positions prosodiques « en-caché » : le cas de l'assonance espagnole¹⁷

Les exemples (20) à (22) montrent les caractéristiques rythmiques de l'espagnol, en termes d'éléments « suprasegmentaux »¹⁸. Dans l'approche formelle des parémies suivantes nous constatons des patrons accentuels (rime riche/pauvre) inusuels : les assonances se basent parfois sur les voyelles entourant une **consonne qui reste invisible à la métrique** à cause de :

- *l'allongement de la voyelle tonique*¹⁹ :

(19) caaja/cercaana ; espino/cuclillo ; hormiguero/duelo ; odrero/toledo ; honra(d)lo/reputa(d)lo²⁰ ; llega/era ; arrima/cobija ; madruga/ayuda ; cuna/dura ; Pedraza /masa ; lobo / lobo.

¹⁷ Les règles de la métrique de l'espagnol décrivent une rime par assonance qui parfois tient compte seulement des éléments vocaliques (voir Bloise Campoy 1946 pour des explications éclairantes).

¹⁸ Les traits suprasegmentaux, de type phonologique (liaison, enchaînement, accent de groupe, allophones...), ne sont souvent pas reproduits par la graphie.

¹⁹ Exemples de Hernando (2010 : 50, 54, 66, 67, 76) : *Cuando el gato se lava la cara, lluvia cercana* [Quand le chat se passe la patte sur la tête, bientôt il y aura tempête] ; *Cuando nace la flor en el espino, canta el cuclillo* [À la mi-mars le coucou est dans l'épinard] ; *A quien destruye un hormiguero, le vendrá duelo* [Qui tue le goéland, la mort l'attend] ; *Soplará el odrero y alborotarse ha Toledo* [Le tanneur gourde à la bouche fait devenir Tolède farouche] ; *Tanto el vencedor es honrado cuanto más el vencido es reputado* [Le vainqueur est toujours loué sur la réputation du vaincu] ; *Cuando junio llega, afila la hoz y prepara la era* [Juin est arrivé, aigüise ta faux et prépare ton pré] ; *El que a buen árbol se arrima buena sombra le cobija* [Pour t'abriter, cherche le bon arbre] ; *A quien madruga, Dios le ayuda* [De bonne heure, tout est l'heur] ; *Lo que se aprende en la cuna siempre dura* [Ce qu'au berceau fut un jour, tu garderas pour toujours] ; *La rosca de Pedraza, gran agujero y poca masa* [La couronne de Pedraza, un gros trou pour peu de pâte] ; *En viendo la oveja al lobo, se queda sin sebo el lomo* [Mouton, le loup approche à grands pas et ton dos n'a plus de gras].

²⁰ Les terminaisons en *-ado* voient souvent disparaître le *-d-* intervocalique. Ce phénomène montre que les traits syntaxiques émergent de la verbalisation et dépendent de la syllabation effective ; nous y revenons.

Ces exemples montrent une rime en position pénultième qui répond bien aux paramètres phonologiques internes de syllabation de l'espagnol ; les consonnes n'affectent pas l'effet de répétition (coup) du *discours répété*²¹ qui implique une complétude énonciative.

- la *diphthongaison agissant sur la position tonique*²², même dans le cas du yod :

(20) huevo/fuero ; mayea/marea ; veas/deseas ; suegra/nuera ; vio/yo.

- l'*accent final peu fréquent en espagnol*²³ :

(21) pelar/remojar ; corral/temporal ; Blas/verás ; disimular/lograr ; caudal/ahorrar²⁴.

Ces constantes rythmiques espagnoles mettent en évidence les patrons accentuels différents du français (accent lexical *vs* accent démarcatif). En effet, si les rimes espagnoles concernent bien plus les positions pénultièmes que les positions finales (exemples 20 et 21 *vs* 22), cela a certes des conséquences dans la **grammaticalisation des périodes intonatives** que nous interprétons en termes d'ouverture ou fermeture énonciative définissable en termes phonologiques. En français c'est l'accent démarcatif qui va marquer la fermeture énonciative, le trait de télicité (complétude effective que seul le sujet d'énonciation peut anticiper). Ainsi, des constantes prosodiques produisent des constructions idiosyncrasiques dans les différentes langues pouvant être engrangées dans la mémoire phonologique pour rendre la cohérence textuelle. Nous considérerons alors que les marqueurs discursifs sont de type pragmatique, étant déterminés par les groupes prosodiques qui composent les discours. Le domaine de la méso-syntaxe à un corrélat dans la macro-syntaxe qui est le domaine des représentations où l'idéologie s'exprime dans un métalangage à rebours de l'expression.

