

LE LATIN DE THOMAS D'AQUIN ET L'INTELLECTUALISATION DE HAVRÁNEK. ENTRE LEXICOGRAFIE COMPUTATIONNELLE ET HISTOIRE DES IDÉES SÉMIOTIQUES

Savina RAYNAUD, Marco PASSAROTTI
Università Cattolica del S. Cuore, Milan

Abstract (En): The paper aims to evaluate the semantic - terminological effect on the use of the Latin lexicon that are shaped by a particular author, given the complete documentation of the occurrences of his lemmata (*Index Thomisticus*), further supplied by on-going syntactic annotation, but nevertheless requiring proper, theoretically well-founded semantic analysis and lexicographic production. We propose linking the rich theoretical tradition which has produced the *Prague Dependency Treebank* (PDT) and the correlated *Tectogrammatical Annotation* (TGA) to the objectives of the *Bicultural Thomistic Lexicon* (LTB). To do this, we will examine and discuss Havránek's contribution to the shaping of standard language, especially through the distinctive, creative procedure he calls intellectualisation. At the same time, we will outline the effect determined by the use – established in all its extensions – of an expression (lexical item or noun phrase). Given that the user – Thomas Aquinas – was an author who profoundly influenced subsequent thought, the creation or codification of a terminological value of the item will be investigated.

Keywords (En): Semantics; terminology; *Prague Dependency Treebank* (PDT); *Tectogrammatical Annotation* (TGA); Bohuslav Havránek; standard language; intellectualisation; *Index Thomisticus* (IT); *Index Thomisticus Treebank* (IT-TB); terminological codification, *suppositio*.

Mots-clés (Fr): Sémantique; terminologie; *Prague Dependency Treebank* (PDT); *Tectogrammatical Annotation* (TGA); Bohuslav Havránek; langue standard; intellectualisation; *Index Thomisticus* (IT); *Index Thomisticus Treebank* (IT-TB); codification terminologique, *suppositio*.

Introduction

Le thème *Usage, norme et codification à l'âge des corpus informatisés*, porte à la présentation de la convergence – que les deux coauteurs de cette contribution considèrent comme une *heureuse* convergence – de deux pistes de recherche, approfondies depuis longtemps, de manière à la fois autonome et en forte corrélation.

À l'âge des corpus informatisés, notre point de départ est *l'usage*. Il nous faut déclarer tout d'abord que cet usage justement, et de façon plus spécifique l'emploi de la langue-objet dont nous nous occupons, est celui du latin – latin médiéval, philosophique et théologique – et qu'il est accompli par un auteur bien précis, Thomas D'Aquin (1225-1274). Il s'agit donc d'un idiolecte, objet d'un traitement informatisé dont nous illustrerons les réalisations et l'état d'élaboration.

Ce *corpus* fait actuellement l'objet d'une ultérieure annotation dont la finalité est lexicographique, comme nous l'expliquerons plus loin.

Il est donc nécessaire de disposer d'une bonne *théorie sémiotique* et *sémantique*, qui ait mûri pendant une période de temps presque aussi longue que celle de la philosophie, mais qui soit en même temps sélectionnée sur la base de

ses capacités à rendre compte d'un langage savant, spécialisé, lointain dans le temps et qui n'appartient plus à une langue vivante.

D'un côté donc, une langue comme le latin, d'une extension spatio-temporelle incomparable (une vingtaine de siècles au moins), médiatrice d'une culture encore actuellement déterminante pour l'Occident ; de l'autre, une théorie linguistique comme celle de l'École de Prague, toute jeune encore (moins de cent ans de vie) et relativement périphérique pour ce qui concerne la langue d'élection (le tchèque). Pourtant, il s'agit d'une théorie qui s'est développée au cœur même de l'Europe et dans le cadre d'une tradition désormais reconnue comme classique, voire comme particulièrement innovante puisqu'elle a traité le thème de la culture linguistique (et avec lui celui de la langue de la culture) selon une perspective linguistico-générale ; ceci, non pas uniquement pour expliquer des données de fait mais également pour planifier ou guider des développements futurs. Sur ce plan, nous chercherons donc des réponses concernant le processus de *codification* – c.-à-d. de la création d'une langue de spécialité – qui a été réalisé dans le latin médiéval de la philosophie et de la théologie, et l'institution, qui en découle, d'une *norme* métalinguistique opérant sur le lexique philosophique et théologique à venir.

1. Les données et leur élaboration

La productivité de l'emploi de l'appareil métalinguistique pragois sur des corpus de langues uniques, même typologiquement très différentes les unes des autres, ainsi que sur des corpus parallèles, est déjà bien documentée. En particulier, pour ce qui concerne le latin philosophique médiéval et plus spécifiquement l'idiolecte de l'Aquinat dont l'indexisation lemmatisée, ou *Index Thomisticus* (IT) était déjà disponible, l'*IT-Treebank* et le *Valency Lexicon* sont désormais en phase de développement avancé. Le projet LTB (Lessico Tomistico Biculturale), ébauché par l'auteur de l'IT, Roberto Busa, a pour but d'analyser la nomenclature complète des textes de l'IT, afin de produire un « dictionnaire de synthèse avec des traductions conceptuelles qui, pour chaque entrée latine du *corpus thomisticum*, dise quelles entrées de plusieurs langues d'aujourd'hui expriment les différents concepts que le latin exprimait alors » (Busa, 2004 : 4).

Index Thomisticus

Initié en 1949 par Père Roberto Busa, l'Index Thomisticus (IT; BUSA, 1974-1980) est le premier corpus digital du latin. L'IT contient l'*opera omnia* de Thomas d'Aquin (soit 118 textes) ainsi que 61 textes d'autres auteurs en relation avec Thomas d'Aquin, pour un total de 11 millions d'occurrences. Le corpus est étiqueté d'un point de vue morphologique et lemmatisé.

Le projet de Treebank de l'Index Thomisticus

C'est ici que se greffe le lien entre l'entreprise pionnière de Père Busa et le développement linguistico-computationnel de l'École de Prague. La thèse de base est que pour rendre compte des valeurs sémantiques du lexique d'un auteur, l'analyse des occurrences des unités prises individuellement ne suffit pas. La

même intégration entre Index et Concordances dans la publication d'origine de l'Index confirme que son auteur était conscient de l'importance de fournir au lecteur le contexte le plus proche de chaque occurrence. En effet, la notion de concordance est de nature syntaxique, et elle représente, par antonomase, n'importe quelle relation syntaxique, y compris celles de complément / régime.

