La valence verbale dans les traditions linguistiques françaises et slaves

P. Pognan

Notre but ne sera pas de donner ici une image la plus complète possible de la valence verbale (ce qui dépasserait de très loin le cadre de ce chapitre), mais de présenter au lecteur les fondements et les filiations dont nous avons besoin pour élaborer une grammaire de la valence verbale dans les langues slaves et tout particulièrement en slovaque. Cette grammaire relative à la valence doit rendre compte simultanément du système linguistique slave et de sa calculabilité dans une vision didactique basée sur le groupe de langues.

Il n'est pas totalement étonnant que Lucien Tesnière soit à l'origine non seulement des questions de valence, mais aussi de la syntaxe en structures de dépendances dont les principes fondamentaux ont été repris dans le langage de programmation Lisp en notation préfixée pour le développement des travaux en intelligence artificielle dans les années 1970 et 1980.

En effet, la connaissance des langues slaves a amené Lucien Tesnière – slavisant spécialiste du russe et du slovène – à percevoir la nature centrale et dominante du verbe, les langues slaves étant essentiellement « verbo-centriques ».

Lucien Tesnière (1893-1954) a fait ses études supérieures aux Langues orientales (à l'époque ENLOV) et à l'Université de Paris, mais aussi auprès des Universités de Leipzig et de Vienne. Il fut également lecteur à l'Université de Ljubljana où il fonda l'Institut français de Ljubljana en 1920.

Il soutient à l'Université de Strasbourg en 1925 sa thèse de doctorat sur *Les formes du duel en slovène* avec pour thèse secondaire un *Atlas linguistique pour servir l'étude du duel en slovène*. C'est vraisemblablement la première étude de géographie linguistique pour les langues slaves. Tesnière s'est beaucoup intéressé au duel dans les langues slaves du Sud-Ouest, mais aussi dans l'aire mésopotamienne (perte du duel en akkadien) et dans l'aire chamito-sémitique avec, en particulier, une bonne connaissance des travaux d'André Basset sur le berbère, langue sans duel contrairement à l'arabe.

Après avoir été titulaire de la chaire de langues et littératures slaves à l'Université de Strasbourg, il succède à Maurice Grammont à la chaire de grammaire comparée des langues indo-européennes à l'Université de Montpellier.

Si l'on connaît bien les *Éléments de syntaxe structurale*, auxquels il travailla de la fin des années 1930 jusqu'en 1950 et publiés *post-mortem* en 1959 chez Klincksieck grâce à ses amis J. Fourquet [1] et F. Daumas, de nombreux travaux (en plus de ceux sur le duel) parsèment la carrière de Lucien Tesnière.

Nous pourrions les regrouper suivant les axes suivants :

- La cartographie linguistique, notamment le projet d'atlas linguistique slave prévu en collaboration avec Antoine Meillet.
- De manière abondante, des travaux concernant la phonétique et la prononciation ainsi que la graphie (nasales en slovène, les voyelles nasales slaves et le parler slovène de Replje, l'accent et le timbre des voyelles en slovène, les signes diacritiques et la transcription phonétique, les diphones « tl » et « dl » en slave, l'opposition morphologique de l'accent dans le substantif russe, « pour prononcer le grec et le latin »).
- Mais aussi un ouvrage sur le poète slovène Oton « Joupantchich » (noté à la française).
- Et une petite grammaire du russe publiée en 1934 chez Didier ainsi qu'un petit vocabulaire russe publié *post-mortem* également chez Didier en 1957.

La valence verbale est un concept linguistique dont la dénomination a été empruntée par Lucien Tesnière à une science, la chimie, ce dont certains auteurs se sont fait l'écho à l'image de Claire-Blanche Benvéniste qui écrit [2] : « Nous devons à L. Tesnière (1959) cet emprunt au vocabulaire de la chimie, suggérant une comparaison entre la structure des verbes et celle des atomes… » ou de Franck Neveu qui en fait une entrée de son dictionnaire [3] :

> « Le terme de valence a été emprunté par Lucien Tesnière (éléments de syntaxe structurale) au vocabulaire de la chimie, où il sert à désigner le nombre de liaisons d'un atome avec d'autres atomes dans une combinaison. Dans le domaine linguistique, il désigne le mode de construction des verbes selon le nombre d'actants qu'ils sont susceptibles de régir. La valence décrit donc un aspect fondamental de la structure actancielle du verbe. »

L'immense ouvrage de Tesnière, préfacé par Fourquet, est divisé en parties : (a) la connexion ; (b) la jonction ; (c) la translation, parties elles-mêmes subdivisées en « livres ». La partie connexion comprend notamment un livre B « Structure de la phrase simple », un livre C « Interrogation et négation », un livre D « Valence », un livre E « Métataxe ».

1. Auteur d'une *Grammaire de l'allemand*, suivant les principes de Tesnière, chez Hachette en 1969.
2. Dans « La complémentation verbale : petite introduction aux valences verbales », publié dans la revue *Tranel (Travaux neuchâtelois de linguistique)*, 37, 47-73, 2002.
3. *Dictionnaire des sciences du langage*, 2e édition revue et augmentée, Armand-Colin, Paris, 2011.