²¹ [Discurso repetido] Terme emprunté à Hernando (2010).

²² Hernando (2010 : 54, 67, 77, 98) ne propose pas de correspondances dans ces cas : *No es por el huevo, sino por el fuero* [C'est le for et non pas l'oeuf] ; *Cuando marzo mayea, mayo marcea* [En mars comme en mars, en mai comme en mai] ; *Pues mi mal deseas, antes cieguas que lo veas* [Puisque tu cherches mon tort, aveugle je te veux plutôt que voyant] ; *Acuérdate suegra que fuiste nuera* [Rappelle-toi marâtre que tu étais bru] ; *Allí estaba quien lo vio, pero no era yo* [Un certain était là pour le voir, mais pas moi pour apercevoir].

²³ (*Ibid.*: 55, 66, 78, 95) : *Cuando las barbas de tu vecino veas pelar, pon las tuyas a remojar* [Si ton voisin se fait raser, ta barbe est prête à mouiller] ; *Sal y salud y cuidado y temporal hunchen el corral* [Sel, santé, soins et averse dans ta basse-cour à verse] ; *Por San Blas, la cigüena verás* [À la Saint-Blaise, les cigognes se plaisent] ; *Disimular buen medio para lograr* [Les bonnes apparences sont toujours les plus payantes] ; *No hay caudal como el ahorrar* [Qui bien épargne, gros y gagne].

²⁴ La confusion des liquides *l* et *r*, comme dans *quiero/abuelo* se produit souvent dans certaines variantes d'espagnol.

4.2. Fonction des phonèmes fluctuants dans les PE et détermination du niveau méso-syntaxique en français

Une étude exhaustive des contextes de liaison en cours, en termes de *types* et *tokens*, dans le cadre du projet PFC (Barreca : 2015)²⁵, permettra de montrer que la prosodie est analysable dans la perspective des valeurs intono-accentuelles fréquentielles significatives. Dans la langue en général, les constructions de tout type (dé/figements) vont au-delà des contextes syllabiques déterminés de façon *illusoire* (Laks 2005) par l'orthographe.

La liaison est un témoin évident de la souplesse caractéristique de la langue française et des phénomènes prosodiques qui se produisent en général dans toutes les langues pour pouvoir les enregistrer dans la mémoire phonologique :

la liaison est un phénomène hétérogène et composite mettant en jeu un petit nombre de formes figées construites par un processus mémoriel stochastique conduit par l'usage » [mais] il faut également postuler un processus plus marginal qui [...] efface les frontières de mots et donne au groupe prosodique une place centrale. (Calderone & Laks 2014 : 74)

La mémoire phonologique collective, en effet, contient un tel nombre de formes figées par le *processus mémoriel*, qu'elle semble inabordable ; l'orthographe joue ainsi un rôle important dans la mémorisation, car elle présente à l'écrit des suites prosodiques, des limites entre les mots qui sont le « support de l'enchaînement prosodique » : « par figement ou effacement des limites de mots, ou par liaison productive entre mots orthographiquement séparés, ils [les deux processus] expriment le même génie propre de la langue : rendre le français oral profondément lié (*ibid.* : 84).

Nous allons voir dans un extrait de langue orale transcrite manuellement et automatiquement comment se manifestent les deux processus dont parlent Laks et Calderone. Le texte transcrit automatiquement reste souvent *profondément lié* et le transcripteur manuel, nous en l'occurrence, doit chercher la trace des clauses prosodiques qui l'organisent à l'écrit. Nous déterminons donc les *locus* des étapes de mémoire phonologique dont un interprète « humain » a besoin pour emmagasiner les significatifs. Par défaut, les énoncés a-significatifs de la transcription de la traduction automatique vont nous servir de base comparative.

Notre procédure est la suivante :

- a) Nous présentons d'abord une notation de type phonologique pour signaler les sous-titres biaisés par rapport à la traduction manuelle proposée ;

²⁵ Communication orale (séminaire sur la liaison du laboratoire Modyco à Paris X, avril 2015).

- b) Nous traitons après comme « types » les énoncés signifiants et comme « tokens » les déviations constatées ;
- c) Nous étiquetons les *tokens* en essayant de ne pas recourir à des catégories grammaticales qui ne peuvent pas décrire des chaînes syllabiques atypiques ;
- d) Nous procédons à un comptage et tirons les conséquences de l'analyse.