Le projet intitulé *Index Thomisticus Treebank* (IT-TB; <http://itreebank.marginalia.it>), conçu par Marco Passarotti et lancé en 2006, a donc pour objectif de réaliser l'annotation syntaxique de l'IT en entier. Actuellement, le corpus de l'IT-TB contient environ 250 000 occurrences pour un total d'approximativement 15 000 phrases extraites du *Scriptum super Sententiis Magistri Petri Lombardi*, de la *Summa contra Gentiles* et de la *Summa Theologiae*.

Nous pouvons citer comme exemple de concordance la section de l'Index Thomisticus Treebank qui comprend l'annotation syntaxique de toutes les phrases qui contiennent au moins une occurrence du lemme « forma » dans le *Commentaire des Sentences de Pierre Lombard*, dans la *Summa Contra Gentiles* et dans la portion de la *Summa Theologiae* comprise entre le début de l'œuvre et la Quaestio 76 de la première partie.

Les treebanks

Les *treebanks*, ou corpus arborés, sont des corpus textuels annotés au niveau syntaxique. Les données peuvent être annotées en fonction de deux cadres grammaticaux principaux, tous deux représentant les structures syntaxiques par des graphiques en arbre. Il s'agit d'une grammaire des structures syntagmatiques (*phrase structure grammar*, *PSG*), utilisée surtout dans les approches génératives (Chomsky, 1957), et d'une grammaire de dépendance (*dependency grammar*, *DG*), initiée par L. Tesnière (1959).

Fondements théoriques de l'IT-TB: FGD

Alors que dans les arbres de la DG, les nœuds sont tous étiquetés par des mots ou des strings vides, dans les arbres de la PSG, seuls les nœuds qu'on appelle les « *leaf nodes* », sont systématiquement étiquetés par des mots ou des strings vides, tandis que les nœuds internes sont étiquetés par des symboles non terminaux.

Le latin est une langue comportant de nombreuses flexions, des constituants discontinus et un ordre des mots modérément libre. La PSG étant considérée comme un cadre bien adapté à la représentation des langues avec peu de flexions, ce sont ces caractéristiques du latin qui nous ont amenés à choisir la DG comme le cadre grammatical le plus approprié pour construire les *treebanks* du latin.

De plus, la DG a récemment suscité un intérêt particulier car il s'agit d'un cadre qui, pour être simple, offre cependant des informations utiles à l'exécution de nombreuses tâches de TALN (Traitement Automatique du Langage Naturel) et à la représentation de la structure prédicat-argument.

Le choix du cadre de représentation (dans ce cas, la DG ou la PSG) ne détermine pas la représentation d'une phrase donnée, vu qu'il peut y avoir plusieurs arbres de dépendance et plusieurs arbres de structures syntagmatiques / de type PSG (tous corrects), même pour des phrases simples. Inversement, la

représentation d'une phrase de façon neutre d'un point de vue théorique est impossible, puisque tout cadre de représentation nécessite une théorie permettant d'en extraire le sens.

Le style d'annotation et le flux de travail de l'IT-TB se fondent sur la Description Fonctionnelle Générative (*Functional Generative Description*, FGD; Sgall et al., 1986), un cadre théorique basé sur la dépendance, développé à Prague, qui a été appliqué et testé de façon intensive lors de la construction du PDT.

PDT est un treebank basé sur la dépendance avec une structure à trois niveaux, ordonnés comme suit :

1. un niveau morphologique, celui de l'annotation morphologique désambiguïsée et de la lemmatisation ;
2. un niveau analytique (qui est le niveau d'annotation actuelle dans le IT-TB) : celui de l'annotation syntaxique superficielle (de surface) basée sur la dépendance ;
3. le niveau « tectogrammatical » (actuellement en cours de réalisation) : l'annotation du sens sous-jacent de la phrase.

Le développement de chaque niveau exige la disponibilité du ou des précédents.

Les niveaux analytique et tectogrammatical décrivent tous deux la structure de la phrase à travers des graphiques arborés de dépendance, appelés respectivement les structures analytiques arborées (*Analytical tree-structure* ou ATSS) et les structures tectogrammatisées arborées (*tectogrammatical tree-structures*, ou TGTSS).

Les principales différences entre les ATSS et les TGTSS sont les suivantes :

- Dans les ATSS, chaque mot et chaque signe de ponctuation d'une phrase est représenté par le nœud d'un arbre de dépendance à racines. Les contours de l'arbre correspondent aux relations de dépendance syntaxique (de surface) telles que Sujet, Objet, etc. (appelées « fonctions analytiques »). La non-projectivité est autorisée¹.
- Les TGTSS décrivent les structures sous-jacentes de la phrase, conçues comme la contrepartie se rapportant aux aspects sémantiques des moyens d'expression grammaticaux (reportés par les ATSS). Les nœuds des TGTSS ne représentent que les mots autosémantiques, étiquetés par des tags à rôle sémantique appelés '(deep) functors', tels qu'Acteur, Patient, Destinataire, Effet et Origine pour les 'participants internes' (les arguments, soit les compléments obligatoires de verbes, de noms et d'adjectifs – même s'ils ne sont pas nécessairement représentés au niveau textuel), ainsi que différents types d'adverbiaux pour les 'modifications libres' (ajouts).

¹ La condition de projectivité, définie officiellement par MARCUS (1965), exige que chaque sous-arborescence de dépendance couvre une région contiguë de la phrase : un mot et ses dépendants directs doivent couvrir une séquence continue dans un ordre linéaire. Si cette condition n'est pas remplie, la non-projectivité a lieu.

Les valeurs des mots à fonction et des marques de ponctuation sont attribuées par les étiquettes des nœuds se rapportant aux mots autosémantiques. Les deux dimensions des TGTs représentent la structure syntaxique d'une phrase (la dimension verticale) et sa structure informationnelle, c'est-à-dire l'articulation sujet-focus (*topic-focus articulation* ou TFA), basée sur l'ordre des mots sous-jacent (la dimension horizontale). Dans la FGD, la TFA gère l'opposition entre ce qui a trait au contexte (*contextual boundness*, soit l'information donnée, sur la gauche) et ce qui n'a pas trait au contexte (*contextual unboundness*, soit la 'nouvelle' information sur la droite). La résolution des ellipses et des anaphores ainsi que les relations coréférentielles sont également représentées en TGTs. La non-projectivité n'est pas autorisée.

