

Université Paris Ouest, Nanterre – La Défense  
Sciences du langage  
Linguistique générale et française

**Représenter la connaissance en linguistique**  
**Observations sur l'édition de matériaux et sur l'analyse syntaxique**

Nicolas Mazziotta

**Mémoire de synthèse**  
en vue de l'obtention de  
**l'Habilitation à diriger des recherches**

[soutenance 20161209 – version 20161219]

Jury :

Prof. Annie Bertin, Université Paris Ouest, Nanterre – La Défense  
Prof. Sylvain Kahane (référent), Université Paris Ouest, Nanterre – La Défense  
Prof. Alain Polguère (rapporteur), Université de Lorraine  
Prof. Lene Schøsler, Københavns Universitet  
Prof. Achim Stein (rapporteur), Universität Stuttgart  
Prof. Pierre Swiggers, Katholieke Universiteit Leuven/Université de Liège



*À mes filles, Jeanne et Margot*



## Remerciements

Qu'il me soit permis de remercier les personnes suivantes :

- Sylvain Kahane, pour l'ensemble des échanges que nous avons eu et qui ont contribué grandement à mon développement scientifique ces dernières années.
- Mesdames et Messieurs les membres du jury, pour l'intérêt qu'ils portent à mon travail : Annie Bertin, Alain Polguère, Lene Schøsler, Achim Stein, Pierre Swiggers.
- Les autres amies, amis et collègues, qui ont contribué intellectuellement à ce mémoire par des discussions et des correspondances stimulantes ; principalement (mais non exclusivement) : Esther Baiwir, Marie-Guy Boutier, Maria Giulia Dondero, Kim Gerdes, Julie Glikman, Céline Guillot, Benoît Habert, Serge Heiden, Jean-Marie Klinkenberg, Alexei Lavrentiev, Christophe Lejeune, Christiane Marchello-Nizia, Timothy Osborne, Sophie Prévost, Pascale Renders, Marie Steffens.
- Les relecteurs assidus qui ont traqué les erreurs dans ma prose.



# Sommaire

---

<b>0. Introduction</b>	<b>1</b>
0.1. Présentation succincte de mes travaux.....	2
0.2. Inscriptions des connaissances.....	3
0.3. Questions traitées.....	7
<b>1. Lire</b>	<b>9</b>
1.1. Lire les éditions de matériaux.....	9
1.2. Lire les schémas syntaxiques.....	16
1.3. Synthèse.....	21
<b>2. Choisir</b>	<b>23</b>
2.1. Choisir les phénomènes représentés.....	23
2.2. Choisir les analyses explicitées.....	28
2.3. Choisir la structure formelle de l'inscription.....	36
2.4. Synthèse.....	42
<b>3. Découvrir</b>	<b>45</b>
3.1. Explorer les faits.....	46
3.2. Élaborer la théorie.....	50
3.3. Évaluer les structures formelles.....	53
3.4. Synthèse.....	63
<b>4. Numériser</b>	<b>65</b>
4.1. Inscription informatique d'éditions de textes.....	66
4.2. Les inscriptions dans SRCMF.....	73
4.3. Synthèse.....	80
<b>5. Conclusion</b>	<b>83</b>
<b>Références</b>	<b>87</b>
<b>Table des matières</b>	<b>95</b>



# 0. Introduction

---

Ce mémoire livre une analyse réflexive de mes propres recherches, anciennes et présentes, à l'aune de mes conceptions actuelles. J'ai tenté de prendre un peu d'altitude par rapport à mes travaux et d'en dégager la cohérence. Je ne prétends toutefois pas atteindre ici un niveau de généralité suffisant à l'universalisme, ni une analyse fondée sur des outils d'épistémologie générale, laissant ce genre d'approche à des spécialistes plus qualifiés que moi. Mon objectif est principalement de transmettre mes démarches dans la dialectique scientifique où je m'insère quotidiennement. Je parle bien de dialectique en me gardant d'emblée de dire que j'ai « inventé » quoi que ce soit : mes découvertes ne sont que la cristallisation de l'appropriation de tendances actuelles et passées de la linguistique, auxquelles mes lectures, mais aussi les personnes exceptionnelles avec lesquelles j'ai eu la chance de travailler m'ont autorisé l'accès.

Mon identité scientifique se caractérise par une profonde interdisciplinarité. On peut résumer mon profil par une interaction constante entre deux domaines du savoir : 1/ l'analyse et la modélisation syntaxique ; 2/ l'édition de texte ou d'analyses linguistiques (annotations et répertoires). Cette section présente de façon très succincte l'essentiel de mon parcours dans ces deux dimensions. Elles sont unies par un fil rouge épistémologique dont je n'ai pris conscience que très récemment : la question de la *représentation de la connaissance* et celle de la *formalisation des analyses linguistiques* forment la clef de voûte de mes travaux. Qu'il s'agisse d'analyser des structures ou de « donner à lire » un texte ou un ensemble de données, la question fondamentale de la nature de ces données se pose sans cesse : au travers d'un processus subjectif d'identification de la pertinence, le chercheur produit de nouvelles « données » dont la principale caractéristique est la manière dont elles sont représentées.

Pour introduire ce travail de synthèse, je présente tout d'abord les grandes lignes de ma recherche de façon très succincte (→0.1). Je donne ensuite le cadre méthodologique dans lequel je traite ce sujet (→0.2). Enfin, je présente la progression du reste du mémoire, en énumérant les questions auxquelles je vais tenter de répondre en me raccrochant à mon parcours personnel (→0.3).

### **0.1. Présentation succincte de mes travaux**

La pratique de l'analyse syntaxique du français, et plus particulièrement des états anciens de la langue, m'a progressivement amené à la question de la formalisation de l'analyse et à celle de sa diagrammatisation. J'ai soutenu à Liège en 2007 une thèse sous la direction de Marie-Guy Boutier. Ce travail a abouti à la publication du volume *Ponctuation et syntaxe dans la langue française médiévale* dans la collection des *Beihefte zur Zeitschrift für romanische Philologie* (Mazziotta 2009). J'y ai analysé de manière dépendancielle toutes les relations syntaxiques dans un corpus de 148 chartes rédigées en français à Liège au 13<sup>e</sup> s. J'ai ensuite participé activement au projet *Syntactic Reference Corpus of Medieval French* (désormais « SRCMF », ANR-DFG 2009-2012 ; v. Prévost/Stein 2013 et Mazziotta 2012b), qui a consisté en l'annotation exhaustive des relations syntaxiques dans un corpus de français médiéval, et je poursuis encore aujourd'hui son exploitation. Ces travaux m'ont poussé à me poser des questions pratiques comme celle de l'encodage de la coordination dans les corpus (Mazziotta 2013b), ou théoriques, comme celle du statut syntaxique des membres coordonnés (en partic. Mazziotta 2010a ; 2015).

C'est en participant à divers congrès de traitement informatisé du langage et de linguistique théorique (p.ex. Mazziotta 2010a, Mazziotta 2013a) que j'ai rencontré Sylvain Kahane, puis amorcé et développé une collaboration fructueuse avec lui dans deux directions complémentaires. Premièrement, en explorant le fonds Tesnière à la Bibliothèque nationale de France, nous avons étudié les idées de l'auteur et ses schémas syntaxiques, ou *stemmas* (notamment Mazziotta/Kahane 2016 ; soumis ; Kahane/Mazziotta 2015b). Deuxièmement, nous avons travaillé sur les structures formelles dédiées à l'encodage des analyses syntaxiques (Kahane/Mazziotta 2015a ; 2015c).

De ce parcours a émergé une réflexion générale approfondie sur la modélisation et, plus particulièrement, la mise en diagramme des relations syntaxiques.

Parallèlement à mon travail sur la syntaxe, je me suis intéressé à l'édition de matériaux linguistiques. Mon travail de thèse a consisté à mettre en relation les occurrences de signes de ponctuation avec les structures syntaxiques dans des documents médiévaux. Pour rencontrer cet objectif, j'ai constitué l'édition numérique des chartes mentionnées ci-dessus. Mon premier post-doctorat à l'École Normale Supérieure de Lyon (2008) m'a permis d'échanger sur le sujet avec les

membres de l'équipe de la Base de français médiéval (Guillot *et al.* 2007). Depuis mes débuts, j'ai réfléchi au statut des transcriptions et à leur encodage dans des formats numériques (Mazziotta 2001 ; 2004 ; 2006c ; 2008b).

Quand je parle ici d'*édition* numérique, je ne me restreins pas aux éditions de texte. J'entends ici la mise à disposition, sous forme informatique, de matériaux linguistiques bruts ou analysés. La formalisation est la condition *sine qua non* de l'exploitation outillée de ce type de ressources, ce dont je me suis rendu compte dès ma recherche doctorale et que j'ai développé de deux façons : 1/ l'annotation ; 2/ le traitement de répertoires structurés. J'ai ainsi partagé mes pratiques en matière d'encodage d'annotations syntaxiques de l'ancien français avec les membres du projet SRCMF, puis nous les avons appliquées (notamment Mazziotta 2010a ; 2010b). Parallèlement, mon ancrage à Liège m'a permis de travailler sur les structures du *Französisches etymologisches Wörterbuch* (Wartburg 1922-2002 ; désormais « FEW ») et celles de l'*Atlas linguistique de la Wallonie* (v. Counet 1990 pour une présentation ; désormais « ALW ») dans une série de travaux où je suggère des moyens pour construire une édition numérique qui rende compte de la structuration multiple de ces ouvrages (Mazziotta 2008a ; 2011b ; Mazziotta/Renders 2010).

Toutes ces démarches d'édition ont en commun qu'elles nécessitent de faire des choix : Que faut-il représenter et comment ? Que faut-il abandonner ? Or, il est clair que ces questionnements ne sont pas spécifiques au travail d'édition : ils interviennent également de manière naturelle quand on utilise des schémas pour représenter les relations syntaxiques que l'on essaye de modéliser. Tel est le liant fondamental qui unit les domaines de la recherche présentés dans ce mémoire et pour lesquels je proposerai des pistes d'analyse communes.

## 0.2. Inscriptions des connaissances

Représenter des analyses syntaxiques, construire des éditions de texte et réfléchir à la forme de répertoires structurés d'analyses linguistiques sont des activités qui me paraissent pouvoir être éclairées par la *théorie du support et de la connaissance inscrite* (Bachimont 2007). J'emprunte à ce cadre la notion d'*inscription des connaissances*, essentiellement dans un but de clarification de l'exposé, auquel cette section fournit un cadre épistémologique général. Je ne développerai pas le débat sur ces notions (que j'espère ne pas trop dénaturer), considérées ici comme purement opératoires.

La recherche scientifique vise à la construction de connaissances pratiques et théoriques, c'est-à-dire de capacités à, respectivement, exercer une activité sur un objet du monde ou produire des énoncés sur un sujet. Les énoncés ainsi produits sont la matérialisation de la connaissance théorique abstraite, qui s'en trouve *inscrite* sur un support. Par exemple, un livre, les enregistrements d'une base

de données, un disque audio, etc. véhiculent des connaissances, sans nécessairement les formaliser, mais en les rendant perceptibles au travers d'une forme matérielle.

Il n'est pas de connaissance pure en dehors de l'esprit, qui n'a d'autre choix que de produire une inscription pour l'exprimer. Inscription et connaissance se trouvent dès lors unies dans une relation d'interdépendance réciproque : l'inscription matérialise une connaissance et y donne accès via un processus d'interprétation. Par exemple, si j'analyse une structure syntaxique et que je produis un diagramme, mon analyse devient accessible par ce diagramme. Chaque inscription se caractérise par un type de *support*, qui conditionne l'accès à la connaissance qu'elle enregistre par ses propriétés matérielles. Ainsi, une édition de texte imprimée sur une feuille de papier ne pourra comprendre simultanément deux transcriptions concurrentes d'un unique caractère dans le même flux d'écriture. Je distinguerai *support/inscription traditionnel(le)* de *support/inscription numérique*. C'est-à-dire, dans le cadre de ce travail, respectivement, les livres (ainsi que le papier) et les fichiers informatiques.

J'emploierai le terme *inscription* à de nombreuses reprises dans la suite de ce travail. Ce terme général englobe toutes les formes de matérialisation de la connaissance, des moins systématisées aux plus rigoureusement ordonnées. Je parlerai principalement (mais pas exclusivement) de connaissances organisées de manière rationnelle. Pour clarifier l'exposé, je propose d'introduire quelques distinctions personnelles. J'aurai fréquemment recours à la notion d'inscription *formelle*, à celle d'inscription *graphique* et à celle d'inscription *diagrammatique*, qui nécessitent une clarification immédiate.

### 0.2.1. Inscription formelle

La locution *inscription formelle* (ou *modélisation*) fait référence à l'expression formalisée, par l'emploi de formules logiques et mathématiques ou d'un langage contraint dont le vocabulaire et la grammaire sont ou peuvent être exhaustivement connus (comme les règles de réécriture de Chomsky 1957) ; je parle fréquemment d'*encodage formel* pour désigner l'opération de formalisation qui consiste à choisir une structure formelle particulière pour enregistrer une connaissance. La formalisation n'est rien d'autre que l'expression explicite des connaissances par un langage dont le vocabulaire et la grammaire constituent un système fermé<sup>1</sup> (Hofstadter 2000 : 98 sv.). Ainsi, le fait de compter des objets unité par unité peut être encodé par une notation algébrique qui rend compte de

---

1 Selon Bachimont (2010 : § 4.1.2) : « formalisme : une écriture telle que la signification associée aux inscriptions est déterminée en fonction de la forme de ces inscriptions ; c'est donc une écriture formelle. Seront des formalismes les langages de programmation, les langages mathématiques et logiques. Ces langages possèdent une syntaxe, et des règles sémantiques explicitent comment la signification varie quand on fait varier la syntaxe. On distinguera du formalisme l'écriture des langues naturelles : pour ces dernières, la sémantique d'une inscription varie de manière imprédictible et arbitraire vis-à-vis de son écriture syntaxique quand cette dernière subit des modifications (changer un mot dans un texte change le sens, mais on ne peut prédéfinir comment). » Voir notre conception de ce qui « donne du sens » aux formalismes ci-dessous (→1.2.2).

cette opération de manière formalisée (1) ; ou encore, la combinaison des constituants immédiats de la phrase est traditionnellement notée par (2). Ces deux notations ont en commun de reposer sur un ensemble fini de composants dont les règles de combinaison sont également limitées.

$$(1) 1 + 1 + 1 + 1 = 4$$

$$(2) S \rightarrow NP VP$$

C'est bien le caractère fermé des conventions et de l'inventaire qui permettent de déterminer que l'on a affaire à une inscription formelle. Par conséquent, un répertoire téléphonique ou un tableau des départs à la gare peuvent également être considérés comme des inscriptions formelles. On pourrait certes y ajouter des noms pour l'un, des destinations pour l'autre, mais cela se ferait toujours sans sortir des limites du système.

### 0.2.2. Inscription graphique et encyclopédie visuelle

La locution *inscription graphique* fait référence à l'expression graphique de connaissances. Une inscription graphique est un signe visuel (Groupe  $\mu$  1992) ; les propriétés sémiotiques de son signifiant sont spécifiques à son support. Les inscriptions graphiques sont constituées d'*entités* qui peuvent être distinguées les unes des autres et éventuellement sémiotisées. Ainsi, le dessin de la fig. 1 est constitué d'une ellipse et de quatre traits, qui peuvent être vus comme la représentation d'un visage pourvu de deux yeux, un nez et une bouche.

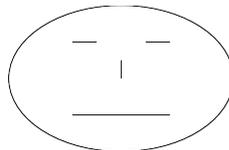


Figure 1 – Inscription graphique d'un visage

À partir de maintenant, je noterai systématiquement les noms des entités graphiques à l'aide de caractères sans empattement<sup>2</sup>. Soit, en m'appuyant sur la forme d'expression de ces entités : trait, ellipse, etc. ; ou, en m'appuyant sur leur valeur pour les distinguer : œil, nez, etc. Notre système cognitif a enregistré au cours de notre vie de nombreuses entités qui nous permettent de catégoriser notre environnement visuel, si bien que les lignes, les flèches et même des objets complexes comme des triangles ou des triskèles peuvent être considérés comme des primitives. Il s'agit de notre *encyclopédie visuelle* (Groupe  $\mu$  1992 : 147-148). La nature spatiale de l'écriture implique qu'elle se manifeste généralement visuellement<sup>3</sup>. Étant donné que je traiterai plutôt de l'écriture comme mode

<sup>2</sup> Le Groupe  $\mu$  (1992) note les entités graphiques entre barres obliques.

<sup>3</sup> V. Harris 1995 : 45 et sv. (résumé dans Mazziotta 2009 : 24) sur la distinction entre *spatial* et *visuel* en ce qui concerne l'écriture.



Dans la fig. 2, on voit en effet que l'inventaire des entités est limité (disques, ellipses, lignes colorées, mots) et que leur mode d'agencement relatif est conventionnel (les disques et les ellipses sont placés sur les lignes et les mots sont placés à proximité de ces entités), de manière à représenter les stations de métro et leur ordre<sup>5</sup>. Comme les parties constitutives de ce système de relations sont identifiables, le diagramme peut être dit *complexe*, d'autant qu'en deçà de deux éléments, la représentation ne peut être conçue comme diagrammatique car il n'y a pas de relation (Chauviré 2008 : 44).

Outre le terme *diagramme*, j'emploierai également les termes de *schéma* et de *schématisation* dans une acception non technique et intuitive, c'est-à-dire sans me prononcer sur la dimension formelle ou non de l'inscription graphique.

### 0.3. Questions traitées

Au travers d'une série de questions, je vais examiner en priorité la manière dont sont inscrites trois types de connaissances que j'ai étudiées dans mes recherches : 1/ des analyses syntaxiques ; 2/ des éditions de textes ; 3/ des répertoires de matériaux linguistiques sous forme d'atlas ou de dictionnaires<sup>6</sup>. Bien que je tenterai de mettre en évidence la similarité des procédés à l'œuvre, je ne fournirai pas une comparaison systématiquement approfondie des différents types d'inscriptions que j'ai utilisés. Je relèguerai certains détails aux notes. Pour les besoins de l'exposé, j'évoquerai parfois des champs que j'ai peu pratiqués, mais j'éviterai de m'appesantir sur des thèmes qui sortent de mes activités effectives de recherche.

Je poserai successivement les questions qui suivent. Jusqu'à la dernière, je traiterai uniquement les inscriptions traditionnelles (c'est-à-dire non informatiques ; →0.2) de ces connaissances.

- Comment **lire** les inscriptions de connaissance linguistique ? (→1) En tant que représentations de la connaissance, les inscriptions sont des signes complexes. Par *lire* ces signes, j'entends comprendre les règles qui président à leur structuration, qui peuvent être plus ou moins cohérentes et plus ou moins explicitées.
- Comment **choisir** la part de la connaissance que l'on désire inscrire ? (→2) Une partie seulement de la connaissance est concrètement représentée dans les inscriptions qui sont censées l'exprimer. Le choix ne porte pas que sur le contenu, mais également sur la forme de leur inscription, qui limite ce qui peut être inscrit.

---

5 La nature graphique du diagramme implique cependant qu'un contenu non formalisé peut s'y combiner sur le même support.

6 J'évoquerai brièvement d'autres inscriptions lors de développements particuliers, comme certains graphiques statistiques et certaines cartes dialectologiques.

- Comment, en se fondant sur les inscriptions existantes, **découvrir** de nouvelles connaissances pour leur donner à leur tour une nouvelle inscription ? (→3) Utiliser les inscriptions de manière dynamique permet de faire émerger la nouveauté. Le processus de découverte peut porter sur la connaissance, mais il peut également porter sur l'inscription elle-même, en particulier au travers de la comparaison entre plusieurs inscriptions scientifiques concurrentes. La comparaison soulève la question des rapports entre inscriptions, c'est-à-dire la question de savoir si une inscription véhicule le même contenu qu'une autre.
- Une fois ces questions traitées pour le support traditionnel, je synthétiserai l'apport du support numérique de manière plus transversale. Que signifie **numériser** la connaissance linguistique ? Voilà le thème qui fera l'objet du dernier chapitre (→4).

# 1. Lire

---

La première étape de la compréhension des inscriptions scientifiques est celle des moyens mis en œuvre pour les rendre compréhensibles, pour permettre de les lire. Cette première étape annonce les deux autres, qui seront traitées dans les chapitres suivants. On ne s'étonnera donc pas de voir évoqués, laissés en suspens, les sujets sur lesquels je me concentrerai par la suite. Je vais traiter la question de l'accès aux inscriptions au travers de ma propre expérience tant à utiliser qu'à concevoir de façon traditionnelle les éditions de textes et de matériaux ainsi que les schémas syntaxiques. Je commencerai par une présentation de la manière dont on lit des textes anciens ou des relevés de matériaux (→1.1). Je parlerai ensuite de l'interprétation des schémas syntaxiques, dont il apparaîtra qu'ils soulèvent en partie les mêmes problèmes (→1.2).

## 1.1. Lire les éditions de matériaux

Par édition, j'entends la mise à disposition à autrui de matériaux linguistiques en vue d'une exploitation. Je parlerai ici exclusivement de l'édition de matériaux manuscrits. Le problème de la lecture des inscriptions est assez simple à expliquer en commençant par évaluer les difficultés de lecture qu'on rencontre face à un manuscrit ou face à une édition de texte ancien (→1.1.1). Sur cette base, on verra que la lecture de répertoires de matériaux relève des mêmes principes (→1.1.2).

### 1.1.1. Lire les textes

Éditer un texte, c'est le comprendre pour le « donner à lire » à un public donné, c'est-à-dire en fournir une autre inscription. Le manuscrit inscrit de manière peu formalisée une connaissance à laquelle l'éditeur doit avoir accès pour projeter sa propre interprétation dans la nouvelle inscription qu'est l'édition de texte. Pour appréhender un manuscrit comme celui dont un extrait est représenté

dans la fig. 3, l'éditeur doit se familiariser avec son écriture, sa langue, le contexte historique de la rédaction du texte, la tradition des copies, les liens intertextuels, etc.

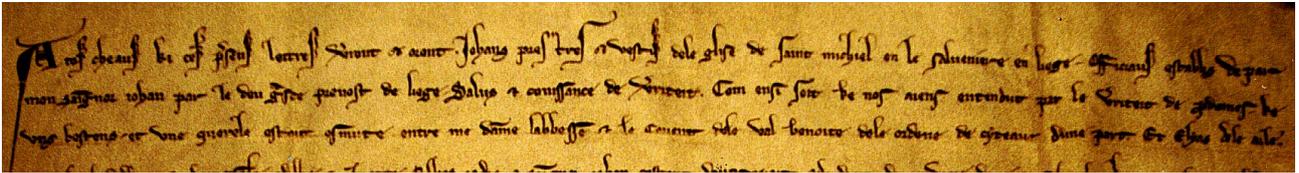


Figure 3 – Fac-similé de Document 1273-05-12 (détail)

Toutes ces compétences font appel à de nombreuses sciences auxiliaires qui définissent la pratique du philologue. Séparer ces aspects est assez artificiel (comprendre le contexte permet de transcrire correctement), mais elle a du sens s'il s'agit de comparer les inscriptions comme je le fais ici. Je ne m'intéresserai donc uniquement à l'activité de déchiffrement. Ce point est à mon avis essentiel pour la suite de l'exposé. Toutefois, comme il me paraît trivial tant qu'on n'entre pas dans les détails approfondis dans les chapitres suivants, je le traiterai rapidement.

Les manuscrits suivent des conventions visuelles qui ne sont évidemment jamais documentées par les scribes et qui relèvent de la tradition censée acquise par les praticiens, ainsi que le public qu'on anticipe comme lecteur<sup>7</sup>. Pour nous, accéder à la forme des lettres est parfois problématique dans la mesure où elles ne sont pas toutes identiques à celles que nous employons actuellement. Nous devons modifier notre encyclopédie visuelle (→0.2.2) de manière suffisamment fiable pour reconnaître les lettres et les mots, qui sont des entités graphiques à proprement parler. Tel est d'ailleurs le but des manuels de paléographie (p.ex. Stiennon 1991). Le travail d'édition des documents étudiés dans mon mémoire de fin d'études (2001) et dans ma thèse (publiée en 2009) m'a amené à faire personnellement l'expérience de ces difficultés. Je livre dans la fig. 4 le début de ma première transcription personnelle, corrigée manuellement par Marie-Guy Boutier ; la transcription repose sur le manuscrit de la fig. 3.

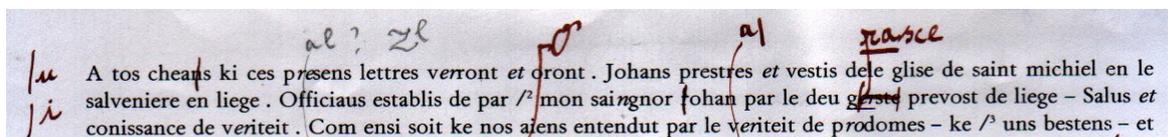


Figure 4 – Transcription de fig. 3 (été 1999)

Faute de montrer l'acquisition progressive de l'encyclopédie visuelle adaptée, les corrections témoignent d'une partie des manques à la compréhension de la lettre du manuscrit. Cette focalisation sur l'activité de déchiffrement ne doit pas nous laisser oublier que cette dernière est dépendante de la compréhension sous tous ses aspects (y compris les dimensions historiques, sociologiques, etc.) : « on reconnaît seulement ce que l'on comprend » (Wilhelm 2015 : 145). Il est d'ailleurs clair que quand je

<sup>7</sup> Parfois, il s'agit littéralement de *tout le monde* (cf. les formules d'adresse des chartes « À tous, présents et à venir, salut » ; v. Guyotjeannin et al. 1993 : 75).

transcrivais \**gerste* pour *grasce*, je ne devais pas comprendre grand-chose au texte. Les mauvaises lectures ne sont pas spécifiques aux textes anciens, bien que les manuscrits les plus modernes nous demandent moins d'efforts, car nous partageons l'essentiel de l'encyclopédie visuelle de leur auteur en matière d'écriture – v. notre édition d'un brouillon de Lucien Tesnière (Mazziotta/Kahane 2016 : § 4).

L'accès aux éditions de texte repose à la fois sur l'encyclopédie du lecteur et sur l'explicitation d'une partie des conventions de transcription. Hormis dans certaines études paléographiques (p.ex. Gilissen 1973), personne ne passe son temps à récapituler la forme de toutes les lettres employées par rapport au manuscrit : seuls les cas problématiques sont traités. Une attention plus grande est généralement portée aux interventions de l'éditeur pour enrichir les matériaux transcrits, qui inscrit la part de sa compréhension qui vient s'*ajouter* à la simple reconnaissance des caractères. Comme il existe des manuels de paléographie, il existe des manuels et des recommandations d'édition (p.ex. Foulet/Speer 1979) qui décrivent les pratiques communes à destination des éditeurs eux-mêmes et de leurs lecteurs, comme l'utilisation d'italiques pour indiquer les résolutions d'abréviations (si l'on désire les marquer<sup>8</sup>) ou l'ajout de majuscules pour identifier les noms propres. Un lecteur formé comprendra immédiatement que les caractères italiques de (3), qui transcrit le début du manuscrit de la fig. 3, sont la figuration d'une abréviation résolue, ce qu'on peut vérifier en se référant à la photographie.

(3) A tos cheaus ki ces *presens* lettres verront *et* oront 'À tous ceux qui verront et qui entendront la présente' (Document 1273-05-12 : 1)

Les éditeurs les plus scrupuleux explicitent même les pratiques supposées partagées, mais de façon attendue, ce sont les conventions les moins habituelles qui sont généralement présentées. Ainsi, déjà dans mon mémoire de fin d'études (2001), je précise que j'ai employé le signe <\_> pour indiquer une séparation de mots<sup>9</sup> introduite par l'éditeur et <-> pour indiquer l'opération inverse. Cette convention est différente de celle de Monfrin (1974), qui emploie <-> dans les deux cas mais marque le second par une note. On retrouve le même genre de démarche d'explicitation de conventions rares dans les éditions effectuées par des romanistes en vue d'études linguistiques (Holtus *et al.* 2003).

Les exemples que je viens de citer concernent l'édition de chartes, qui sont des originaux pour lesquels on ne collationne pas les documents. Je n'ai personnellement jamais élaboré d'édition critique de texte médiéval<sup>10</sup>, mais j'en donnerai toutefois un simple exemple. La fig. 5 reproduit un extrait

8 La question de la partie du manuscrit qui est représentée dans la transcription sera traitée sous →2.1.

9 Le terme *mot* est employé ici avec une acception générale non technique.

10 Pour des raisons de cohérence de l'exposé, je ne développe pas ici l'édition que Sylvain Kahane et moi-même avons effectué au départ de deux brouillons autographes de Lucien Tesnière (Mazziotta/Kahane 2016 : § 4). Il s'agit d'un texte moderne qui a posé très peu de problèmes d'analyse de la tradition.

d'une édition de *La Chanson de Roland* (Segre 2003 : 213), pourvue de son apparat critique (en bas de la page).

Morz est Rollant, Deus en ad l'anme es cels.  
 Li emperere en Rencesvals parvient;  
 Il nen i ad ne veïe ne senter,  
 2400 Ne voide tere nē alne ne plein pied  
 Quē il n'i ait o Franceis o paien.  
 Carles escriet: «U estes vos, bels niés?  
 U est l'arcevesque e li quens Oliver?  
 U est Gerins e sis cumpainz Gerers?  
 405 † U est Otes e li quens Berengers,  
 Ive e Ivorie, que jo aveïe tant chers?  
 Que est devenuz li Guascuinz Engeler,  
 Sansun li dux e Anseïs li fiers?  
 U est Gerard de Russillun li veïlz,  
 2410 Li .XII. per, que jo aveïe laisét?»  
 De ço qui chelt, quant nul n'en respundiét?  
 «Deus! dist li reis, tant me pois esmaïer  
 Que jo ne fui a l'estur cumencer!»

2397 **Morz**: petite lettre initiale, sans indication d'un début de  
 laisse.

2398 Les deux s de **rencesvals** sont grattés.

2400 **alne plein**, avec les dernières lettres de **alne** et le **p** de **plein**  
 retouchés, et, au-dessus de la ligne, entre les deux mots, **lli**,  
 du réviseur. Corr. M etc. etc.

2408 **li bers**, corr. H etc. (d'après V4 2573 P 2728 T 2063 L 1588;  
 et cf. 105, 796; en outre dans la ChR *ber* est très souvent à  
 l'assonance *é*, jamais à l'assonance *ie*).

2412 **ema...er**, avec une ou deux lettres grattées; et au-dessus de  
 e, un s ajouté par le réviseur. Br II, 251, lit *enra...er*, qui cor-  
 responderait à *enragier* de V4 2576 C 4195 V7 T 2068 L 1591.  
 Mais P 2734 a *esmaïer* comme dans la conjecture acceptée  
 ici.

Figure 5 – Fac-similé d'un extrait d'une édition de *La Chanson de Roland* (Segre 2003: 213)

Je ne vais pas expliquer en détails la manière dont il faut lire ce genre de présentation, qui sert à enregistrer les variantes jugées dignes d'intérêt qu'on rencontre dans les différents manuscrits ainsi que les remarques sur leur matérialité. Comme on le voit, chaque élément des notes mentionne une ligne (identifiée par des chiffres) et les mots concernés (en gras). Les deux premiers (2397, 2498) mentionnent des faits matériels (grattages, caractères retouchés), le troisième et le cinquième (2408, 2412) mettent en relation un fait matériel et les variantes auxquelles il correspond. Le quatrième (2400) mentionne des corrections dans d'autres manuscrits. Je voudrais simplement souligner que ce genre d'inscription transmet notamment les liens qu'on a pu *reconstruire* entre les matériaux : une connaissance relationnelle en somme. Il ne s'agit plus d'enregistrer des phénomènes observés, mais du résultat d'une analyse qui n'est pas observable avant d'être inscrite.

Pour comprendre ces éditions, il faut donc non seulement s'appropriier l'encyclopédie relative aux faits graphiques qu'elles transcrivent (les mots), mais également les règles qui président à la disposition du texte. Ce dernier, contraint par le support « papier », doit exprimer des relations entre les leçons – c'est-à-dire les témoignages des différents manuscrits. Il le fait sous une forme arbitraire et

délimitée. Si l'on reprend l'exemple ci-dessus, une partie de l'apparat est réglé par un nombre limité de conventions strictes d'agencement : position du numéro de ligne, emploi de lettres pour identifier les sources, etc. Or, la finitude de l'inventaire des entités et des règles qui président à leur organisation définit ce que j'entends par *inscription formelle* (→0.2.1). Mais il faut remarquer qu'une partie seulement du contenu des notes est formalisé : les rapports entre les manuscrits et leurs leçons sont exprimés en langue naturelle.