Dans cette procédure *down-top*, pour déterminer les « types », il est nécessaire de réaliser tout le travail d'annotation préalablement. On comprend alors pourquoi la détermination du niveau méso-syntaxique d'interprétation des signifiants et le concept de PE, qui lui est associé, sont indispensables dans notre approche. Le travail sur la classification syllabique des parémies a été donc une condition *sine qua non*. Il faut préciser aussi que, comme conséquence de ce choix méthodologique, la détermination des « types » et leur étiquetage se précise davantage au fur et à mesure de l'analyse ; les étiquettes catégorielle que nous présentons dans la suite sont donc chronologiquement postérieures à l'annotation complète du corpus.

4.2.1. « Types catégoriels » énonciatifs :

- ✓ Cluster syllabique (unité lexicale nucléaire),
- ✓ Cluster syllabique étendu (unité lexicale faisant partie d'une PE),
- ✓ Éléments ortho-phonologiques (positions syntaxiques graphiques incohérentes),
- ✓ Fermeture syllabique a-signifiante (PE déviante),
- ✓ Forme dérivée (lemme en contexte syntaxique énonciatif).

4.2.2. Symboles utilisés pour l'annotation des PE

⇒ introduit le texte sous-titré automatiquement (sous-titres)
xxx xxx xx souligne le sous-titrage déviant (transcription manuelle)

xx¹xxx² xx³... xxx⁴ PE déviantes + nombre de syllabes (sous-titres)

° phénomènes de dissimilation de diphtongues (transcription manuelle)

x-x phénomènes d'assimilation syllabique (transcription manuelle)

pauses prosodiques/début d'énoncé

! non correspondance ortho-phonologique (sous-titres)

ç^x / ?^x syllabes manquantes / syllabes ajoutées (sous-titres)

/x/ liaison

La vidéo dont nous retranscrivons les sous-titres est disponible sur Internet ; il s'agit d'un document sur la médecine traditionnelle chinoise dont nous avons sélectionné deux extraits, le début (exemples 22 à 48), et un extrait vers la fin (exemples 49 à 52). Chaque exemple est retranscrit manuellement, directement sur le texte entendu, et accompagne le sous-titrage automatique copié sur Internet. Un astérisque explicite la terminologie employée pour étiqueter les 5 « types » catégoriels au cours de l'analyse. Le but final de l'analyse est de trouver des pistes pour implémenter les systèmes de traduction automatique avec des « macros » et de justifier la pertinence des PE et du niveau méso-syntaxique dans l'analyse textuelle.

4.3. Texte oral annoté : sous-titrage nternet/transcription manuelle

LA MÉDECINE TRADITIONNELLE CHINOISE

<https://www.youtube.com/watch?v=-0cgREA2io8>

Début du document :

- (22) #l'acupuncture et la moxibustion de la médecine traditionnelle chinoise ⇒ l'acupuncture et la moxibustion de la médecine traditionnelle chinoise
- (23) #sont/**t/** une connaissance et une pratique traditionnelle concernant la vie-et la maladie
 ⇒ tou¹s/**s/ont**² une connaissance et une pratique traditionnelle
con¹cern²e mon³ /n/a⁴vis⁵la⁶ ma⁷la⁸die¹⁰
 (23.1) Token : forme dérivée
 (23.2) Token : forme dérivée

Ce premier « type », **forme dérivée**, qui apparaît dans le « token » déviant correspond à un lemme à morphologie signifiante. La confusion entre les valeurs prosodiques et syntaxiques des consonnes de liaisons /t/ et /s/ provoquent une « grammaticalisation » déviante sujet+verbe dans le sous-titrage (23.1). Dans (23.2) les nasales sont déplacées et confondues (concernant/mon) et donnent lieu à un signifié bizarre.

Comme on peut le voir, *les positions syntaxiques/prosodiques n'ont pas été reconnues par la traduction automatique.*

- (24) #formées sur le sol de l'ancienne culture chinoise
 ⇒ formé ! sur le sol de l'ancienne culture Chinoise
 Token : éléments ortho-phonologiques

Le deuxième « type » que nous annotons, **éléments ortho-**

phonologiques, répond à une *incohérence orthographique dans la correspondance sens/graphie*.

- (25) #elles sont basées sur le concept /t/holistique d'unité entre l'homme et la nature
 ⇒ #le¹ so²lo³ et⁴ sur le concept po¹li²ti³que⁴ d'une/n/3 unité entre l'homme et la nature
 (25.1) Token : forme dérivée
 (25.2) Token : fermeture énonciative a-signifiante

Le troisième « type », **fermeture énonciative a-signifiante**, correspond à une partie de PE non reconnue. Les instances de mots non reconnus deviennent *des entités repérables par leur position syntagmatique dans le groupe énonciatif (PE) fermé qui contient des segments phonétiques communs aux deux séquences*.