Dans le cadre du projet IT-TB, nous avons choisi le style d'annotation de la PDT pour des raisons à la fois linguistiques et structurelles. En ce qui concerne ces premières, le latin et le tchèque partagent certaines propriétés importantes, comme par exemple le fait d'être des langues riches en flexions, avec des expressions discontinues, un ordre des mots modérément libre et un degré de synonymie et d'ambiguïté des désinences élevé. Ces deux langues ont trois genres (masculin, féminin, neutre), des cas ayant plus ou moins le même sens et pas d'article. Pour ce qui concerne les raisons structurelles, la structure à trois niveaux du PDT est idéale autant pour nos besoins actuels que dans la perspective de développement de nouvelles ressources linguistiques et d'instruments pour le latin. En effet, l'annotation analytique en PDT n'est pas pensée pour être un niveau indépendant mais plutôt comme une étape technique vers l'annotation tectogrammaticale. L'étroite relation existant entre la structure générale du flux de travail d'annotation en PDT et une solide base théorique comme la FGD nous permet de considérer chaque niveau d'annotation pris individuellement comme une partie d'un cadre général, fonctionnant grâce à une perspective fonctionnelle et ayant comme but la compréhension du sens sous-jacent des phrases.

Le Valency Lexicon

La valence est définie par le nombre de compléments obligatoires requis par un mot : ces compléments sont généralement appelés les "arguments", tandis que ceux qui ne sont pas obligatoires sont appelés "ajouts". Même si une valence peut être attribuée à différentes parties du discours (notamment des verbes, des noms et des adjectifs), ce sont surtout les verbes qui ont retenu l'attention des scientifiques, à tel point que la notion de valence coïncide souvent avec celle de valence du verbe.

Alors que Karl Bühler peut être considéré comme le pionnier de la théorie moderne de la valence², Lucien Tesnière est généralement reconnu comme son

² Dans la *Sprachtheorie*, Bühler écrit que « die Wörter einer bestimmten Wortklasse eine oder mehrere *Leerstellen* um sich eröffnen, die durch Wörter bestimmter anderer Wortklassen ausgefüllt werden müssen » (BÜHLER, 1934 : 173) ; « les mots d'une classe déterminée ouvrent autour d'eux une ou plusieurs *places vides*, qui doivent être remplies par des mots d'autres classes de mots déterminées » (BÜHLER, 2009 : 291). Cependant, d'un point de vue théorique, on ne peut passer sous silence les noms de Peirce et de Frege en tant que fondateurs de la conception valentielle : v. RAYNAUD (2012 : 65-70).

vrai fondateur. Tesnière considère la valence comme une qualité quantitative des verbes, puisque seuls les verbes agissent à la fois sur la quantité et la qualité (c'est à dire les noms et les adverbes) de leurs arguments obligatoires ; reprenant une métaphore empruntée au théâtre, Tesnière classe les dépendants en deux groupes : les actants (arguments) et les circonstants (ajouts) : « Le nœud verbal [...] exprime tout un petit drame. Comme un drame en effet, il comporte obligatoirement un procès, et le plus souvent des acteurs et des circonstances. Transposés du plan de la réalité dramatique sur celui de la syntaxe structurale, le procès, les acteurs et les circonstances deviennent respectivement le verbe, les actants et les circonstants » (TESNIÈRE, 1959 : 102).

Le lexique de valence *Index ThomisticusTreebank Valency Lexicon* (MCGILLIVRAY & PASSAROTTI, 2009 ; IT-VaLex) est un ensemble d'entrées lexicales verbales auquel on a ajouté des cadres de valence et de sous-catégorisation. L'IT-VaLex est étroitement lié à l'IT-TB, vu qu'il s'agit d'un lexique de valence « corpus-driven », induit automatiquement par le niveau d'annotation de l'IT-TB. La procédure automatique permet non seulement d'étendre ce travail au *Latin Dependency Treebank* du latin classique (grâce aux lignes communes d'annotation ; BAMMAN, PASSAROTTI, CRANE, RAYNAUD, 2007), mais également de mettre à jour le lexique au fur et à mesure que le *Treebank* augmente.

Il est possible de naviguer dans l'IT-Valex grâce à une interface en ligne, par entrée lexicale ou bien en fonction du nombre et de l'ordre de surface des arguments, qui sont liés à leurs *fillers* lexicaux. (<http://itreebank.marginalia.it/itvalex>).

Le lexique thomiste biculturel

Le projet IT-TB fait partie d'un projet plus ample intitulé « Lexique Thomiste Biculturel » (*Lessico Tomistico Biculturale*, ou LTB). Le LTB a pour but de construire un nouveau lexique de Thomas d'Aquin à travers une comparaison empirique avec les résultats obtenus par l'IT. En effet, les entrées des lexiques déjà disponibles de Thomas d'Aquin sont systématiquement biaisées par les critères de sélection des exemples adoptés pour décrire les différents sens des lemmes. Cette limitation peut maintenant être dépassée en exploitant l'IT et, en particulier, le IT-TB.

Nous avons jugé qu'une exposition soignée de la constitution du corpus était indispensable. L'élaboration des données (§1) est néanmoins une condition trop importante pour la traiter à la hâte. Aucun problème ne peut être examiné et résolu sans prendre en charge les données, aucune finalité ne peut être poursuivie sans préparer soigneusement les moyens qui lui seront nécessaires.

Jusqu'à présent, notre intention a été de présenter l'architecture des méthodes et des finalités de l'élaboration des données dans son ensemble. L'analyse sémantique n'en est qu'à ses débuts avec jusqu'ici le lemme „forma“. Nous choisirons plus loin d'analyser le lemme „suppositio“.

2. L'exploitation des données. De la syntaxe à la sémantique, la technique du clustering

Le *Clustering* est une technique permettant de passer du traitement syntaxique au traitement sémantique des données.

Il s'agit d'une méthode d'apprentissage non supervisée (*machine learning*) ayant pour objectif de trouver une structure dans un ensemble de données, catégorisées ou non. Les objets sont organisés en groupes (les clusters) dont les membres ont certaines caractéristiques en commun et diffèrent des membres des autres clusters.

Les techniques de clustering de mots suivent généralement une procédure en deux temps :

1. classification : chaque occurrence du mot est représentée par une observation dans une matrice et la similitude ou la dissimilitude de deux observations est évaluée.
2. clustering : un certain algorithme de clustering est appliqué, afin de regrouper les occurrences qui se ressemblent.