### 1.1.2. Lire les répertoires de matériaux linguistiques

Le *Französisches etymologisches Wörterbuch* (Wartburg 1922-2002 ; désormais « FEW ») retrace l'histoire du vocabulaire galloroman ; l'*Atlas linguistique de la Wallonie* (voir la présentation de Counet 1990 ; désormais « ALW ») décrit la phonétique, la morphologie et le lexique des parlers belgoromans. J'ai examiné de près le travail de mes collègues rédacteurs de ces ouvrages pour en comprendre au mieux l'organisation. Pour présenter ces travaux plus loin (→2.2.2. b), il est nécessaire que je dise dès à présent un mot de comment lire ces répertoires.

J'ai constaté que la démarche de déchiffrement des fiches de répertoires linguistiques est tout à fait similaire à celle du déchiffrement des manuscrits. Les deux ouvrages sont fondés sur la récolte de données linguistiques auprès de locuteurs (ALW) ou dans des dictionnaires, des glossaires et des atlas linguistiques (FEW). Ces données sont inscrites sous la forme de fiches ; la fig. 6 (p. 13) présente un échantillon des fiches qui consignent les matériaux de l'ALW<sup>11</sup>.

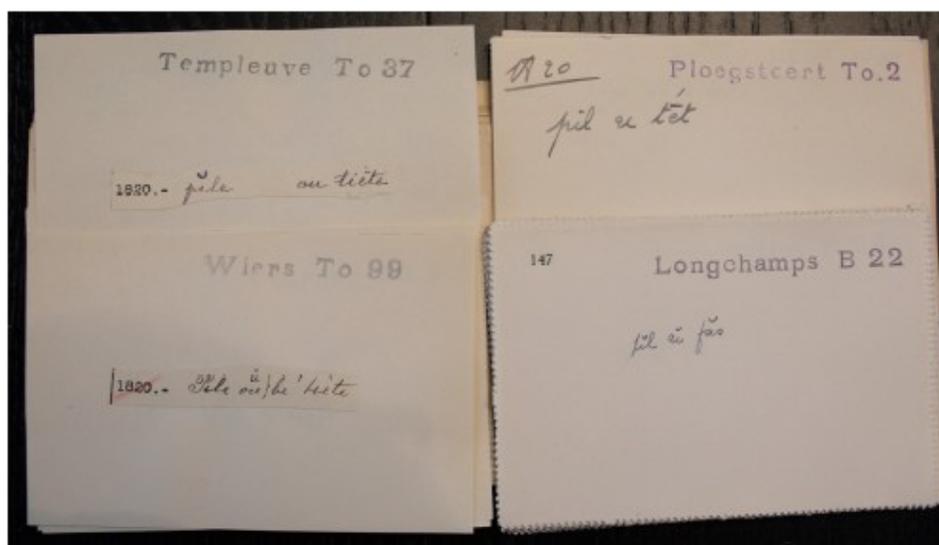


Figure 6 – Exemple de fiches d'enquête de l'ALW (Baiwir/Renders 2013 : 32)

11 Je laisse ici de côté le fait que cette inscription repose déjà sur l'interprétation d'une donnée orale : l'enquêteur a transcrit sur sa fiche ce qu'il entendait. Compte tenu de la manière dont les volumes de l'ALW sont rédigés depuis des années (les rédacteurs ne retournent pas dans les communes, faute de locuteurs), je considère que ces fiches sont la donnée primordiale sur la base de laquelle les notices sont rédigées.

Les notations obéissent à des conventions de notation (alphabet phonétique issu du système de Rousselot et orthographe proposée par Feller 1905 pour transcrire les patois wallons). Il faut bien entendu maîtriser ces conventions, par un processus d'apprentissage similaire à celui évoqué ci-dessus pour lire les écritures anciennes, mais il faut en outre faire face à la variété des notations employées pour représenter les mêmes sons (Baiwir/Renders 2013 : § 2.1 ).

#### 124. ROUGEOLE (carte 50)

Q.G. 725 «la rougeole».

Les mots qui désignent la rougeole s'emploient tous au pluriel: les uns, en effet, réfèrent explicitement aux vésicules rouge rosé qui caractérisent cette maladie; les héritiers de \*RÛBEÛLA, quant à eux, ont connu, en même temps que le passage du singulier au pluriel, une réinterprétation du contenu sémantique du mot: d'abstrait qu'il était à l'origine, il est devenu, tout comme les autres, un mot concret, qui ne désigne plus la maladie elle-même, mais ses symptômes.

Ces continuateurs de \*RÛBEÛLA (A) occupent une aire étendue (délimitée par un trait continu sur la carte), qui ne déborde qu'au sud-ouest l'aire proprement wallonne. Parmi eux, la forme \*wêroûles (A II') de l'extrême est se caractérise par son hétérogénéité (les autres formes étant en r-v-), ce qui lui a valu d'être rattachée à VARIOLA (Haust; Wartburg); elle résulte en réalité du télescopage d'une forme \*rèvioûle, représentant \*RÛBEÛLA, par une forme \*wêroûle, continuant VARIOLA; v. pour plus de détails l'article cité en bibliographie.

Au sud et à l'ouest, \*RÛBEÛLA a fait place à des dérivés de RÛBEUS, respectivement 'rougettes' (B 1) et 'rougeurs' (B 2). \*rouvète (C), attesté en deux points picards, doit peut-être s'analyser comme un dérivé de 'rouvir' 'rougir'.

Quant à \*wahiètes... (D), mot de sens plus général puisqu'il s'emploie pour diverses maladies éruptives (cf. not. 125 VARIOLE, B; not. 127 VARICELLE, C), l'enclave qu'il forme dans l'aire de \*RÛBEÛLA s'explique vraisemblablement par un retrait de ce type lexical, lié à son histoire à l'est du domaine; v. l'article cité ci-dessous.

◆ M.-G. BOUTIER, 'Les noms belgoromans de la rougeole. Une étude de géographie linguistique', DW 21-22, 209-35.

A.<sup>1</sup> [de \*RÛBEÛLA, par voie populaire.] I. 'les' \*rovioules, -oûles, \*rouvi(y)oules, -oûles, -oles, -eules...: rÛvyûl Ni 26 (arch., ou A'), 93; Na 23, 59 (ou -ûl), 112 | -ûl Ni 61 (vieux; v. B 2); Na 1, '13, 22, 44, 49, 59 (ou -ul), 79, '96, 99 (ou ri-), 116, D '5, 7, '9, 38; Ma 1 (sg.), '31, '38, 39, 43, '44; Ne 16 | rÛvyûl Ph 16 | rÛvyûl Ch 33, 61 (importé; v. B 2); Na 107, 109, 130, '136; Ph 33, '43, 53, 61; D 36 | -viyûl Ph 37 (rÛ-), 54 | -ûl Ph 42 | -ûl Na 69, 135; Ar 1; Ma 24, '28, '32, 40 | -vyûl Ch '54; Ph 15 (sg.; ou A'), 45, 69 (ou -viyûl), 79 (id.), 81, 84 | -viyûl Ph 69, 79 | -vyûl Th 64, 72 | -vyûl Th '67 | -vyûl Th 73.

II. a. 'les' \*rèvioules, -oûles, ... \*rê-, \*ré-, \*ri-...: rÛvyûl Ni '25, 80, '97, 98; Na 6; W '36 | -ûl Ni 6, '9; Na 84, 129; Ph '49; D '32 (ou ri-), 34 (id.), '63, 64; W 13; H 8 (ou rê-), '42, 49, 50, 67, 68, '73, '77, '79; L '33, 94, 106, 114 (ou rê-), 116; Ve 1-8, '15, 24-31, 35, 40 (à Francheville); My; Ma 2-9, '10, 12, '13, '15, '18, 19, 20, '49 | lè- W '45 | rÛvyûl Ni 45, '102; Na '20 («raivioules»), 30 | -ûl Ni '29; W '39 | -ûl Ni 85; W 1 (ou B 2), '8, '9, 10, 30, '42, '56, 59, 63; H 2, 8 (ou rê-), 21, '26, 27, '28, 37, '39; L 1-7, '8, 14, '17, 29, '32 (ou rê-), 35-45, '50, 61-87, '100, '108, 113, 114 (ou rê-) | rê- Ni 19; W 3, '32, 35, '52, 66 | rÛvyûl Ni 17 | rÛvyûl Ni '18 | -ûl Na 19 | rÛvyûl, rê- Ni 28; Na '8; L 19, '32 (ou rê-), 101 | rÛvyûl H '35 | rÛvyûl H 1 | ri- Na 99 (ou rê-), 101, '103, '120, 127; D '10, 15, '16, 25, 30, '32 (ou rê-), 34 (id.), 40, '45, 46, '54, '55 (-ûl), '56, 58, 68, '71, 72, 73 (-ûl), '80, 81, 84, '85, '92, 94, 96, '99, '100, 101 (sg.), '102, '103, '104, '108, '109; H 38, '45, 46, 53, 69, '71; Ve 37; Ma 29 (ou \*wêroûles), 35, 36, '37, 46, '47, '48, '50 (id.), '54; Ne '2, 4, '5, '6, '8, 9 | -vyûl Ar 2 || b. 'les' \*rèwioules, \*rê-: rÛvyûl Ni 20 | rê- Ni '30, '51 | rê- Ni 20 (Ard., sg.); W 21.

Figure 7 – Début de la notice « rougeole » dans ALW 15

Les textes de l'ALW et du FEW livrent des matériaux analysés (donc une construction abstraite non observable dans les matériaux initiaux). Ils sont structurés en répertoire de données selon une logique qui n'est pas celle du discours naturel (fig. 7 ; Boutier 1997 : not. 124). Les règles typographiques et la mise en page ont pour but d'inscrire les relations qu'on a pu reconstruire entre les formes. Ainsi chaque notice de l'ALW comprend un tableau qui explique leur histoire au travers de

leur groupement hiérarchisé – ce tableau est accompagné d’une carte si le rédacteur l’a jugé nécessaire (je développe ce point ci-dessous ; →2.2.2. b). De la même manière, un article du FEW synthétise les aboutissements d’un étymon dans la Galloromania. Les notices de l’ALW comme les articles du FEW ont en commun de montrer comment il faut comprendre les rapports entre les formes qui ont été relevées sur les fiches transcrites. Il s’agit à nouveau d’une connaissance relationnelle.

C’est, à l’échelle du mot, le même principe que celui que je viens d’évoquer concernant l’apparat des éditions critiques (même si les buts et les méthodes sont évidemment complètement différents). L’analyse n’est pas linéaire : le tableau de formes de la notice « rougeole » (v. fig. 7, colonne de droite) a une organisation qui peut être représentée de façon arborescente (fig. 8) quand on comprend les conventions de rédaction<sup>12</sup>. Le sens ‘rougeole’ est exprimé par des lexèmes issus de quatre étymons différents (premier cadre en partant du haut). Les formes issues de lat. *\*rubeola* sont héréditaires (A) ou sont des emprunts au français (A’), etc.

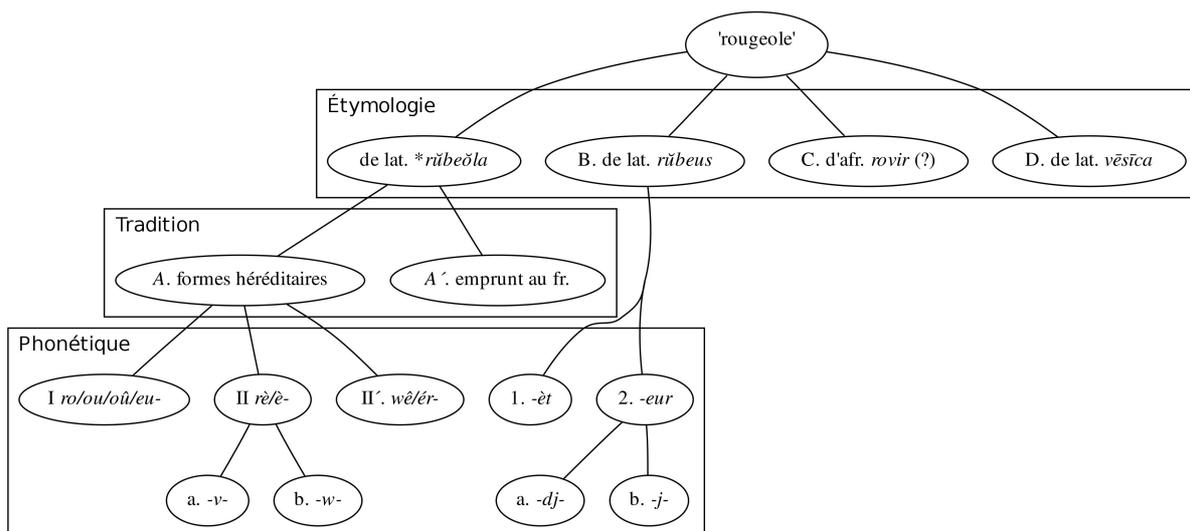


Figure 8 – Représentation arborescente de la structure de la notice « rougeole » de l’ALW (Mazziotta 2011b)

Pour inscrire cette structuration de façon explicite dans le tableau des formes, le rédacteur de la notice se sert de conventions similaires à celles qu’on rencontre dans toutes les autres notices : hiérarchie indiquée par des lettres et des chiffres, succession strictement définie de certaines formes, etc. (Boutier 2008). En somme, il intègre les données et leur interprétation dans un système fermé, que je qualifierai à nouveau de *formel* (→0.2.1). La notice formalise l’analyse proposée par le rédacteur. Cela a notamment pour conséquence qu’il serait possible de l’informatiser.

<sup>12</sup> Cela nous renvoie à la question de la transformation des inscriptions en d’autres inscriptions, qui sera traitée surtout d’un point de vue informatique (→4.2.2. b).

La stabilité de la formalisation est cependant parfois problématique, car il est marquant que les conventions qui président à la lecture de ces textes demeurent parfaitement implicites. Le cas du FEW est emblématique, d'autant que les conventions changent tacitement au fur et à mesure de la rédaction de ce gigantesque ouvrage. D'autres que moi ont davantage travaillé sur ces articles et les conventions propres à la lecture de leur contenu, au point de produire de véritables modes d'emploi (pour le FEW, v. Büchi 1996 ; Renders 2015).

## 1.2. Lire les schémas syntaxiques

Une fois mise de côté la question de la provenance des données analysées (qui relève de la section précédente), il faut constater que l'analyse syntaxique se distingue de l'activité de transcription par le fait qu'elle est tout entière une construction du linguiste. Les caractères transcrits sont des phénomènes, alors que les constructions syntaxiques sont des concepts. La multiplicité des approches n'en est que plus grande. Quand j'ai travaillé sur la coordination (notamment Mazziotta 2012a ; 2013b ; 2015), j'ai voulu comparer la manière dont cette construction était analysée par différents linguistes utilisant des modèles dépendanciels. La fig. 9 donne un aperçu de différentes options descriptives que j'ai rencontrées, sous la forme d'inscriptions graphiques – que je nomme *schémas* dans un premier temps, ne me prononçant pas encore sur leur degré de formalisation (→0.2.3).

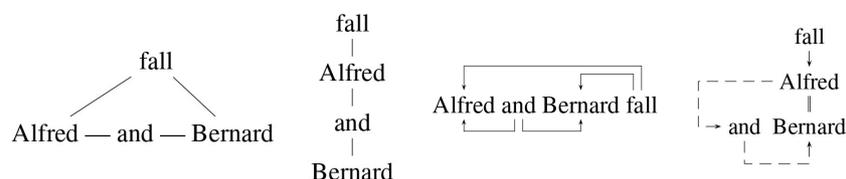


Figure 9 – Schématisations de ang. *Alfred and Bernard fall* ; de gauche à droite : Tesnière, Mel'čuk, Hudson, Kahane

Sans entrer dans les détails de la structuration graphique des schémas repris ici, n'importe quel linguiste un peu familier des modèles dépendanciels a acquis l'encyclopédie visuelle (→0.2.2) nécessaire à leur compréhension partielle. Il voit que les rapports entre les mots, qui sont indiqués par des formes en mention<sup>13</sup>, sont représentés par des entités graphiques comme des traits ou des flèches et peut comprendre comment les groupes de mots sont constitués. On peut aisément tirer des conclusions sur la structure des rapports entre les termes coordonnés et la conjonction de coordination. De ce point de vue, la démarche est tout à fait similaire à celle qui consiste à lire des manuscrits ou des éditions : l'éducation, c'est-à-dire la modification de l'encyclopédie visuelle joue un rôle prépondérant.

Lorsque les schémas présentent des particularités peu communes, il faut se renseigner pour arriver à les comprendre (enrichir notre encyclopédie visuelle). C'est le cas de la place du *et* entre les

<sup>13</sup> Les mots sont des entités graphiques qui représentent les mots du discours par un mécanisme autonymique. Il s'agit de métalangage.

traits horizontaux du stemma de Tesnière (fig. 9) : le *et* n'a pas le même statut que les autres mots car il fonctionne comme un indice grammatical de la relation qui les unit. C'est également le cas des traits doubles et discontinus dans le schéma de Kahane, qui livre lui-même (2012) la clef de lecture de ces traits : les différentes sortes de traits indiquent des relations syntaxiques de types différents et les traits discontinus et doubles ne figurent pas des dépendances classiques.

Tous les linguistes ne prennent pas la peine d'explicitier systématiquement les règles d'interprétation de leurs schémas. Ainsi, celui de la fig. 10 (Lits 2009 : 73) n'est assorti d'aucune notice explicative au sein de l'ouvrage et son interprétation est laissée à l'intuition du lecteur.

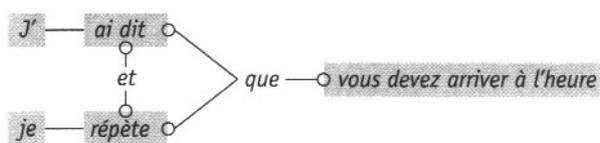


Figure 10 – Schéma syntaxique dans Lits 2009 : 73

D'autres au contraire prennent la peine d'entrer dans le détail de règles d'interprétation, en particulier lorsque ils font figure de pionniers et s'adressent souvent à des enfants. Comme lorsque l'on édite un texte en suivant des conventions personnelles, il serait malaisé de compter principalement sur l'encyclopédie du public. Ce fait m'a particulièrement marqué quand je me suis occupé des travaux du grammairien américain Stephen W. Clark (1847 ; voir Mazziotta, sous presse) et de Lucien Tesnière (1959 ; voir Mazziotta 2014c ; 2016c). J'y viens dans un premier temps (→1.2.1) Je montrerai ensuite l'importance de rendre explicite les critères qui président à l'encodage d'analyses syntaxiques dans des arbres dépendanciels (→1.2.2).

### 1.2.1. Lire les diagrammes des pionniers : Clark (1847) et Tesnière (1959)

Je vais montrer ici que les schémas de deux auteurs qu'on peut considérer comme de véritables précurseurs de l'inscription graphique de la syntaxe méritent le statut de *diagramme*, c'est-à-dire, dans ma conception, de schémas formalisés (→0.2.3).

À ma connaissance, Clark est le premier à avoir employé systématiquement des inscriptions graphiques pour représenter l'analyse syntaxique en disposant les mots selon une configuration qui ne reproduit pas l'ordre linéaire (Brittain 1973 ; Mazziotta, sous presse : § 1). Pour représenter ses analyses, Clark utilise des entités graphiques bulles agrégées les une aux autres (fig. 11).



Figure 11 – Schéma à bulles de *The king of shadows loves a shining mark*

Chaque bulle contient un mot et la configuration des bulles explicite les rapports syntaxiques. La première grammaire de Clark (1847 : 19) contient la fig. 12, qui exprime graphiquement l'inventaire des entités et les règles qui président à leur combinaison. Les moyens graphiques sont limités dans un système pratiquement fermé dont le caractère formel est clair ( $\rightarrow 0.2.1$ )<sup>14</sup>.

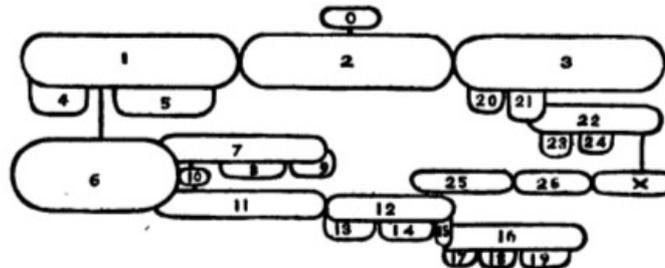


Figure 12 – Inscription graphique des règles de Clark (1847: 19)

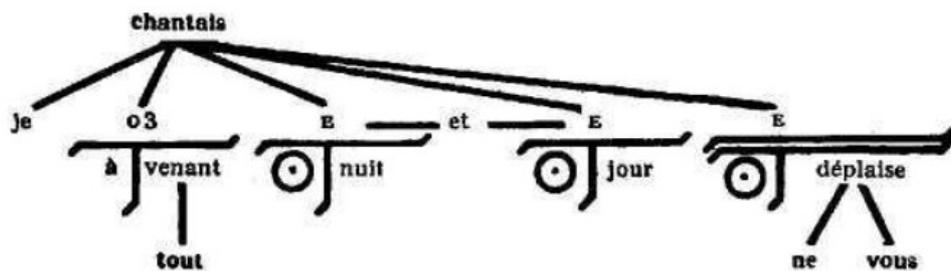
Ce diagramme est assorti d'une légende, où il explique à quelles positions se trouvent les mots analysés, qui sont ainsi disposés selon leur fonction (ang. *office* selon les termes de Clark). Le sujet prend la position étiquetée 1, le verbe (*predicate*) la position 2, l'objet la position 3, leurs différents compléments (appelés *adjuncts*) sont agrégés verticalement, et ainsi de suite. Toutes les règles ne seront pas décrites ici (pour une étude exhaustive, voir Mazziotta, sous presse), mais le fait qu'elles existent est capital. La compréhension des diagrammes n'est pas laissée à l'appréciation intuitive du lecteur : elle est contrainte de façon explicite. Ces règles font le pont entre la théorie grammaticale, qui nous instruit sur la manière de reconnaître un sujet, un complément, etc., et les diagrammes, qui permettent d'exprimer le résultat d'une analyse, c'est-à-dire une connaissance élaborée.

Lire les diagrammes de Clark, c'est d'abord construire une partie de notre encyclopédie visuelle de façon à ce que les entités bulles (et tous leurs adjuvants) acquièrent une signification spécifique à l'analyse syntaxique selon son modèle grammatical. Acquérir la part de l'encyclopédie visuelle qui correspond aux diagrammes de Clark ne suffit pas à les construire : il faut également comprendre son raisonnement linguistique.

Quant à Lucien Tesnière, dès qu'il commence à travailler ses *Éléments de syntaxe structurale* (1959), il a l'impression qu'il est en train de révolutionner la linguistique (Mazziotta/Kahane, soumis : § 4). Il prend soin de documenter la logique de ses diagrammes (qu'il nomme *stemmas*) dès les premières lettres qu'il écrit à ses collègues pour s'entretenir avec eux de ses travaux<sup>15</sup>. La fig. 13 est un exemple de stemma classique.

<sup>14</sup> À de très rares exceptions près, Clark n'introduit pas d'autres entités sans en définir clairement l'emploi.

<sup>15</sup> Voir Mazziotta/Kahane, soumis : § 4.2. Je ne m'étendrai pas ici sur les stemmas primitifs que Tesnière a dessinés entre 1930 et 1934 : ils seront abordés de manière contrastive lorsque je répondrai à la question de la comparaison d'inscriptions concurrentes ( $\rightarrow 3.3.2. b$ ).

Figure 13 – Stemma de *Je chantais à tout venant nuit et jour, ne vous déplaise* (Tesnière 1959: 639)

Les stemmas sont constitués de traits entre des mots et d'un très petit nombre de symboles spéciaux. Tesnière consacre toute une série de paragraphes (1959 : ch. 2 et ch. 3) à la définition des règles fondamentales de construction du stemma pour en donner la signification : le stemma est une représentation de la « hiérarchie des connexions », dont l'expression la plus simple figure dans le stemma de la fig. 14. La position relative des deux mots et le trait qui les lie expriment un rapport de dépendance ou *parle régit Alfred*.

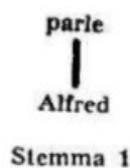


Figure 14 – Stemma simple (Tesnière 1959 : 2.7, stemma 1)

Les règles sont progressivement affinées au fur et à mesure qu'il introduit des notions supplémentaires : *jonction* (c'est-à-dire, à peu près, « coordination » ; Tesnière 1959 : 2<sup>e</sup> partie) et *translation* (mécanisme de changement de classe de mot ; Tesnière 1959 : 3<sup>e</sup> partie). Comme l'inventaire des entités et la manière dont elles se combinent sont fixés<sup>16</sup>, on a également affaire à un système fermé (et donc formel ; →0.2.1). À l'instar de Clark, Tesnière donne les moyens à son lecteur de produire lui-même des stemmas, car la manière dont il exprime les règles établit également un lien explicite entre sa théorie et les diagrammes stemmatiques (Mazziotta 2014c). Ainsi, pour comprendre les stemmas et en générer, il faut notamment savoir que les mots sont associés à une classe qui régit naturellement les mots d'autres classes (le nom régit l'adjectif, le verbe régit le nom et l'adverbe, etc.).

Lire les stemmas de Tesnière actuellement, c'est donc modifier son encyclopédie visuelle de manière à comprendre la signification subtile des rapports représentés par les traits, auxquels la plupart des adultes éduqués assignent une valeur de liaison entre d'autres entités. Cette transformation se fait en rapport avec la théorie, de la même manière qu'on l'a évoqué pour les diagrammes de Clark.

16 Comme Clark, Tesnière déroge assez peu aux règles de ses stemmas sans expliquer comment il procède.

### 1.2.2. Lire les arbres dépendanciel

Les diagrammes de Tesnière et de Clark se sont en fin de compte peu répandus<sup>17</sup>, au sens où ils n'ont été utilisés intensément que par leur inventeur et les élèves qui fréquentaient leur cours. Pour représenter des relations de dépendance, les schémas arborescents et les graphes ont quant à eux connu un essor bien plus grand (c'est d'ailleurs avec les graphes que j'ai le plus travaillé, notamment dans le cadre de SRCMF ; voir notamment →2.2.1 ; →2.2.2. a). Pour beaucoup, le cadre dépendanciel trouve en effet une expression intuitivement satisfaisante au travers d'un encodage qui fait correspondre les arêtes de la structure formelle qu'on nomme *arbre*<sup>18</sup> et les relations de dépendance syntaxique. Graphiquement, ces arêtes sont représentées par des traits ou des flèches (cp. les représentations de Hudson et de Mel'čuk dans la fig. 9). Il y a ainsi correspondance entre arêtes, traits/flèches et relations syntaxiques. Les conventions de base sont très répandues, même en dehors de la linguistique, pour représenter graphiquement des rapports entre objets. De ce fait, elles sont rarement explicitées par les auteurs, qui ne font pas face, comme c'était le cas de pionniers, à des esprits vierges.

Toutefois, même si elles permettent de comprendre l'encodage des constructions syntaxiques, ces simples conventions ne suffisent pas à comprendre comment cet encodage est réalisé en rapport avec la théorie linguistique. En effet, les structures arborescentes sont des formes pures, qui n'ont aucune signification intrinsèque.

Dans une version préparatoire de Kahane/Mazziotta 2015a, j'avais pensé à cet exemple : imaginons qu'on veuille construire une inscription arborescence dépendancielle de la structure syntaxique de (4) (les mots entre parenthèses sont ignorés).

(4) (the) beauty of (the) sun

Il faut définir ce que l'on entend par *dépendance syntaxique*, c'est-à-dire expliciter les règles qui permettent de repérer les dépendances et de les hiérarchiser. Les structures comme les arbres sont contraintes par nature : elles sont constituées d'entités abstraites qui peuvent être rapportées à des axiomes mathématiques qui en délimitent la valeur. Toutefois, si la définition de *dépendance* se limite à « relation orientée entre deux mots », la transposition dans une structure arborescente n'est pas contrainte davantage et les neuf solutions de la fig. 15 sont autant de représentations valables du réseau de dépendances de (4).

17 Je ne parle pas ici de la descendance de ces diagrammes : les deux auteurs ont eu un retentissement phénoménal – les arbres des approches dépendancielles descendent des stemmas de Tesnière et le système Reed/Kellogg est hérité de celui de Clark (Florey 2006).

18 C'est-à-dire, en simplifiant, une structure pourvue d'une racine et où chaque nœud ne peut avoir qu'un nœud parent (voir Mel'čuk/Miličević 2014 : 296).

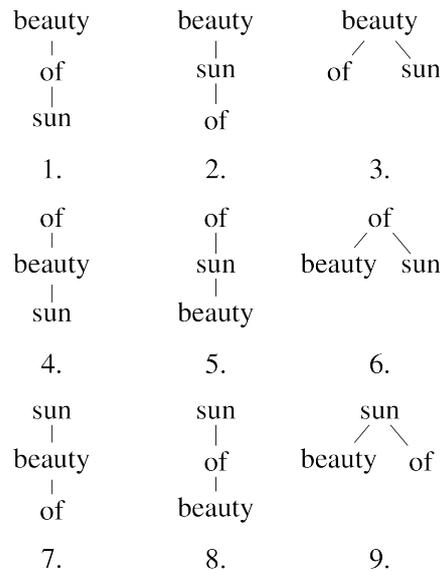


Figure 15 – Options de formalisation dépendancielle du syntagme *(the) beauty of (the) sun*

L'intuition se heurte à la plupart de ces configurations, mais elle ne peut justifier seule le choix de l'une ou l'autre option.

J'ai mentionné que Clark et Tesnière prenaient tous deux la peine d'expliquer comment leurs diagrammes étaient construits en les rapportant à leur théorie linguistique. Il faut procéder de même ici et expliciter comment : 1/ on détecte une dépendance entre deux mots ; 2/ on oriente cette dépendance (je ne traiterai pas ici la classification des dépendances). Ce genre de démarche trouve son expression la plus aboutie chez Mel'čuk (2009 : 25-40), qui pose explicitement le problème. Il mobilise des critères proprement linguistiques pour inscrire les dépendances dans un arbre. Je ne mentionnerai qu'un seul d'entre eux : le critère de la *valence passive*, selon lequel, dans un rapport de dépendance, c'est le mot qui contraint le plus la distribution de la structure qu'il forme avec l'autre mot qui est le *gouverneur*. En suivant Mel'čuk, l'application de ce critère à (4) génère automatiquement le diagramme 1 de la fig. 15, car la distribution de *of sun* est davantage contrainte par *of* que par *sun*. L'application d'autres critères (parfois implicites et plus sémantiques), comme ceux des *Stanford Universal Dependencies* (v. p.ex. de Marneffe/Manning 2008 et notre analyse dans Kahane/Mazziotta 2015a) générerait plutôt le diagramme 2. Le choix des critères détermine donc l'analyse et, partant, l'inscription diagrammatique y correspondant. Il est clair que ces choix sont nécessairement guidés par une intuition et par des convictions. D'une certaine manière, ils se heurtent à toute objection.

### 1.3. Synthèse

Dans ce chapitre, j'ai abordé la question de la lecture des manuscrits et des éditions de textes (→1.1.1) et de celle des fiches et des répertoires de matériaux linguistiques (→1.1.2). J'ai ensuite montré l'importance des conventions de lecture dans les diagrammes syntaxiques anciens (→1.2.1) et

les arbres dépendanciels modernes ( $\rightarrow$ 1.2.2). Peu importe le type d'inscription de connaissance linguistique, le travail de lecture est toujours un travail de décryptage, qu'il s'agisse de lire la lettre d'un manuscrit, de comprendre la logique d'une édition de texte, celle d'un répertoire structuré ou les conventions d'un schéma syntaxique. Comprendre le flux de l'écriture manuscrite, c'est avant tout être capable de reconnaître les lettres qui forment des mots. Éditer ces matériaux, c'est rendre compte de cette compréhension, mais également inscrire l'analyse des relations entre ces mots, par exemple en hiérarchisant des leçons ou en décrivant l'histoire du vocabulaire.

Quand les inscriptions représentent des connaissances relationnelles de façon formalisée, elles reposent sur un inventaire limité d'entités graphiques et de règles configurationnelles liées à l'épistémologie de la discipline concernée (ecdotique, lexicologie ou syntaxe). Telle est la condition pour que l'on puisse parler d'*inscription formelle* ( $\rightarrow$ 0.2.1) et de *diagramme* ( $\rightarrow$ 0.2.3). Il n'est pas nécessaire que les inscriptions correspondent à des structures définies algébriquement par ailleurs pour qu'on puisse parler d'inscription formelle. Au niveau de l'opération de lecture, les inscriptions formelles ont un avantage majeur : puisque leurs règles peuvent être explicitées, on peut les comprendre de manière exhaustive.