- (26) #guidés par la théorie des méridiens et des points d'acupuncture
 ⇒ guidés par la théorie des mu¹su²mans³ les⁴ tit⁵res⁶ les⁷ plus⁸
po⁹pu¹⁰ laires¹¹
 Token : cluster syllabique étendu
- (27) #utilisant des aiguilles et # des moxins #comme outils et matériaux principaux
 ⇒ utilisant le¹s é²lus 3de⁴puis⁵ 200x65⁷ #comme outils et matériaux principaux
 Token : cluster syllabique étendu
- (28) #par la poncture# ou le réchauffement de parties spécifiques du corps
 ⇒ #dans¹ la² con³fu⁴sion⁵ ou le réchauffement de partie 1
spécifique 1 du corps
 (28.1) Token : cluster syllabique étendu
 (28.2) Token : éléments ortho-phonologiques

Le quatrième « type », **cluster syllabique étendu**, correspond à une instance *bien construite au niveau de la méso-syntaxe mais incohérente par rapport au cadre macro-syntaxique*. Ce genre de « tokens » pourrait aider à déterminer des contextes préalables à la traduction automatique. Le terme « étendu » dans ce nouveau « type » fait référence à sa position par rapport à la séquence modèle marquée d'une ouverture /fermeture énonciative (#xxx#), il est souvent difficile de déterminer les correspondances biaisés sauf par la détermination totale de la période intonative. De là la pertinence du niveau méso-syntaxique d'analyse, qui offre de nouvelles voies à la catégorisation grammaticale contrastive.

- (29) #l'acupuncture-et la moxibustion# permettent-de-traiter les douleurs# soigner les maladies-et promouvoir la santé

⇒ le¹cul²tu³rel⁴le⁵si⁶li⁷ci⁸um⁹##les¹li²fes³ty⁴le⁵ç¹ç²
²lors⁸##que¹les²na³zis⁴ç¹ç²ç³ç⁴ç⁵voir la santé

(29.1) Token : cluster syllabique étendu

(29.2) Token : forme dérivée

(29.3) Token : fermeture énonciative a-signifiante

(30) #le concept/t/ holistique d'unité entre l'homme et la nature

⇒ #le concept [les¹ mous²tiques³] d'unité entre l'homme et la nature

Token : cluster syllabique étendu

(31) provient de la philosophie chinoise ancienne

⇒ ç¹ç²ç³²⁶ la philosophie chinoise ancienne

Token : forme dérivée

(32) #les anciens pensaient que l'homme #se trouvait entre le ciel et la terre

⇒ [les anciens penser l-que l'homme]# se trouvait entre le ciel et la terre

Token : éléments ortho-phonologiques

(33) #dans le monde naturel #le yin représente la terre la lune# et la nuit

⇒ dans le monde ac¹tu²el³ # nous⁴ ne⁵ pro⁶po⁷sons⁸ que⁹ la
¹⁰plai¹¹gnante¹²##[i] a¹³nu¹⁴i¹⁵

(33.1) Token : cluster syllabique étendu

(33.2) Token : forme dérivée

(33.3) Token : cluster syllabique étendu

(34) #tandis que le yang représente le ciel, le soleil et le jour

⇒ tandis que le yang représente le ciel [bleu ?¹] le soleil ç¹ç²ç³
³²⁷

(34.1) Token : cluster syllabique

(34.2) Token : cluster syllabique étendu

Le cinquième « type », **cluster syllabique**, correspond aux rares occasions où des « mots lexicaux » sont mal interprétés ou inventés par l'interprète virtuel, comme c'est le cas dans cette occurrence. Les signes + ou – qui suivent les signes ? ou ç indiquent le nombre de positions phonologiques manquantes.

(35) #à l'intérieur du corps humain le mouvement vers le haut # représente le y°ang

⇒ à l'intérieur du corps humain le mouvement vers le haut
 nous¹ pro²po³sons⁴ que⁵ [(es)-ans⁶]

Token : forme dérivée

²⁶ Ce symbole indique une suite syllabique non identifiée ni interfaçable.

²⁷ Le système de traduction interprète les syllabes mais pas les positions syntaxiques, ce qui montre bien que le sujet énonciateur « construit de la syntaxe » dans l'énoncé à un niveau intermédiaire entre la micro- et la macro-syntaxe.