Le clustering hiérarchique est une méthode spécifique d'analyse par cluster qui a pour but de construire des hiérarchies de clusters. Elle utilise deux stratégies principales :

- la stratégie par agglomération (de bas en haut) : chaque observation commence dans son propre cluster, puis des paires de clusters sont unies au fur et à mesure que l'on remonte la hiérarchie ;
- la stratégie par division (de haut en bas) : toutes les observations commencent dans un cluster, et des divisions sont effectuées de façon récursive au fur et à mesure que l'on descend dans la hiérarchie.

L'application des techniques de clustering aux données textuelles afin d'organiser les différentes significations des mots se fonde sur l'hypothèse théorique que les mots utilisés dans des contextes similaires ont tendance à avoir des sens similaires ou proches (l'hypothèse distributionnelle ou *Distributional Hypothesis* ou DH de HARRIS, 1954). Comme la DH, la notion de contexte de situation de Firth souligne la nature dépendante du contexte que présente le sens, comme le souligne cette célèbre citation : « You shall know a word by the company it keeps » (FIRTH, 1957 : 11).

L'analyse du comportement contextuel d'un mot dans le but d'organiser son sens est en étroite relation avec la désambiguïsation du sens d'un mot.

La désambiguïsation du sens d'un mot doit être nettement séparée de la différenciation des sens d'un mot. La première attribue à un mot un ensemble de significations extrait d'un répertoire de sens déjà existant, habituellement proposé par un dictionnaire ou par d'autres ressources artisanales. La deuxième est une méthode qui n'utilise aucun répertoire de sens : les clusters n'ont aucune définition ou étiquette de sens qui leur soit associée, mais ils sont étiquetés de façon (semi-) automatique à travers un commentaire qui décrit la signification sous-jacente du mot cible dans les contextes qui partagent les mêmes caractéristiques distributionnelles.

Les contextes utilisés pour le clustering peuvent être de deux types différents :

- Les « contextes à *header* », ayant un mot cible (head) qui joue un rôle de focalisation du contexte ;
- Les « contextes sans *header* », qui ne contiennent pas de mot cible particulier et qui sont utilisés pour déterminer le texte analysé pris globalement, et non un mot spécifique.

La similarité existant entre plusieurs contextes peut être elle-même considérée de deux façons différentes :

La similarité de premier ordre est l'approche la plus intuitive : deux contextes partageant un pourcentage élevé de mots sont probablement semblables et les résultats sont basés sur le nombre de mots qui correspondent.

La similarité de deuxième ordre remplace les contextes par quelque chose qui les représente mais qui offre cependant une base de mesure de la similarité plus riche. Par exemple, un mot peut être remplacé par ses traits sémantiques proposés par une ressource lexicale comme WordNet³.

3. Langue populaire ou commune, langue standard et intellectualisation

Mais comment associer à chacun des mots-clefs de Thomas d'Aquin – 'forma' pour commencer – sa valeur sémantique, identifiée à partir du dépouillement intégral de ses contextes d'occurrence et traduite en prêtant une attention toute particulière au contexte philosophique et théologique dans lequel ces mots sont placés, ainsi qu'au hiatus culturel de près de sept siècles les séparant ? La notion d'*intellectualisation* de Havránek, un Pragoise de la première heure⁴, arrive tout à fait à propos. « By *intellectualization* of the standard language [...] we understand its adaptation to the goal of making possible precise and rigorous, if necessary abstract, statements, capable of expressing the continuity and complexity of thought, that is, to reinforce the intellectual side of speech. » (HAVRÁNEK 1932 : 147). En effet, c'est presque toujours un terme déjà disponible dans la langue standard et progressivement soumis à un processus sémiotique d'« intellectualisation »⁵ qui devient un terme technique, l'expression de la conceptualité propre d'une doctrine philosophique. Ce terme, à son tour, poursuit son immersion dans l'histoire intellectuelle soit de la langue-matrice soit des langues qui en accueillent l'héritage culturelle : plus précisément, la langue de la philosophie qui était la langue latine, héritière de la langue grecque, est devenue le français philosophique, l'allemand, l'italien, etc.

Comme l'exprime le terme d'intellectualisation lui-même, la perspective est celle d'une sémiotisation de niveau avancé: il s'agit de l'attribution de valeurs intellectuelles ultérieures par rapport à celles de base, celles qui sont courantes, et de la reconnaissance de ces valeurs.

³ Cf. <https://wordnet.princeton.edu/>. Sur le WordNet latin cf. MINOZZI (2008).

⁴ Bohuslav Havránek (1893-1978) a été professeur à Brno puis à l'Université Charles de Prague; de 1952 à 1965, il a été directeur de l'Institut de la langue tchèque de l'Académie des Sciences de Tchécoslovaquie. Membre de l'Académie des Sciences, il a aussi été président du Cercle linguistique de Prague et rédacteur en chef de la revue *Slovo a slovesnost*.

⁵ Cf. DOVETTO (2014).

1929 est une année mémorable pour le Cercle Linguistique de Prague et sa neuvième thèse concerne de très près notre sujet : *Importance de la linguistique fonctionnelle pour la culture et la critique des langues [slaves]*.

Toujours dans ce premier numéro des *Travaux du Cercle Linguistique de Prague*, Havránek publie le premier de trois de ses écrits concernant notre thème : *Influence de la fonction de la langue littéraire sur la structure phonologique et grammaticale du tchèque littéraire* (HAVRÁNEK, 1929a).

‘Langue littéraire’, ‘tchèque littéraire’ : nous nous limiterons ici à mentionner le problème délicat que pose la traduction d’un syntagme tel que *spisovná čeština*, traduit parfois par tchèque littéraire, tchèque standard, etc. Bien comprendre la valeur de *spisovná* est tout aussi crucial si on élargit l’extension du syntagme en remplaçant *čeština* par *jazyka* : langue écrite, littéraire ou standard ? Nous renvoyons à un article publié très récemment⁶.

En 1932 fut publié un essai bien plus ample, *Emploi et culture de la langue standard*, en tchèque (HAVRÁNEK, 1932), tandis qu’en 1942, toujours en tchèque, sera publié un article intitulé *La différence fonctionnelle de la langue standard* (HAVRÁNEK, 1942). Comme nous pouvons le voir, c’est ‘langue standard’ qui, dans les traductions, a été le plus souvent choisi.

Tout comme nous nous sommes occupés jusqu’ici d’un travail d’annotation et d’interprétation focalisé sur la *textualité* d’un auteur, nous rendrons compte à présent d’un travail de description et d’explication qui concerne un *système linguistique* et ses stratifications. Cette différence ne comporte cependant pas de divergence, mais plutôt une certaine complémentarité : les possibilités de choix qui se dessinent à partir d’un état donné de la langue fournissent un contrepoint aux choix effectués par un auteur particulier.