Si les auteurs des inscriptions se chargent parfois de nous expliquer comment interpréter les entités et leurs configuration, il faut souvent se baser sur des adjuvants externes et sur une opiniâtre démarche de comparaison pour arriver à comprendre. Les inscriptions ne sont compréhensibles que par le biais d'un savoir traditionnel que le lecteur a intégré *a priori* ou par le recours à un référentiel qui en explicite les règles de lecture, les rendant du même coup plus formelles.

Puisqu'il est légitime de formaliser la connaissance, la question qui se pose est de savoir comment différentes causes contraignent son inscription, allant des choix personnels aux contraintes des formes choisies pour les encoder. Ces choix font l'objet du chapitre suivant.

## 2. Choisir

---

Dans le chapitre précédent, je me suis concentré sur la lecture des inscriptions. Au détour des exemples, j'ai été amené à montrer que la compréhension se basait souvent sur l'acquisition de conventions de représentation. Dans ce chapitre, je me focaliserai davantage sur la production des inscriptions de connaissances linguistiques en examinant les principaux choix qui doivent être effectués et ce, à trois niveaux. Tout d'abord, le niveau des phénomènes représentés : quelle partie de ce qui a été observé doit être inscrite et quels sont les contraintes qui limitent cette démarche (→2.1) ? Ensuite, quelles analyses doivent être explicitées dans l'inscription (→2.2) ? En troisième lieu, si l'on se concentre sur les inscriptions formelles, en quelle mesure le choix de la structure qui sert à encoder la connaissance limite-t-elle le contenu effectivement inscrit (→2.3) ?

### 2.1. Choisir les phénomènes représentés

La connaissance en linguistique se fonde sur l'étude de phénomènes observables qui sont enregistrés dans des inscriptions. Pour traiter la question du choix des phénomènes inscrits, je vais revenir plus en profondeur sur la transcription des sources, dont je n'ai parlé jusqu'à présent que de façon très superficielle (→1.1.1). J'ai peu pratiqué la transcription de fiches de répertoires d'analyses linguistiques (→1.1.2) ; je vais donc me focaliser ici sur l'activité pratique de transcription des manuscrits médiévaux – mais j'ai la conviction que les enjeux et les difficultés ne sont pas très différents. Je traiterai successivement deux points : celui de la recherche de la pertinence (→2.1.1) et celui des contraintes que la présentation enrichie des sources impose quant au choix des phénomènes inscrits (→2.1.2).

### 2.1.1. Recherche de la pertinence

Comme je m'en suis aperçu en éditant mes premiers documents (→1.1.1), la ponctuation originale est généralement évacuée par toutes les éditions traditionnelles. Ces dernières comportent à la place une ponctuation modernisée qui facilite la lecture et témoigne de la manière dont l'éditeur a interprété le texte (Wilhelm 2015 : 145). Ces éditions inscrivent le texte d'une manière qui empêche l'accès à cet élément original. La cause de l'absence de la ponctuation dans les éditions classiques est assez simple : si l'on excepte quelques travaux pionniers (Roques 1952, Marchello-Nizia 1978), la plupart des philologues ont jugé cette ponctuation trop pauvre ou trop inconsistante pour être compréhensible (voir notamment les remarques de Monfrin 1974 : lxiv ; Naïs 1979). Cette conception n'a changé que très récemment<sup>19</sup>. Il en résulte une première prise de position triviale : il ne faut pas éliminer ce que l'on veut étudier. Les études préparatoires à ma thèse (Mazziotta 2007a ; 2007b) et à celle d'Alexei Lavrentiev (2009a), qui a lui aussi étudié ce domaine, reposent donc toutes sur des éditions qui préservent la ponctuation originale.

Il ne suffit toutefois pas de décider qu'on conservera certains caractères pour que plus aucune question ne se pose. Quand l'éditeur transcrit ses matériaux, il produit une inscription comprenant une représentation des faits qu'il ne veut pas réduire sur la base de la connaissance qu'il a construite en accédant à la matière du manuscrit. Comme je l'écrivais plus haut (→1.1.1), une manière de transcrire le début du manuscrit de la fig. 16 (qui reprend la fig. 3) est (5) (qui reprend (3)).

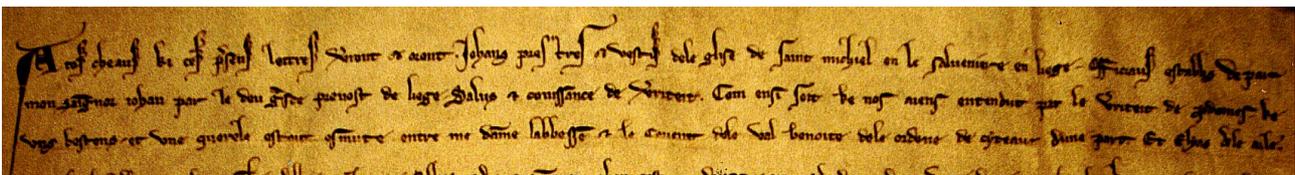


Figure 16 – Fac-similé de Document 1273-05-12 (détail) (reprise de fig. 3)

(5) A tos cheaus ki ces *presens* lettres verront *et* oront ‘À tous ceux qui verront et qui entendront la présente’ (Document 1273-05-12 : 1)

On voit que la transcription comporte plusieurs occurrences de <o>, que l'éditeur a voulues rigoureusement identiques. Saussure raisonnait déjà à partir de cette démarche qui consiste à regrouper des lettres individuelles différentes sous une unique représentation (1967 : 165). L'activité pratique de transcription consiste ainsi à associer des caractères individuels comme faisant partie d'une même classe au sein de notre encyclopédie, à réduire une réalité complexe à un inventaire limité d'unités (Mazziotta 2009 : 25-29). Sous cet aspect, il s'agit d'une formalisation, au sens associé à

<sup>19</sup> Voir, encore cette année, l'ouvrage collectif Fasseur/Rochelois 2016.

*l'inscription formelle* (→4). Il s'agit en outre d'une recherche de forme au sens purement linguistique, c'est-à-dire d'une recherche de *pertinence*.

L'évaluation de la pertinence est particulièrement délicate lorsqu'on parle de de ponctuation médiévale. Dans une de mes premières études sur la ponctuation (Mazziotta 2011a, terminée en 2003), j'ai effectué la transcription d'une charte comportant un grand nombre et une grande variété de signes de ponctuation. Parmi eux figurait celui qu'on appelle *punctus elevatus* ou *comma* (Parkes 1992), qui a la forme d'un point surmonté d'une virgule droite ou d'une virgule incurvée – la fig. 17 (tirée de Mazziotta 2009 : annexes) donne un exemple de chacune de ces configurations.

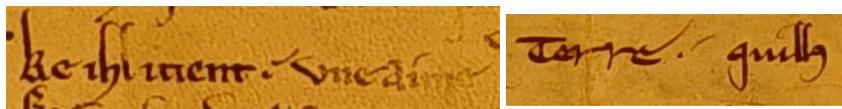


Figure 17 – Fac-similé comprenant «↗» (gauche ; Document 1236-05 : 6) et «↘» (droite ; Document 1236-12-15 : 2)

La principale question qui se pose à l'éditeur qui transcrit ce genre de signe est la suivante : si l'on peut observer une différence entre ces deux formes d'expression<sup>20</sup>, est-il pertinent de les reproduire dans l'édition, sous la forme des entités «↗» et «↘», c'est-à-dire de formaliser cette différence ? Dans le cas que je présente, la réponse est positive : la distribution de «↗» et celle de «↘» sont différentes (Mazziotta 2011a : § 1.2.2). Pour arriver à cette conclusion, cependant, j'ai dû observer cette différence et l'évaluer, ce qui n'aurait pas été possible en travaillant sur une édition qui aurait fusionné les deux signes sous la forme «↗».

Face à l'objet concret qu'est le manuscrit, on pourrait avoir la tentation de choisir naïvement de « tout garder au cas où », mais j'ai bien dû me faire à l'idée que ce genre de démarche était souvent illusoire (voir domageable, cf. Wilhelm 2015). L'accès au texte est fondé sur une encyclopédie visuelle (→ 0.2.2). Comme l'école de Trèves (Holtus *et al.* 2003, critiquée notamment par Wilhelm 2015), j'ai moi-même tenté de reproduire les différentes largeurs des espaces entre les mots dans les chartes que j'étudiais. La transcription (6) comprend le signe «↗» pour indiquer une petite espace.

(6) et por ce ke ce foit ferme et ftable auonof a ce letref pendut noftre faeal 'Et pour que ce soit établi de manière ferme et permanente, nous avons pendu notre sceau à cette lettre'  
(Document 1236-05 : 11)

Cette petite espace est visible sur la fig. 18.

<sup>20</sup> Le problème est similaire à celui de la distinction entre les sons et les phones dans l'approche de Mel'čuk/Milićević 2014 : 9.

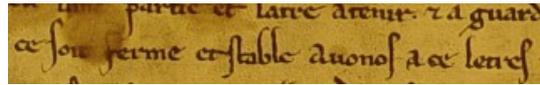


Figure 18 – Fac-similé d'un passage comprenant une petite espace (Document 1236-05 : 12)

Cette démarche fait émerger un point très sensible du processus de transcription des sources : quand on essaye de reproduire des distinctions que l'on ne comprend pas complètement, le risque d'y projeter nos conceptions modernes est grand. Pour le cas précis des différences d'espacement entre les mots, faute de mesures précises (comme on peut le faire dans une démarche paléographique ; Gilissen 1973), je m'en suis remis à une impression. J'étais nettement enclin à repérer des espaces plus étroites entre les prépositions *à*, *de*, *en* et leur régime que partout ailleurs. En somme, je me suis *imaginé* un fonctionnement spécifique à ces petites espaces et je l'ai inscrit dans l'édition comme une donnée<sup>21</sup>. Je pense que les perturbations de ce type sont constantes lorsqu'il s'agit de relever des données nouvelles.

L'équilibre est donc toujours difficile : faut-il se contenter du *peu* ? peut-on se justifier d'en faire *trop* ?

### 2.1.2. Contraintes liées à l'enrichissement

Une autre raison qui pousse à réduire est liée au processus d'enrichissement et à certaines contraintes du support. Ce peut être pour des raisons ergonomiques. Ainsi, on élimine généralement la ponctuation ancienne pour se libérer l'espace nécessaire à l'inscription d'une ponctuation moderne. Sinon, on se heurte à des problèmes d'ergonomie. Ainsi, dans les travaux préparatoires de ma thèse, je proposais (Mazziotta 2004 : 797) de se doter des moyens de produire des éditions aux allures de (7), où les virgules modernes se mêlent aux points anciens, au milieu de la ligne (cela n'est possible que si les répertoires d'entités anciennes et modernes sont clairement différenciés).

(7) Je, Fastrés de Berlos, et je, Eustasses, se fiuz, · prometons, · por bien et por amor a doien · de Saen Johan de port le glize Sen Johan, ke [...] 'Moi, Fastré de Berloz, et moi, Eustache, son fils, promettons, pour le bien et pour l'amour du doyen de l'église Saint-Jean, représentant l'église Saint-Jean, que [...]' (Document 1241 : 1)

La lisibilité se trouve fortement compliquée par la présence de deux couches de ponctuation : l'ancienne, qu'on a refusé de réduire, et la nouvelle, qui constitue un enrichissement.

Au-delà des contraintes ergonomiques, il y en a d'autres qui sont plutôt de l'ordre de l'impossibilité matérielle. Reprenons l'exemple de la résolution d'abréviations que j'ai évoqué ci-dessus (→1.1.1) : il existe des conventions qui permettent de rendre compte du résultat de l'analyse de

<sup>21</sup> Le même genre d'incertitude se présente quand il est question d'évaluer la taille des lettres. C'est une des raisons qui m'ont poussé à écarter prudemment les « capitales » de la question de ma thèse, qui n'étaient parfois que des variantes à peine plus grandes des lettres « minuscules ».

l'éditeur (la mise en italiques ou encore l'emploi de parenthèses). Je peux produire différentes transcriptions du mot abrégé de la fig. 19 (qui reprend un détail de la fig. 3), où une sorte de crochet au-dessus du <þ> remplace la chaîne de caractères <re>.



Figure 19 – Fac-similé d'un mot abrégé (Document 1273-05-12)

(8.a) þfenf 'présentes' (dans ce contexte)

(8.b) *presens*

(8.c) þ*re*fenf

Dans (8.a), l'abréviation n'est pas résolue : le procédé est simplement reproduit. Par contre, dans (8.b), l'abréviation a été remplacée par les lettres qu'elle sert à éluder. La matérialité du support empêche de reproduire simultanément l'abréviation et sa résolution de façon lisible et ergonomique : il faut choisir. Je n'ai jamais croisé d'éditions proposant des transcriptions comme (8.c), où le procédé abrégatif et la résolution de ce dernier figureraient côte à côte. Mais contrairement à ce que j'ai montré à propos de la ponctuation, il n'y a pas que les habitudes de lecture qui empêchent ce genre d'initiative. Ainsi, certains procédés abrégatifs consistent en des déplacements de lettres en exposant, comme <q̇> pour <qui> (Mazziotta 2008b : § 2.1.3) et il serait dans ce cas impossible de juxtaposer une abréviation et sa résolution sans déplacer ces lettres (et donc perdre le procédé abrégatif). Par ailleurs, certains procédés abrègent des lettres qui ne sont pas contiguës dans la forme résolue (Mazziotta 2008b : § 2.2.2), comme <ch̄l̄r> pour <cheualier>.

Enrichir la transcription par la résolution de l'abréviation a donc pour conséquence la réduction des procédés originaux. De la même manière, on voit que dans (8.a), j'ai conservé la forme <f>, mais pas la distinction entre ce que j'ai identifié comme sa première occurrence et sa seconde occurrence. Je ne le montre pas ici, mais cette forme <f> s'oppose à la forme <s> dans le reste de la transcription (on voit des exemples de ce <s> dans la fig. 3 ci-dessus). J'indique ainsi que je représente deux entités que je peux distinguer. Je n'ai pas réduit la différence de *ductus* qui les oppose (mais j'ai réduit les variations dans leurs réalisations). Ce faisant, je me prive de la possibilité de représenter de façon directe une autre partie de ma connaissance des matériaux : selon un certain point de vue, <s> et <f> sont équivalents. Sans rompre la logique du flux de l'écriture, il est matériellement impossible d'inscrire cette information, qui reste implicite, à moins de créer une édition synoptique (Careri *et al.* 2001) qui n'est autre qu'une juxtaposition d'inscriptions concurrentes.

## 2.2. Choisir les analyses explicitées

S'il faut choisir la part d'observable à inscrire, la question se pose également pour les constructions mentales issues de l'analyse. En prenant l'exemple des classes de fonctions syntaxiques, je montrerai comment la description est inévitablement réduite par la sélection d'une partie de ce que l'analyste conçoit (→2.2.1). Je montrerai ensuite que les analyses (syntaxiques et autres) peuvent prendre corps dans les inscriptions au moyen d'entités discrètes par lesquelles elles sont, comme je le dirai, *réifiées* (→2.2.2).

### 2.2.1. Limiter l'inventaire des classes

L'éducation grammaticale traditionnelle m'a inculqué une très grande variété de distinctions et de classements des mots et des rapports qu'ils entretiennent. Qu'elle soit considérée comme une connaissance inconsistante à réformer (Wilmet 2016) ou comme un mal nécessaire, cette constellation de termes et d'analyses fonde le socle de mes raisonnements de linguistique. Face à une structure comme *de Pierre* dans (9), je pense simultanément que j'ai affaire à un « complément déterminatif subjectif », à un « syntagme prépositionnel en position de complément du syntagme nominal » (Laenzlinger 2003), à un « actant subjectal d'un gouverneur nominal » (Miličević/Mel'čuk 2014), à un « substantif transféré en adjectif » (Tesnière 1959), etc.

#### (9) La réaction de Pierre

Bon nombre de ces analyses, issues d'écoles de pensée différentes, peuvent être comparées et partiellement assimilées les unes aux autres. Par ailleurs, je distingue clairement la différence entre *le* (déterminant article, tête du syntagme « déterminantiel », etc.) et le groupe *de Pierre* eu égard à leurs fonctions respectives. Cependant, à chaque fois que je l'ai pratiquée concrètement, l'analyse syntaxique m'a enclin à limiter mes analyses pour deux raisons : 1/ le système conceptuel employé pour travailler ne mobilise pas les apports de toutes les écoles de pensée ; 2/ certaines distinctions intuitivement accessibles sont jugées inutiles ou insuffisamment délimitées<sup>22</sup>.

En annotant les chartes dont j'ai comparé la structuration syntaxique à la ponctuation, j'ai élaboré moi-même la liste des espèces de ce que j'ai appelé les « constituants immédiats de la phrase ». Je ne vais pas développer le débat sur la définition de la phrase dans les corpus que j'ai construits ou auxquels j'ai participé (voir notamment Mazziotta 2009 : 77-86 ; Lavrentiev 2010) : il est assez similaire à celui sur la définition des éléments qui la composent, sur lesquels je me concentre ici. Dans une conception « plate » de l'analyse syntaxique, j'ai distingué le verbe (que j'ai appelé *prédicat*) et ses arguments syntaxiques, dont j'ai limité le nombre à six : sujet (S1), deux sortes de

<sup>22</sup> Je laisse ici de côté les cas d'appartenance des unités à plusieurs classes (v. Prévost 2011 : 70-72).

régimes (R2 et R3), « adjets »<sup>23</sup> (A4) et deux sortes de circonstants (C5 et C6). Cette classification est résumée dans la fig. 20, où les termes sont définis par leurs propriétés syntaxiques (marquées par <@>) énumérées de façon succincte.

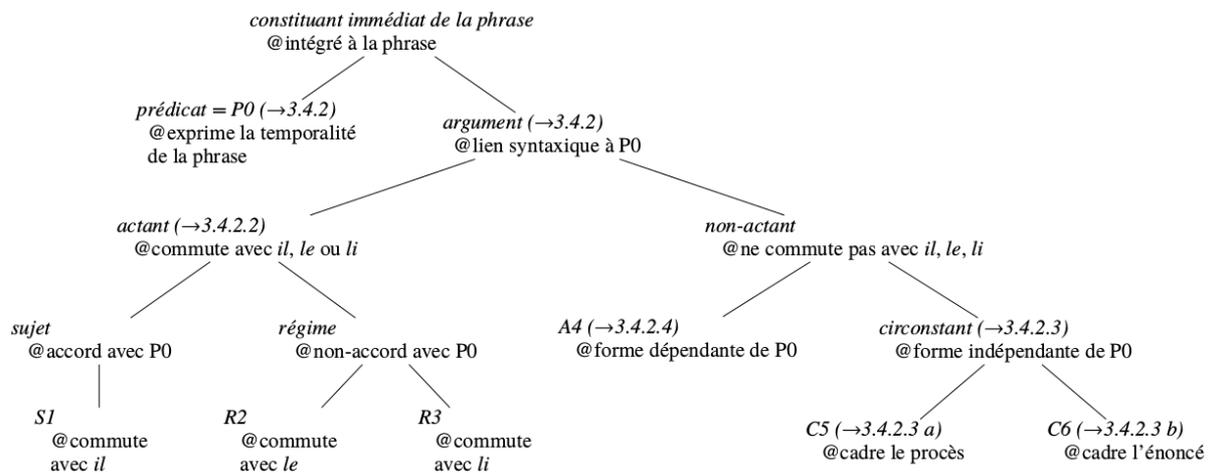


Figure 20 – Hiérarchie conceptuelle des dépendants du verbe (Mazziotta 2009 : 102)

Ce n'est pas le lieu de décrire intégralement la logique du classement et de la critiquer. Je veux attirer l'attention sur le fait que les classes de la taxinomie sont limitées à six. Un bon nombre de distinctions traditionnelles sont absentes : pas de distinction particulière du sujet « impersonnel », des différentes espèces de compléments « circonstanciels », des « compléments d'agent », etc. Toutes ces classes, que je peux reconnaître indépendamment du jugement que je pose sur leur intérêt scientifique, sont purement et simplement assimilées à l'une ou l'autre des six que j'ai retenues et définies.

Bien que fondant l'analyse sur un nombre beaucoup plus élevé de classes, l'équipe du projet *Syntactic Reference Corpus of Medieval French* (SRCMF) a procédé exactement de la même manière<sup>24</sup> pour annoter la syntaxe d'un grand nombre de phrases en français médiéval (Stein/Prévoist 2013 ; Mazziotta 2012b). Le travail collectif que nous y avons effectué a rendu extrêmement claire la raison essentielle qui nous a poussés à limiter le nombre de classes : dans nos travaux préparatoires, nous avons rapidement constaté que faute de consensus sur la définition des unités, il était impossible de produire une annotation cohérente. Les longues et passionnantes réunions du projet SRCMF ont presque toutes été consacrées à la délimitation de classes permettant d'effectuer un travail d'une homogénéité acceptable. Il en résulte que nous n'avons par exemple pas distingué les compléments « datifs » d'autres compléments proches du verbe. La fiche consacrée à la définition des éléments que

<sup>23</sup> Dans la classification des compléments du verbe, les adjets sont requis (ce qui les distingue des circonstants), mais non régis (donc, de forme libre) et cumulables (Lazard 1994 : 68-80).

<sup>24</sup> Voir Bischof *et al.* 2012 pour l'inventaire des classes employées et les critères qui ont servi à leur repérage.

nous avons nommés *compléments* (Bischof *et al.* 2012 : Cmpl<sup>80</sup>) donne ainsi les trois exemples (10), (11) et (12) (italiques en rouge dans le texte).

(10) afr. et distrent a çax dedenz qu'il *lor* aprestassent lor armes 'Et ils dirent à ceux qui se trouvaient à l'intérieur qu'ils *leur* apprêtent leurs armes' (SGraalIVQuesteM : p. 243)

(11) afr. *Desuz un pin* en est li reis alez 'Le roi s'en est allé *sous un pin*' (RolMoign : p. 36 v. 165)

(12) afr. et puis fu li marchis ochis *de Haussassis* 'Et puis le marquis fut tué *par Haussassis*' (RobClariL : p. 39)

L'inventaire fini de classes et l'explicitation de leur combinatoire correspond à une formalisation de l'analyse (voir les exemples de théories syntaxiques sous →1.2), qui inscrit une partie seulement de nos connaissances sur les constructions. Partant, cette inscription ne comprend que la partie jugée suffisamment cohérente pour être transmise.

En outre, les contingences (temps, financement) impliquent toujours qu'au terme de projets de ce type, ses membres regrettent de n'avoir pu inscrire des connaissances qui étaient en fin de compte suffisamment structurées pour être formalisées. Comme le corpus devient la « donnée » d'études ultérieures, ces choix ont des conséquences importantes. Pour SRCMF, nous regrettons par exemple de ne pas avoir identifié les impératifs et les énoncés interrogatifs, qui auraient été facilement repérables et qui seraient actuellement d'un grand intérêt pour étudier les tendances syntaxiques associées à l'« oral représenté » (Marchello-Nizia 2012). Ainsi l'étude de Glikman/Mazziotta 2013 ne distingue pas les structures impératives des autres, mais essaye tout de même d'évaluer le taux relatif de présence du sujet dans la narration et l'oral représenté. Il ne faut pas voir en ces regrets un jugement inutilement négatif sur le travail effectué, mais une invitation à poursuivre le travail d'inscription et à l'enrichir.

### 2.2.2. Réification des éléments de l'analyse

Outre le choix de l'inventaire des classes qui servent à étiqueter les unités dans l'analyse, se pose la question du statut des unités à étiqueter. J'ai déjà parlé ci-dessus de la démarche de reproduction (réduite) de caractères d'écriture, c'est-à-dire de phénomènes observables *a priori* (→2.1). Je vais m'occuper ici des éléments qui constituent l'analyse syntaxique (→a) et lexicologique (→b), c'est-à-dire la *construction mentale* inscrite par le linguiste. En inscrivant ces éléments de façon à les rendre observables, on peut dire que le linguiste les *réifie*.

### a. Réification et analyses syntaxiques

En repartant des diagrammes de Clark et de Tesnière, dont j'ai expliqué succinctement les conventions de lecture (→1.2.1), on peut voir que leurs diagrammes n'ont pas la même structure. Je reprends dans la fig. 21 les diagrammes des fig. 11 et 14.

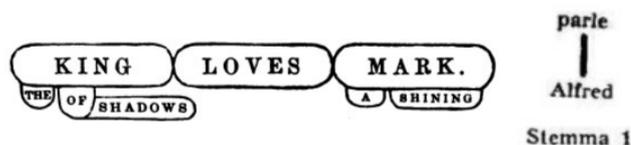


Figure 21 – Diagrammes de Clark (gauche, cf. fig. 11) et de Tesnière (droite, cf. fig. 14)

En utilisant ses stemmas, Tesnière a l'intention de rendre visibles les relations syntaxiques entre les mots, auxquelles il donne le statut d'objet indépendant des mots qui les contractent (1959 : ch. 1, § 5). Rendre visible les relations, c'est en l'occurrence choisir de les représenter par une entité discrète, dont on peut dire qu'elle la *réifie*. Dans le stemma de la fig. 21, on peut dire que la relation entre *Alfred* et *parle* est réifiée par un trait, contrairement à ce qu'on observe chez Clark, où c'est l'agencement relatif des bulles qui exprime la relation qui unit les mots<sup>25</sup>. Les relations ne sont donc pas réifiées. Toutefois, la logique configurationnelle permet d'exprimer le même genre d'analyse que chez Tesnière : si *shining* est placé sous *mark*, on peut dire que le second régit le premier.

Dans d'autres cas, le choix de ce qui est réifié conditionne le contenu qui peut être inscrit. J'ai bien dit que les taxinomies proposées dans ma thèse et dans SRCMF relevaient d'une conception « plate » de la phrase, où le sujet et tous les compléments sont au même niveau. Il s'agit de la description dépendancielle classique depuis Tesnière. Prenons (13) : en suivant cette approche « plate », on construit un arbre dépendanciel comme celui de la fig. 22, qui est tirée de SRCMF<sup>26</sup>.

(13) afr. Des lors te toli li anemis la veue. 'Dès cet instant, le Diable t'a dépouillé de ta vue.'

(SGraalIVQuesteM : 190a, 1)

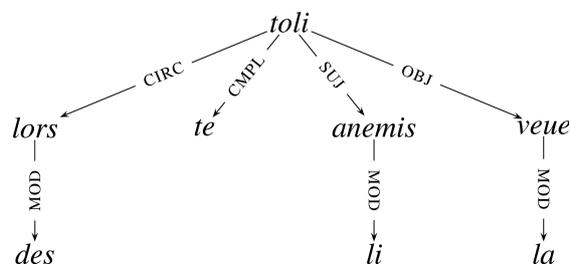


Figure 22 – Arbre dépendanciel de afr. *Des lors te toli li anemis la veue* (cf. Mazziotta 2012b : 704)

<sup>25</sup> Pour une analyse formelle de cette structure, voir Kahane/Mazziotta, soumis.

<sup>26</sup> Les étiquettes sur les flèches correspondent aux noms des dépendances syntaxiques, abrégés : *circ(onstant)*, *c(omplément)*, *suj(et)*, *obj(et)*, *mod(ifieur)*. Voir Bischof et al. 2012.

Les relations syntaxiques sont réifiées par des flèches étiquetées. La lecture classique de ce genre de schéma ne reconnaît pas d'autre fonction sémiotique à ces dernières. Cette conception ne rend pas compte des hiérarchies des constituants, qui pourrait représenter structurellement si certains arguments sont plus ou moins éloignés du verbe. La fig. 23 montre quant à elle une inscription de la même construction qui se fonde sur une analyse en constituants immédiats<sup>27</sup>.

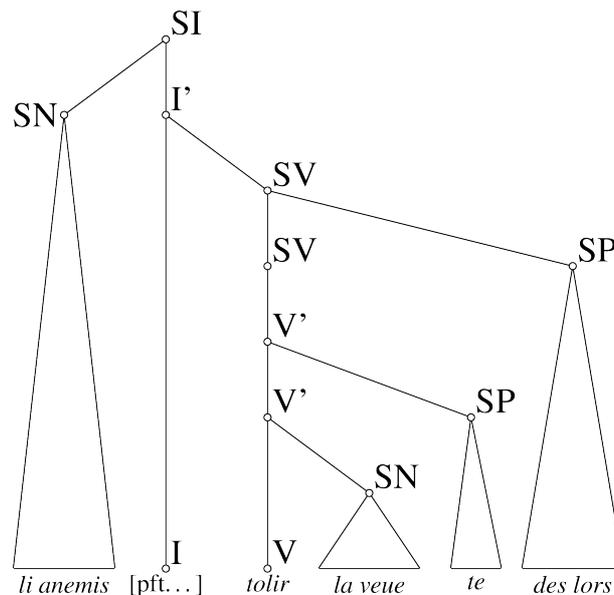


Figure 23 – Arbre X-bar de afr. *Des lors te toli li anemis la veue* (ordre de surface non respecté)

On voit comment la réification de chaque construction par un nœud étiqueté permet d'exprimer qu'à chaque niveau de projection de V (puis de I), un complément unique est agrégé au groupe formé au niveau précédent. Cette réification rend compte de la « stratification » de la structure, selon les termes de Kahane 1997 (développés dans Kahane/Mazziotta 2015c). Par contre, les relations de dépendance ne sont pas réifiées de manière directe par un objet unique.

### b. Réification et analyses lexicologiques

La réification d'une analyse dans une inscription se fait d'une manière qui est conforme à l'épistémologie et à la technique associée à la discipline. Je vais prendre ici l'exemple des glossaires de textes médiévaux et des études lexicales pour montrer qu'une démarche d'analyse différente aboutit également à une réification. Pour « donner à lire » le texte, il est de bon ton que l'éditeur donne au lecteur les moyens de comprendre son lexique. Cela se fait généralement au moyen d'un glossaire<sup>28</sup>. Dans mon mémoire de fin d'études (2001), qui comprenait une transcription imitative et une édition réduisant davantage les spécificités graphiques des documents transcrits, le glossaire est fondé sur cette

<sup>27</sup> Il s'agit d'une analyse simplifiée d'une prétendue « structure profonde », à laquelle je n'adhère pas, mais que je livre ici comme exemple. J'y laisse de côté la question de l'ordre des constituants.

<sup>28</sup> Dont les « bonnes pratiques » actuelles (Möhren 2015) imposent qu'il soit quasi exhaustif et élaboré à l'aide du recours systématique à la documentation lexicographique disponible.

dernière. Ainsi, les formes transcrites <fauoir> et celles transcrites <sauoir> dans la transcription imitative sont rangées sous la même entrée, étant donné qu'elles actualisent le même lexème<sup>29</sup> (Polguère 2003 : ch. 3) en dépit de leurs différences. C'est là une pratique courante<sup>30</sup>. Cependant, la construction du glossaire ou de toute étude lexicale nécessite de rechercher un commun dénominateur encore plus abstrait. Je vais prendre ici un exemple tiré de ma pratique (Mazziotta 2006a) pour illustrer mon propos.

(14) afr. [...] quant li miliares *coroit* par [7] m et cc et xl trois ans 'Quand l'année 1243 était en cours' (Document 1243-08-15 : 6, je souligne)

À l'époque de mon relevé, il s'agissait de la première attestation du sens technique du verbe français *courir* 'se dérouler' (14). Cependant, ce qui m'intéresse ici, ce n'est pas tant qu'il s'agisse d'une première attestation, mais les opérations mentales nécessaires pour le justifier et la manière dont cela est inscrit. L'étude lexicologique ne peut se satisfaire d'un travail sur les mots-formes. On remarquera facilement que *coroit* est une forme fléchie du lexème afr. *corir*. Voilà une première étape qui permet d'assigner un type abstrait à *coroit*. Ce faisant, on réduit les spécificités de la forme observée pour ne plus considérer que le lexème, envisagé comme « métasigne » (Polguère 2003 : 72). La réduction va ici de pair avec l'identification du lexème qui est commun à toute une série de formes fléchies.

***courir*** verbe intr. [A. Une période de temps] court « A. se déroule ».

En tesmongnage de ce, li doing je mes lettres saelees de mon sael, [6] que furent faites le jor del Assumption Nostre Damme, quant li miliares *coroit* par [7] m et cc et xl trois ans (1243, LgSLa 1243b : 5)

TLF(i) s.v. *courir* I. A. 2. b) a) : « se dérouler de façon continue » ; historique : rien pour ce sens. — Von Wartburg in FEW 2, 1565b, CURRERE I 1 a ; Gdf 2, 337b ; TL 2 : pas ce sens. Néanmoins, GdfC 9, 224a donne « s'écouler » (Deschamps). — Ø DocHainR, DocFlandrM. DocHMarne 66, 12 : d'après le glossaire, *l'an de l'Incarnation nostre Signor que corroit [13] par mil et dous cenx et cinquante seis* (1256) ; voir les collocations avec des militaires comme actant 1 : DocHMarne 6, 13 : *Ce fu fait en l'an que li miliaire corroit mil et douz cenx et trente sex ans, en mois [14] d'auost* (1236). Le problème est le même pour retrouver les attestations de *courir* dans DocVosL 3, 5 1238, ainsi que HoltusLux 4, 7 1242 et DocAubeC 11, 14 1253.