- (36) alors que la relative stabilité vers le bas représente le vin
 ⇒ alors que la relative stabilité vers le bas ?+1que représente
pour ?+2 vous ?+3
 Token : forme dérivée
- (37) et le yang régissent la nature ainsi que l'intérieur du corps humain#
 ⇒# et le yang régissent la nature ainsi que l'intérieur du corps humain
- (38) #si l'équilibre# yin/yang# entre la nature et l'homme
 ⇒ qui¹ l'équilibré² d(e)³ vie⁴ par⁵ con⁴tre la nature et l'homme
 Token : fermeture énonciative a-signifiante
- (39) ou de l'homme lui-même est brisé # la maladie survient
 ⇒ ou tre¹ le Yé²men³ il⁴ faut⁵ es⁶sa⁷yer⁸# la maladie survient
 Token : fermeture énonciative a-signifiante
- (40) l'acupuncture et la moxibustion considèrent que
 ⇒ l'acupuncture et la moxibustion considère que
 Token : éléments d'ortho-phonologie
- (41) l'âme humaine est/t/ un petit univers# il y a douze mois dans une/n/ année
 ⇒ lui-m¹ême² était³/t/ un petit univers # il y a douze mois dans une/n/ unique
 (41.1) Token : forme dérivée
 (41.2) Token : cluster syllabique
- (42) #correspondant aux douze méridiens du corps humain
 ⇒ correspondant en¹ tout² cinq³ in⁴diens⁵ corps humain
 Token : cluster syllabique étendu
- (43) #le corps humain a 365 points d'acupuncture # correspondant aux 365 jours
 ⇒ c⁻¹c⁻²c⁻³c⁻⁴1¹365 ré¹pon¹c⁻¹du¹ # correspondant de 365 jours
 (43.1) Token : forme dérivée
 (43.2) Token : cluster syllabique étendu
- (44) #les méridiens et les collatéraux
 ⇒ les mairies⁵ et les collatéraux
 Token : cluster syllabique
- (45) #sont les passages de communication interne/externe # et haut/bas du corps
 ⇒ c¹c²c³ passages de communication interne/ externe il¹
com²pare³le⁴ corps
 (45.1) Token : forme dérivée
 (45.2) Token : fermeture énonciative a-significative
- (46) #ce sont les théories des méridiens et des collatéraux
 ⇒ qui¹ sont les théories des méridiens et collatéraux
 Token : forme dérivée

- (47) ainsi que des points d'acupuncture des thérapies de l'acupuncture #et de la moxibustion de la médecine traditionnelle chinoise
 ⇒ ainsi que les points d'acupuncture de lit¹té²ra³ture⁴ lors⁵ de⁶
la⁷ 4⁸ème⁹ # il¹⁰ an¹¹nonce¹² é¹³di¹⁴tion¹⁵ de la médecine traditionnelle chinoise
 Token : fermeture énonciative a-signifiante
- (48) ces théories expérimentent la relation entre le ciel, la terre et l'homme [02:15] ⇒ ces théories est¹-ce²qu'^une relation entre le ciel la terre [02 :16]
 Token : forme dérivée

Deuxième extrait :

- (49) la matière première utilisée pour la moxibustion est l'armo^oise#
 ⇒ la matière première utilisée pour la mo¹ti²va³tion⁴ est au¹
moins² cer³tain⁴
 (49.1) Token : cluster syllabique
 (49.2) Token : cluster syllabique
- (50) #elle pousse dans la plupart des régions de Chine
 ⇒ elle pousse dans la plupart des régions du¹ Chi²li³
 Token : cluster syllabique étendu
- (51) #c'est une plante avec un parfum particulier# depuis l'antiquité les chinois croient qu'elle /1/ a une capacité de dissiper les facteurs pathogènes#
 ⇒ c'est une plante avec un parfum particulier #depuis l'antiquité les chinois croient qu'elle a la capacité de dissiper les facteurs pathogènes
 Token : cluster syllabique étendu
- (52) #les cônes et les bâtonnets de moxins sont faits avec des feuilles d'armoise desséchées et# transformées en filaments
 ⇒ Les cônes et les bâtonnets de Montréal sont plus¹
nom¹breuses³plus⁴ d'un⁵ mois⁶ se⁷ suivent⁸il⁹ # transforme!¹ en
fil²m³
 (52.1) Token : cluster syllabique
 (52.2) Token : forme dérivée
 (52.3) Token : forme dérivée

Nous comptabilisons maintenant les *types* et les *tokens* obtenus.