Nous pouvons lire en effet dans l’article de HAVRÁNEK (1929a) le passage suivant : « la marque différentielle principale d’une langue littéraire ne consiste pas dans un caractère conservateur » (TCLP, 1 : 107).

« Les différences du lexique [...] dans toute langue littéraire évoluée sont surtout les suivantes : une langue littéraire possède des termes pour des notions ignorées d’une langue populaire, en particulier pour les idées générales et abstraites, elle possède des mots différenciés avec plus de précision et de spécialisation et des mots au sens unique plus souvent [ils ont une tendance aux mots concepts], mais par contre elle se voit limitée en matière d’expression fortement affective par la censure qu’exercent l’intelligence et les convenances sociales.

⁶ Raynaud (2014 : 2). *Spisovná čeština* était la première partie du titre de l’œuvre de 1932. Au premier plan, on y trouve la langue tchèque [čeština]. *Spisovna* a été traduit en anglais par *Standard [Czech]* et en allemand par *Literatursprache* (en français, *langue littéraire*). Ceci pose bien évidemment un problème d’interprétation, puisque les deux traductions ne sont pas équivalentes. Nous jugeons préférable la proposition de Tomáš Hoskovec, actuellement Président du Cercle Linguistique de Prague, qui traduit *spisovná čeština* par *tschechische Gemeinsprache* (en français, *langue standard tchèque*). L’adjectif tchèque *spisovný* dérive de la racine *spis*, nom déverbal provenant de *psát* “écrire” ; la traduction allemande est donc plus littérale.

Les raisons de ces différences sont claires: *elles résident dans la fonction spéciale d'une langue littéraire*. Étant l'objet d'exigences accrues par rapport à celles d'une langue populaire, chargée d'exprimer la civilisation ainsi que la vie intellectuelle, les résultats de la pensée philosophique et religieuse, scientifique, politique et sociale, juridique et administrative, et ce non seulement pour des fins pratiques, mais en vue d'un enseignement technique et de la codification, la langue littéraire a vu son lexique s'étendre considérablement et s'intellectualiser. Le choix plus minutieux et plus sévère des mots [...] est déterminé par le mode d'expression linguistique » (1929a : 108) :

- manifestations écrites
- langage monologué et continu, destiné au public le plus étendu ;
- l'expression linguistique doit être relativement complète, se trouvant moins complétée par la connaissance de la situation et par des procédés de communication autres que linguistiques, mais on a la possibilité de mieux réfléchir aux éléments linguistiques et de les choisir.

« Les différences fonctionnelles dans le lexique d'une langue littéraire ne font pas qu'étendre le vocabulaire [...] ; souvent elles créent *certaines catégories de mots* (de dénomination), par ex. la substantivation des actions verbales » (HAVRÁNEK, 1929a : 109).

« La *tendance aux phrases-jugements* – parallèle à la tendance aux mots concepts – entraîne dans les langues littéraires, et principalement dans les langues écrites, *la prédominance de la phrase à deux parties normalisée et clairement divisée au point de vue formel en sujet et prédicat* » (1929a : 114).

Enfin, « la structure des phrases est *plus fermée et plus complexe* dans une langue littéraire que dans une langue courante et populaire » (1929a : 116).

Et voilà la conclusion, qui semble pensée pour nous : « une langue littéraire évoluée, dans le domaine et à l'époque en cause, est l'intermédiaire de la culture universelle (par ex. le latin au Moyen Âge » (1929a : 120).

À propos de l'essai de 1932, nous désirons seulement rappeler quelques thèses centrales qui en sont extraites et qui concernent le thème de notre discussion, c'est-à-dire l'usage et la norme : « la norme d'une langue standard n'est pas formée simplement de l'usage. Elle se forme, c'est-à-dire qu'elle naît et se développe à la base de différentes tendances et est sujette à différentes interventions, par lesquelles elle diffère de la norme d'une langue populaire. La théorie linguistique, entre autres facteurs, est donc intervenue et peut intervenir encore dans le développement d'une langue standard.

La norme d'une langue standard diffère de celle d'une langue populaire par une plus grande différenciation fonctionnelle et stylistique, par un plus haut degré de prise de conscience de la norme, ainsi que par son caractère coercitif plus prononcé. Tout cela est lié à une exigence de stabilité plus grande.

L'exigence de stabilité est une marque immanente de la norme en général. » (HAVRÁNEK, 1983)

La perspective fonctionnelle émerge nettement : « Même dans la langue populaire, le choix des ressources linguistiques dans les énoncés particuliers est gouverné par le *but de l'énoncé* : il est dirigé par sa fonction [...] La différence réside dans le fait que les fonctions de la langue standard sont mieux développées

et différenciées de façon plus précise [...] dans la langue standard, il y aura toujours un bon nombre de ressources d'usage spécial qui seront l'apanage d'un groupe spécial [...] Le domaine de l'expression *professionnelle pratique* est presque entièrement réservé à la langue standard, et celui de l'expression *scientifique* lui est réservé dans sa totalité. Enfin, c'est la langue standard qui normalement constitue la base régulière des manifestations du *langage poétique*. » (*ibid.*)

Nous sommes d'avis que la remarque de Havránek, que nous proposons ci-dessous, est tout à fait en ligne avec la décision prise dans le cadre de l'IT-TB de faire précéder une annotation syntaxique à l'attribution de valeurs sémantiques : « *on ne peut pas évaluer des termes isolés* séparés de leur emploi fonctionnel et des combinaisons automatisées dont ils peuvent faire part ou prendre la signification automatisée dans un contexte et une fonction donnés pour l'unique signification » (*ibid.*).

Havránek porte donc son attention avec beaucoup d'avance (nous sommes en 1932) sur la terminologie⁷, dans la perspective de sa création et pas seulement dans une perspective de réflexion et/ou de régulation :

« Dans la création des *terminologies* des différents domaines, on a souvent recours aux linguistes. Il faut admettre que dès le XIX^e siècle, et plus récemment depuis la création de la République tchécoslovaque, les linguistes ont accompli un travail très respectable. »

La confirmation et l'explication du rôle joué par l'école linguistique pragoise et sa reconnaissance publique émergent de l'importante contribution de J. V. NEUSTUPNÝ et J. NEKVAPIL (2003)⁸, qui soulignent également l'apport réalisé par la suite, à l'âge classique, par František Daneš⁹.