→ Première attestation et datation du sens spécifique — non daté par la lexicographie historique — dans DocHMarne, pour peu qu'on ne se limite pas à dépouiller le glossaire. Dans les recueils de chartes dépouillés, nous n'avons trouvé ce sens attesté que dans l'eschatocole<sup>6</sup>, avec comme actant 1 un substantif dénotant une période de temps.

Figure 24 – Fac-similé de l'article *courir* dans Mazziotta 2006a

En exprimant cette lexie de manière concrète par la vedette, comme dans la fig. 24, le philologue qui rédige un glossaire et le lexicographe qui produit un article de dictionnaire font exactement la même opération que le syntacticien pour expliciter l'analyse. Ce dernier emploie par exemple un trait pour

29 Plus précisément, il s'agit de deux graphies associée à différentes occurrences d'un même mot-forme (Polguère 2003 : 87 sv.).

30 Les éditions qui se fondent sur la transcription imitative et qui n'identifient pas ces correspondances lexicales ne sont pas toujours bien accueillies (v. p.ex. la critique de Wilhelm 2015).

exprimer une relation, eux dotent leur métasigne d'une forme d'expression discrète. On peut dire que tous *réifient* une part de l'analyse sous la forme d'une entité visuelle distincte.

Les grandes entreprises que sont l'ALW et le FEW (→1.1.2) emploient des techniques tout à fait similaires à celles que je viens de présenter et ce à différentes échelles. J'ai déjà brièvement introduit la philosophie de ces inscriptions de connaissances lexicales : il s'agit de repérer les liens qui unissent les mots-formes (ALW) et les lexies (FEW) traitées. Comme je l'ai illustré, les données sont présentées selon des conventions parfois complexes qui inscrivent une connaissance qu'on peut dire « relationnelle » (→0.2.3). C'est sur la construction de ces connaissances relationnelles que j'aimerais axer la suite de cette section. Je commencerai par examiner les structures du FEW, dont l'objectif est de décrire l'histoire du vocabulaire galloroman. J'ai montré (Mazziotta 2011b, Mazziotta/Renders 2010) que le processus de classification par identification de communs dénominateurs successifs était le fondement de la démarche de rédaction du FEW. Les lexies relevées sont regroupées en vertu de leurs similarités ; selon Eva Büchi (Büchi 1996 : 105-116), sept critères organisationnels se combinent : 1/ le critère de l'étymon immédiat ; 2/ le critère de la transmission ; 3/ le critère phonétique ; 4/ le critère morphologique ; 5/ le critère sémantique ; 6/ le critère géolinguistique ; 7/ le critère chronologique. Le classement peut être conçu comme multidimensionnel. La matière de chaque article est organisée de manière spécifique par son rédacteur<sup>31</sup>, de façon à expliquer au mieux l'histoire des mots. Compte tenu du support traditionnel (c'est-à-dire non informatique ; →0.2), on n'a d'autre choix que de déployer les articles dans un espace bidimensionnel. Le dictionnaire prend la forme d'un texte, où un jeu de symboles réifie la hiérarchie des différents critères (Renders 2015).

La structure formelle dans laquelle l'article est encodé correspond à peu près à un arbre (nous verrons ce que cela implique ci-dessous ; →2.3.1). À titre d'exemple, l'organisation observée dans l'article *basiolum* de la refonte du FEW (Nissille 2008) peut être inscrite dans le schéma de la fig. 25 ci-dessous. Pareille inscription rend évidente la similarité entre la réification d'éléments de l'analyse syntaxique et celle d'éléments de l'analyse lexicologique. La fig. 25 ne contient pas l'inscription de lexèmes, mais uniquement des étiquettes explicitant les critères de regroupement.

31 Les critères de l'étymon direct et de la transmission sont généralement prééminents en raison de la philosophie même du FEW (Büchi 1996 : 106).

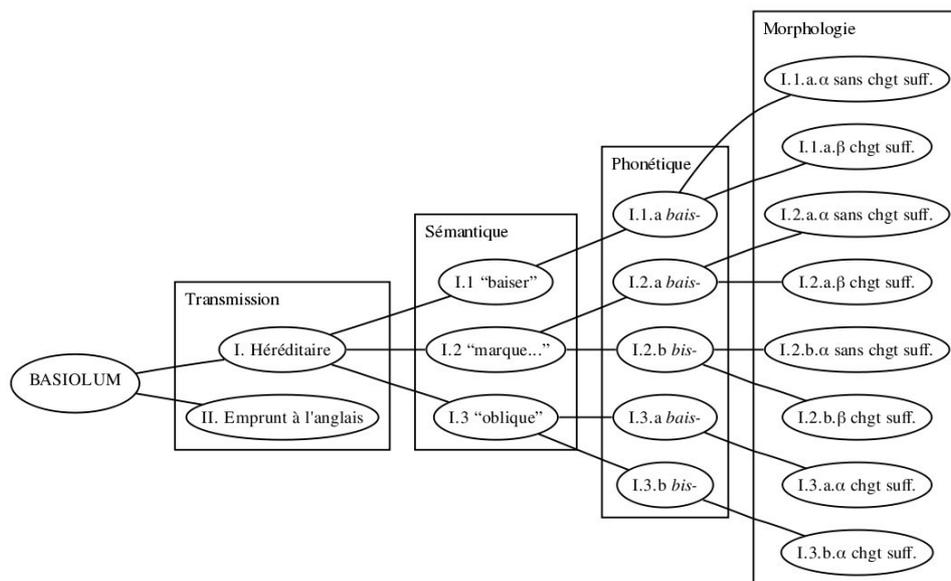


Figure 25 – Organisation de l'article *basiolum* de la refonte du FEW (Mazziotta 2011b : § 3.1.1.a)

Il faut comprendre chacun de ces regroupements comme une abstraction, un vocable étymologique dépourvu de forme graphique, mais qu'on peut appréhender conceptuellement comme le regroupement d'une partie des aboutissements d'un étymon. Chaque nœud de l'arbre représente en fait, de façon réduite, l'ensemble des lexies qui satisfont les critères correspondants. Par rapport au processus de réification, cette démarche est très similaire à celle de l'analyse en constituants immédiats, où chaque construction est réifiée par un nœud : chaque nœud de l'arbre de la fig. 25 réifie un regroupement.

Par contre, les liens entre les différents niveaux de la hiérarchie, qui sont réifiés par des traits dans la fig. 25 sont exprimés par la configuration de la mise en page dans la fig. 26 (p. 36). Qu'il s'agisse de construire un diagramme syntaxique ou un article de dictionnaire, l'analyse qui prend une forme explicite et discrète se trouve réifiée dans l'inscription produite. La connaissance construite se mêle ainsi à la représentation des données, ce qui rend possible leur exploitation concomitante. Mais sous quelles formes les unités et leur rapports doivent-elles être encodées ? Telle est la question du choix de la structure formelle.

<p><b>I.1</b> “baiser”</p> <p><b>I.1.a</b> <i>bais-</i></p> <p><b>I.1.a.α. Sans changement suffixal</b> Afr. <i>baisol</i> m. “baiser” (980, PassionA = Bartsch; Stengel), <i>baisuel</i> (wall. 4<sup>e</sup> q. 12<sup>e</sup> s., SBemCantH<sup>3</sup>).</p> <p><b>I.1.a.β. Avec changement suffixal [+ -ellu]</b> Bress. <i>b ā γ ē y</i> m. “baiser” ALLR 672 p 123.</p> <p><b>I.2</b> “marque laissée sur la croûte du pain à l'endroit où celui-ci en a touché un autre lors de la cuisson au four ; baiseure”</p> <p><b>I.2.a</b> <i>bais-</i></p> <p><b>I.2.a.α. Sans changement suffixal</b> Norm. <i>baiseul</i> m. “baiseure” DT, Condés. <i>bāiseūl</i>, ALENçon <i>baiseul</i>, Manche <i>b ē z æ</i> ALN 1043 p 16, <i>b ē z ū</i> ibid. p 15, IlleV. <i>b ē z æ l</i> ALBRAMms p 35, Mayenne <i>b ē i z æ</i> ibid. p 85, Vosgesmér. <i>b ā ž æ</i> ALFC 328 p 10, HSAône id. (ALFC 328 p 22, 38),</p>
---

Figure 26 – Fac-similé du début de l'article *basiolum* de la refonte du FEW (Nissile 2008)

Dans une pareille inscription, ce sont les nœuds qui réifient les distinctions jugées pertinentes alors que le reste des matériaux est intégralement réduit, mais si l'on observe l'article rédigé (v. fig. 26 pour un extrait), on voit comment la numérotation et les titres au sein de l'inscription textuelle représentent les mêmes regroupements que les nœuds de la fig. 25.

### 2.3. Choisir la structure formelle de l'inscription

Ce n'est pas la formalisation qui réduit ou qui force à l'approximation, mais les choix liés à l'épistémologie propre de la discipline à l'œuvre (Bachimont 2007 : 92, n. 8). Néanmoins, le choix d'une structure formelle ou d'une autre contraint ce qui peut être encodé (et donc inscrit). Je commencerai par ce point (→2.3.1). Je présenterai ensuite la notion de *polygraphe*, une structure formelle qui permet d'encoder de façon ergonomique des analyses qui ne pourraient être inscrites facilement au moyen d'autres structures moins souples (→2.3.2).

#### 2.3.1. Limites des structures formelles

L'arbre qui décrit l'organisation de la notice de *basiolum* ci-dessus (fig. 25 sous →2.2.2. b) est facile à rapprocher des diagrammes syntaxiques. Au niveau du contenu, il exprime des catégories et des analyses qui émergent des phénomènes étudiés non par observation, mais par analyse. Au niveau de la forme, sa structure naturelle contraint la manière dont l'encodage de l'analyse est réalisé. La correspondance entre les objets de la structure formelle et ce qu'elle exprime, conventionnelle et cantonnée à un inventaire et à des règles organisationnelles énumérables, en font une inscription formelle à proprement parler (→0.2.1). Sur le plan de l'économie des ramifications, on aurait éventuellement pu intervertir les critères sémantique, phonétique (aboutissements de la voyelle

initiale) et morphologique (observation d'un éventuel changement de suffixe), puisque les embranchements se répètent. C'est le rédacteur qui a choisi la position des critères et qui considère que cette dernière explique au mieux l'histoire des lexies issues de *basiolum*. Pourtant, si les embranchements se répètent, il est probable qu'il y a bien quelque chose qui unit toutes ces formes « sans changement suffixal » ou « avec changement suffixal ». Ce n'est pas la critique du classement qui m'intéresse ici, mais son caractère inévitablement réducteur. Inscrire l'analyse sur le support papier ne donne guère d'autre possibilité que d'exploiter une structure arborescente.

On touche ici un problème fondamental de la formalisation : en choisissant la structure formelle dans laquelle l'analyse est encodée, on se doit de définir comment l'encodage et la théorie linguistique s'articulent. Mais il ne s'agit pas que de choix de conventions de correspondance entre la structure de l'arbre et la théorie linguistique. Le choix de la structure formelle est lui-même déterminant. Je vais revenir à l'analyse syntaxique pour poursuivre, mais mon argumentation pourra être transposée aux répertoires lexicaux.

J'ai dit jusqu'à présent que les membres de SRCMF (→2.2.1) avaient choisi d'encoder les constructions au moyen d'arbres, mais cela n'est pas tout à fait exact : c'est plutôt de *graphes* dont il faut parler<sup>32</sup>. Un arbre ne permet pas à un nœud d'avoir plusieurs parents, alors qu'un graphe est simplement une structure pourvue de nœuds reliés deux à deux<sup>33</sup>. Or, comme le français moderne, l'ancien français possède des « pronoms relatifs », dont on dit traditionnellement qu'ils assument à la fois la fonction de subordonnant et une autre fonction à l'intérieur de la subordonnée. Dans un cadre dépendancier, cela peut être analysé comme une double dépendance (Hudson 2010 : § 11.5.6), ou en décomposant le pronom relatif en deux « mots » dans l'analyse (Tesnière 1959). Toutefois, si l'on encode l'analyse dans une structure purement arborescente, il faut choisir. Si l'on suit l'avis de Mel'čuk (2009 : 67 sv.) à cet égard, sans contester la double fonction du relatif, ce dernier ne doit pas être encodé comme un subordonnant : il se trouve en position de dépendant par rapport au verbe de la relative (fig. 27).

---

32 Il y a d'autres raisons que celles évoquées ci-après pour lesquelles SRCMF utilise des graphes plutôt que des arbres. En outre, l'interface du logiciel d'annotation de *NotaBene*, qui a servi à encoder les analyses, se sert d'une représentation arborescente complexe qui est convertie en graphe *a posteriori* (v. →4.2.1 concernant l'interfaçage).

33 Pour une présentation adaptée au linguiste, v. Mel'čuk/ Milićević 2014 : 294-298.

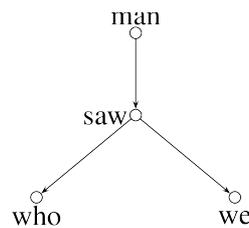


Figure 27 – Arbre dépendanciel contenant une relative selon Mel’čuk (2009 : 68)

On voit donc que dans ce cas, la structure formelle oblige le linguiste à réduire la connaissance qu’il y inscrit. En utilisant des graphes pour encoder la double fonction du relatif, SRCMF choisit au contraire de ne pas opérer cette réduction.

L’analyse de la coordination pose des difficultés similaires (Mazziotta 2012a ; 2015). Dans la fig. 28, je reprends la fig. 9. On voit que le deuxième diagramme en partant de la gauche (celui de Mel’čuk) est un arbre, mais que tous les autres sont des graphes. Alors que le diagramme de Mel’čuk traite la conjonction comme n’importe quel autre nœud de l’arbre, les autres lui assignent un statut particulier.

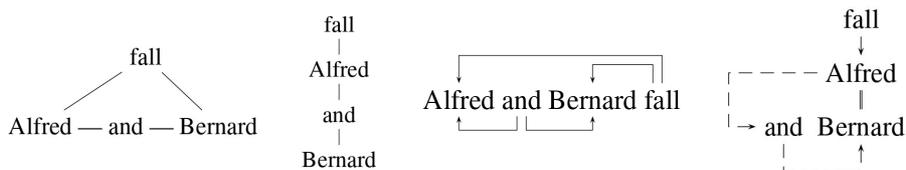


Figure 28 – Schématisations de ang. *Alfred and Bernard fall* ; de gauche à droite : Tesnière, Mel’čuk, Hudson, Kahane

Il y a donc différentes manières d’encoder l’analyse, mais la structure formelle choisie vient toujours avec son lot de contraintes.

### 2.3.2. L’apport du polygraphe

Pour montrer l’importance du choix de la structure formelle, je vais à présent introduire en quelques paragraphes la structure que Sylvain Kahane et moi-même avons appliquée depuis peu à l’analyse syntaxique : le polygraphe (Kahane/Mazziotta 2015a ; 2015c ; soumis). On verra qu’elle est moins contraignante que l’arbre ou le graphe.

#### a. Dépendance ou stratification

Le syntacticien qui choisit un modèle arborescent ou un graphe où les nœuds encodent les mots et où les arêtes encodent les relations ne peut pas vraiment encoder en même temps la stratification (→2.2.2. a). Ainsi, dans SRCMF, il n’y a aucun moyen d’indiquer que, dans (15), le déterminant *li* s’agrège au nom *papes* avant que l’apposition *Innocens* ne soit associée à ce groupe considéré dans son ensemble.

(15) afr. li papes Innocens estoit apostoiles de Roume 'Le pape Innocent était apôtre de Rome'  
(RobClariL : p. 1)

Une autre convention de correspondance entre la structure arborescente et la théorie linguistique, comme celle de l'ACI, où les nœuds représentent soit des mots, soit des constructions, aurait permis d'encoder ce genre d'information (cf. fig.23 ci-dessus). Toutefois, pareille structuration de l'analyse a comme inconvénient de ne pouvoir rendre compte de l'existence d'une *tête* (c'est-à-dire d'un *gouverneur*) que par des moyens détournés : alors qu'un arbre dépendancier montre clairement, par l'orientation des arêtes, quels sont les gouverneurs et les dépendants, les têtes des grammaires en constituants sont identifiées comme telles par la comparaison de leur étiquette catégorielle avec celles de leur entourage. Ainsi, en relisant la fig. 23 ci-dessus, on peut comprendre que V est la tête de V' et de SV parce que tous ces symboles font partie de la même structure et ont la lettre V en commun<sup>34</sup>. Dans une structure comme un arbre, à moins d'opter pour une signification différente des branches en fonction de leur contexte<sup>35</sup>, il faut choisir entre une approche dépendancielle et une approche stratifiée.

### b. Solution du polygraphe

Le polygraphe résout ce dilemme : il s'agit d'une structure formelle dont les arêtes peuvent avoir d'autres arêtes comme sommets (Kahane/Mazziotta 2015c : § 3.1). On peut observer ce type de structure dans la fig. 29 (volontairement abstraite), où les traits entrent en contact avec d'autres traits.

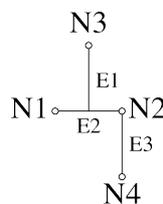


Figure 29 – Exemple abstrait d'un polygraphe

Cet objet formalise une conception : les traits réifient simultanément des relations et la construction qui est associée à cette relation. Je vais à présent montrer concrètement comment cette structure formelle peut être employée en linguistique. Pour ce faire, je vais définir la manière dont la structure formelle est rendue signifiante. Dans Kahane/Mazziotta 2015a, nous avons choisi de nous servir de cette structure pour exprimer deux types de contraintes. Étant donné que ces analyses sont nouvelles, je vais les illustrer quelque peu.

Premièrement, les contraintes de présence, que nous nommons *dépendances*, sont représentées par des traits pleins. Comme une arête peut être le sommet d'une autre arête, on peut laisser le

34 C'est une manière de noter le principe d'*endocentricité* de X-barre, selon lequel une tête projette ses propriétés combinatoires à tous les niveaux supérieurs du syntagme (Laenzlinger 2003 : 37).

35 Comme on le verra, ce genre de démarche complexifie singulièrement l'encodage et la lecture (→3.3.2. a).

gouverneur d'une relation non déterminé et faire d'une construction entière le dépendant d'un autre mot (16.a).

(16.a) ang. Mary talks to Peter.

(16.b) ang. \*Mary talks to.

(16.c) ang. \*Mary talks Peter.

(17.a) Marie vient après Noël.

(17.b) Marie vient après.

(17.c) \*Marie vient Noël.

(18.a) afr. le message de la roïne "Le message de la reine"

(18.b) afr. \*le message de

(18.c) afr. le message la roïne

Le polygraphe permet ainsi de distinguer trois configurations pour analyser les prépositions (fig. 30), selon qu'elles ont un régime obligatoire (16.a) ou facultatif (17.a), ou qu'elles sont elles-mêmes facultatives (18.a). Le polygraphe (a) de la fig. 30 présente une stratification de l'analyse.

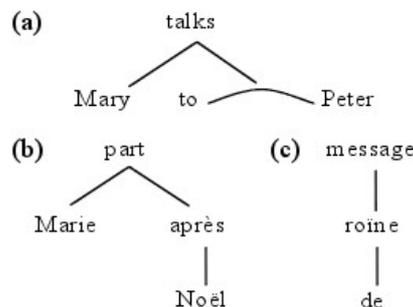


Figure 30 – Exploitation du polygraphe pour encoder les prépositions (Kahane/MazziottaDepling: 188)

Deuxièmement, les contraintes de valence passive, selon laquelle c'est le terme qui contraint le plus la distribution d'un groupe de mots qui en est le gouverneur ( $\rightarrow$ 1.2.2), sont encodées par les arêtes dont la source est une autre arête et nous les nommons *spécifications*<sup>36</sup>. Elles sont figurées par des flèches dont le trait est discontinu. Dans (18.a), le choix de *de* demeure contraint par la relation de « complément génitif » (on ne pourrait avoir *par* ou *por*), ce qui est illustré par la fig. 31.

36 La notion de *spécification* est empruntée à Alain Lemaréchal (1989 et 1997). Selon lui, les spécifications sont des marques grammaticales qui précisent les relations à la fois sémantiquement et syntaxiquement. Par exemple, la marque de l'accusatif en latin est une spécification de la fonction *objet* : sa présence indique prototypiquement le *patient* et elle précise la relation entre le terme nominal et le verbe. Les spécifications ne sont pas toujours obligatoires (comme *de* dans (18.a)), mais quand elles sont présentes, elles ont une forme contrainte par la relation qu'elles précisent.

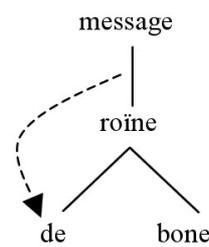


Figure 31 – Encodage d'une spécification dans un polygraphe (Kahane/MazziottaDepling : 189)

Ce type d'objet peut être employé pour décrire de manière plus intuitive certaines constructions qu'on ne peut modéliser que de façon très lourde par des structures arborescentes. Je donnerai à ce propos l'exemple de la déclinaison en ancien français (Mazziotta 2013a, révisé dans 2014b). En me fondant sur le critère de la valence passive pour analyser les phrases simples (19) et (20), j'avais obtenu les arbres de la fig. 32 (les flèches en pointillés expriment des relations morphologiques).

(19) afr. *Li chevaliers conduit le cheval [...]* 'Le chevalier conduit le cheval.' (SGraalIVQuesteM, p. 149, 10)

(20) afr. *La reïne vient a Galaad.* 'La reine vient à Galaad.' (SGraalIVQuesteM : p. 21, 5)

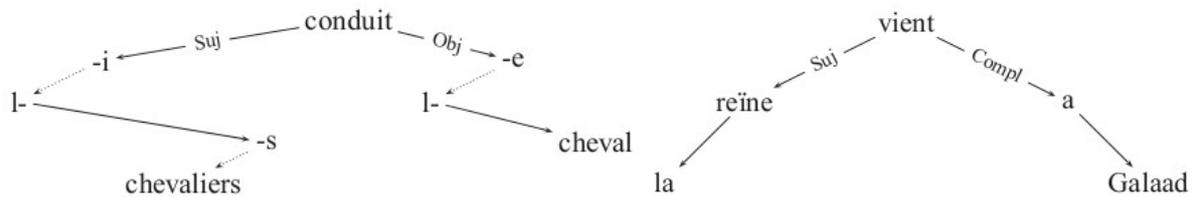


Figure 32 – Analyses arborescentes du marquage des dépendants du verbe en ancien français (Mazziotta 2014b)

Sans expliquer tout le raisonnement qui mène à ces arbres, les grammaires enseignent (Buridant 2000) que certaines classes de noms employés comme sujets en afr., ainsi que leur article, peuvent prendre une forme spécifique : en l'occurrence, *chevaliers* a un *-s* final et le *-i* de *li* marque le sujet (sinon, la forme est *le*). Il est connu depuis longtemps (Schøsler 1984) que ce marquage est facultatif et compte de nombreuses exceptions, mais les marquages indus sont extrêmement rares. En revanche, nombre de noms féminins en *-e* et leur article *la* ne varient pas du tout. En examinant d'autres attestations et en suivant le critère de la valence passive, j'en suis venu à déterminer que la structure syntaxique changeait en fonction de la classe du nom et du genre du déterminant. Comme on me l'avait fait remarquer à l'époque, malgré la rigueur de cette analyse, elle heurte singulièrement l'intuition et complique l'interfaçage entre l'analyse syntaxique et l'analyse sémantique. À présent, armé des polygraphes, je pourrais produire les diagrammes de la fig. 33.

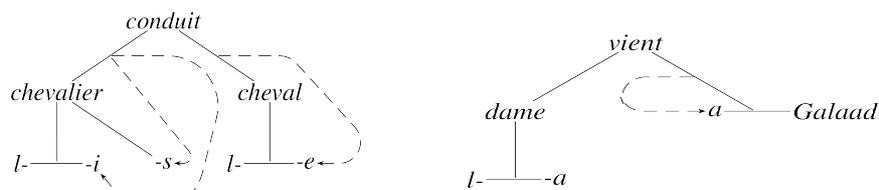


Figure 33 – Inscription polygraphique des dépendants du nom en ancien français

On voit que les deux diagrammes positionnent de façon identique les termes nominaux et que le contraste entre les contraintes distributionnelles est indiqué par les relations de spécifications. L'avantage de cette solution est que les fonctionnements similaires restent visibles.

Au travers de notre proposition pour donner du sens au polygraphe, on a pu constater que la sélection d'une structure formelle plus souple que l'arbre ou que le graphe permet d'encoder des relations syntaxiques de manière plus libre. On doit cette liberté uniquement à la structure formelle sélectionnée, ce qui prouve à quel point la sélection du formalisme est un enjeu capital de la description.

## 2.4. Synthèse

Dans ce chapitre, j'ai montré l'importance de poser des choix dans la construction d'inscriptions de connaissances linguistiques. Ces choix portent sur les phénomènes représentés lorsqu'il s'agit d'éditer des matériaux (→2.1). Il concernent la pertinence de ce qui est retenu dans l'inscription (→2.1.1) et sont contraints par la manière dont sont inscrits des matériaux enrichis plutôt que des matériaux bruts (→2.1.2). Ces choix portent sur la représentation de l'analyse quand il s'agit de sélectionner les classes conceptuelles qui la délimitent (→2.2.1), de même que quand il s'agit de déterminer la part de l'analyse qui sera *réifiée*, c'est-à-dire inscrite par des entités visuelles discrètes (→2.2.2). Enfin, ces choix portent sur les structures formelles de l'inscription, qui contraignent naturellement l'encodage par les règles qui gouvernent leur définition (→2.3.1). Certaines structures, comme les *polygraphes*, ont l'intérêt d'être plus souples que les arbres ou les graphes (→2.3.2).

On ne peut donc éviter de choisir : quiconque construit une inscription de connaissance en linguistique doit se limiter à la représenter de manière partielle, à la *réduire*. Or, ce qui n'est pas inscrit échappe à toute tentative d'exploration ultérieure. En inscrivant les données ou l'analyse, nous limitons *de facto* l'exploitation qui pourra en être faite. Étant donné que les choix correspondent à une posture scientifique qui s'insère dans un contexte de recherche particulier, il ne sera jamais possible de contenter tout le monde. Il est évident que c'est le « point de vue qui précède l'objet » (Saussure 1967 : 23) et que nos catégories distordent inévitablement ce que nous observons. Ce simple fait

justifie que les mêmes matériaux et les mêmes analyses fassent l'objet de multiples inscriptions successives, destinées à des publics variés.

Même si les inscriptions sont partielles, elles permettent des manipulations qui mènent à la découverte de nouvelles connaissances, processus qui fait l'objet du chapitre suivant.



### 3. Découvrir

---

Jusqu'à présent, j'ai envisagé les enrichissements de matériaux en examinant la manière dont ils conditionnent l'accessibilité aux inscriptions ( $\rightarrow 1$ ) et les choix nécessaires pour produire ces dernières ( $\rightarrow 2$ ). J'ai considéré les enrichissements comme l'inscription des résultats de l'analyse. Dans ce chapitre, je voudrais montrer comment les inscriptions peuvent être manipulées pour faire émerger de nouvelles connaissances. J'ai beaucoup parlé de diagrammes dans les chapitres précédents et l'essentiel de mon travail de recherche fait usage de ce type d'inscription. Pour cette raison, le présent chapitre traitera surtout de *l'utilisation dynamique de diagrammes*. En guise de préambule, il est utile de préciser ce que l'on entend par là.

J'ai défini le diagramme comme l'inscription formelle graphique d'un réseau de relations ( $\rightarrow 0.2.2$ ) et l'on a vu que l'opération de réification, qui discrétise la connaissance dans le diagramme, était capitale ( $\rightarrow 2.2.2$ ). On mentionne souvent une caractéristique proprement peircienne du diagramme : sa capacité à être employé de manière dynamique pour ajouter un surcroît de sens qui fait avancer le raisonnement (Chauviré 2008 : 62). Selon Peirce, le diagramme permet un raisonnement qui ne se contente pas de l'observation de prémisses, mais qui ajoute du contenu pour mener à une nouvelle interprétation créative (Chauviré 2008 : 36-42). Étendant les conceptions de Peirce, Tylén *et al.* (2014) extraient de cette capacité à être manipulé les caractéristiques définitoires du diagramme. Selon eux, un diagramme peut : 1/ être déformé pour devenir plus conforme au réseau de relations qu'il représente ; 2/ être utilisé pour faire émerger de nouvelles relations par expérimentation concrète ; 3/ être découpé pour se focaliser sur une partie seulement de l'information représentée par abstraction des éléments non pertinents. Ces caractéristiques potentielles sont, je pense, communes à tous les signes représentant un réseau de relations. Or, toute relation est une

construction mentale pouvant être réévaluée, commentée, interprétée, etc. En d'autres termes, le diagramme peut être manipulé par une utilisation *dynamique* (Giardino 2011)<sup>37</sup>.

Comment les diagrammes peuvent-ils être employés de façon concrète ? D'une manière somme toute assez proche de celle dont j'ai procédé au chapitre précédent, je vais envisager les manipulations visant à mettre au jour des analyses spécifiques (→3.1), puis je me concentrerai sur celles qui permettent d'élaborer la théorie (→3.2) ou d'évaluer les structures formelles (→3.3).

### 3.1. Explorer les faits

Dans cette section, je présente comment la connaissance des faits peut se développer grâce à l'exploitation de diagrammes. Les cas examinés portent sur l'examen de connaissances qui ont déjà été filtrées pour être inscrites d'une manière spécifique à leur exploitation : les cartes linguistiques (→3.1.1) et les analyses statistiques (→3.1.2).

#### 3.1.1. Cartes linguistiques

Comme je l'ai mentionné ci-dessus (→1.1.2), les notices de l'ALW comprennent un tableau des formes qui est accompagné de cartes si le rédacteur l'a jugé opportun. L'ALW suit une logique similaire à celle du FEW pour construire les tableaux de formes (→2.2.2. b) et il se sert de cartes linguistiques pour faire émerger l'histoire des variétés qui constituent son objet. Dans mes travaux récents (Mazziotta 2016a), j'ai étudié ces cartes comme des inscriptions diagrammatiques et j'ai évalué un procédé qui m'a semblé d'une importance capitale pour leur élaboration : la réduction des mots-formes à des symboles. Le processus est tout à fait similaire à celui qui consiste à regrouper des lexies en fonction de critères et réifier ces regroupements comme des objets distincts.

À titre d'exemple, prenons la carte de la fig. 34 (p. 47), qui accompagne la notice onomasiologique « tour de reins » (Boutier 1997 : notice 108). La légende nous indique comment interpréter les symboles : chacun d'entre eux correspond à ce que l'on appelle un *type*, c'est-à-dire une unité lexicale abstraite qui neutralise les différences entre des unités qu'on considère comme partiellement équivalentes<sup>38</sup>. En l'occurrence, le trait horizontal représente le type 'tour de reins', qui fédère les formes *toûrs di rins*, *tour du rins*, etc. Le disque est lui-même l'abstraction qui fédère *clike*, *clinke*, etc. En somme, la carte inscrit une version réduite des données en donnant aux types une forme graphique explicite (en les réifiant ; →2.2.2. b). Les procédures de typisation ne sont pas systématiques, mais chaque carte peut être vue comme un diagramme du fait de la finitude de son

37 Bruno Bachimont (2007 : 32) fait remarquer que la discrétisation de la connaissance et la manipulation des inscriptions fondent l'« essence du numérique ». Je pense qu'il s'agit également d'une caractéristique inhérente aux diagrammes.

38 On pourrait avoir des types de lexies, mais en ce qui concerne l'ALW, on a affaire à des types de mots-formes.

système : la légende définit l'inventaire des symboles et leur disposition est contrainte par le fond de carte. C'est grâce à cette démarche que la géographie linguistique met en évidence l'histoire des unités linguistiques : l'utilisation d'entités différentes permet d'appréhender la forme de leur regroupement<sup>39</sup>. Dans la carte ci-dessous, on voit que la zone où figurent les traits horizontaux est pénétrée par la zone où figurent les disques et que de nombreux points comportent les deux symboles. Cette observation mise en lien avec une étude historique laisse l'auteur de la notice conclure que 'tour reins' a été remplacé par *clike* là où ce dernier apparaîtrait. Sans typisation (donc sans réduction), il serait impossible de visualiser quoi que ce soit.