4.4. Résultats

Dans le tableau des résultats ci-dessous nous rapprochons les PE relevées dans l'analyse des sous-titres déviants de celles de la classification syllabique des parémies. Dans les deux cas, le

cadre de référence est l'énoncé, mais dans le cas de la traduction automatique, il a été amputé de son contexte macro-syntaxique sémantisant, car celui-ci nécessite pour être interprété une configuration complexe tenant compte de constantes phonologiques externes à l'énoncé, comme le montre le travail contrastif réalisé dont nous comptabilisons les occurrences des PE déviantes et leur classification par types catégoriels :

Catégorisation des PE déviantes : « types »	Exemples : « tokens »	Total
Clusters syllabiques (configuration interne)	35.1 - 42.2 - 45 - 50.1 - 50.2 - 53.1 : 15,9%	6
Clusters syllabiques étendus (configuration externe)	27 - 28 - 29.1 - 30.1 - 31 - 34.1 - 34.3 - 35.2 - 44.2 - 51 - 52 : 24,4%	11
Formes dérivées (configuration interne/ externe)	24.1 - 24.2 - 26.1 - 27 - 30.2 - 32 - 34.2 - 36 - 37 - 42.1 - 43 - 44.1 - 46.1 - 47 - 49 - 53.2 - 53.3 : 38,6%	17
Éléments ortho-phonologiques	25 - 29.2 - 33 - 41 : 0,9%	4
Fermeture énonciative a-significative	26.2 - 30.3 - 39 - 40 - 46.2 - 48 : 15,9%	6
		44

Tableau des rapports énoncés -type /occurrences (tokens) du texte

Ces résultats indiquent que le pourcentage le plus haut de *tokens* sur un total de 44 énoncés déviants correspond respectivement aux *formes dérivées* (38,6%) et aux *clusters syllabiques étendus* (24,4%). Les autres résultats ne semblent pas représentatifs et se révèlent même redondants : les éléments ortho-phonologiques déviants ne sont que la transcription d'un effet de syntaxe orale non reconnu ; les clusters syllabiques et les fermetures énonciatives a-significatives indiquent que les items « lexicaux » ne sont pas reconnus, car un système de traduction automatique ne dispose pas des classifications emmagasinées dans la mémoire phonologique d'un locuteur non virtuel. Nous concluons donc que, comme nous l'attendions, seules deux catégories indissociables rendent compte de la syntaxe discursive orale :

- l'une se définissant par sa syntaxe interne (constitution)/externe (position) : **formes dérivées**,
- l'autre par des éléments externes (position syntagmatique) : **clusters syllabiques étendus**.

Les résultats de notre travail montrent donc qu'un système de traduction automatique ne reconnaît pas le domaine de la méso-syntaxe où s'expriment les constantes phonologiques qui sont à la base de la génération des formes dérivables. Ce niveau d'analyse est alors pertinent pour révéler quel est le rôle du sujet énonciateur ; c'est aussi le domaine des explications métalinguistiques par référence à un macro-cadre d'instanciation discursive. Ce rôle est idéo-logique et sert à construire la cohérence discursive.

Conclusions et perspectives : traitement des icônes idéologiques dans les discours

Nous avons relevé des récurrences dans les éléments syllabiques de segments marqués d'une ouverture/fermeture énonciative (phraséologies) et noté leur absence dans les séquences médianes de la traduction automatique. Le caractère lexical des premières et l'absence de signifiants reconnaissables dans certains segments transcrits dans les deuxièmes, nous ont poussé à analyser les caractéristiques formelles des séquences signifiantes dans le discours d'un locuteur non virtuel, quand celles-ci ne peuvent pas être déterminées par leur constitution récurrente. Nous avons trouvé, dans la sémiologie, une des réponses possibles à la question par une courte incursion dans les principes qui conforment le signe linguistique depuis Saussure : la linéarité et l'arbitraire de sa configuration syllabique. Nous avons découvert dans l'analyse des parémies que des patrons rythmiques divergents, comme le sont ceux de l'espagnol et du français, donnent pourtant lieu à des clusters signifiants de même nature sémiologique (la récurrence d'éléments, le geste intonatif d'ouverture/fermeture intonative anticipant les réactions de l'entendant et les représentations qui s'expriment dans les paramètres temporel et spatial des séquences prosodiques). La notion « d'étendue » dans un espace énonciatif limité (micro-syntaxe) nous a permis de comparer des séquences syllabiques par le caractère phonologique de leur constitution. Des éléments « en caché » viennent alors s'ajouter à la simple interprétation des formes écrites que le « syllabeur » virtuel ne semble pas pouvoir interpréter. En rapprochant l'interprétation manuelle et virtuelle sous la même étiquette métalinguistique par une notation de type métrique nous avons repéré des opérateurs qui correspondent à des instances lexicales modulables dans des clauses prosodiques (PE). Les récurrences de ces instances lexicales à l'intérieur de ces segments ont été traitées en termes de *types* (catégories générales) et de *tokens* (occurrences) à l'instar du traitement informatique de la morphologie. Cette procédure a été possible grâce à une prémisse qui « sémantise » un niveau d'analyse linguistique, la méso-syntaxe rendant compte des représentations