⁷ La terminologie et la terminographie sont devenues entre-temps (cf. WÜSTER, 1931 ; 1973) des champs structurés de recherche et d'action, avec des sociétés savantes qui leur sont consacrées : « La création d'une association européenne de terminologie – voici ce qu'on lit dans le site de AET (<http://eaf-aet.net/fr/historique/>) – faisait partie des recommandations du rapport final du projet Pointer (Proposals for an Operational Infrastructure for Terminology in Europe), publié début 1996. Ce projet a été conduit par plus de quarante organisations de terminologie de toute l'Europe et cofinancé par la Commission européenne dans le cadre de son Programme de plan d'action multilingue (MLAP).

L'Association européenne de terminologie (AET) a ainsi vu le jour à Kolding, au Danemark, le 3 octobre 1996, en tant qu'organisation professionnelle à but non lucratif, visant à rassembler un maximum de personnes, physiques ou morales, intéressées par la terminologie.

L'AET se propose de réunir tous les individus et institutions européens qui s'intéressent à la terminologie ou sont actifs dans ce domaine. Bien entendu, le mot "européen" s'entend ici dans son acception la plus large et ne se limite pas aux pays de l'Union européenne. Des accords de coopération sont prévus, permettant à d'autres institutions, réseaux ou associations, qu'ils soient européens ou non, d'y participer. C'est notamment le cas avec le Centre international d'information sur la terminologie (Infoterm), le Réseau panlatin de terminologie (Realiter), le Réseau Lexicologie Terminologie Traduction (LTT) et l'Association européenne pour les ressources linguistiques (Elra). »

⁸ Cf. en particulier p. 167-201 : Part VI : Theories of Language Management.

⁹ Cf. DANEŠ (1987) et (1988).

4. Études de cas

Le temps des conclusions, des nouveautés à conquérir serait maintenant venu. Mais comment peut-on conclure sans avoir correctement posé des prémisses? Et comment pouvait-on sous-entendre ce qui n'est bien évidemment pas connu, sinon par les experts de chaque discipline convoquée, c.-à-d. la linguistique computationnelle d'une part et la sémantique théorique de l'autre?

Jusqu'à présent, nous avons donc seulement récapitulé deux prémisses, *de facto* et *de jure* : des prémisses effectivement en cours, en tant qu'œuvres déjà réalisées ou encore en pleine réalisation, mais en même temps proposées pour leur justification d'un point de vue épistémologique et leur précision d'un point de vue méthodologique. Quelles conclusions peut-on tirer, en les mettant en relation ? Ou, autrement dit, comment peut-on inférer quelque chose à partir des données accessibles, grâce à une théorie adéquate?

En effet, c'est presque toujours un terme déjà disponible dans la langue standard et progressivement soumis à un processus sémiotique d'« intellectualisation » qui devient un terme technique, l'expression de la conceptualité propre d'une doctrine philosophique. Ce terme, à son tour, poursuit son immersion dans l'histoire intellectuelle soit de la langue-matrice soit des langues qui en accueillent l'héritage culturelle : plus précisément, la langue de la philosophie qui était la langue latine, héritière de la langue grecque, est devenue le français philosophique, l'allemand, l'italien philosophique, etc., donnant parfois lieu à de véritables transpositions interprétatives.

Ces transpositions interprétatives ont déjà été effectuées par l'auteur de l'IT lui-même : Roberto Busa (1979 et 1981) qui a, par exemple, proposé de rendre *ordo* et *ratio seminalis* par *structure* et *code génétique* respectivement. À d'autres occasions, les solutions adoptées sont plus paresseuses, des calques voire même carrément des emprunts *sine glossa*. Nous aurions aimé illustrer la question délicate des traductions “paresseuses”¹⁰, à partir du cas de la “famille lexicale” de ‘*supponere* / *suppositio* / *suppositum*’, mots-clefs de la sémantique philosophique médiévale, même avant Saint Thomas¹¹.

4.1. Suppono, suppositum, suppositio

Dans l'Index Thomisticus

Le verbe “suppono, -ere” apparaît 3073 fois dans l'IT.

Parmi celles-ci, 2054 sont des formes du participe parfait.

Le participe parfait couvre donc environ deux tiers du total des occurrences.

¹⁰ Cf. it *supposizione*, angl. *supposition*, fr. *supposition* dans GOUBIER (2009). Et la remarque finale de GOUBIER (2009): 532 montre bien le risque inhérent au fait de ne pas thématiser la “spécialisation” intervenue dans la valeur sémantique d'un terme : “Après tout, il s'agit de *supposer*”.

¹¹ Cf. MAIERÙ (1972 : 217-321) ; L.M. De Rijk, *La supposizione naturale: una pietra di paragone per i punti di vista filosofici*, in: R. Fedriga e S. Puggioni, *Logica e Linguaggio nel medioevo*, LED, Milan, 1993, pp. 185-220; Jan Pinborg, *Logica e semantica nel medioevo*, Boringhieri, Turin, 1984 ; Ph. Boehner, *Medieval logic, an Outline of Its Development*, Manchester University Press, Manchester, 1952 ; cf. également <http://www.ontology.co/supposition.htm>, avec ses 242 entrées bibliographiques.

Ce n'est pas un hasard si la forme la plus fréquente est « *suppositum* » (859 occurrences). La seconde est « *supposito* » (518).

Le lemme « *suppositio* » apparaît 477 fois dans l'IT. Plus de la moitié des occurrences est représentée par la forme ablative au singulier « *suppositione* » (248).

Dans l'Index Thomisticus Treebank

Dans l'IT-TB, les lemmes « *suppono*, -ere » et « *suppositum* » ont été séparés.

Le lemme « *suppositum* » apparaît 50 fois.

Son modificateur attributif le plus fréquent est le lemme « *natura* » (5) : par exemple dans « *suppositum naturae* ».

Le lemme dont il dépend le plus fréquemment est « *sum* » (9), suivi de « *significo* » (7).

Le lemme « *suppone*, -ere » apparaît 42 fois.

Le lemme qui dépend le plus souvent de lui est « *sum* » (7), suivi par « *volo* » (4).

Le lemme « *suppositio* » apparaît 24 fois.

Ses modificateurs attributifs¹² les plus fréquents sont les lemmes « *aeternitas* », « *propono* » et « *voluntas* » (2).

Le lemme dont il dépend le plus fréquemment est « *sum* » (9), suivi de « *necessitas* » et « *habeo* » (3).

Les données actuellement disponibles dans le IT-TB ne suffisent pas à fournir une évaluation solide des relations lexico-syntaxiques tissées dans les lemmes en question.