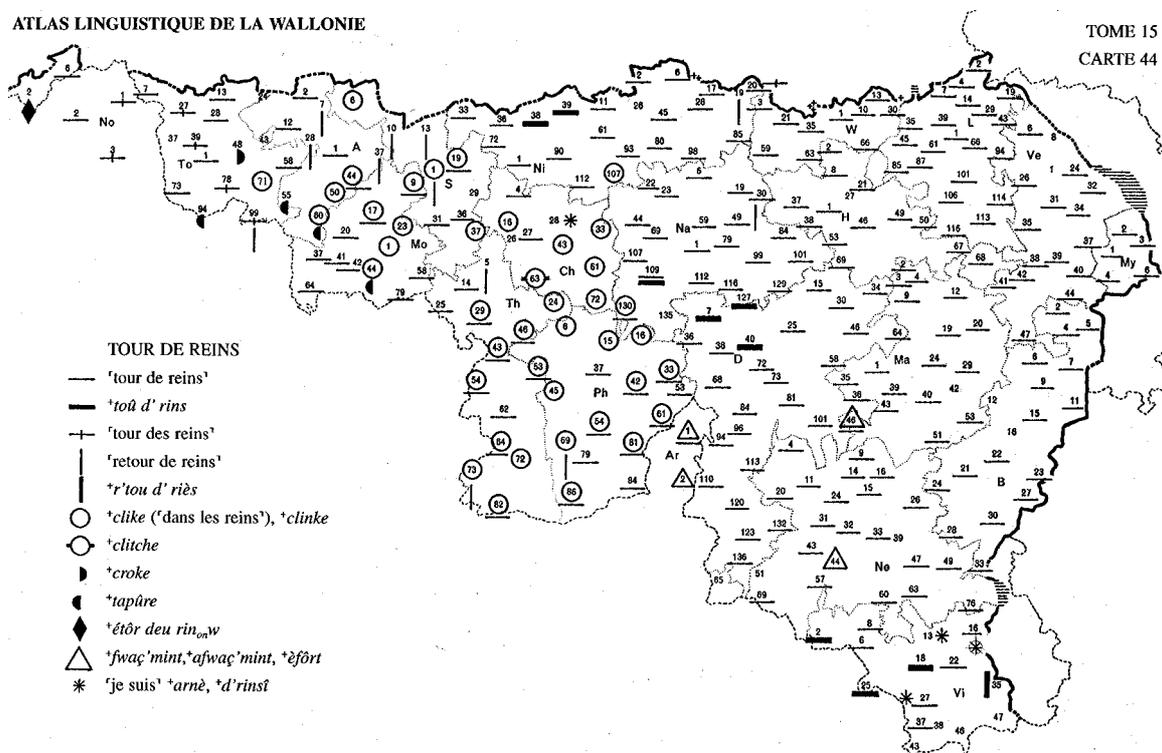


Figure 34 – Carte « tour de reins » (Boutier 1997 : notice 108)

### 3.1.2. Analyses statistiques et plans factoriels

Les diagrammes statistiques sont également conçus pour que leur lecteur y trouve un surcroît de sens. Je vais montrer plus loin dans cette sous-section quels diagrammes j'ai employés et comment je les ai manipulés. Cependant, j'aimerais tout d'abord rappeler quelques notions de base par rapport à l'usage des statistiques. Il est connu et évident pour tout chercheur qui recourt aux méthodes statistiques que les procédures de traitement des données ne sont pas utilisables si l'objet est trop complexe. Dans le cadre d'une étude statistique, les individus sont définis selon un ensemble de

39 On peut dire que plusieurs symboles contigus forment une « sur-entité » (Groupe  $\mu$  1992 : 148-152) qui distingue la zone qu'ils occupent du reste de l'aire géographique représentée. Cognitivement, le processus n'est pas toujours efficace : il dépend du nombre et des qualités des entités employées (v. notamment Bertin 2005).

variables qui permettent de les différencier ou d'établir leur similarité (Howell 1998). Pour reprendre l'exemple de ma thèse (Mazziotta 2009 : ch. 4 sv.), j'ai essayé de déterminer quelles structures syntaxiques attirent le plus la ponctuation ou, au contraire, la rejettent. Pour ce faire, j'ai décrit les caractéristiques de chaque construction par des variables comme : *fonction syntaxique*, *présence d'une conjonction/préposition*, *nature du gouverneur*, mais également *présence d'un signe de ponctuation initial* et *présence d'un signe de ponctuation finale*. Je passe le détail de la définition de ces individus. Il apparaît déjà au travers des variables mentionnées qu'elles décrivent l'individu de manière *directe*. Observons (21) (transcription imitative ; →1.1.1).

- (21) **jo nicholes firef de rumignjf · et de florinef · fai fauoir a tos ceiaz ki or funt et a uenir funt · que [...]** 'Moi, Nicolas, seigneur de Rumillies et de Florennes, fais savoir à tous ceux qui sont [= existent] à présent et qui seront que [...]' (Document 1236-07 : 1)

En utilisant les variables citées ci-dessus pour décrire *nicholes* et *sires de rumignis et de florines*, on construit la tab. 1 (0 indique une absence et 1 indique une présence).

<i>Forme</i>	<i>Fonction</i>	<i>Conjonction/préposition</i>	<i>Nature</i>	<i>Ponct. initiale</i>	<i>Ponct. finale</i>
nicholes	apposition	0	nom	0	0
sires de ...	apposition	0	nom	0	1

Table 1 – Description tabulaire de constructions

Comme on le voit, les variables sélectionnées ne disent rien du gouverneur commun des deux constructions (*jo*). Elles ne disent pas non plus que la seconde contient deux compléments coordonnés (*de rumignis et de florines*), qui plus est avec un signe de ponctuation devant la conjonction. On pourrait facilement accroître l'inventaire des informations qui « manquent ». On pourrait sans doute les ajouter en multipliant les colonnes. Indépendamment du fait qu'il est impossible de rendre compte des rapports indirects entre toutes les constructions (notamment à cause de la récursivité des structures), un trop grand nombre de variables a généralement comme conséquence que de nombreux cas s'en trouvent sous-représentés. Cela les rend peu exploitables par des méthodes statistiques. Ce n'est qu'en réduisant la description à un petit nombre de variables que j'ai pu employer des tests de significativité pour déterminer si différentes fonctions syntaxiques étaient associées ou non à la présence d'un signe de ponctuation à l'initiale ou à la finale. La réduction peut d'ailleurs être drastique, au point de limiter la description de chaque individu à deux variables pour évaluer la force de leur corrélation (Howell 1998 : 165-166). Quand, dans Glikman/Mazziotta 2013, nous avons essayé de voir quelles fonctions syntaxiques étaient caractéristiques de l'oral dit « représenté » en ancien français (Marchello-Nizia 2012), nous n'avons conservé que deux variables : une représentant les fonctions et

une autre représentant le type discursif où on les rencontre. En somme, le repérage des tendances (une connaissance nouvelle, inscrite de façon traditionnelle dans un texte scientifique) émerge de la réduction.

Toutes les études statistiques que j'ai menées par la suite ont évidemment suivi le même principe de réduction. Certaines d'entre elles (Mazziotta 2014a ; 2016b) ont mené à la construction de diagrammes statistiques appelés *plans factoriels*, qui relèvent de ce qu'on appelle « analyse factorielle des correspondances ». Il s'agit de méthodes heuristiques qui visent à mettre en évidence les liens les plus saillants entre les modalités définissant l'échantillon étudié<sup>40</sup>. La fig. 35 montre (d'une manière un peu cryptique que je vais expliquer partiellement) quelles sont les propriétés syntaxiques des lexèmes exprimant le désespoir en français médiéval : *chetif*, *dolent*, *las* et *maleüré*.

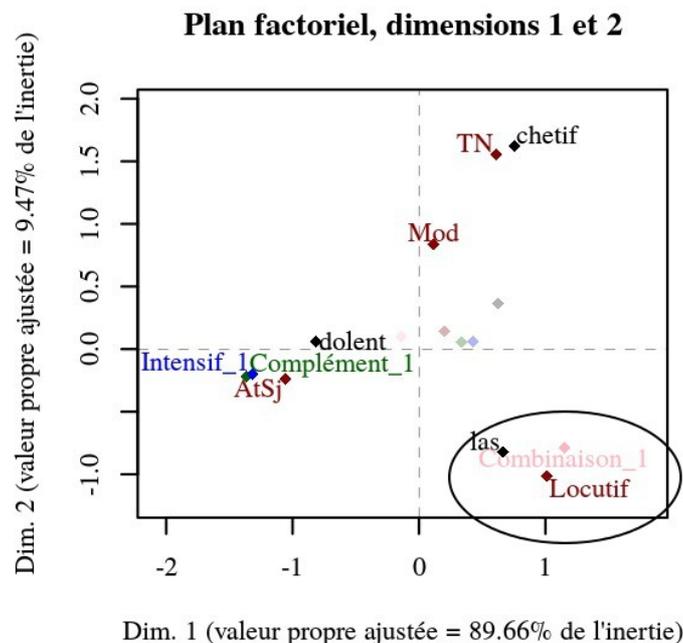


Figure 35 – Analyse factorielle des correspondances multiples des lexèmes exprimant le désespoir en français médiéval (Mazziotta 2016b : § 2.2 ; ellipse ajoutée ici)

À l'exception de l'ellipse en bas à droite, dont je vais reparler, ce type de graphique est généré par une série de traitements statistiques de manière entièrement formalisée. Il s'agit donc d'un diagramme (→0.2.3). L'objectif des calculs est de projeter sur un plan les tendances qu'ont certaines modalités (valeurs de variables) à apparaître ensemble pour le même individu, ou, au contraire, à s'exclure les unes les autres. En simplifiant fortement : les points les plus proches les uns des autres qui sont

<sup>40</sup> Voir Cibois 2000 pour une introduction accessible aux calculs nécessaires et Lebart *et al.* 1998 pour une description mathématique très complète de ces méthodes. Lebart/Salem 1994 introduit ces techniques de façon adaptée aux linguistes.

éloignés de l'origine (coordonnée [0,0]) indiquent que les modalités s'attirent. Par exemple, on lit dans le cadran inférieur droit du plan que le point étiqueté par le lexème *las* est proche du point étiqueté *Locutif* et de celui étiqueté *Combinaison\_1*. Suivant les conventions choisies par l'expérimentateur, le premier signifie que le mot se comporte comme un « locutif » (c'est-à-dire, *grosso modo*, une interjection →3.2.2) et le second signifie « qui apparaît en cooccurrence avec d'autres lexèmes exprimant le désespoir ». On peut verbaliser la conclusion qui s'ensuit comme ceci : « En français médiéval, *las* a tendance à être employé comme locutif avec d'autres lexèmes exprimant le désespoir. » Je ne vais pas exposer comment extraire toutes les informations contenues dans le graphique. Il est déjà clair que ce type d'inscription diagrammatique exprime une connaissance réduite d'une manière qui fait apparaître l'information. L'ellipse a été ajoutée manuellement et symbolise l'identification d'un regroupement. Cette altération du diagramme, rendue possible par sa substance graphique, montre comment il est possible de manipuler l'inscription pour en faire ressortir de manière non formalisée ce qu'on y juge pertinent.

À nouveau, c'est la réduction qui a rendu possible l'émergence de nouvelles connaissances. Le lecteur attentif aura en outre remarqué que le graphique ne comporte pas de point étiqueté *maleüré*, alors que j'annonçais ce lexème dans la liste de ceux qui sont intégrés à l'étude. Ce n'est pas accidentel. Les calculs préalables qui déterminent ce qui doit être représenté sur le plan sont chargés d'éliminer de l'inscription tous les points qui sont trop peu informatifs. Les traitements mathématiques sous-jacents à la création du diagramme contribuent à la réduction.

### 3.2. Élaborer la théorie

Les cas que je viens d'aborder montrent comment les données peuvent être explorées et les inscriptions manipulées ou comparées pour améliorer la connaissance des faits. Dans cette section, je me concentre sur le niveau d'abstraction supérieur, en montrant comment l'utilisation d'inscriptions diagrammatiques permet de justifier la théorie (→3.2.1) et d'objectiver les alternatives descriptives (→3.2.2).

#### 3.2.1. Démontrer

Voyons tout d'abord comment l'utilisation dynamique de diagrammes permet de dégager une nouvelle théorie. Lors de l'exploration du fonds Tesnière (BnF NAF28026), Sylvain Kahane et moi avons trouvé une fiche qui témoigne de la manière dont l'auteur a élaboré une partie du concept de « mot-phrase » (voir notre étude Mazziotta/Kahane 2016) en se servant de moyens purement graphiques. Il est clair que la fiche dont je parle est liée à une réflexion théorique approfondie plutôt

qu'au traitement d'une question particulière : elle se trouve dans un ensemble de documents consacrés aux « mots-phrases ». Tesnière analyse la réponse « Non, pas tout à fait », issue du dialogue (22).

(22) – As-tu fini tes devoirs ?

– Non, pas tout à fait.

La fig. 36 montre assez clairement quel a été son raisonnement.

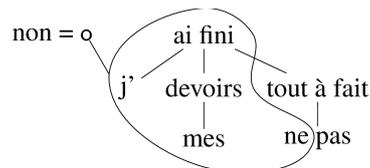


Figure 36 – Stemma de *Non pas tout à fait* (Tesnière, BnF NAF28026, B67, f367), version 1 (Mazziotta/Kahane 2016 : 79)

On voit qu'un trait courbe entoure une partie des mots du stemma qui correspond à (23) : *j', ai fini, devoirs, mes et ne*.

(23) Je n'ai pas tout à fait fini mes devoirs.

Le diagramme mélange le système stématique et les notations algébriques classiques pour montrer que *non* est équivalent à la portion du stemma qui est entourée. Cela a deux conséquences.

Premièrement, cela autorise Tesnière à produire le stemma de la fig. 37.



Figure 37 – Stemma de *Non pas tout à fait* (Tesnière, BnF NAF28026, B67, f367), version 2 (Mazziotta/Kahane 2016 : 89)

L'équivalence du mot *non* à l'ensemble formé par *j', ai fini, devoirs, mes et ne* fait que Tesnière place ce mot tout en haut de la structure syntaxique. Il aboutit donc à la conclusion que *non* se comporte comme un verbe à cet égard.

Deuxièmement, cela l'amène à dire :

Le seul élément qui ne puisse pas manquer dans les phrasillons est la notion verbale, qui fait que ces mots sont les équivalents de phrase entière. D'autre part si tous les autres éléments peuvent manquer, ils ne sauraient manquer tous à la fois sans quoi nous n'aurons plus un phrasillon mais un simple verbe. Nous aboutissons ainsi à définir le phrasillon comme un mot comportant, à l'état inanalysé, outre une notion verbale fondamentale et obligatoire, au moins un actant ou circonstant, et le plus souvent plusieurs. Ainsi défini, ce phrasillon est susceptible de former phrase à lui tout seul [...]. (Tesnière, BnF NAF 28026, bte 42, chemise 1, § 183 ; édité dans Mazziotta/Kahane 2016)

Je pense que c'est la mise en diagrammes, puis leur comparaison qui a permis à Tesnière de se forger une idée théorique du fonctionnement des « mots-phrases » et de démontrer le bien-fondé de sa

conception : ils contiennent la notion verbale. Mettre côte à côte les deux stemmas a des allures de démonstration.

### 3.2.2. Tester et sélectionner

Au travers d'un prolongement moderne de cette notion de « mot-phrase », je vais montrer à présent comment les diagrammes syntaxiques sont également des outils pour la sélection entre des hypothèses alternatives. Avec Sylvain Kahane, nous avons montré comment l'analyse de Tesnière quant à la classe des « mots-phrases » pouvait être révisée (Kahane/Mazziotta 2015b). Nous avons ainsi proposé de distinguer les *prédicatifs* (mots équivalents à une proposition, comme *oui* et *non*) des *locutifs* (mots équivalents à un énoncé complet, comme *aïe !*, *chut !*). Durant les travaux préparatoires à l'article, j'ai essayé de construire des diagrammes pour représenter les relations entre les locutifs et les autres mots. Comme il correspond à un énoncé complet, le locutif a une visée illocutoire (un effet pragmatique attendu par son énonciateur, comme donner un ordre ou poser une question ; v. Searle/Vanderveken 1985 : 37-40, 51-62) et cette visée illocutoire est figée dans le lexique. En conséquence, les locutifs ne sont jamais subordonnés et ne peuvent pas changer de visée illocutoire. Partant de ce principe, j'ai réfléchi à l'analyse que nous ferions de *est-ce que... ?*, qui est bien un locutif, puisqu'il ne peut servir qu'à poser une question. J'ai donc produit les diagrammes de la fig. 38 pour analyser l'énoncé (24).

(24) Est-ce que Jean dort ?

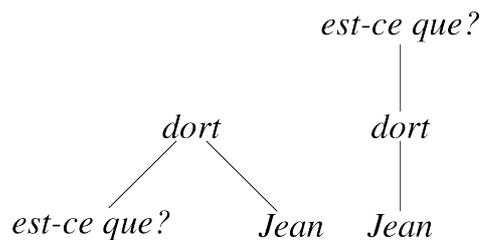


Figure 38 – Analyses alternatives du locutif *est-ce que... ?*

Il a fallu choisir entre ces diagrammes celui qui correspondait à notre approche théorique. Seul celui de droite est acceptable quand on considère le rapport entre diagramme et théorie : les propriétés des locutifs ne sont pas respectées dans l'autre diagramme.

On visualise ainsi que : 1/ *est-ce que... ?* est tout en haut de la hiérarchie syntaxique ; 2/ il a un dépendant. En conséquence, *est-ce que... ?* est capable de régir un verbe fini (un prédicatif selon notre approche). Cela signifie qu'il existe bien des locutifs qui ont cette propriété. Aucun des diagrammes de la fig. 38 ne démontre réellement quoi ce que soit, mais leur comparaison permet d'objectiver des différences et de sélectionner la meilleure analyse<sup>41</sup>.

41 J'ai suivi exactement la même démarche en analysant la construction *Las moi !* 'Pauvre de moi !' en français médiéval

Dans cette sous-section et la précédente ( $\rightarrow$ 3.2.1), j'ai montré que les diagrammes étaient des outils qui, au départ de l'analyse de faits, permettent de tirer des conclusions au niveau de la théorie linguistique avec laquelle ils sont mis en rapport.

### 3.3. Évaluer les structures formelles

Pour le linguiste, plus la recherche s'éloigne des matériaux, plus elle devient abstraite. J'ai montré ci-dessus comment les inscriptions pouvaient aider à générer des analyses nouvelles de ces matériaux ( $\rightarrow$ 3.1), mais aussi, au niveau d'abstraction supérieur, comment elles aidaient à élaborer la théorie linguistique ( $\rightarrow$ 3.2). Dans cette dernière section du chapitre, je monte encore d'un niveau d'abstraction en me concentrant sur les structures formelles dédiées à la modélisation syntaxique. J'aborde successivement comment la manipulation de diagrammes permet de changer le formalisme ( $\rightarrow$ 3.3.1) et comment jauger la différence entre inscriptions diagrammatiques ( $\rightarrow$ 3.3.2).

#### 3.3.1. Changer le formalisme

Par sa nature graphique, le diagramme autorise l'utilisateur à sortir du système formel pour laisser libre cours à sa créativité. Si les ajouts sont également formalisables, cela peut mener à étendre le système initial ( $\rightarrow$ a). Cette même propriété nous autorise à transformer progressivement les diagrammes en changeant petit à petit la manière dont les relations qu'ils expriment sont représentées. Ce genre de démarche permet de justifier que certaines structures formelles négligées jusqu'à présent peuvent être exploitées en syntaxe ( $\rightarrow$ b).

##### a. Enrichir le système formel

Prenons un exemple banal d'une utilisation dynamique d'un stemma de Tesnière. On voit dans la fig. 39 (p. 54) qu'il a reporté en rouge l'ordre linéaire des mots de la phrase latine dont le stemma est tracé en noir ; l'ordre linéaire de la phrase française correspondante figure en vert. Par cette simple diagrammatisation, Tesnière compare l'ordre des mots en latin et en français (il démontre le contraste de manière fine et détaillée). Mais ce n'est pas tout : les flèches de couleurs obéissent à des règles strictes qui se combinent à celles du stemma. D'une certaine manière, l'inscription formelle du stemma se voit ajouter une autre inscription formelle qui ne contredit pas les règles de la première, mais en ajoute. Ainsi, au niveau du diagramme pris dans son ensemble, on assiste à un accroissement de l'inventaire des entités visuelles et les conventions qui président à leur configuration. Le système formel a été enrichi.

---

(Mazziotta 2016d : § 4.1.4).

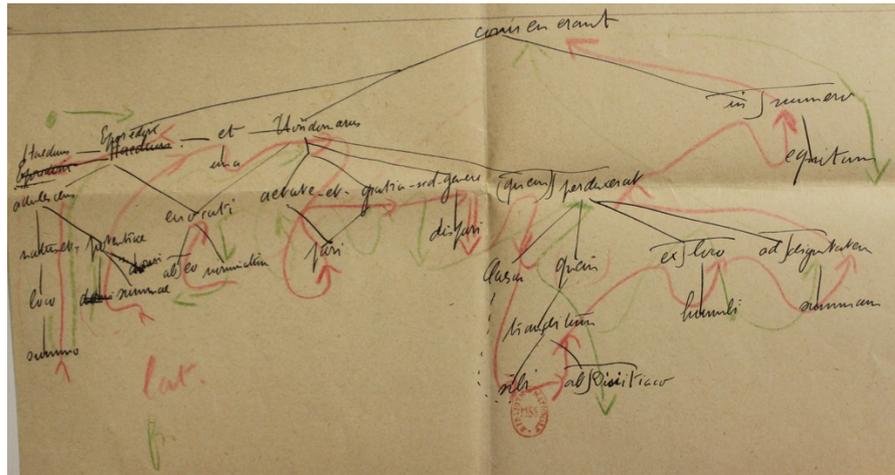


Figure 39 – Emploi dynamique d'un stemma (Tesnière, BnF NAF28026, B63, f68)

Parce qu'il annonçait avant la lettre la notion de polygraphe dont j'ai parlé ci-dessus (→2.3.2), un autre stemma enrichissant le système de Tesnière m'a particulièrement marqué<sup>42</sup>. Il est reproduit dans la fig. 40.

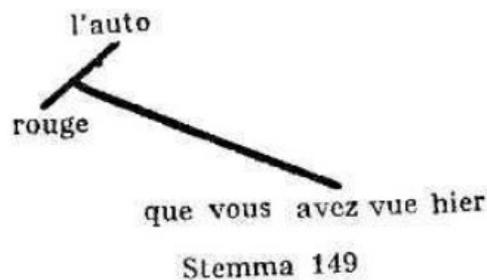


Figure 40 – Stemma 149 (Tesnière 1959 : 65/25)

Tesnière associe à ce stemma le commentaire suivant (1959 : ch. 65, § 25) :

Pour représenter ces nuances, qui mériteraient une analyse plus poussée qu'on ne peut la faire ici, il est possible (mais non nécessaire dans la pratique courante des stemmas) d'adopter des traits de connexion plus longs pour les subordonnés plus éloignés et de faire partir ces traits de connexion, non pas du nœud lui-même, mais d'un point du trait de connexion qui unit le substantif à un de ses subordonnés proches. C'est par ce procédé que, dans le membre de phrase *l'auto rouge que vous avez vue hier*, il est possible de faire apparaître sur le stemma que le subordonné *que vous avez vue hier* détermine moins le substantif *l'auto* lui-même que le groupe *l'auto rouge*. En effet ce que vous avez vu hier, ce n'est pas *l'auto*, c'est *l'auto rouge* (v. st. 149).

Le type d'inscription graphique de la fig. 40 ne correspond ni au dessin conventionnel d'un arbre, ni à celui d'un graphe, ni à celui des stemmas auxquels nous sommes habitués. Les contraintes de ces structures ne sont apparemment pas respectées : un trait a un autre trait comme sommet, alors que les

<sup>42</sup> Il y a d'autres stemmas qui enrichissent le système tesnérien ; parmi eux, les stemmas 215 (qui représente plusieurs dépendants d'un nom situés sous lui à une distance variable), 296 (qui comporte une accolade) mériteraient sans doute une analyse poussée.

arêtes des arbres, les graphes et les stemmas n'ont que des mots comme sommets. On a plutôt affaire à un polygraphe ( $\rightarrow$ 2.3.2), introduit par Tesnière des décennies avant que nous ne proposons le terme<sup>43</sup>.

Qu'avons-nous vu dans ce paragraphe ? En explorant les brouillons inédits de Tesnière, j'ai observé comment il manipulait ses propres stemmas. Il n'y a pas de flèches colorées dans les stemmas de Tesnière auxquels on est habitués, mais on pourrait sans peine articuler la chaîne exprimant l'ordre des mots au reste de la structure formelle des stemmas sans « casser » leur fonctionnement. De même, il n'y a qu'un seul stemma où un trait part d'un autre trait, mais cet ajout à la structure des stemmas ne remet pas en cause les conventions qu'on leur connaît par ailleurs. Tesnière enrichit ainsi implicitement les structures formelles qu'il utilise. La question qui se pose à ce stade est celle de l'explicitation et de la justification des enrichissements de ce type. On va voir comment cette démarche peut être faite par le biais d'un raisonnement fondé sur les diagrammes par l'exemple de l'élaboration de la notion de *polygraphe*.

### b. Élaborer de nouvelles structures

La notion de *polygraphe* ( $\rightarrow$ 2.3.2) a trouvé sa justification au travers de la manipulation de diagrammes. Il est fondamental de souligner d'emblée que les inscriptions graphiques qui suivent ne consistent pas qu'en des illustrations de la théorie : c'est bel et bien par elles que le concept de *polygraphe* a pu être élaboré. C'est la comparaison de diagrammes qui nous a fait prendre conscience qu'une structure formelle différente de celles que nous avons l'habitude de pratiquer pouvait être envisagée.

Tout d'abord, si l'on considère (25), l'analyse simplifiée en constituants (non X-barre) peut être inscrite sous la forme d'une parenthésiation (26) ou sous celle d'un diagramme arborescent classique (fig. 41), où les traits représentent des relations *partie-tout*, qui unissent les constructions à leurs constituants immédiats.

(25) Mary loves red cars

(26) (S Mary<sup>N</sup> (VP loves<sup>V</sup> (NP red<sup>Adj</sup> cars<sup>N</sup>)))

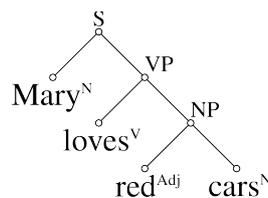


Figure 41 – Arbre de constituants (Kahane/Mazziotta 2015c : 154)

43 On trouve des diagrammes dont la configuration est celle des polygraphes, c'est-à-dire où les traits rejoignent d'autres traits, chez d'autres pionniers (notamment Nida 1966 ou dans les diagrammatisations de la pensée de Gustave Guillaume ; p.ex. Soutet 2005 : 291) et dans certaines approches modernes Van Raemdonck 2009.

On conviendra aisément que le diagramme de Venn de la fig. 42 représente exactement la même hiérarchie de constituants que l'arbre de la fig. 41 (v. Lima 2014 sur la variété des représentations graphiques des structures arborescentes). Les deux diagrammes ont la même signification.

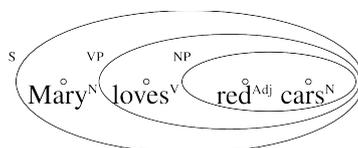


Figure 42 – Arbre de constituants représenté par un diagramme de Venn

Les entités qui les composent et la manière dont elles sont agencées servent à inscrire la même analyse syntaxique. Du point de vue de la signification de ces diagrammes, les différences qui les opposent sont contingentes et non pertinentes. Dans ce genre d'inscription, les bulles réifient simultanément les constructions et le rapport qui les lie à leurs constituants immédiats, alors que dans la fig. 41 les deux éléments sont réifiés par des entités différentes (disques et traits). Dans la fig. 42, on peut dire que la relation *partie-tout* est *déréifiée*.

Dès que l'on réalise que chacune des bulles de la fig. 42 peut aussi être la figuration du lien qui unit les éléments qu'elle contient, on peut facilement représenter ce lien par un trait qui relie ces éléments au lieu d'une bulle qui les entoure. La fig. 43 montre comment la même structure peut être diagrammatisée en utilisant des disques étiquetés pour réifier les mots et des traits étiquetés pour réifier simultanément les constructions et le lien qui les unit à leurs parties, en faisant courir les traits entre chacun des disques.

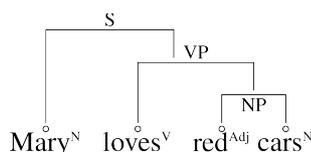


Figure 43 – Arbre de constituants représenté par une hiérarchie de traits (Kahane/Mazziotta 2015c : 157)

On voit notamment que le trait étiqueté *NP* relie le disque *red* au disque *car* et que le trait étiqueté *S* court du disque qui correspond à *Mary* au trait qui correspond au syntagme verbal. Selon les règles qui présideraient à ce type d'inscription graphique formalisée, les traits peuvent donc relier des disques et d'autres traits. Du point de vue de l'organisation des structures syntaxiques représentées, ces dernières sont stratifiées (Kahane 1997 ; →2.2.2. a) : l'agrégation de mots se fait par couches successives qui sont inévitables dans une analyse en constituants immédiats classique.

Ainsi est née la notion de *polygraphe*, par l'observation de la dérémification de la relation *partie-tout* dans différentes inscriptions équivalentes de la même connaissance. C'est la manipulation de diagrammes qui fait émerger la notion. On comprend que la notion d'*équivalence* entre les diagrammes

a joué là un rôle essentiel. Une question reste donc à approfondir : celle de l'enjeu du repérage des similitudes et des différences entre les diagrammes.

### 3.3.2. Contraster les diagrammes

Comme cela apparaît dans l'exposé des étapes qui ont mené à la notion de *polygraphe*, contraster des diagrammes inscrivant des connaissances similaires de façon différente permet de constater que le contenu peut demeurer équivalent malgré les variations de forme d'expression. Je parle dans un premier temps de l'inscription de connaissances identiques dans des diagrammes différents ( $\rightarrow$ a), avant de venir à celle du repérage des véritables différences de contenu dans des cas qui pourraient sembler similaires ( $\rightarrow$ b).

#### a. Comprendre l'équivalence

Qu'est-ce qui fait que deux diagrammes différents peuvent être considérés comme représentant la même connaissance ? Pour commencer à répondre à cette question, je vais prendre un exemple issu de ma recherche doctorale. Pour des raisons techniques et des contraintes contingentes, j'ai élaboré deux sortes de schémas pour travailler sur mon corpus de thèse. La fig. 44 et la fig. 45 représentent des relations et des groupements d'observables tirés de (27) (le *et* initial n'est pas représenté dans le premier diagramme).

(27) et aura li glise le deus et li voweiz le tierce [...] 'L'église aura la deuxième [partie] et l'avoué la troisième' (Document 1244-01-19 : 6)

La fig. 44 présente l'analyse d'une manière adaptée à l'inscription traditionnelle.

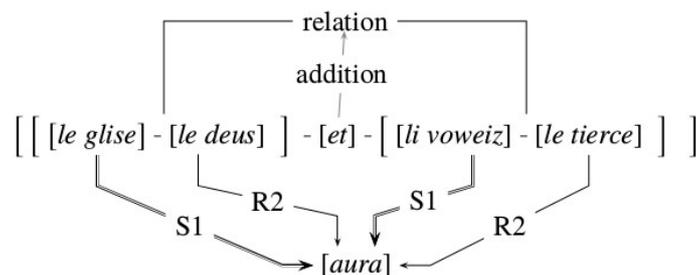


Figure 44 – Représentation de la structure de *aura le glise le deus et li voweiz le tierce* (Mazziotta 2009 : 143)

Je me limite ici au minimum d'explications pour rendre la fig. 44 compréhensible. Les flèches indiquent des relations dépendanciennes et les spécifications (suivant une logique qui préfigurait les polygraphes ;  $\rightarrow$ 2.3.2. b), alors que les crochets réifient les ensembles de mots et indiquent une relation de coïnclusion. Le diagramme est assez chargé du fait que les deux logiques se combinent. Dans le texte de mon étude, j'introduis les conventions de manière progressive, si bien que le lecteur est formé

à lire relativement facilement ce type de diagramme, où l'agencement et la nature des associations sont rendus explicites.

Toutefois, la fig. 44 est dessinée manuellement et le temps nécessaire à sa confection est très important. Or, ma thèse reposant sur des dizaines de milliers de constructions syntaxiques encodées, j'avais besoin de pouvoir en visualiser de nombreuses rapidement. J'ai donc opté pour des moyens de génération automatique des diagrammes (sur la dimension informatique de ce genre de travail, v. →4.2.2. b) sans vraiment me soucier de parfaire leur ergonomie.