iconiques phonologiques. Ce domaine de catégorisation métalinguistique entre la micro- et la macro-syntaxe est un terrain pour l'analyse des unités lexicales, et en conséquence pour l'élaboration de catégories grammaticales toujours plus fines. L'apprentissage/acquisition des langues, l'analyse textuelle et la traductologie devraient pouvoir en profiter grâce à la création de répertoires lexicaux générés par contraste. L'ouverture de nouvelles voies d'étude par le décodage de la méso-syntaxe, avec les outils de la phonologie et de la sémiologie, se révèle un domaine riche pour la catégorisation et le repérage des unités linguistiques toujours en création. On peut alors parler d'une (poly)phon(ologie de l'énonciation car, comme le dit Batoul (2006 : 10), « parler n'est [...] pas un acte neutre, et s'approprier de manière réfléchi et autonome le discours à travers son énonciation est [...] un acte éminemment politique ».

Références bibliographiques

- Anscombre, J.-C., Ducrot, O. (1997), *L'argumentation dans la langue*, Mardaga, Liège.
- Ashraf, M., Miannay, D. (1995), *Dictionnaire des expressions idiomatiques françaises*, Librairie générale française, Paris.
- Avanzi, M. (2006), « Propositions pour une analyse de la structure interne des périodes narratives », *Cahiers of French Language Studies*, 12/2, p. 6-19.
- Batoul, J.-W. (2006), « Glissement de sources énonciatives et polyphonie », in Batoul, J.-W., *Spectateurs en dialogue*, Presses de l'Ifpo, Damas, p. 215-228.
- Bizzarri, H. O. (2004), *El refranero castellano en la Edad Media*, Laberinto, Madrid.
- Blanche-Benveniste, C. (2010), *Approches de la langue parlée en français*, Paris, Ophrys.
- Bloise Campoy, P. (1946), *Diccionario de la rima*, Aguilar, Madrid.
- Calderone, B. et Laks, B. (2014), « La liaison en français contemporain : approches lexicales et exemplaristes », in Soum-Favaro, Chr., Coquillon, A. et Chevrot, J.-P. (éds) (2014), *La liaison : approches contemporaines*, Peter Lang, Berne, p. 61-85.
- Collet, G. (2000), *Langage et modélisation scientifique*, CNRS, Paris.
- Coursil, J. (2000), *La fonction muette du langage*, Ibis rouge, Guadeloupe.
- Culioli, A. (1999), *Pour une linguistique de l'énonciation*, Ophrys, Paris.
- Desclés, J.-P. (2003), « Représentations cognitives, schèmes prédicatifs et schèmes énonciatifs », in *Parcours énonciatifs et parcours interprétatifs, théories et applications*, Ophrys, Paris, p. 21-46.
- Dufaye, L. (2009), *Théorie des opérations énonciatives et modélisation*, Ophrys, Paris.
- Encrevé, P. (1988), *La liaison avec et sans enchaînement. Phonologie tridimensionnelle et usages du français*, Seuil, Paris.
- Fabre, G. (2015), « Les dissymétries lexicales provoquées par les contacts de langues », *Colloque « La fabrique des mots »*, Centre culturel international de Cerisy, 20-27 juin 2015.