La théorie de l'intellectualisation mériterait, à ce point, d'être appelée à contribution afin d'expliquer le rôle de l'élément formatif préfixe fréquemment utilisé dans la terminologie logico-ontologique : *sub-* (*suppositio*, *subiectum*, *substantia*) et son équivalent grec *hypò*, ainsi que celui des verbes *ponere*, *iacere*, *stare* fréquemment utilisés dans la terminologie logico-grammaticale : nous pensons à *adpositio*, *expositio*, *impositio*, *praepositio*, *propositio* ; à *subjectum*, *objectum*, *adjectivum* ; à *substantivum* et à la réception très large qu'ils ont connue dans les langues modernes, et pas uniquement dans les langues romanes.

Selon nous, une aide formidable provient de la théorie humboldtienne de la forme interne et de son avancement par Anton Marty, qui trouve donc son origine dans la philosophie du langage développée à Prague, une fois encore. Mais il s'agit d'une autre sujet de discussion.

Conclusion

À la lumière de la perspective fonctionnelle du structuralisme pragois, on peut tracer les contours précis du défi qui s'annonce désormais : c'est-à-dire non

¹² « Attribute is a sentence member which depends on a noun [...] and which, as a rule, specifies more closely, or "determines" its meaning. » (HAJIČ et al. 1999 : 55).

seulement celui de la corrélation des données linguistiques (les *corpus* structurés) avec des théories métalinguistiques, mais également celui d'inférer, à partir des données textuelles disponibles, les intentions communicatives de l'auteur, et de chercher ensuite les traces d'une codification terminologique, diffusée de façon plus ou moins large et de manière plus ou moins avisée d'un point de vue sémantique.

L'analyse linguistique, aussi précise soit-elle, ne peut donc pas entièrement satisfaire le devoir herméneutique ; celui-ci doit être conduit en mettant en relation la focalisation de l'enquête lexicale effectuée à chaque fois (dans notre exemple, *suppositio*) avec le système textuel d'appartenance dans son entier, son apport de doctrine, et l'intertextualité contemporaine et successive. L'esprit géométrique et l'esprit de finesse sont invités à faire alliance.

C'est ce que nous espérons réaliser lors de la suite de ce projet, c.-à-d. au cours de la conférence convoquée à Paris par le Département des Sciences du langage de la Faculté des Sciences Humaines et Sociales de l'Université Paris Descartes/COMUE Sorbonne Paris Cité fin novembre 2015, selon l'appel à communication sur *Usage, norme et codification à l'âge des corpus informatisés & de la communication par les nouvelles technologies*, à propos d'*Une intellectualisation à renouveler : entre suppositio et référence, de l'Index Thomisticus jusqu'au LTB*.

BIBLIOGRAPHIE

- BAMMAN David; PASSAROTTI Marco; CRANE Gregory; RAYNAUD Savina (2007), *Guidelines for the Syntactic Annotation of Latin Treebanks*, «Tufts University Digital Library».
- BÜHLER Karl (1934), *Sprachtheorie: die Darstellungsfunktion der Sprache*, Jena: Gustav Fischer. Version française éditée par Didier SAMAIN et Janette FRIEDRICH (2009), *Théorie du langage. La fonction représentationnelle*, Agone, Marseille.
- BUSA Roberto (1974-1980), *Index Thomisticus: sancti Thomae Aquinatis operum omnium indices et concordantiae, in quibus verborum omnium et singulorum formae et lemmata cum suis frequentiiis et contextibus variis modis referuntur quaeque / consociata plurium opera atque electronico IBM automato usus digessit Robertus Busa SJ*, Frommann-Holzboog, Stuttgart-Bad Cannstatt. En ligne : www.corpusthomisticum.org.
- BUSA Roberto (1979), *ORDO dans les œuvres de St. Thomas d'Aquin* in *ORDO 11 Coll. Internazionale Lessico Intellettuale Europeo*, Roma 7/9 Genn. 1977, Ed. Ateneo-Bizzarri, Roma, 1979, p. 59-184.
- BUSA Roberto (1981) *Per S. Tommaso 'ratio seminalis' significa 'codice genetico': problemi e metodi di lessicologia e lessicografia tomistica*, in : *Atti dell'VIII Congresso Tomistico Internazionale*, vol. 1, Città del Vaticano, p. 437-451.

- BUSA Roberto (2004) *Procedure per microanalisi di voci significanti "cose od oggetti"*, *LdM 13-1*, CAEL, Gallarate.
- CHOMSKY Noam (1957), *Syntactic Structures*, The Hague / Paris, Mouton.
- DANEŠ František (1987), Values and attitudes in language standardization, in CHLOUPEK Jan ; NEKVAPIL Jiří (éds.), *Reader in Czech Sociolinguistics*, Amsterdam, J. Benjamins, p. 206-245.
- DANEŠ František (1988), Sprachkultur, in : AMMON Ulrich ; DITTMAR Norbert ; MATTHEIER KLAUS J. (éds.), *Sociolinguistics. An International Handbook of the Science of Language and Society*, Berlin ,Walter de Gruyter, p. 1697-1703.
- DOVETTO Francesca M. (2014), Sul metalinguaggio. A proposito di una recente pubblicazione, in : ORIOLES Vincenzo ; BOMBI Raffaella ; BRAZZO Marica (éds.), *Metalinguaggio. Storia e statuto dei costrutti della linguistica*, Roma, Il Calamo, p. 67-75.
- GOUBIER Frédéric (2009), La teoria della supposizione e le sue cronologie semantiche, *Rivista di Filosofia Neo-Scolastica* CI, 4, p. 501-532.
- HAIČ Jan ; PANEVOVÁ Jarmila ; BURÁNOVÁ Eva ; UŘEŠOVÁ Zdeňka ; BÉMOVÁ Alla (1999), *Annotations at analytical level: Instructions for annotators* (English translation by Z. Kirschner). Technical report. Prague, UFAL MFF UK. <http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/en/a-layer/pdf/a-man-en.pdf>
- HAVRÁNEK Bohuslav (1929a), Influence de la fonction de la langue littéraire sur la structure phonologique et grammaticale du tchèque littéraire. *Travaux du Cercle linguistique de Prague* 1. p. 106-120 ; réimpr. dans *A Prague School Reader in Linguistics*, Bloomington & London, Indiana University Press, 1964¹, 1967³, p. 252-269.
- HAVRÁNEK Bohuslav (1929b), Funkce spisovného jazyka, in : *Sborník přednášek pronesených na Prvém sjezdu československých profesorů filozofie, filologie a historie v Praze 3.7.dubna 1929*, Praha, Stálý přípravný výbor sjezdový 1929, p. 130-138.
- HAVRÁNEK Bohuslav (1932), Úkoly spisovného jazyka a jeho kultura, in : HAVRÁNEK Bohuslav ; WEINGART Miloš (sous la direction de), *Spisovná čeština a jazyková kultura* [Le tchèque standard et la culture de la langue], Praha, Melantrich, pp. 32-84 ; extrait et traduit par Paul L. GARVIN : Emploi et culture de la langue standard, dans *La norme linguistique* (1983), Textes colligés et présentés par Édith BEDARD et Jacques MAURIS, Conseil Supérieur de la langue française, Québec, ISBN 2-551-05243-2 <http://www.cslf.gouv.qc.ca/bibliotheque-virtuelle/publication-html/?tx_igccpplus_pi4%5bfile%5d=publications/pubf101/f101app.html#iii> tr. angl. partielle, The Functional Differentiation of the Standard Language, dans VACHEK Josef (ed., 1983), *Praguiana. Some Basic and Less Known Aspects of the Prague Linguistic School*, Prague, Academia, p. 143-164.
- HAVRÁNEK Bohuslav (1938), Zum Problem der Norm in der heutigen Sprachwissenschaft und Sprachkul, in : *Actes du quatrième Congrès International de Linguistes* (Copenhague 1936), Kopenhagen: 151-156; réimpr. dans VACHEK Josef (ed., 1964), *A Prague School Reader in Linguistics*, Bloomington, Indiana University Press, p. 413-20; une version légèrement adaptée réimpr. dans SCHARNHORST, Jürgen – ISING, Erika (hrsg.,