Figure 45 – Représentation de la structure de *et aura le glise le deus et li voiez le tierce* (Mazziotta 2009 : annexes)

Ainsi, la fig. 45 a été construite en employant uniquement des boîtes et des labels, qui réifient les constructions, mais laissent les relations de dépendance exprimées de façon très indirecte. La lecture en devient plus compliquée. Les relations sont indiquées par la coïncidence de deux ou plusieurs unités dans la même boîte, l'orientation des dépendances est implicite et liée aux propriétés théoriques des unités, identifiées par des étiquettes : p.ex., si *P0* est coïncidant avec quoi que ce soit d'autre, c'est *P0* qui gouverne la relation. L'économie que représente l'emploi exclusif de boîtes rend le diagramme apparemment simple, mais dans les faits beaucoup plus difficile à lire. Certaines relations ne sont pas représentées par des éléments distincts. Il faut donc apprendre qu'une boîte étiquetée *S1* (c'est-à-dire *sujet*) dans une boîte étiquetée *args* qui se trouve dans une boîte *groupe* est considérée comme coïncidant avec les éléments de même niveau que le *groupe* ; *li glise* est donc le sujet de *aura*. Comprendre cette logique est une chose, lire les diagrammes en l'appliquant en est une autre...

Parce qu'il est possible d'expliquer comment la fig. 45 représente une analyse du même type que la fig. 44, on voit par cet exemple que deux diagrammes aussi différents inscrivent en réalité exactement la même connaissance. Les différences entre les deux inscriptions se situent au niveau de ce qui est réifié au point de vue graphique (→2.2.2) et de leur configuration. Il y a équivalence formelle, dans la mesure où l'on pourrait écrire des règles de conversion entre les deux diagrammes pour transformer l'un en l'autre sans la moindre perte d'information. Cependant, même si les deux diagrammes véhiculent la même connaissance, les différences visuelles qui les distinguent ont un impact ergonomique et sémiotique majeur. C'est une question importante, dans la mesure où l'équivalence entre inscriptions concurrentes est parfois difficile à concevoir, si bien que l'on accorde

une importance démesurée sur le plan conceptuel à des différences qui se situent sur le plan de l'expression.

La prise de décision primordiale concerne le choix de ce qui est réifié et de ce qui ne l'est pas. Je vais poursuivre en expliquant comment l'on peut proposer des formalisations différentes du concept de *translation* chez Tesnière, que je vais commencer par présenter brièvement. Dans certains cas, l'emploi d'un outil grammatical spécialisé (préposition, conjonction, désinence,...) permet de contourner les règles de combinaison des espèces de mots (Tesnière 1959 : 3<sup>e</sup> partie). Dans (28), *livre* est un nom et *Alfred* l'est également.

(28) le livre d'Alfred

Or, selon Tesnière un nom ne peut régir un autre nom si ce dernier ne subit pas une transformation spéciale qui lui permet, généralement grâce à une forme grammaticale, de se comporter comme un adjectif. Tesnière nomme cette opération *translation*. Il la représente dans ses stemmas sous la forme d'un <T> stylisé autour duquel se placent l'outil grammatical utilisé (dit *translatif*), le terme qui subit la translation (dit *transférénde*) et le symbole de la partie du discours résultante. Dans la fig. 46, qui inscrit l'analyse de *le livre d'Alfred*, on voit que le symbole <A> (adjectif) surplombe la barre horizontale du <T>, tandis que le translatif *d'* et le transférende *Alfred* figurent de part et d'autre de la hampe, dont la partie inférieure est inclinée du côté du translatif.

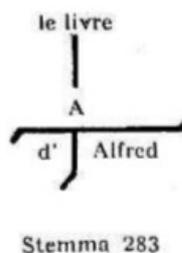


Figure 46 – Stemma contenant une translation (Tesnière 1959 : stemma 283)

Plusieurs formalisations alternatives de la translation sont possibles, principalement parce que le dispositif graphique de Tesnière se prête à plusieurs interprétations. L'inscription diagrammatique est ici plus éloignée des diagrammes de graphes et d'arbres classiques, ou même de polygraphes, et les chercheurs ont proposé des modélisations alternatives pour rendre compte de la translation dans une structure formelle moderne. Assez naturellement, Sylvain Kahane (Kahane/Osborne 2015 : § 4.9.2) a produit la même analyse que moi (Mazziotta 2014c) et nous avons proposé ensemble de réinterpréter les stemmas en nous servant de polygraphes (Kahane/Mazziotta 2015a : § 5). Selon nous, on peut modéliser la translation de façon similaire à une connexion ou à une jonction (→1.2.1) : au moyen d'une arête. Soit :

1. La translation est encodée au moyen d'une arête contenant les nœuds correspondant au translatif (étiquetés comme tel) et au transférende.
2. L'étiquette de l'arête encodant la translation correspond au résultat de la translation.

Si l'on produit une diagrammatisation de ce genre de formalisation pour l'analyse de (28), on obtient la fig. 47 (j'ai volontairement omis les étiquettes de parties du discours et d'identification du translatif), qui est un polygraphe de structure assez simple, où chaque construction du linguiste est représenté par un trait. L'inscription exploite le potentiel du polygraphe à représenter des relations de contraintes de présence ( $\rightarrow$  2.3.2. b), mais selon les conceptions de Tesnière, pour qui la translation conditionne nécessairement les capacités combinatoires du terme qui la subit.

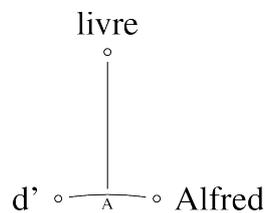


Figure 47 – Inscription polygraphique de la translation

Cette manière de décrire la translation chez Tesnière n'est pas la seule acceptable. Ainsi, Timothy Osborne (Kahane/Osborne 2015 : § 4.9.1) propose quant à lui une inscription de la translation qui, une fois transposée dans nos cadres d'analyse, réifie l'opération de translation sous la forme d'un disque indépendant et son rapport au translatif et au transférende au moyen de traits. Il obtient de la sorte l'arbre de la fig. 48.

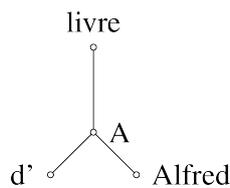


Figure 48 – Inscription arborescente de la translation

Osborne produit une inscription diagrammatique. S'appuyant sur cette dernière, il remarque également que la translation introduit une structuration proche de l'analyse en constituants immédiats dans le modèle tesniérien.

On voit donc mieux ici que pour les autres concepts précédemment traités à quel point la démarche est tributaire de l'interprétation. La conversion de l'arbre d'Osborne en polygraphe peut se faire en appliquant une règle simple qui fusionne les réifications ( $\rightarrow$  2.2.2. a) du résultat de la translation et ses rapports aux termes qui y participent en un seul objet. Cette opération n'engendre aucune perte de contenu et est réversible. Les deux formalisations sont donc équivalentes. On pourra

avoir l'impression que les objets réifient des choses différentes et s'échiner à défendre les présupposés théoriques de l'une ou de l'autre, cette équivalence demeurera indubitable.

Il n'en reste pas moins que les formalisations concurrentes, si elles sont acceptables et équivalentes, peuvent néanmoins être hiérarchisées en fonction de critères comme l'économie du modèle, la cohérence sémiotique ou l'exploitabilité technique. Ainsi :

1. Le polygraphe utilise moins d'objets conceptuels pour exprimer la même connaissance (critère de l'économie).
2. Le diagramme du polygraphe distingue nettement les analyses, sous forme de traits, des mots, sous forme de disques étiquetés, ce qui n'est pas le cas de l'arbre d'Osborne (critère de la cohérence sémiotique).
3. L'arbre d'Osborne est plus facile à traiter avec les outils informatiques largement disponibles actuellement (critère technique).

La comparaison de deux formalisations concurrentes mais inscrivant la même connaissance met donc ici en évidence que le débat à propos du choix de l'une ou de l'autre est sans doute plus fructueux s'il est mené en dehors de la question de la connaissance qu'elles expriment. Cette dernière est strictement identique dans les deux cas.

#### **b. Comprendre la différence**

La comparaison permet de mettre en évidence des similarités, mais aussi des différences. Il est tentant, face à des diagrammes différents produits par le même auteur, de chercher à repérer ce qu'ils ont d'équivalent et de négliger leurs divergences. Voyons cela à nouveau au travers d'un exemple qui a surgi au gré de l'exploration du fonds Tesnière à la BnF. Sylvain Kahane et moi y avons retrouvé les premiers brouillons des stemmas de Lucien Tesnière et relu son article de 1934 : « Comment construire une syntaxe » (v. Swiggers 1994 pour une synthèse des apports de cet article). Le fait le plus marquant est que les conventions de construction des stemmas ont changé entre cet article et les *Éléments...*

Si l'on compare les stemmas classiques (comme celui de la fig. 13, repris ici dans la fig. 49) aux stemmas primitifs de la fig. 50, on voit immédiatement que la logique de structuration n'est pas identique. Vu qu'il s'agit de diagrammes élaborés par le même auteur, on pourrait s'attendre à retomber dans le cas de figure précédent, où il est clair que deux inscriptions concurrentes de la même connaissance sont équivalentes au niveau formel (→3.3.2. a).

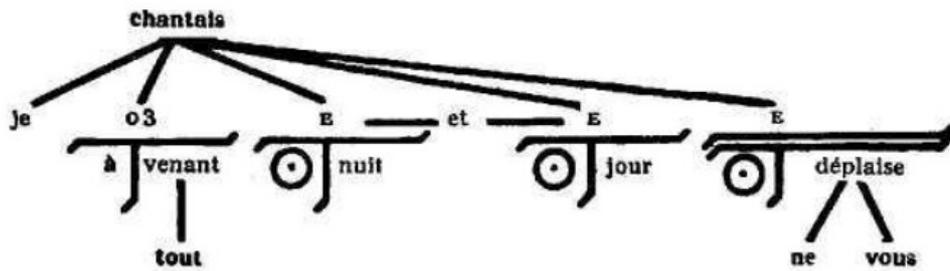


Figure 49 – Stemma de *Je chantais à tout venant nuit et jour, ne vous déplaise* (reprise de la fig. 13)

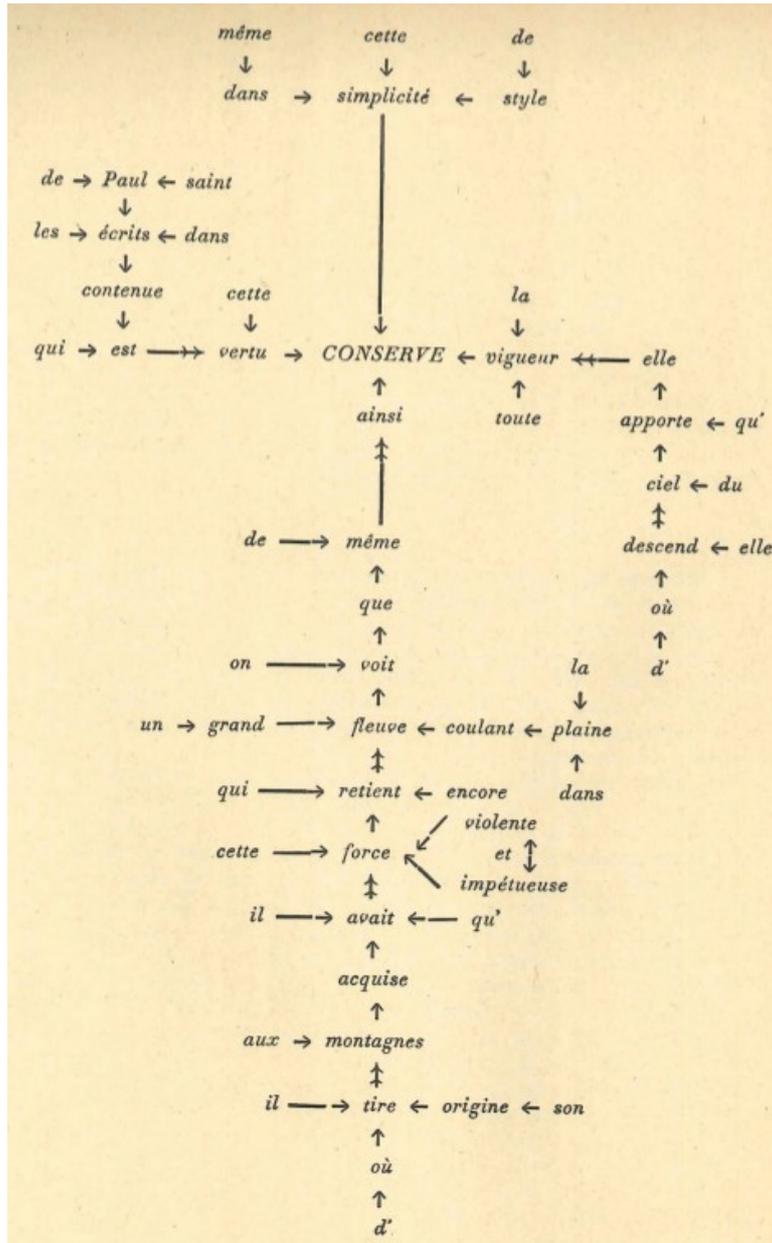


Figure 50 – Premier stemma d’une phrase française par Tesnière (1934 : 224)

Évaluons le détail des différences (Mazziotta 2016c). Au-delà du fait que l’inventaire des entités mobilisées n’est pas identique, on met principalement en évidence deux contrastes<sup>44</sup>. Le premier

44 Outre les deux contrastes décrits ci-dessous, la conjonction de subordination n’est pas traitée similairement et les nucléus sont absents des stemmas primitifs. Je ne m’étendrai pas sur ces questions.

contraste est en effet purement sémiotique. L'exploitation de l'espace n'est pas similaire : la hiérarchie des stemmas classiques procède d'une sémiotisation de l'espace, alors que les stemmas primitifs présentent plutôt une notation symbolique de la hiérarchie. Ainsi, les règles qui président à la construction des stemmas classiques impliquent que la connexion et la jonction sont exprimées par la configuration relative des mots et des traits : un trait entre deux mots qui ne sont pas situés à la même altitude exprime une connexion, alors qu'un trait entre deux mots situés au même niveau sur l'axe vertical exprime une jonction. Par comparaison, pour les stemmas primitifs, la hiérarchie est représentée par différents types de flèches, c'est-à-dire des moyens symboliques plutôt que configurationnels. Il n'en demeure pas moins que les rapports inscrits restent des connexions et des jonctions quel que soit le diagramme considéré.

Le second contraste met le stemma en relation avec la théorie : la théorie de la translation n'est pas intégrée aux stemmas de 1934. On voit dans le groupe en haut à gauche que *les écrits de Paul* ne représentent pas explicitement une relation particulière entre *de* et *Paul*, alors que les stemmas classiques mobilisent une entité spécifiquement pensée pour inscrire leur interaction syntaxique. Il y a donc bien une différence de contenu entre les deux types de diagrammes, puisque les stemmas les plus anciens n'inscrivent pas la translation. S'il est donc possible de poser des règles de correspondance entre les stemmas primitifs et les stemmas classiques pour les convertir les uns dans les autres, ces règles se heurtent à un obstacle irréductible : une part de la connaissance n'est pas représentée dans les stemmas primitifs. C'est là que se situe le plus fort contraste.

Au travers des analyses de cas que je viens de passer en revue, il est apparu que l'évaluation des différences et des points communs entre les diagrammes se fait à deux niveaux distincts : celui de la structure formelle et graphique et celui de l'information que cette dernière véhicule. Si l'on peut expliquer comment retrouver la même information dans deux inscriptions fondées sur des structures différentes, on peut dire que ces deux inscriptions sont équivalentes. On peut les convertir l'une dans l'autre sans perte d'information.

### 3.4. Synthèse

Comme on l'a vu dans ce chapitre, nombre de démarches menant à la découverte de nouvelles connaissances sur les faits reposent sur la manipulation d'inscriptions diagrammatiques, qu'il s'agisse de diagrammes syntaxiques, de cartes ou encore de graphiques statistiques ( $\rightarrow$ 3.1.1 et  $\rightarrow$ 3.1.2). Les cas observés ici, fondés sur l'exploitation d'inscriptions déjà constituées, ont la particularité de réduire encore davantage la connaissance inscrite pour focaliser l'analyse sur la part jugée pertinente (ce qui ramène aux choix exposés sous  $\rightarrow$ 2). Si les manipulations de diagrammes permettent de mieux

comprendre les faits, elles sont aussi capitales pour élaborer la théorie : elles offrent les moyens de démontrer (→3.2.1) et d'objectiver des alternatives (→3.2.2). De cette manière, l'élaboration conceptuelle prend appui sur la connaissance inscrite et rendue manipulable. Cet enrichissement conceptuel peut également avoir lieu au niveau le plus formel. Les deux questions que j'ai traitées ici sont celles de la découverte de nouvelles structures, et en particulier celle du *polygraphe* (→3.3.1. b) et de l'utilisation du raisonnement formel pour comparer la connaissance inscrite dans des diagrammes apparemment différents (→3.3.2. a). Il en est ressorti que les opérations effectuées sur les diagrammes démontrent les véritables différences qui les caractérisent, tout en faisant ressortir leurs points communs (→3.3.2. b).

Pour la plupart, les découvertes sur la base d'inscriptions diagrammatiques reposent en fin de compte sur deux opérations complémentaires : l'enrichissement et la comparaison. On a vu que bien que formels, les diagrammes autorisaient l'adjonction d'éléments extérieurs à leur logique de structuration, de façon non formalisée (comme les ellipses ajoutées sur les plans factoriels ; →3.1.2) ou en modifiant le système formel par des ajouts qui lui sont compatibles (→3.3.1. a). La comparaison est également primordiale, dans la mesure où elle donne les moyens non seulement de développer les théories linguistiques (→3.2), mais également de manipuler les formalismes (→3.3.1. b).

Si toutes ces manipulations sont possibles pour des inscriptions traditionnelles (c'est-à-dire non informatiques ; →0.2), les possibilités sont décuplées pour les inscriptions qui se déploient sur un support informatique. Comment ce dernier affecte-t-il toutes les opérations que j'ai explorées jusqu'à présent : *lire*, *choisir* et *découvrir* ? Dans le but de traiter cette question, le prochain chapitre se concentre sur l'apport des inscriptions informatiques.

## 4. Numériser

---

Bien que je ne l'aie que très évasivement évoqué, un aspect dominant d'une partie des travaux que j'ai détaillés ci-dessus est leur lien avec le numérique. Ce dernier se situe au cœur de mon travail concret ; notamment :

- Les chartes que j'ai transcrites ont été encodées dans un corpus électronique, de même que les textes que j'ai étudiés par la suite.
- La question de la formalisation des analyses syntaxiques de SRCMF s'est posée dès qu'il a fallu les encoder concrètement.
- La réflexion que j'ai menée sur le FEW et l'ALW vise à leur informatisation et donc à la transposition fidèle de leur organisation.

En ce qui me concerne, la plupart des réflexions dont je fais part dans les chapitres précédents ont le numérique pour origine. Car comme le dit Robert Martin dans son analyse de la « rétroconversion » du *Trésor de la langue française* en une version informatisée (2003 : 14), le numérique force à expliciter de manière univoque. En effet, inscrire une connaissance dans un medium informatique dans le but de l'y manipuler pousse à la formalisation<sup>45</sup> ; et la formalisation procède de l'explicitation. J'en suis venu à me questionner sur les fondements épistémologiques de mes pratiques en voulant me servir des ordinateurs, en concevant des *inscriptions informatiques*. Il n'est pas inutile que je commence par définir cette locution.

La locution *inscription informatique* s'oppose à *inscription traditionnelle* et désigne une forme spécifique de l'inscription *formelle* : il s'agit d'une inscription spécialement conçue pour être interprétée par des moyens informatiques. Je distingue donc l'*encodage informatique* de l'encodage

---

45 À moins de s'en tenir à une numérisation naïve qui ne tient pas compte des spécificités du médium numérique (v. à ce propos Bertin 2009 sur les dictionnaires).

formel selon les mêmes critères. Je ne m'estime pas compétent pour entrer dans une discussion approfondie sur le statut philosophique des inscriptions informatiques (Bachimont 2007). Ces dernières permettent l'archivage et la conversion de la connaissance en de multiples formes, mais elles posent également des problèmes très complexes en ce qui concerne l'identité documentaire. Ainsi, les possibilités de conversion des formats informatiques génèrent une quantité désordonnée et incontrôlable d'inscriptions concurrentes des mêmes connaissances (Salaün 2008 : § 4) qui oblige à repenser l'*invariant* qui fonde l'unité *texte* (Pédaque 2 2006). Par ailleurs, l'ajout d'annotations multiplie également les versions enrichies fondées sur un même texte. Je me garderai donc de faire davantage que rendre compte de la manière dont les inscriptions informatiques sont employées pour être *lues*, conçues sur la base de *choix* et exploitées pour *découvrir*, reprenant les thèmes des chapitres précédents. Pour cette raison, j'opérerai des raccourcis fréquents. Par exemple, je pourrai parler de la conversion d'une inscription informatique en inscription graphique même si cette dernière apparaît sur un écran d'ordinateur et demeure encodée sous un format qui lui est tout aussi spécifique. Je parlerai également de conversion d'un encodage informatique du texte en une inscription textuelle traditionnelle, même si cette dernière prend « concrètement » la forme d'un fichier sur un disque dur.

Dans ce chapitre, je vais simplement me limiter à présenter les deux principaux chantiers sur lesquels j'ai effectivement conçu des inscriptions informatiques : l'édition de chartes (→4.1) et l'annotation des constructions syntaxiques de SRCMF (→4.2). Le troisième type d'objet sur lequel j'ai travaillé, les répertoires structurés de matériaux, soit n'a pas encore donné lieu à une inscription informatique concrète (ALW, FEW), soit s'est limité à des manipulations de données dont rendre compte n'enrichirait pas vraiment ce mémoire. Je vais tenter de présenter mes recherches dans le domaine de la manière la moins technique possible, mais on conviendra que l'illustration concrète de mon propos sera profitable et je ne pourrai faire l'économie d'une terminologie adaptée.

#### **4.1. Inscription informatique d'éditions de textes**

Le premier projet conséquent sur lequel j'ai travaillé est l'encodage des chartes dont j'ai étudié la ponctuation (Mazziotta 2009). Ce travail a été effectué à l'aide du langage nommé XML (*eXtensible Markup Language*) et de conventions idiosyncrasiques. Le développement des techniques informatiques d'encodage a pris un essor phénoménal avec la venue de la norme XML à la fin des années 1990. Étant donné l'importance de cette norme pour l'édition numérique, je me permets d'introduire ce langage en quelques mots, en utilisant les exemples les plus simples possible. Ceci fera l'objet de la première sous-section (→4.1.1). La manière dont ce langage peut être utilisé pour se sortir de certaines impasses qui affectent l'édition traditionnelle sera traitée ensuite (→4.1.2).

Certaines difficultés demeurent cependant irrésolues malgré un net progrès par rapport aux inscriptions traditionnelles, qui ne font pas appel au numérique. Quels que soient les moyens mis en œuvre dans les projets présentés, on verra que les inscriptions ne peuvent éviter complètement les choix réducteurs et les difficultés de lecture (→4.1.3).

#### 4.1.1. Introduction à XML

On se référera aux documents techniques adéquats pour lire une définition formelle du langage<sup>46</sup> : je me limite ici à ce qui est nécessaire à la compréhension des exemples qui suivront. XML encode les données sous la forme de chaînes de caractères, parmi lesquelles les caractères << et >> (respectivement plus petit que et plus grand que ou encore chevrons) ont un statut particulier. En effet, les caractères qui se trouvent entre ces chevrons sont nommés *balises* et structurent le reste du texte. Cod. 1 est un exemple de structuration à l'aide de balises XML.

```
Le mot <mention>chat</mention> compte quatre lettres.
```

Code 1 – Exemple simple d'utilisation d'XML

On voit que le mot *chat* est encadré par <mention> à sa gauche et </mention> à sa droite. Ce sont, respectivement, une balise ouvrante et une balise fermante (la fermeture étant indiquée par le </ derrière le caractère <>). On dit que la chaîne de caractères *chat* est contenue dans l'*élément* nommé *mention*. On comprend intuitivement que les balises servent à délimiter des empan de texte et à leur assigner une étiquette. Pour reprendre les termes utilisés jusqu'ici, elles *réifient* une classification (→2.2.2).

La structure formelle sous-jacente à l'encodage XML est celle d'un arbre (qui ressemble à un arbre de constituants). Les éléments peuvent donc être contenus les uns dans les autres et former une hiérarchie. On voit dans cod. 2 (p. 68 ; l'utilisation des retours de chariot n'a pas de valeur et ne fait que simplifier la lecture) que chacune des lettres du mot *chat* est entourée de balises nommées *consonne* ou *voyelle* et deux consonnes ont été regroupées dans un élément nommé *digramme*.

---

46 XML est défini en ligne par une recommandation technique exhaustive <<https://www.w3.org/TR/REC-xml/>>.

```

Le mot
<mention>
  <digramme>
    <consonne>c</consonne>
    <consonne>h</consonne>
  </digramme>
  <voyelle>a</voyelle>
  <consonne>t</consonne>
</mention>
compte quatre lettres.
Code 2 – Exemple simple de hiérarchie

```

Ainsi, l'élément nommé *mention* contient trois éléments : un élément nommé *digramme*, un élément nommé *consonne*, un élément nommé *voyelle*. L'élément nommé *digramme* contient à son tour deux éléments *consonne*, alors que les deux autres contiennent un caractère. La mise en page de l'exemple aide à visualiser, par un jeu sur l'indentation des lignes, la hiérarchie des éléments. Il est toutefois déjà clair que l'ergonomie n'est pas optimale.

Par ailleurs, comme on le voit dans cod. 3, d'autres informations, appelées *attributs* sont parfois ajoutées dans les balises ouvrantes selon une forme conventionnelle que je laisse ici appréhender de manière intuitive : il est dit que l'élément a un attribut nommé *identifiant* dont la valeur est « 1 ». La différence entre les éléments et les attributs est que ces derniers ne contiennent pas d'autres entités structurantes, mais uniquement un nom et une valeur.

```

Le mot <mention identifiant="1">chat</mention> compte quatre lettres.

```

Code 3 – Exemple simple d'attribut XML

Les noms donnés aux éléments et aux attributs, la sémantique associée à ces noms et les contraintes de hiérarchie ne sont pas fixés par le langage, mais déterminés par les utilisateurs, qui disposent des moyens techniques pour formaliser leurs choix<sup>47</sup>. L'exemple que je viens de donner montre un balisage qui classe une partie des caractères. Si l'on ajoute ce genre de balisage, c'est pour pouvoir ensuite manipuler l'inscription XML pour extraire des informations ou la transformer en une autre inscription. Ainsi, un programme pourrait produire une page web où tous les mots balisés *mention* sont mis en italiques ; un autre pourrait extraire toutes les consonnes des mots en mention.

On perçoit intuitivement les portes ouvertes par ce genre d'encodage. L'utilisation de la norme XML a permis de revoir en profondeur la manière dont les inscriptions éditoriales pouvaient être conçues. Le principal avantage des outils qui se sont développés est d'affranchir l'inscription des

47 Par exemple, les « schémas » XSD ; <<https://www.w3.org/XML/Schema>>.

contraintes spatiales inhérentes au support traditionnel. Rien n'empêcherait en effet d'encoder des analyses non destinées à apparaître dans une édition traditionnelle. Rien n'empêcherait d'encoder de multiples analyses concurrentes ou des interprétations différentes en vue de n'en afficher qu'une partie, selon les besoins. Voyons à présent comment j'ai exploité concrètement ce genre de potentiel.

#### 4.1.2. Inscription unique, vues multiples et diffraction des unités

Rappelons le problème du choix qui s'est posé ci-dessus au sujet des abréviations et de la ponctuation (→2.1). Avec des moyens traditionnels, à moins de recourir à une édition synoptique, il n'est pas possible d'inscrire de façon ergonomique et lisible à la fois les caractères originaux et leur interprétation. On ne peut représenter la ponctuation ancienne et la ponctuation moderne simultanément et l'on ne peut représenter une abréviation et sa résolution de façon concomitante. Dès lors, comment satisfaire à la fois le littéraire qui n'a que faire de ces éléments originaux et le linguiste ou le sémioticien qui s'est donné pour but de les étudier ?

XML offre une réponse apparemment simple à ce genre de question, car il permet de baliser l'inscription informatique de manière à encoder les différences conceptuelles, pour ensuite générer des inscriptions traditionnelles qui ne correspondent qu'à une partie seulement de ce qui est encodé<sup>48</sup>. Voyons cela concrètement. La fig. 51 est un exemple d'abréviation du mot *chevaliers*.



Figure 51 – Exemple d'abréviation (Document 1260-02-29 : 3)

Cod. 4 donne un exemple d'encodage<sup>49</sup> d'une abréviation et de sa résolution (le code *&tilde;* représente le tilde).

```
<w>
  <norm>chevaliers</norm>
  <dipl>ch<expan>eua</expan>l<expan>ie</expan>rs</dipl>
  <fac>
    <mdv_abbr>chl&tilde;r</mdv_abbr>s
  </fac>
</w>
```

Code 4 – Encodage d'une abréviation et de sa résolution (d'après Lavrentiev 2005 : 3)

On voit que les trois éléments contenus dans l'élément *w* consistent en une transcription normalisée (*norm*), en une transcription plus proche de l'original (*dipl*) et en une transcription fac-similaire (*fac*),

48 L'équipe de la Base de français médiéval a beaucoup travaillé sur le sujet (Lavrentiev 2009b ; Marchello-Nizia *et al.* 2015).

49 Qui s'inspire d'une des premières tentatives de la Base de français médiéval pour encoder les abréviations originales.

qui tente de reproduire au mieux les entités graphiques de la source transcrite. Ce genre d'inscription peut, comme je l'ai dit, être traité par des programmes informatiques pour produire une inscription traditionnelle (peu importe son format), dont la forme serait *une seule* de celles données sous (29).

(29.a) chevaliers

(29.b) *cheualiers*

(29.c) ch̄lrs

Pour produire (29.a), on peut écrire un programme qui demande à l'ordinateur de ne pas tenir compte du contenu des éléments *facs* et *dipl*. Pour produire (29.b), on demande à l'ordinateur d'ignorer les éléments *norm* et *facs* et de mettre en italiques le contenu de *expan*, qui marque une abréviation résolue. (29.c) est produit suivant le même principe d'exclusion, mais en remplaçant les codes spéciaux par des caractères. Chacune de ces possibilités ne représente qu'une portion de la connaissance que l'inscription informatique encode apparemment complètement. (29.a), (29.b) et (29.c) sont autant de vues différentes d'une unique inscription. Elles sont adaptées à des publics différents : le linguiste/sémioticien travaillera sur (29.c), ce qui ne gênera en rien le littéraire, qui aura (29.a) à sa disposition.

Malgré les avantages que je viens de décrire, il faut cependant noter que les vues multiples ne sont pas convertibles les unes dans les autres : contrairement aux inscriptions concurrentes d'une même analyse (→3.3.2. a), aucune règle formalisable ne permet de prouver leur équivalence. Or, si les éléments qui correspondent à ces vues sont associés par leur insertion dans le même élément nommé *w*, qui représente l'unité conceptuelle du mot, il faudrait déployer des moyens considérablement plus élaborés pour que les caractères soient mis en correspondance de la même manière. On se heurte à un phénomène complexe de *diffraction* des connaissances à propos des caractères (Mazziotta 2006b), car rien dans l'encodage ne montre que <u> dans (29.b) et <v> dans (29.a) représentent une unique entité dans le manuscrit. Par ailleurs, la relation entre les moyens abrégatifs et les lettres qu'ils remplacent est également opaque.

#### 4.1.3. Complexité de l'inscription et ergonomie

Une manière de résoudre le problème qui vient d'être décrit est de mobiliser les mêmes techniques au niveau du caractère et d'employer un système de référence pour encoder les abréviations et leur résolution. La diffraction a alors lieu à un autre niveau de granularité (Lavrentiev 2009a : § 5.1) et l'unité du mot est préservée. J'ai ainsi proposé un encodage assez complexe (cod. 5).

```

<w>
  ch
  <expan abbr='#a1'>
    e
    <c>
      <norm>v</norm>
      <facs>u</facs>
    </c>
    a
  </expan>
  l
  <expan abbr='#a1'>ie</expan>
  <abbr xml:id='a1'><g ref='#titulus' /></abbr>
  r
  s
</w>

```

Code 5 – Encodage d'une abréviation et de sa résolution  
(d'après Mazziotta 2008b)

Il est un peu compliqué de lire ce type d'encodage et je ne vais pas détailler ici son organisation. Ce qu'il y a à observer ici, c'est que même si l'inscription est plus satisfaisante, elle n'est pas ergonomique du tout, tant au point de vue de sa lecture qu'au point de vue de sa saisie manuelle.