- Fernández-Echevarría, M.-L. (2014a), « Clasificación de unidades fraseológicas francesas según su peculiaridad silábica », *Paremia*, 23, p. 91-99.
- Fernández-Echevarría, M.-L. (2014b), « Unités linguistiques et périodes énonciatives en français/espagnol », *Annales de l'université de Craïova*, 24/1, p. 211-224.
- Fernández-Echevarría, M.-L. (2015), « Criterios para el estudio contrastivo de unidades lingüísticas españolas y francesas a caballo entre coligación y colocación », *Paremia* (sous-presse).
- Forgas i Berdet, E. (1996), *Los ciclos del pan y del vino en las paremias hispanas*, Publicaciones del ministerio de agricultura, pesca y alimentación, Madrid.
- González Rey, M. I. (coord.) (2012), *Unidades fraseológicas y TIC*, Instituto Cervantes, Madrid.
- Haillet, P.P. (2012) *Pour une linguistique des représentations discursives*, Paris, De Boeck
- Hernando, L. A. (2010), *El refrán como unidad lingüística del discurso repetido*, Escolar y Mayo, Madrid.
- Herslund, M. (2012), « A quoi bon le mot ? Réflexions sur sémantique et lexicologie », in Begioni, L. et Bracquenier, Chr. (dir.) *Sémantique et lexicologie des langues d'Europe, théories méthodes et applications*, PUR, Rennes, p. 9-13.
- Kahane, S. et Pietrandrea, P. (2009), « Les parenthétiques comme "Unités Illocutoires Associées" », *Linx*, 61, p. 49-70.
- Lacheret-Dujour, A. et Victorri, B. (2002), « La période intonative comme unité d'analyse pour l'étude du français parlé : modélisation prosodique et enjeux linguistiques », *Verbum*, 24/1-2, p. 55-73.
- Laks, B. (1997), *Phonologie accentuelle : métrique, autosegmentalité et constituance*, CNRS, Paris.
- Laks, B. (2005), « La liaison et l'illusion », *Langages*, 158, p. 101-126.
- Loiseau, S. (2012), « Théories de la fréquence linguistique et interprétations des faits quantitatifs en sémantique », *Actes du 3^{ème} Congrès mondial de linguistique française*, Lyon, 4-7 juillet 2012.
- Maingueneau, D. (2012), *Les phrases sans texte*, Armand Colin, Paris.
- Martin, Ph. (1979), « Une théorie syntaxique de l'accentuation en Français », in Fonagy, I. et Léon, P. (dirs), *L'accent en français contemporain*, Marcel Didier, Ottawa, p. 1-12.
- Perrin, L. (2012), « L'énonciation des proverbes », in Anscombe, J.-C., Darbord, B. et Oddo, A. (dirs), *La parole exemplaire*, Armand Colin, Paris, p. 53-66.
- Pop, L. (2014), « Segmentations linéaires, hiérarchiques et "profondes" », *Studia Universitatis Babeş Bolyai. Philologia*, LVIX/4, p. 51-68.
- Renaud, R. (2005), *Temps, durativité, télélicité*, Peeters, Louvain-Paris.
- Robert, S. (2003), « L'épaisseur du langage et la linéarité de l'énoncé », in Ouattara, A. (éd.) *Parcours énonciatifs et parcours interprétatifs. Théories et applications*, Ophrys, Paris, p. 255-274.
- Rossi, M. (1979), « Le français, langue sans accent ? », in Fonagy, I. et Léon, P. (dirs), *L'accent en français contemporain*, Marcel Didier, Ottawa, p. 13-51.
- Roulet, E. (1995), « Étude des plans d'organisation syntaxique, hiérarchique et référentiel du dialogue : autonomie des interrelations du dialogue », *Cahiers de linguistique française*, 17, p. 122-140.

- Rousseau, A. (2012), « La théorie sémantique de Gottlob Frege et son application à l'évolution du sens », in Begioni, L. et Bracquénier, Chr. (dir.) *Sémantique et lexicologie des langues d'Europe, théories méthodes et applications*, PUR, Rennes, p. 17-33.
- Saussure, F. (1972 [1915]), *Cours de linguistique générale*, Éditions Payot, Paris.
- Schapira, Ch. (2012), « La tautologie dans l'énoncé parémique », in Anscombre, J.-C., Darbord, B. et Oddo, A. (dirs), *La parole exemplaire*, Armand Colin, Paris, p. 21-39.
- Tollis, F. (2008), *Signe mot et locution entre langue et discours. De Gustave Guillaume à ses successeurs*, Lambert Lucas, Limoges.
- Tutin, A. (2013), « La phraséologie transdisciplinaire des écrits scientifiques : des collocations aux routines sémantico-rhétoriques », in Grosman F. et Tutin, A. (éds), *L'écrit scientifique : du lexique au discours*, PUR, Rennes, p. 27- 43.
- Sevilla Muñoz, J., Crida Álvarez, C. (2013), « Las paremias y su clasificación », *Paremia*, 22, p. 105-114.
- Vignaux, G. (2013), *Le discours, acteur du monde*, Ophrys, Paris.

Corpus

- Desalmand, P. et Stalloni, Y. (2014), *Proverbes*, Chêne, Paris.
- Hernando, L. A. (2010), *El refrán como unidad lingüística del discurso repetido*, Escolar y Mayo, Madrid.
- <http://cvc.cervantes.es/lengua/refranero/>
- <https://www.youtube.com/watch?v=-0cgREA2io8>