- 1976). *Grundlagen der Sprachkultur. Beiträge der Prager Linguistik zur Sprachtheorie und Sprachpflege*, Berlin, Akademie Verlag, p. 74-85.
- HAVRÁNEK Bohuslav (1942), K funkčnímu rozvrstvení spisovného jazyka. [On the functional stratification of literary language.] *Časopis pro moderní filologii* 28: 409-416.
- HORÁLEK Karel ; TYL Zdeněk (1954). Soupis prací akad. Bohuslava Havránka (1918-1952), in : *Studie a práce lingvistické*, Praha, Nakl. Československé akademie věd, 1, p. 529-551.
- MAIERÙ Alfonso (1972), *Terminologia logica della tarda scolastica*, Roma, Edizioni dell'Ateneo.
- MARCUS Solomon (1965), Sur la notion de projectivité, *Mathematical Logic Quarterly*, 11(2), p. 181-192.
- MCGILLIVRAY Barbara ; PASSAROTTI Marco (2009), The Development of the Index Thomisticus Treebank Valency Lexicon, in : *Proceedings of LaTeCH-SHELT&R Workshop 2009*, Athens, March 30, 2009, p. 43-50.
- MINOZZI Stefano (2008), La costruzione di una base di conoscenza lessicale per la lingua latina: Latinwordnet, in : SANDRINI Giuseppe (éd.), *Studi in onore di Gilberto Lonardi*, Verona, Fiorini, p. 243-258.
- NEUSTUPNÝ Jiří Václav ; NEKVAPIL Jiří (2003), Language Management in the Czech Republic, *Current Issues in Language Planning*, vol. 4, no. 3 & 4, p. 181-366; <http://ukonline-web.uni-koeln.de/remarks/d5134/rm2159688.pdf> ; réimpr. dans *Language Planning and Policy in Europe*, Clevedon – Buffalo – Toronto, Multilingual Matters Ltd, vol. 2, p. 16-201.
- ORIOLES Vincenzo ; BOMBI Raffaella ; BRAZZO Marica (éds., 2012), *Proceedings of the First Workshop on the Metalanguage of Linguistics. Models and Applications* (University of Udine – Lignano, March 2-3, 2012), Roma, Il Calamo ("Lingue, Linguaggi, Metalinguaggio" 11).
- PASSAROTTI Marco (2012), *From Treebanks to Lexical Entries. Clustering the Index Thomisticus*, in RADIMSKÝ Jan (éd.), *Actes du 31^e Colloque International sur le Lexique et la Grammaire*, České Budějovice, Université de Bohême du Sud à České Budějovice (République Tchéque), p. 143-147.
- PETR Jan ; TYLOVÁ Milena (1985). *Bohuslav Havránek: bibliografický soupis vědeckých prací s přehledem k jeho činnosti*, Praha UK.
- RAYNAUD Savina (2012), Queries and Predicate - Argument Relationship, in *Semantics – Advances in Theories and Mathematical Models*, Dr. Muhammad Tanvir Afzal (ed.), ISBN: 978-953-51-0535-0, InTech, En ligne: <<http://www.intechopen.com/books/semantics-advances-in-theories-and-mathematical-models/queries-and-predicate-argument-relationship>>
- RAYNAUD Savina (2014), The Prague Linguistic Circle, its founder Vilém Mathesius and its commitment to developing linguistic culture from linguistic theory, Dossier d'HEL, SHESL, 2014, *Linguistiques d'intervention. Des usages socio-politiques des savoirs sur le langage et les langues*, p. 8, en ligne : <<https://halshs.archives-ouvertes.fr/halshs-01115236>>.
- SGALL Petr, HAJČOVÁ Eva ; PANEVOVÁ Jarmila (1986), *The Meaning of the Sentence in its Semantic and Pragmatic Aspects*, D. Reidel, Dordrecht, NL.

- TESNIÈRE Lucien (1959), *Éléments de syntaxe structurale*, Editions Klincksieck, Paris.
- Thèses (1929), *Travaux du Cercle linguistique de Prague*, I: 1-7-29.
<http://crecleco.seriot.ch/textes/theses29.html>
- VALLINI Cristina (éd., 2000), *Le parole per le parole. I logonimi nelle lingue e nel metalinguaggio*. Atti del Convegno di Napoli, Istituto Universitario, Orientale, 18-20 dic. 1997, a cura di C. Vallini. Presentazione di T. De Mauro, Roma, Il Calamo.
- WÜSTER Eugen (1931), *Internationale Sprachnormung in der Technik, besonders in der Elektrotechnik*. (Die nationale Sprachnormung und ihre Verallgemeinerung). Berlin, VDJ; 2. erw. Aufl. Bonn, Bouvier 1966.
- WÜSTER Eugen (1973), *Benennungs- und Wörterbuchgrundsätze, Ihre Anfänge in Deutschland*. In: *Muttersprache, Zeitschrift zur Pflege und Erforschung der deutschen Sprache*. Nr. 1 Y 4146 F.