C'est une critique sérieuse, mais elle ne prend pas en compte un élément essentiel : les inscriptions informatiques ne sont pas nécessairement faites pour être saisies ou lues par des humains sans interface appropriée (c'est-à-dire sans une forme de médiation par l'ordinateur). Le texte de cod. 5 est illisible, mais il ne devrait pas être considéré comme une inscription traditionnelle. Malheureusement, on fait souvent face à un blocage pratique de taille : il est vrai que la génération de vues différentes est un processus relativement trivial, mais comment faire si aucun logiciel n'a été développé pour la saisie ? Cette limitation est d'autant plus récurrente que chaque projet a ses propres attentes<sup>50</sup>. On est donc forcé à prendre en compte la critique et à chercher une manière de procéder qui rende la saisie possible. C'est ce que différents projets ont tenté de faire à l'époque (Mazziotta 2004, Glessgen 2005, Lavrentiev 2009b). Je vais prendre un exemple (cod. 6 ; je numérote les lignes pour faciliter la discussion) tiré de mon corpus pour illustrer la solution commune qui a été trouvée. Nous avons tous les trois procédé plus ou moins de la même manière : en mêlant à XML des conventions idiosyncrasiques.

50 À ce propos, les problèmes de la saisie d'une édition de matériaux oraux sont similaires. Certains projets renoncent purement et simplement à la transcription phonétique en raison de leur intérêt plus marqué pour la syntaxe. D'autres profitent de logiciels spécialisés, comme PRAAT, pour encoder la correspondance entre les sons enregistrés et leur transcription de manière parallèle et non diffractée (<<http://www.praat.org/>>).

```

1 <texte>
2 <part/>
3 <recto>
4 <l/>
5 <w id="1"><c>J</c>akemes</w>
6 <mponct id="470">,</mponct>
7 <w id="2">li</w>
8 <w id="3">p[ro]°uo$t</w>
9 <mponct id="471">,</mponct>
10 <ponct id="428">*</ponct>
11 <w id="4"><c>J</c>ohans</w>
12 <mponct id="472">,</mponct>
13 <w id="5">li</w>
14 <w id="6"><c>d</c>oi`n$</w>
    [...]

```

Code 6 – Encodage de saisie (Document 1236-05 : 1)

Je suis conscient du caractère cryptique de cette inscription, que je vais commenter (à nouveau, les explications de détails sont limitées ici au minimum nécessaire à mon propos). Examinons tout d'abord la ligne 8. On voit qu'elle comporte la chaîne de caractères <p[ro]°uo\$t>, qui correspond à la graphie modernisée *provost* 'prévôt'. Telle qu'elle figure dans l'inscription, cette chaîne comporte trois informations remarquables, qu'on ne peut comprendre qu'en ayant connaissance des conventions de saisie :

1. Les caractères <ro> sont la résolution d'une abréviation, mais rien n'est dit de la forme originale de cette abréviation. Cela est indiqué par les crochets.
2. Le caractère original <u> correspond à <v> dans une graphie modernisée. Cela est indiqué par le caractère <°> qui précède.
3. Le caractère <\$> figure une <f> originale, qui correspond à <s> dans une graphie modernisée. Il s'agit d'une règle automatique.

Par ailleurs, on voit aux lignes 9 et 10 que la ponctuation originale (ligne 10, élément *ponct*) est encodée après la ponctuation moderne (ligne 9, élément *mponct*).

Cela saute aux yeux : la proposition présentée dans cod. 6 n'est qu'un pis-aller. D'une part, l'objectif d'ergonomie est loin d'être atteint et, d'autre part, certains éléments comme la forme originale de l'abréviation, ont dû être réduits. On aurait toutefois tort de s'en priver car la connaissance inscrite grâce à ces techniques est quantitativement supérieure à celle qui peut l'être par des méthodes traditionnelles, même si la saisie demande un peu d'entraînement.

Nous venons de voir jusqu'à quel point les inscriptions informatiques permettaient d'encoder des connaissances complexes à propos des caractères des manuscrits. L'ordinateur ouvre certes des portes, mais il est loin de fournir la panacée. Une des pierres d'achoppement les plus importantes est l'ergonomie, qui, on le devine, devient plus cruciale encore lorsque ce sont des hiérarchies complexes qui doivent être saisies, puis exploitées. La sous-section suivante traitera des moyens mis en œuvre dans le cadre de SRCMF, où il a fallu fournir une inscription informatique de constructions syntaxiques, donc de hiérarchies.

## 4.2. Les inscriptions dans SRCMF

Le deuxième projet important pour lequel j'ai manipulé des inscriptions informatiques est celui du *Syntactic Reference Corpus of Medieval French* (SRCMF ; v. notamment →2.2.1). Comme je l'ai évoqué plus haut, il s'agit d'un projet d'annotation exhaustive de la syntaxe d'une sélection de textes médiévaux. Ces textes sont issus de deux corpus préexistants : la *Base de français médiéval* (BFM ; Guillot et al. 2007) et le *Nouveau corpus d'Amsterdam* (NCA ; Kunstmann/Stein 2007b). Le travail d'annotation a été financé pendant trois ans par l'Agence nationale de la recherche et la Deutsche Forschungsgemeinschaft (2009-2012), mais l'enrichissement de la ressource continue encore aujourd'hui. On a vu que les inscriptions informatiques reposent sur des formats spécialement conçus pour être traités par des logiciels (parfois trop peu nombreux à pouvoir les manipuler ; →4.1.3). Certains formats sont idiosyncrasiques, d'autres sont courants<sup>51</sup>. Comprendre ces inscriptions en ouvrant naïvement les fichiers qui les contiennent est parfois possible pour un humain<sup>52</sup>, parfois assez difficile et parfois tout bonnement impossible<sup>53</sup>. On ne peut cependant jamais dire que les inscriptions informatiques dont je parle ici sont optimisées pour être lues sans la médiation de l'ordinateur.

Nous allons voir ci-dessous quels moyens ont été déployés par l'équipe dont j'ai fait partie pour travailler sur ce corpus dont la base est un ensemble de ressources disponibles par ailleurs (→4.2.1). Ce sera l'occasion d'aborder les mécanismes d'interfaçage, qui permettent des manipulations diverses, et les conversions variées, qui multiplient les inscriptions « dérivées » (→4.2.2).

### 4.2.1. Référencement et interfaçage

J'ai rappelé que SRCMF exploitait des documents issus de corpus préexistants : la BFM et le NCA. Les annotations sont ainsi appliquées à des textes qui ont déjà une inscription informatique par

---

51 Comme XML et son « dialecte » TEI (<<http://www.tei-c.org>>), RDF (<<https://www.w3.org/RDF/>>), JSON (<<http://www.json.org/>>) et les bases de données relationnelles supportant une version du langage de requête SQL.

52 XML et JSON sont initialement prévus pour être « human readable », même si cet objectif n'est pas toujours atteint.

53 Par leur recours à des formats compilés, les bases de données relationnelles sont généralement faites pour être lues par des ordinateurs uniquement.

ailleurs. Pour le dire de manière très simple, les fichiers contenant les textes ne sont pas les mêmes que les fichiers contenant les annotations (on parle d'annotation « débarquée » ; Habert 2005 : 31-33). Voilà une propriété importante des inscriptions informatiques : elles rendent possible le *référencement*, c'est-à-dire la formalisation des relations entre inscriptions. Ce référencement est à la fois un avantage incontestable pour l'inscription de connaissances complexes et une difficulté majeure en termes d'ergonomie. Voyons cela concrètement.

### a. Difficultés ergonomiques

Cod. 7 reproduit l'inscription informatique de la BFM du texte (30).

- (30) Prologue [...] Ici cumencet amiable cançon e spiritel raisun d'iceol 'Prologue [...] Ici commence l'agréable chanson et le discours spirituel à son propos [= celui de saint Alexis]' (AlexisS<sup>2</sup> : 1 sv.)

```
<w xml:id="w19_0001">Prologue</w>
[...]
<w xml:id="w19_0002">Ici</w>
<w xml:id="w19_0003">cumencet</w>
<w xml:id="w19_0004">amiable</w>
<w xml:id="w19_0005">cançon</w>
<w xml:id="w19_0006">e</w>
<w xml:id="w19_0007">spiritel</w>
<w xml:id="w19_0008">raisun</w>
<w xml:id="w19_0009">d'</w>
<w xml:id="w19_0010">iceol</w>
```

Code 7 – Inscription XML du début de la *Vie de saint Alexis*

Il est possible de la lire avec de l'entraînement et un minimum de connaissance des conventions d'encodage (cf. →4.1.1). En parallèle, l'inscription informatique du début de l'annotation correspondante est présentée dans (cod. 8).

```
doc:w19_0001_6:hierarchy <http://www.srcmf.org/annotations/alexis#_1270041812.09>.
doc:w19_0002_6:hierarchy <http://www.srcmf.org/annotations/alexis#_1270041832.56>.
doc:w19_0003_6:hierarchy <http://www.srcmf.org/annotations/alexis#_1270041835.24>.
doc:w19_0004_6:hierarchy <http://www.srcmf.org/annotations/alexis#Tom%20%-
%20%1/aLexis_01_1316175294.92>.
doc:w19_0005_6:hierarchy <http://www.srcmf.org/annotations/alexis#_1270893912.88>.
doc:w19_0006_6:hierarchy <http://www.srcmf.org/annotations/alexis#_1270893864.78>.
doc:w19_0007_6:hierarchy <http://www.srcmf.org/annotations/alexis#Tom%20%-
%20%1/aLexis_01_1316175302.97>.
doc:w19_0008_6:hierarchy <http://www.srcmf.org/annotations/alexis#_1272714351.71>.
doc:w19_0009_6:hierarchy <http://www.srcmf.org/annotations/alexis#Tom%20%-
%20%1/aLexis_01_1316175329.17>.
```

Code 8 – Inscription RDF du début de l'annotation SRCMF de la *Vie de saint Alexis*

On remarque à chaque ligne la présence d'une forme qui commence par *doc:w19*. Il s'agit en fait de références qui correspondent au contenu des attributs nommés *xml:id* dans cod. 7. Même en comprenant le principe, il est difficile d'exploiter le référencement pour un humain. L'inscription est incomplète, mais même présentée de manière exhaustive à quelqu'un qui connaît les conventions d'encodage, elle demeurerait cryptique. En outre, comme les mots sur lesquels l'annotation porte ne sont pas représentés, un constant va-et-vient serait nécessaire entre l'inscription informatique du texte et celle des annotations pour comprendre ces dernières. L'inscription est, il faut l'avouer, complètement *illisible*, au sens où elle ne permet pas d'accéder à la connaissance comme je le présentais dans le premier chapitre (→1).

### b. Importance de l'interface

L'enjeu des outils informatiques appropriés est de rendre lisible la connaissance ainsi inscrite. Pour ce faire, on peut utiliser le référencement pour agréger les inscriptions dans une seule structure formelle et, partant, les représenter de manière conjointe : les inscriptions fusionnent en une nouvelle inscription informatique qui peut être manipulée (comme le sont les diagrammes que j'ai présentés ci-dessus ; →3.2). C'est exactement de cette manière que j'ai conçu le logiciel expérimental dédié à l'encodage des analyses syntaxiques, que j'ai nommé *NotaBene* (Mazziotta 2010a ; 2010b). *NotaBene* permet de manipuler les inscriptions. Initialement, le logiciel a été conçu pour permettre l'ajout de couches d'annotations différentes, agrégées en une seule inscription au sein de l'interface (2010b). La fig. 52 est une capture d'écran de la fenêtre d'annotation du logiciel où sont agrégés cod. 7 et cod. 8.

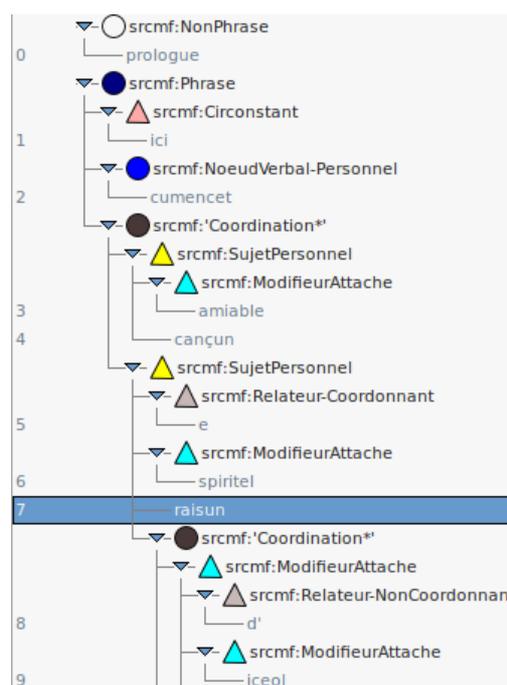


Figure 52 – Fenêtre d'annotation de *NotaBene* (neuf premiers mots de la *Vie de saint Alexis*)

Ce qui m'intéresse ici, c'est que le texte est représenté dans la même inscription que les analyses (associées sur la figure à des symboles géométriques), alors que ces deux ensembles de données sont initialement inscrits dans deux fichiers séparés. L'agrégation des inscriptions est la condition *sine qua non* à la manipulation des annotations, car ces dernières, détachées du texte, ne sont par elles-mêmes que des constructions syntaxiques sans réalisation lexicale.

La conséquence directe de la nécessité d'une médiation est que cette dernière est liée à la génération d'inscriptions concurrentes d'une même connaissance. En effet, si les formats ne sont pas utilisables par les humains qui sont censés s'en servir pour inscrire de la connaissance, ils ne peuvent être employés que par des machines. Il faut donc donner aux machines les moyens de représenter les inscriptions de façon lisible, ce qui ne peut se faire que par une transformation de leur forme. La fig. 52 montre l'interface de *NotaBene*, qui se charge de produire une inscription graphique du rapport entre les textes et les annotations syntaxiques, et de donner les moyens de lire ces dernières. Cette inscription créée dynamiquement sur la base de deux autres est similaire à une inscription traditionnelle comme celles des nombreux exemples cités dans les chapitres qui précèdent. Toutefois, cette manière de concevoir le résultat de l'agrégation est simpliste et s'en tient à un point de vue strictement visuel. Il y a plus : l'interface donne les moyens d'inscrire l'analyse et facilite les manipulations.

#### 4.2.2. Manipulations

Cette sous-section traite brièvement des manipulations d'inscriptions dans l'interface de *NotaBene* (→a). L'exportation des inscriptions dans le but de les utiliser sans ordinateur ou avec d'autres logiciels est ensuite abordée (→b).

##### a. Manipulations dans *NotaBene*

Il y a une différence fondamentale entre l'interface présentée dans la fig. 52 et les inscriptions diagrammatiques traditionnelles : si l'on peut annoter ou comparer ces dernières, il n'est que rarement possible de les manipuler et de les corriger sans avoir à les recopier. L'interface d'annotation permet justement cela, ouvrant ainsi la porte à de multiples possibilités, comme la comparaison assistée en vue de corrections. La comparaison d'annotations fonctionne exactement de la même manière qu'exposé ci-dessus au sujet de la sélection entre différentes analyses syntaxiques (→3.2.2) : elle consiste à mettre côte à côte deux inscriptions concurrentes pour permettre à l'analyste de déterminer laquelle il préfère. La fig. 53 montre comment une erreur dans l'annotation de (31) est mise en évidence par *Notabene*.

(31) afr. Li rois pense que par folie, Sire Tristan, vos aie amé ‘Le roi pense que c’est par folie, Sire Tristan, que je vous ai aimé’ (TristBérM<sup>4</sup> : 20)

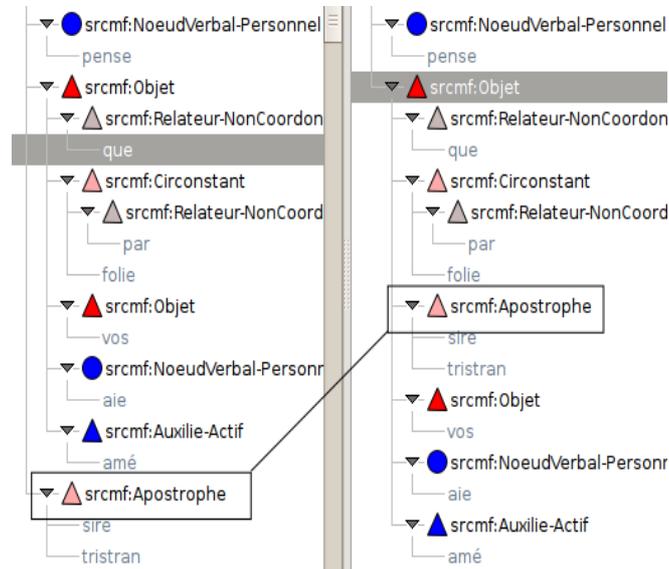


Figure 53 – Interface de comparaison entre deux ensembles d’annotations (boîtes ajoutées manuellement ; Mazziotta 2010LAW: 145)

On voit qu’à droite la construction identifiée comme une « apostrophe » (vocatif) a été placée au mauvais niveau, puisqu’elle figure dans la subordonnée (le contexte dialogique ne laisse aucun doute). En utilisant l’interface, l’inscription peut être modifiée directement en tirant les conclusions que la comparaison assistée rend évidente.

Les manipulations rendues possibles par l’inscription informatique des connaissances ne se limitent évidemment pas à des interfaces d’édition : les logiciels transforment les données de manière variée. Je vais à présent en donner quelques exemples.

### b. Exportations

J’ai présenté deux inscriptions différentes de l’analyse syntaxique dans ce chapitre : celle de l’interface de *NotaBene* et celle du format d’encodage dans lequel l’analyse est enregistrée. Toutefois ces formats ne sont pas les seuls. Il faut distinguer deux types de transformations : 1/ celles qui génèrent des inscriptions observables par l’humain ; 2/ celles qui génèrent des inscriptions destinées à être « lues » et manipulées par des logiciels. Je mentionnerai ici trois exemples qui me semblent particulièrement illustratifs.

Un exemple du premier type est livré dans la fig. 54 : *NotaBene* produit des fichiers d’un format particulier qui permettent de produire des diagrammes syntaxiques.

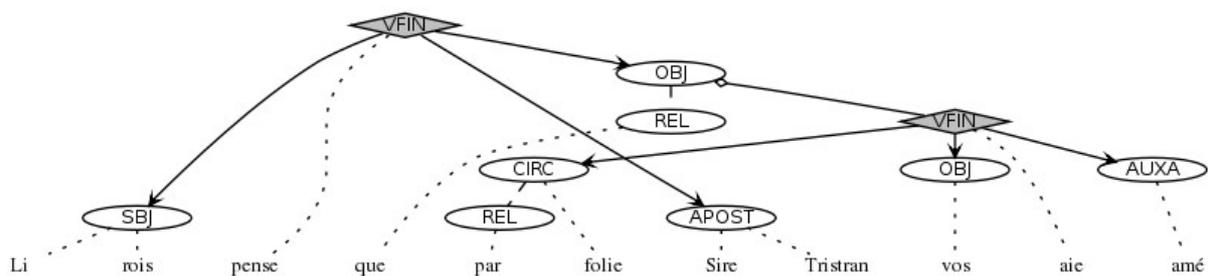


Figure 54 – Exportation générée par *Graphviz* au départ du format d’exportation DOT de *NotaBene*

Je ne détaillerai pas ici les conventions de lecture associées à ces diagrammes. Il faut principalement noter que les conventions de représentation et les entités mobilisées dans ce diagramme diffèrent encore de celles de tous les diagrammes présentés ci-avant. Ce diagramme est en effet issu de ma recherche pour produire automatiquement des inscriptions lisibles sans réduction de la connaissance syntaxique inscrite. Il n’est qu’un essai parmi d’autres, où j’ai changé les conventions de façon expérimentale. Les possibilités ouvertes sont innombrables.

Un premier exemple de conversion destinée à la machine est l’exportation au format *TigerXML*<sup>54</sup> en vue de l’exploitation de SRCMF par extraction de structures syntaxiques<sup>55</sup> (v. Rainsford 2015). Cette exportation est rendue possible au moyen d’une fonctionnalité que j’ai implémentée directement dans *NotaBene*. *TigerXML* est le langage d’encodage des inscriptions destinées à être traitées par le logiciel *TigerSearch*. Il ne serait pas utile de montrer l’inscription informatique obtenue pour constater qu’elle est peu lisible. Par contre, je voudrais souligner que *TigerSearch* est associé à un langage qui permet d’interroger le corpus, c’est-à-dire de formuler des requêtes dans un langage formalisé. Il s’agit d’une procédure de découverte qui est rendue possible par l’inscription informatique (et qui ne figure donc pas sous →3). Cod. 9 est un exemple de requête de ce type. Elle signifie « trouver tous les sujets qui sont réalisés par le mot-forme *rois*. »

```
[cat="SjPer"] >L [word="rois"]
```

Code 9 – Requête *TigerSearch* (Rainsford 2015 : 15)

Comme les résultats de l’extraction sont destinés à être exploités par des humains, leur diagrammatisation s’impose. L’interprétation de la requête par le logiciel mène à la génération d’une liste de diagrammes qui ressemblent à la fig. 55.

54 <<http://www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/tigersearch.html>>.

55 Rainsford 2015 donne par ailleurs un mode d’emploi pour créer des concordances au départ des résultats des requêtes *TigerSearch*. Je ne développerai pas ce point.

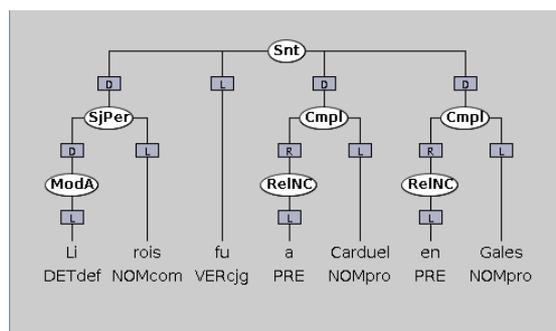


Figure 55 – Diagramme TigerSearch (Rainsford 2015: 5)

Nouvelle inscription, nouvelles conventions de lecture. On voit intuitivement à quel point chaque logiciel, chaque environnement de travail diagrammatise de manière différente les mêmes connaissances. Cette observation amène une question : les diagrammes variés générés par les logiciels sont-ils toujours équivalents ? C'est cette question que je vais envisager au travers du troisième exemple.

Mon troisième et dernier exemple de conversion au départ des fichiers de SRCMF est la génération, toujours par le biais de *NotaBene*, de fichiers au format CoNLL (*Computational Natural Language Learning*)<sup>56</sup>. Ce format connaît un franc succès et de nombreux logiciels lui sont dédiés, notamment des logiciels d'analyse syntaxique automatique (ang. *parsers*). Stein (2014 : 2880) fait usage d'un de ces logiciels pour montrer comment SRCMF peut être employé en vue d'analyser automatiquement de nouveaux textes. Ce faisant, à l'aide de *parsers*, il découvre une nouvelle connaissance en exploitant des diagrammes (→3), même si ces derniers ont une inscription informatique. Il fait cependant remarquer :

Since the CoNLL format does not support orthogonal relations such as coordination, the *Notabene* tool exports coordinations as complex categories, e.g. *coord1\_Obj*. Of these labels only the function (e.g. *Obj*) was retained for our experiments. Hence coordination is marked by (a) dependence on the same word and (b) the presence of the relator (RelC) in the second structure. (Stein 2014 : 2880)

Le formalisme sous-jacent au format CoNLL est arborescent (« CoNLL format does not support orthogonal relations »). En conséquence la coordination n'est pas prise en compte de manière satisfaisante et l'expérimentateur n'a d'autre choix que : 1/ soit de l'éliminer de ses investigations (c'est le cas ici) ; 2/ soit l'inscrire de manière compatible avec la structure arborescente (comme celle suggérée par Mel'čuk ; cf. 1.2.2). On retombe dès lors sur le problème des contraintes posées par le modèle formel choisi (→2.3) et de la réduction afférente à la plupart des expériences concrètes d'exploration des données (→3.1). On multiplierait facilement les exemples de ce type<sup>57</sup>.

56 Voir <<http://www.conll.org/>>.

57 L'exemple des plans factoriels abordé sous →3.1.2, où il était question de traitements statistiques, illustre déjà implicitement la question des modifications subies par les inscriptions au gré de leur exploitation informatique.

Lorsque j'ai conçu et développé le logiciel *NotaBene*, j'ai donc constaté la variété des manipulations possibles et le grand nombre de formats vers lesquels les inscriptions pouvaient être converties. Malgré leur ancrage sur un support numérique, il demeure indubitable que les inscriptions produites se prêtent tout aussi bien à une évaluation selon les critères des chapitres précédents que les inscriptions traditionnelles : peut-on les lire ? quels choix (réducteurs) ont été opérés ? que nous permettent-elles de découvrir ?

### 4.3. Synthèse

Dans ce chapitre, j'ai montré concrètement comment je me sers d'inscriptions informatiques dans mes recherches. Le support numérique résout partiellement certains problèmes posés par le support traditionnel (non informatique). Néanmoins, en dépit des possibilités de l'ordinateur, les limites des inscriptions peuvent toujours être posées dans les mêmes termes que pour les éditions traditionnelles, car elles sont liées à l'ergonomie et aux choix à opérer pour réduire la connaissance inscrite. Cela est apparu très clairement quand j'ai parlé de l'édition de manuscrits (→4.1). Un langage comme XML (→4.1.1) permet d'encoder une inscription unique que les logiciels peuvent diffracter en différentes visualisations, produisant de la sorte des inscriptions adaptées aux besoins variés des chercheurs (→4.1.2). Pour aboutir à un résultat acceptable tant du point de l'ergonomie que de celui de la richesse de l'inscription informatique, on recourt à des méthodes de saisie perfectibles (→4.1.3). La difficulté est d'autant plus grande quand il s'agit d'encoder des annotations syntaxiques, ce qu'on ne peut faire sans une médiation d'un logiciel adéquat (→4.2). Dans le cas de SRCMF, l'annotation débarquée, qui disjoint l'annotation de ce sur quoi elle porte au niveau de l'inscription informatique, est traitée par le logiciel *NotaBene*, qui crée une vue dynamique agrégeant les ressources en une seule inscription (→4.2.1). Cette procédure est obligatoire pour autoriser des manipulations subséquentes. Ces dernières ont en commun qu'elles mènent à la génération d'autres diagrammes fondés sur des conventions spécifiques, parfois en réduisant les données, de manière à générer de nouvelles connaissances (→4.2.2).

Les inscriptions informatiques libèrent donc d'une partie des contraintes liées au support et il est capital de prendre en considération leur potentiel pour les employer à bon escient. Faute de quoi, on ne fera que reproduire des pratiques traditionnelles et subir leurs limites (Bertin 2009 : § 3). Néanmoins, pour novatrices et libératrices qu'elles soient, les pratiques numériques continuent à poser constamment les questions de la lecture, du choix et de la découverte telles qu'elles ont été posées dans les chapitres qui précèdent. Ces questions qui fondent nos recherches sur la base d'inscriptions, qu'elles soient informatiques ou non, sont les mêmes. C'est pourquoi le premier enjeu de la

numérisation de la connaissance est indépendant du médium numérique, mais relève de l'épistémologie de la science à l'ouvrage. Ce passage salvateur de Robert Martin, pionnier de l'informatisation des dictionnaires, nous met en garde dans son introduction de *Sémantique et automate* (2001 : 11) :

Mais pour éviter toute confusion, il nous faut d'entrée nous démarquer nettement d'une attitude qui assimile « mutation technologique » et contenus scientifiques et qui peut aller jusqu'à faire croire que l'informatique n'est plus désormais un outil mais le « cadre théorique de la recherche » : c'est proprement absurde. À chaque science sa théorisation et ses modèles. Si, par une mémoire vertigineusement accrue, par la puissance des tris et des calculs, par l'optimisation qu'elle procure et par les validations qu'elle peut faciliter, l'informatique est un auxiliaire indispensable, c'est totalement s'égarer que d'en confondre la théorie avec celle des autres disciplines. Ainsi pour la lexicologie : informatisez un mauvais dictionnaire, le résultat sera un mauvais dictionnaire informatisé.

Il n'empêche qu'en libérant les inscriptions du support papier, on s'autorise à les penser selon des structures qui dépassent le potentiel expressif d'une simple arborescence déployée en deux dimensions. Cela a un impact très important sur l'accessibilité aux données (la manière dont elles sont lues). Pour les dictionnaires, par exemple, la notion d'*ordre alphabétique* en devient obsolète, car la recherche dans de tels ouvrages peut se faire par interrogations de champs précis (v. p.ex. Steinlin *et al.* 2004). On extrait ainsi les articles correspondant à une partie du discours ou à une combinatoire syntaxique particulière (pour peu que ces informations aient été encodées). L'inscription traditionnelle n'avait d'autre choix que de noyer ces points d'entrée dans la microstructure des articles.



## 5. Conclusion

---

Ce mémoire est parti de la « simple » lecture des inscriptions ( $\rightarrow 1$ ) aux processus de découverte qu'elles autorisent ( $\rightarrow 3$ ) sur la base de la forme réduite de la connaissance qu'elles rendent accessible ( $\rightarrow 2$ ), puis il s'est focalisé sur le numérique et ses enjeux ( $\rightarrow 4$ ), ne cachant pas la dimension transversale de ce sujet.

Au travers de ma propre expérience, j'ai montré que les inscriptions utilisées en linguistique peuvent en grande partie être considérées comme des inscriptions formelles. Les règles qui président à leur lecture sont soit explicitement formulées par les auteurs, soit implicitement acquises par leurs utilisateurs, mais dans de nombreux cas, on peut les décrire ( $\rightarrow 1.3$ ). Il en résulte que la linguistique est un terrain de déploiement important pour les inscriptions que je nomme *diagrammes* : des inscriptions à la fois graphiques et formelles. Pour construire tant les inscriptions moins formalisées que sont les éditions de texte que les diagrammes, nous sommes amenés à faire des choix, contraints par notre recherche de la pertinence, par ce que nous décidons de réifier et par les structures formelles qui encodent notre analyse. Partant, ces choix inévitables réduisent la partie de la connaissance qui est inscrite ; ils justifient que les mêmes matériaux puissent être traités à de multiples reprises, en fonction des décisions prises et du point de vue adopté ( $\rightarrow 2.4$ ). Quand nous réifions une partie de la connaissance dans un diagramme, nous nous pourvoyons d'un outil puissant pour l'explorer et faire émerger de nouvelles connaissances à propos des faits, de la théorie et même des structures formelles qui sous-tendent l'inscription ( $\rightarrow 3.4$ ). *Lire*, *choisir* et *découvrir* sont les trois opérations constantes qui définissent nos démarches. Cela est clair lorsque nous travaillons sur des inscriptions traditionnelles. Cela le reste quand nous nous servons d'inscriptions informatiques. Ces dernières, dont il nous faut apprendre à tirer le meilleur parti, ouvrent certes le champ des possibles, mais elles demeurent tout de même évaluables à l'aune des mêmes opérations ( $\rightarrow 4.3$ ).

Au-delà des éléments mis en avant dans le paragraphe qui précède, je profite de ce bilan pour en venir à une conclusion transversale qui me paraît importante : celle des niveaux du travail du linguiste. Il est connu et enseigné que le linguiste travaille sur la langue et sur le langage (Martin 2002 : 56-59) : les faits observés construisent de manière inductive la théorie. Ou bien, inversement, la théorie est évaluée en étant appliquée aux faits. L'existence de ces deux niveaux d'abstraction dans l'épistémologie de la discipline est acceptée. J'ai moi-même tenté de montrer, notamment dans le deuxième et le troisième chapitre, que l'on devait opérer des choix à ces deux niveaux, ainsi que l'on pouvait y découvrir de nouvelles connaissances. Cependant, j'ai également montré que *choisir* et *découvrir* étaient des opérations qui concernaient également les structures formelles grâce auxquelles la connaissance est inscrite ( $\rightarrow 2.3$  ;  $\rightarrow 3.3$ ). Dès lors que l'on décide de faire usage de formalismes pour décrire le langage et les langues, il devient nécessaire d'évaluer les formes pures. Certes, l'étude de ces formes pour elles-mêmes relève plutôt des mathématiques, mais l'examen de leur adéquation pour formaliser les analyses linguistiques est tout à fait pertinente au sein des sciences du langage.

C'est cette idée que je privilégie dans mes travaux en cours, qui se développent dans deux directions complémentaires. La première direction est celle de l'analyse sémiotique approfondie des inscriptions graphiques. J'ai montré à quel point les diagrammes étaient des outils essentiels à la connaissance en linguistique. Je me concentre à présent (2016a ; 2016c) sur la définition rigoureuse des *règles* graphiques qui structurent les diagrammes : comment décrire les configurations spatiales des stemmas ( $\rightarrow 1.2.1$  et *passim.*) et des cartes linguistiques ( $\rightarrow 3.1.1$ ) ? comment déterminer la proportion de ce qui est bel est bien formalisé dans une inscription graphique ? À quel moment une innovation graphique peut-elle entrer dans le système formel et l'enrichir ( $\rightarrow 3.3.1$ ) ?

La seconde direction est celle de l'utilisation concrète des diagrammes polygraphiques ( $\rightarrow 2.3.2$ ) dans mes recherches. La structure est neuve tout en étant annoncée depuis longtemps par des pionniers ( $\rightarrow 3.3.1$ . a). Une exploitation rigoureuse, qui tienne compte de ses propriétés formelles, dévoilera pleinement son potentiel. J'ai montré que je pouvais l'appliquer à certaines constructions du français moderne et du français médiéval ( $\rightarrow 3.3.1$ . b). Il y a tout intérêt à étendre son utilisation aux constructions essentielles de ces langues, dont je ferai ainsi émerger le système organisationnel d'une manière renouvelée. On conçoit également l'avantage qu'il y aura à travailler aux logiciels qui permettront de transposer les polygraphes sur un support numérique ( $\rightarrow 4$ ), ce qui promet d'ouvrir de nouvelles pistes en linguistique outillée.

Au terme de ce travail de synthèse, j'espère avoir réussi à présenter de manière utile ce qui fonde la cohérence de ma recherche : un questionnement constant sur les *enjeux de la représentation*

---

*de la connaissance en linguistique.* J'espère que le lecteur aura pu, d'une manière ou d'une autre, inscrire ses propres pratiques dans cette réflexion sur un thème qui, en fin de compte, nous unit



# Références

---

- AlexisS<sup>2</sup> = Storey, C. (éd.). <sup>2</sup>1968. *Saint Alexis*. Genève : Droz.
- Bachimont, B. 2007. *Ingénierie des connaissances et des contenus. Le numérique entre ontologies et documents*. Paris : Lavoisier.
- Bachimont, B. 2010. « Le numérique comme support de la connaissance : entre matérialisation et interprétation », G. Gueudet/L. Trouche. *Ressources vives. Le travail documentaire des professeurs en mathématiques*. Rennes : Presses Universitaires de Rennes et INRP.
- Baiwir, E./Renders, P. 2013. « Les atlas linguistiques sont-ils des corpus ? », *Corpus 12*, 27-37.
- Bertin, A. 2009 « Grammaire et dictionnaire », *Linx 60*, 31-46.
- Bertin, J. <sup>4</sup>2005 (1967). *Sémiologie graphique. Les diagrammes – Les réseaux – Les cartes*. Paris : EHESS.
- Bischof, B.-B./Glikman, J./Mazziotta, N./Prévost, S./Stein, A. 2012. *Guide d'annotation du Syntactic Reference Corpus of Medieval French (SRCMF)*. <<http://srcmf.org/fiches>>
- Boutier, M.-G. 1997. *Atlas linguistique de la Wallonie. Tome 15. Le corps humain et les maladies (2<sup>e</sup> partie)*. Liège : Université de Liège.
- Boutier, M.-G. 2008. « Cinq relations de base pour traiter la matière géolinguistique : réflexions à partir de l'expérience de l'Atlas linguistique de la Wallonie », *Estudis Romànics 30*, 301-310.
- Brittain, R. C. 1973. *A Critical History of Systems of Sentence Diagramming in English*. Austin : University of Texas [Thèse inédite].
- Büchi, E. 1996. *Les structures du « Französisches Etymologisches Wörterbuch »*. *Recherches métalxicographiques et métalxicologiques*. Tübingen : Niemeyer.
- Buridant, C. 2000. *Grammaire nouvelle de l'ancien français*. Paris : Sedes.
- Careri, M./Fery-Hue, F./Gasparri, F./Hasenohr, G./Labory, G./Lefèvre, S./Leurquin, A.-F./Ruby, C. 2001. *Album de manuscrits français du XIII<sup>e</sup> siècle. Mise en page et mise en texte*. Roma : Viella.
- Chauviré, C. 2008. *L'œil mathématique. Essai sur la philosophie mathématique de Peirce*. Paris : Kimé.
- Chomsky, N. 1957. *Syntactic structures*. The Hague : Mouton.
- Cibois, P. <sup>5</sup>2000. *L'analyse factorielle. Analyse en composantes principales et analyse des correspondances*. Paris : PUF.
- Clark, S. W. 1847. *Science of English language. A practical Grammar: in which words, phrases, and sentences are classified according to their offices, and their relations to each other. Illustrated by a complete system of diagrams*. New York : A. S. Barnes & Co.
- Counet, M.-T. 1990. *Mots et choses de Wallonie. Aspect du lexique dialectal de nos régions. Catalogue d'exposition – SLLW et BDW*. Liège : Société de Langue et Littérature Wallonne.
- de Marneffe, M.-C./Manning, C. 2008. « The Stanford Typed Dependencies Representation », *Coling 2008*:

- Proceedings of the workshop on Cross-Framework and Cross-Domain Parser Evaluation*, 1-8.  
 Document 1236-07, juillet. Archives de l'État à Liège, Cathédrale Saint-Lambert à Liège.  
 Document 1241-04-25, 25 avril. Archives de l'État à Liège, Collégiale Saint-Jean-en-Isle.  
 Document 1243-08-15, 15 août. Archives de l'État à Liège, Cathédrale Saint-Lambert à Liège.  
 Document 1244-01-19, 19 janvier. Archives de l'État à Liège, Cathédrale Saint-Lambert à Liège.  
 Document 1260-02-29, 29 février. Archives de l'État à Liège, Collégiale Saint-Jean-en-Isle.  
 Document 1273-05-12, 12 mai. Archives de l'État à Liège, Abbaye du Val-Benoît à Liège.  
 Fasseur, V./Rochelois, C. 2016. *Ponctuer l'œuvre médiévale. Des signes au sens*. Genève : Droz.  
 Feller, J. 1905. « Règles d'orthographe wallonne », *Bulletin de la Société de langue et littérature wallonnes* 41(2), 45-96.  
 Florey, K. B. 2006. *Sister Bernadette's Barking Dog: The Quirky History and Lost Art of Diagramming Sentences*. Hoboken : Melville House Publishing.  
 Foulet, A./Speer, M. B. 1979. *On Editing Old French Texts*. Lawrence : Regents Press of Kansas.  
 Giardino, V. 2011. « Vers une classification diagrammatique », *Visible* 8, 185-201.  
 Gilissen, L. 1973. *L'expertise des écritures médiévales recherche d'une méthode, avec application à un manuscrit du XI<sup>e</sup> siècle, le lectionnaire de Lobbes, codex Bruxellensis 18018*, Gand : Story-Scientia.  
 Glessgen, M.-D. 2005. « Editorische, lexikologische und graphematische Erschließung altfranzösischer Urkundtexte mit Hilfe von TUSTEP. Stand der Arbeiten », K. Gärtner/G. Holtus (éds), *Überlieferungs- und Aneignungsprozesse im 13. und 14. Jahrhundert auf dem Gebiet der westmitteldeutschen und ostfranzösischen Urkunden und Literatursprachen*. Trier : Kliomedia, 91-107.  
 Glikman, J./Mazziotta, N. 2013. « Représentation de l'oral et structures syntaxiques dans la prose de la *Queste del saint Graal* (1225-1230) », D. Lagorgette/P. Larrivée (éds). *Représentations du Sens Linguistique* 5. Chambéry : Éditions de l'Université de Savoie, 43-64.  
 Groupe  $\mu$ . 1992. *Traité du signe visuel*. Paris : Seuil.  
 Guillot, C./Marchello-Nizia, C./Lavrentiev, A. (2007). « La base de français médiéval (BFM) : états et perspectives », *Kunstmann/Stein* 2007a : 143-152.  
 Guyotjeannin, O./Pycke, J./Tock, B.-M. 1993. *Diplomatique médiévale*. Louvain-la-Neuve : Brepols.  
 Habert, B. 2005. *Instruments et ressources électroniques pour le français*. Paris : Gap/Ophrys.  
 Harris, R. 1995. *Signs of writing*. Routledge : London/New-York.  
 Heiden, S./Pincemin, B. 2008. *9es journées internationales d'analyses statistiques des données textuelles. Lyon, 12-14 mars 2008*. Lyon : Presses universitaires de Lyon.  
 Hofstadter D. 2000 (1979). *Gödel Escher Bach. Les brins d'une guirlande éternelle*. Paris : Dunod.  
 Holtus, G./Overbeck, A./Völker, H. 2003. *Luxemburgische Skriptastudien. Edition und Untersuchung der altfranzösischen Urkunden Gräfin Ermesindes (1226-1247) und Graf Heinrichs V. (1247-1281) von Luxemburg*. Tübingen : Niemeyer.  
 Howell, D. C. 1998. *Méthodes statistiques en sciences humaines*, Paris/Bruxelles : De Boeck et Larquier.  
 Hudson, R. 2010. *An introduction to Word grammar*. Cambridge : Cambridge University Press.  
 Kahane S. 2012. « De l'analyse en grille à la modélisation des entassements » S. Caddeo/M.-N. Roubaud/M. Rouquier/F. Sabio (éds). *Penser les langues avec Claire Blanche-Benveniste*. Presses de l'université de Provence, 101-116.  
 Kahane, S. 1997. « Bubble trees and syntactic representations », T. Becker./H.-U. Krieger (éds). *Proceedings of the 5th meeting of Mathematics of Language (MOL 5), Saarbrücken*, 70-76.  
 Kahane, S./Mazziotta, N. 2015a. « Dependency-based analyses for function words. Introducing the polygraphic approach. », J. Nivre/E. Hajičová, *Proceedings of Depling 2015. International conference on dependency linguistics*. Uppsala : Uppsala University, 181-190.  
 Kahane, S./Mazziotta, N. 2015b. « Quel classement pour les "marqueurs discursifs", "mots-phrases" et autres

- “inserts”. Prédicatifs et locutifs », *Travaux de Linguistique : Revue Internationale de Linguistique Française* 71, 7-42.
- Kahane, S./Mazziotta, N. 2015c. « Syntactic polygraphs. A formalism extending both constituency and dependency », M. Kuhlman/M. Kanazawa/G. M. Kobele (éds). *Proceedings of the 14th Meeting on Mathematics of Language (MOL 2015)*. Association for Computational Linguistics, 152-164.
- Kahane, S./Mazziotta, N. soumis. « Encoding constituency in a dependency-based representation: introducing dependency polygraphs and their formal properties » [titre provisoire], *Journal of language modelling*.
- Kahane, S./Osborne, T. 2015. « Translators’ introduction », *Tesnière 2015* : xxix-lxxiv.
- Kunstmann, P./Stein, A. (éds). 2007a. *Le Nouveau Corpus d’Amsterdam. Actes de l’atelier de Lauterbad, 23-26 février 2006*. Stuttgart : Steiner.
- Kunstmann, P./Stein A. 2007b. « Le Nouveau Corpus d’Amsterdam », *Kunstmann/Stein 2007a* : 9-27.
- Laenzlinger, C. 2003. *Initiation à la Syntaxe formelle du français: Le modèle Principes et Paramètres de la Grammaire Générative Transformationnelle*. Berne : Peter Lang.
- Lavrentiev, A. 2005. *Manuel d’encodage XML-TEI étendu des transcriptions de manuscrits dans le projet BFM-Manuscrits. Normes utilisées à la date du 14 décembre 2005*. <[http://bfm.ens-lsh.fr/IMG/pdf/BFM-Mss\\_Encodage-XML.pdf](http://bfm.ens-lsh.fr/IMG/pdf/BFM-Mss_Encodage-XML.pdf)> [périmé]
- Lavrentiev, A. 2009a. Tendances de la ponctuation dans les manuscrits et incunables français en prose, du XIII<sup>e</sup> au XV<sup>e</sup> siècle. Volume 1. Lyon : École normale supérieure Lettres et Sciences Humaines [Thèse de doctorat]. <<https://halshs.archives-ouvertes.fr/halshs-00846762>>
- Lavrentiev, A. 2009b. Tendances de la ponctuation dans les manuscrits et incunables français en prose, du XIII<sup>e</sup> au XV<sup>e</sup> siècle. Volume 2. Annexes. Lyon : École normale supérieure Lettres et Sciences Humaines [Annexe inédite à Lavrentiev 2009a].
- Lavrentiev, A. 2010. « La ‘phrase’ en français médiéval : une réalité ou une reconstruction artificielle ? », F. Neveu et al. (éds), *Actes du 2e Congrès Mondial de Linguistique Française, La Nouvelle Orléans, 12-15 juillet 2010*. Institut de Linguistique Française, 277-289.
- Lazard, G. 1994. *L’actance*. Paris : P.U.F.
- Lebart, L./Morineau, A./Piron, M. 1998. *Statistique exploratoire multidimensionnelle*. Paris : Dunod.
- Lebart, L./Salem, A. 1994. *Statistique textuelle*. Paris : Dunod.
- Lemaréchal, A. 1989. *Les parties du discours*. Paris : P.U.F.
- Lemaréchal, A. 1997. *Zéro(s)*. Paris : P.U.F.
- Lima, M. 2014. *The book of trees. Visualizing branches of knowledge*. New York : Princeton Architectural Press.
- Lits, M. 2009. *Le petit Grevisse. Grammaire française*. Bruxelles : De Boeck/Duculot.
- Marchello-Nizia, C. 1978. « Ponctuation et “unités de lecture” dans les manuscrits médiévaux ou : je ponctue, tu lis, il théorise », *Langue française* 40, 32–44.
- Marchello-Nizia, C. 2012. « L’oral représenté : un accès construit à une face cachée des langues ‘mortes’ », C. Guillot/B. Combettes/A. Lavrentiev/E. Oppermann-Marsaux/S. Prévost. *Le changement en français. Etudes de linguistique diachronique*. Bern/Berlin/Bruxelles : Peter Lang, 247-264.
- Marchello-Nizia, C./Lavrentiev, A./Guillont-Barbance, C. 2015. « Édition électronique de la *Queste del saint Graal* », *Trotter 2015* : 155-176.
- Martin, R. 2002. *Comprendre la linguistique. Épistémologie élémentaire d’une discipline*. Paris : PUF.
- Martin, R. 2003. *Sémantique et automate. L’apport du dictionnaire informatisé*. Paris : PUF.
- Mazziotta, N. 2001. *Chartes originales françaises de l’abbaye du Val-Benoît datant de 1259 à 1297*. Liège : Université de Liège [Mémoire de maîtrise non publié].
- Mazziotta, N. 2004. « Le texte dans tous ses états. Philosophie d’encodage du projet *Khartès* » A. Dister/G. Purnelle/C. Fairon (éds). *Le poids des mots. Actes des 7es Journées internationales d’Analyse statistique des Données Textuelles*. Louvain-la-Neuve, 793-803.

- Mazziotta, N. 2006a. « L'apport des chartes originales antérieures à 1271 conservées au dépôt des Archives de l'État de Liège (Belgique) pour la datation du lexique français », E. Buchi, *Actes de la Journée d'étude « Lexicographie historique française : autour de la mise à jour des notices étymologiques du Trésor de la langue française informatisé »* (Nancy/ATILF, 4 novembre 2005). Nancy : ATILF.
- Mazziotta, N. 2006b. « Objectifs, méthodes et outils du projet *Khartès* », A. Mardale/M. Soufflard/G. Vecherand (éds). *9<sup>e</sup> Atelier des Doctorants en Linguistique. 17 & 18 octobre 2005*. Paris, 188-195.
- Mazziotta, N. 2006c. « Proposition pour l'édition des sources primaires CCFM », communication orale aux 2<sup>es</sup> *Journées du Consortium pour les Corpus de Français Médiéval, Lyon, France*. <<http://orbi.ulg.ac.be/handle/2268/17233>>
- Mazziotta, N. 2007a. « Inconstance ou consensus ? Ponctuation et protase dans les chartes écrites en français à Liège avant 1271 » D. Trotter (éd.), *Actes du XXIV<sup>e</sup> Congrès International de Linguistique et de Philologie Romanes, Aberystwyth, 1<sup>er</sup>-6 août 2004*. Tübingen: Niemeyer, t. 4, 225-238.
- Mazziotta, N. 2007b. « Ponctuation et parties du discours diplomatique dans les actes émanés de la cour allodiale de Liège entre 1260 et 1290 », A. Lavrentiev (éd.). *Actes de la journée d'étude sur les systèmes graphiques de manuscrits médiévaux et incunables français : ponctuation, segmentation, graphies*. Lyon, ENS-LSH, 6 juin 2005. Chambéry: Université de Savoie, 67-80.
- Mazziotta, N. 2008a. « Exploitation de l'Ontology Web Language pour la rédaction des notices de l'Atlas Linguistique de la Wallonie », Heiden/Pincemin 2008 : 823-835.
- Mazziotta, N. 2008b. « Traiter les abréviations du français médiéval. Théorie de l'écriture et pratiques d'encodage », *Corpus* 7, 191-218.
- Mazziotta, N. 2009. *Ponctuation et syntaxe dans la langue française médiévale. Étude d'un corpus de chartes originales écrites à Liège entre 1236 et 1291*. Tübingen : Niemeyer.
- Mazziotta, N. 2010a. « Building the Syntactic Reference Corpus of Medieval French Using NotaBene RDF Annotation Tool », *Proceedings of the 4th Linguistic Annotation Workshop (LAW IV)*. Association for computational linguistics, 142-146.
- Mazziotta, N. 2010b. « Logiciel NotaBene pour l'annotation linguistique. Annotations et conceptualisations multiples », *Recherches qualitatives. Hors-série « Les actes »* 9, 83-94.
- Mazziotta, N. 2011a. « Étude linguistique de la ponctuation d'une charte originale en langue française (Liège 1272) ». *Orbis : Bulletin international de documentation linguistique* 42.
- Mazziotta, N. 2011b. « L'informatisation du Französisches Etymologisches Wörterbuch. Concepts pour une approche modélisée commune à l'Atlas Linguistique de la Wallonie ». *Zeitschrift für romanische Philologie* 127, 36-62.
- Mazziotta, N. 2012a. « Approche dépendancielle de la coordination des compléments du verbe en ancien français », F. Neveu/V. Muni Toke/P. Blumenthal/T. Klinger/P. Ligas/S. Prévost/S. Teston-Bonnard (éds). *3<sup>e</sup> Congrès Mondial de Linguistique Française*. Paris : Institut de linguistique française, 187-199.
- Mazziotta, N. 2012b. « Le Syntactic Reference Corpus of Medieval French : Structure, outils et exploitation. », A. Dister/D. Longrée/G. Purnelle (éds). *11<sup>es</sup> Journées internationales d'Analyse statistique des Données Textuelles (JADT)*. Liège/Bruxelles : Université de Liège/Facultés universitaires saint Louis, 701-713.
- Mazziotta, N. 2013a. « Grammatical Markers and Grammatical Relations in the Simple Clause in Old French », E. Hajičová/K. Gerdes/L. Wanner (éds). *Proceedings of the Second International Conference on Dependency Linguistics (DepLing 2013)*. Prague : Charles University in Prague, Matfyzpress, 207-216.
- Mazziotta, N. 2013b. « Traitement de la coordination dans le Syntactic Reference Corpus of Medieval French (SRCMF) », E. Casanova Herrero/C. Calvo Rigual (éds). *Actes du XXVI<sup>e</sup> Congrès de Linguistique et Philologie Romanes*. Berlin: De Gruyter, t. 7, 229-238.
- Mazziotta, N. 2014a. « Analyse exploratoire des propriétés sémantiques et syntaxiques des verbes "psychologiques" (classe P) dans *Les Verbes Français* », É. Née/J.-M. Daube/M. Valette/S., Fleury (éds). *Actes des 12<sup>es</sup> Journées internationales d'Analyse statistique des Données Textuelles*. Paris, 473-484.

- Mazziotta, N. 2014b. « Marquage casuel des dépendants nominaux du verbe en ancien français: approche dépendancielle », F., Neveu/P. Blumenthal/A. Gerstenberg/J. Meinschaefer/S., Prévost (éds). *4e Congrès Mondial de Linguistique Française. Berlin, Allemagne, 19-23 Juillet 2014*. Paris: ILF, 327-339.
- Mazziotta, N. 2014c. « Nature et structure des relations syntaxiques dans le modèle de Lucien Tesnière », *Modèles Linguistiques* 69, 123-152.
- Mazziotta, N. 2015 « Description dépendancielle de la coordination de compléments du verbe en ancien français (13<sup>e</sup> s.) : Différencier la coordination juxtapositive de la coordination appositive », A. Carlier/M. Goyens/B. Lamiroy (éds). *Le français en diachronie. Nouveaux objets et méthodes*. Bern : Peter Lang, 405-430.
- Mazziotta, N. 2016a. « Approche sémiotique de la cartographie linguistique: La représentation symbolique des types », communication orale non publiée à *Des dialectes galloromans aux français régionaux Cartographier et interpréter les données, Strasbourg 27 mai 2016*.
- Mazziotta, N. 2016b. « Étude statistique de la syntaxe des lexèmes *chetif*, *dolent*, *las* et *maleiré* en français médiéval », D. Mayaffre/C. Poudat/L. Vanni/V. Magri/P. Follette (éds). *JADT 2016 - Proceedings of the 13th International Conference on Statistical Analysis of Textual Data*. Nice : CNRS-Université Nice Sophia Antipolis, 473-483.
- Mazziotta, N. 2016c. « La logique de l'espace dans les stemmas de Lucien Tesnière », communication non publiée au *28<sup>e</sup> Congrès International de Linguistique et de Philologie Romanes (CILFR), Roma*.
- Mazziotta, N. 2016d. « Microsyntaxe et macrosyntaxe des lexèmes *chaitif*, *dolent*, *las* et *maleuré/maleureus* en français médiéval. De l'adjectif au locutif », F. Neveu/G. Bergounioux/M.-H. Côté/J.-M. Fournier/L. Hriba/S. Prévost (éds). *Actes du 5e Congrès Mondial de Linguistique Française (CMLF)*, Paris : Institut de linguistique française.
- Mazziotta, N. 2016e. « Variation de la ponctuation dans une charte originale en double expédition écrite en français à Liège au XIII<sup>e</sup> siècle », *Fasseur/Rochelois 2016* : 299-316.
- Mazziotta, N., sous presse. « Drawing Syntax before Syntactic Trees : Stephen Watkins Clark's Sentence Diagrams (1847) », *Historiographia Linguistica*.
- Mazziotta, N./Kahane, S. 2016. « Le "mot-phrase" dans les conceptions syntaxiques de Lucien Tesnière », *Bulletin de la Société de Linguistique de Paris* 111(1), 71-107.
- Mazziotta, N./Kahane S., soumis. « L'émergence de la syntaxe structurale de Lucien Tesnière », *Actes du colloque « Entre vie et théorie : la biographie des linguistes dans l'histoire des sciences du langage »*, 3 et 4 décembre 2015, Amiens.
- Mazziotta, N./Renders, P. 2010. « Vers un enrichissement raisonné de la rétroconversion du Französisches Etymologisches Wörterbuch (FEW) », A. Dykstra/T. Schoonheim (éds). *Proceedings of the 14th EURALEX International Congress*. Ljouwert/Leeuwarden, Netherlands: Fryske Akademy, 1026-1032.
- Mel'čuk, I. 2009. « Dependency in natural language », *Polguère/Mel'čuk 2009* : 1-110.
- Mel'čuk, I./Milićević, J. 2014. *Introduction à la linguistique. Volume 3*. Paris : Hermann.
- Möhren, F. 2015. « L'art du glossaire d'édition », *Trotter 2015* : 397-437.
- Monfrin, J. 1974. « Introduction », J.-G. Gigot. *Chartes en langue française antérieures 1271 conservées dans le département de la Haute-Marne*. Paris : C.N.R.S., xi-lxxx.
- Naïs, H. 1979. « La ponctuation dans le manuscrit B de Villehardouin », N. Catach (éd.). *La ponctuation. Recherches historiques et actuelles. Fascicule deux*, Besançon : C.N.R.S./ HESO, 45-55.
- Nida, E. 1966. *A Synopsis of English Syntax*. London/The Hague/Paris : Mouton and Co.
- Nissille, C. 2007. « Basiolum », *Refonte du FEW*. <<http://www.atilf.fr/few/basiolum.pdf>>
- Parkes, M. B. 1992. *Pause and effect. An introduction to the history of punctuation in the West*. Cambridge : Scholar Press.
- Pédaque 2, R. T. 2006. « Le texte en jeu, permanence et transformation du document », R. T. Pédaque. *Le document à la lumière du numérique*. Caen : C&F éditions.
- Peirce, C. S. 1931-1935. *Collected papers of Charles Sanders Peirce*. Cambridge : Harvard University Press.

- Polguère, A. 2003. *Lexicologie et sémantique lexicale. Notions fondamentales*. Montréal : Presses de l'Université de Montréal.
- Polguère, A./Mel'čuk, I. (éds). 2009. *Dependency in linguistic description*. Amsterdam- Philadelphia : Benjamins.
- Prévost, S. 2011. Français médiéval en diachronie : du corpus à la langue. Lyon : École normale supérieure Lettres et Sciences Humaines [Mémoire d'habilitation à diriger des recherches]. <<https://halshs.archives-ouvertes.fr/tel-00667107v1>>
- Rainsford, T. 2015. « Using the Syntactic Reference Corpus of Medieval French in TIGERSearch (SRCMF) » <<http://srcmf.org/docs/tutorial.pdf>>.
- Renders, P. 2015. *L'informatisation du Französisches Etymologisches Wörterbuch. Modélisation d'un discours étymologique*. Strasbourg : Éditions de linguistique et de philologie.
- RobClariL = Lauer, P. (éd.). 1924. *Robert de Clari. La conquête de Constantinople*, Paris : Champion.
- RolMoign = Moignet, G. (éd.). 1969. *La Chanson de Roland*. Paris : Bordas.
- Roques, M. 1952. « Le manuscrit fr. 794 de la Bibliothèque Nationale et le scribe Guiot », *Romania* 73, 177-199.
- Salauin, J.-M. 2008 « Web, texte, conversation et redocumentarisation », Heiden/Pincemin 2008 : 27-30.
- Saussure, F. de. 1967. *Cours de linguistique générale. Publié par Charles Bailly et Albert Sechehaye, avec la collaboration de Albert Riedlinger. Édition critique préparée par Tullio de Mauro*, Paris, Payot.
- Schøsler, L. 1984. *La déclinaison bicasuelle de l'ancien français : son rôle dans la syntaxe de la phrase, les causes de sa dispration*, Odense : Odense University Press.
- Searle, J. R./Vanderveken, D. 1985. *Foundations of illocutionary logic*. Cambridge : Cambridge University Press.
- Segre. 2003. *La chanson de Roland. Édition critique. Traduite de l'italien par Madeleine Tyssens*. Genève : Droz.
- SGraallVQuesteM = Marchello-Nizia, C./Lavrentiev, A. (éds). 2012. *La quête del saint Graal*. <[http://catalog.bfm-corpus.org/graal\\_cm](http://catalog.bfm-corpus.org/graal_cm)>.
- Soutet, O. 2005. *Linguistique*. Paris : PUF.
- Stein, A. 2014. « Parsing Heterogeneous Corpora with a Rich Dependency Grammar. », N. Calzolari *et al.* (éds). *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavik : European Language Resources Association.
- Stein, A./Prévost, S. 2013. « Syntactic annotation of medieval texts: the *Syntactic Reference Corpus of Medieval French* (SRCMF) », P. Bennett/M. Durrell/Scheible, S./R. Whitt (éds). *New Methods in Historical Corpora*. Tübingen : Narr.
- Steinlin, J./Kahane, S./Polguère, A./El Ghali, A. 2004. « De l'article lexicographique à la modélisation objet du dictionnaire et des liens lexicaux », *Actes de EURALEX'2004, Lorient, 177-186*. <[http://olst.ling.umontreal.ca/pdf/js\\_elx2004.pdf](http://olst.ling.umontreal.ca/pdf/js_elx2004.pdf)>.
- Stiennon, Jacques. <sup>2</sup>1991. *Paléographie du moyen âge*. Paris : Armand Colin.
- Swiggers, P. 1994. « Aux débuts de la syntaxe structurale : Tesnière et la construction d'une syntaxe », J. Orešnik (éd.). *Mélanges Lucien Tesnière. Actes du colloque international Lucien Tesnière, Linguiste européen et slovène (1893-1993). Ljubljana, 18-20 novembre 1993*. Ljubljana : Faculté des Lettres et de Philosophie de l'Université de Ljubljana, 209-219.
- Tesnière, L. 1959. *Éléments de syntaxe structurale*. Paris : Klincksieck.
- Tesnière, L. 2015. *Elements of structural syntax, translated by Timothy Osborne and Sylvain Kahane*. Amsterdam/Philadelphia : Benjamins.
- TristBérM<sup>4</sup> = Muret, E./Defourques, L. M. (éds). <sup>4</sup>1947. Paris : Champion.
- Trotter, D. (éd.). 2015. *Manuel de la philologie de l'édition*. Berlin/Boston : de Gruyter.
- Tylén, K./Fusaroli, R./Bjørndahl, J. S./Rączaszek-Leonardi, J./Østergaard, S./Stjernfelt, F. 2014.

- 
- « Diagrammatic reasoning. Abstraction, interaction, and insight », *Pragmatics & Cognition* 22(2), 264-283.
- Van Raemdonck, D. 2009. « “Je pense donc... je dis.” », *Linx* 61, 31-47
- Wartburg, W. von. (éd.). 1922–2002. *Französisches Etymologisches Wörterbuch. Eine darstellung des galloromanischen sprachschatzes*. Leipzig/Bonn/Bâle : Schroeder/Klopp/Teubner/Helbing und Lichtenhahn/Zbinden.
- Wilhelm, R. 2015. « L'édition de texte – entreprise à la fois linguistique et littéraire », *Trotter* 2015 : 131-151.
- Wilmet, M. 2016. *Il y a grammaire et grammaire*. Bruxelles : Académie royale de Belgique.



# Table des matières

---

<b>Sommaire</b>	<b>vii</b>
<b>0. Introduction</b>	<b>1</b>
0.1. Présentation succincte de mes travaux.....	2
0.2. Inscriptions des connaissances.....	3
0.2.1. Inscription formelle.....	4
0.2.2. Inscription graphique et encyclopédie visuelle.....	5
0.2.3. Inscription diagrammatique.....	6
0.3. Questions traitées.....	7
<b>1. Lire</b>	<b>9</b>
1.1. Lire les éditions de matériaux.....	9
1.1.1. Lire les textes.....	9
1.1.2. Lire les répertoires de matériaux linguistiques.....	13
1.2. Lire les schémas syntaxiques.....	16
1.2.1. Lire les diagrammes des pionniers : Clark (1847) et Tesnière (1959).....	17
1.2.2. Lire les arbres dépendanciels.....	20
1.3. Synthèse.....	21
<b>2. Choisir</b>	<b>23</b>
2.1. Choisir les phénomènes représentés.....	23
2.1.1. Recherche de la pertinence.....	24
2.1.2. Contraintes liées à l'enrichissement.....	26
2.2. Choisir les analyses explicitées.....	28
2.2.1. Limiter l'inventaire des classes.....	28
2.2.2. Réification des éléments de l'analyse.....	30
a. Réification et analyses syntaxiques.....	31
b. Réification et analyses lexicologiques.....	32
2.3. Choisir la structure formelle de l'inscription.....	36
2.3.1. Limites des structures formelles.....	36
2.3.2. L'apport du polygraphe.....	38
a. Dépendance ou stratification.....	38
b. Solution du polygraphe.....	39
2.4. Synthèse.....	42

<b>3. Découvrir</b>	<b>45</b>
3.1. Explorer les faits.....	46
3.1.1. Cartes linguistiques.....	46
3.1.2. Analyses statistiques et plans factoriels.....	47
3.2. Élaborer la théorie.....	50
3.2.1. Démontrer.....	50
3.2.2. Tester et sélectionner.....	52
3.3. Évaluer les structures formelles.....	53
3.3.1. Changer le formalisme.....	53
a. Enrichir le système formel.....	53
b. Élaborer de nouvelles structures.....	55
3.3.2. Contraster les diagrammes.....	57
a. Comprendre l'équivalence.....	57
b. Comprendre la différence.....	61
3.4. Synthèse.....	63
<b>4. Numériser</b>	<b>65</b>
4.1. Inscription informatique d'éditions de textes.....	66
4.1.1. Introduction à XML.....	67
4.1.2. Inscription unique, vues multiples et diffraction des unités.....	69
4.1.3. Complexité de l'inscription et ergonomie.....	70
4.2. Les inscriptions dans SRCMF.....	73
4.2.1. Référencement et interfaçage.....	73
a. Difficultés ergonomiques.....	74
b. Importance de l'interfaçage.....	75
4.2.2. Manipulations.....	76
a. Manipulations dans NotaBene.....	76
b. Exportations.....	77
4.3. Synthèse.....	80
<b>5. Conclusion</b>	<b>83</b>
<b>Références</b>	<b>87</b>