Ce n'est que dans le livre D, chapitre 97 [4], que Tesnière aborde la définition de la valence :

> « 3. – On peut ainsi comparer le verbe à une sorte d'atome crochu susceptible d'exercer son attraction sur un nombre plus ou moins élevé d'actants, selon qu'il comporte un nombre plus ou moins élevé de crochets pour les maintenir dans sa dépendance. Le nombre de crochets que présente un verbe et par conséquent le nombre d'actants qu'il est susceptible de régir, constitue ce que nous appellerons la valence du verbe. »

Cette définition de la valence est permise par l'étude de la structure de la phrase simple que Tesnière conduit au sein du livre B au cours des chapitres 48 à 52, puis 56. Le point essentiel, point de départ absolument décisif, est la mise en exergue du nœud verbal dont nous pourrons montrer ultérieurement qu'il constitue le niveau le plus élevé d'une structure arborescente en dépendances.

Nous soulignerons de l'exposé de Tesnière les points qui nous semblent essentiels :

> « Chapitre 48 : § 1. – Le nœud verbal, que l'on trouve au centre de la plupart de nos langues européennes… comporte… § 2. – … respectivement le verbe, les actants et les circonstants. »

> « § 9. – Nous avons vu que le verbe est au centre du nœud verbal et par conséquent de la phrase verbale. Il est donc le régissant de toute la phrase verbale. § 13. – Quant aux actants et aux circonstants, ce sont les subordonnés immédiats du verbe. »

> « § 3. – Le verbe exprime le procès. »

> « § 4. – Les actants sont les êtres ou les choses qui… participent au procès. § 6. – [Ils] sont toujours des substantifs ou des équivalents de substantifs. § 7. – Les circonstants expriment les circonstances de temps, lieu, manière, etc. dans lesquelles se déroule le procès. § 8. – [Ils] sont toujours des adverbes… ou des équivalents d'adverbes [5]. »

Tesnière éprouve le besoin bienvenu de distinguer entre eux les trois actants [6] par une dénomination particulière qui, au demeurant, subsiste dans la linguistique actuelle :

> « Chapitre 51 : § 3. – Nous dirons : prime actant, second actant, tiers actant [7]. »

À partir d'une remarque presque anodine, où Tesnière constate que : « § 13. – … du point de vue structural, et non du point de vue sémantique, le sujet est un complément

4. L. Tesnière, *Éléments de syntaxe structurale*, p. 238.
5. *Ibid.*, p. 102 et 103.
6. Les verbes à trois actants répondent à deux types bien définis : les verbes de « dire » et de « don », *ibid.* chap. 106 : les verbes « trivalents », § 3, p. 255).
7. *Ibid.*, p. 108.

comme les autres [8] », il développe contre l'opposition sujet – prédicat (et donc d'une manière seconde, implicite, contre les constituants immédiats) une critique qui nous semble parfaitement juste et n'avoir jamais été entendue ou très peu :

> « Chapitre 49 : § 2. – La grammaire traditionnelle s'efforce de retrouver dans la phrase l'opposition logique entre le sujet et le prédicat, le sujet étant ce dont on dit quelque chose, le prédicat ce qu'on en dit [9]. »
>
> « § 13. – L'opposition du sujet et du prédicat empêche ainsi de saisir l'équilibre structural de la phrase, puisqu'elle conduit à isoler comme sujet un des actants, à l'exclusion des autres, lesquels se trouvent rejetés dans le prédicat pêle-mêle avec le verbe et tous les circonstants. C'est là accorder à l'un des éléments de la phrase une importance disproportionnée, qu'aucun fait strictement linguistique ne justifie. »
>
> « § 6. – … Dans aucune langue, aucun fait proprement linguistique n'invite à opposer le sujet au prédicat [10]. »

Il ne faut pas aller loin pour vérifier la validité de cette affirmation. Il suffit d'avoir une structure impersonnelle en russe (*соседа забыло деревом* – m.à.m. : le voisin [accusatif] 'processus de tuer au passé' un arbre [instrumental] / *дорогу занесло снегом* – m.à.m. : la route [accusatif] 'processus de couvrir au passé' la neige [instrumental] ; dans ces structures avec un verbe au neutre singulier, il n'y a pas de prime actant et il n'y a pas – et pour cause – d'accord verbe-sujet) ou une voix ergative en géorgien obtenue dès que le verbe passe à l'aoriste (ბავშვი თამაშობს ბურთს : l'enfant joue au ballon [présent] / ბავშვმა ითამაშა ბურთი : le garçon a joué au ballon [aoriste], სტუდენტი წერს წერილს : l'étudiant [nominatif] écrit [présent] une lettre [accusatif] / სტუდენტმა დაწერა წერილი : l'étudiant [მა : ergatif] a écrit [aoriste] une lettre [ი : nominatif] [11]). D'une manière générale, l'analyse syntaxique d'une langue slave et surtout l'analyse automatique qui peut en résulter se fera aisément avec des actants et des circonstants tous dépendants du prédicat verbal. Nous le verrons plus loin, cette solution est adoptée de manière massive par les linguistes slaves.

La première partie de la syntaxe structurale est consacrée aux connexions. Tesnière montre que les connexions structurales unissent un terme supérieur à un terme inférieur en un rapport de dépendance. Il en découle que « l'étude de la phrase, qui est l'objet propre de la syntaxe structurale est essentiellement l'étude de sa structure, qui n'est autre que la hiérarchie de ses connexions [12] ». Cette hiérarchie de connexions est définie par le fait « qu'un subordonné ne peut dépendre que d'un seul régissant [et qu'au] contraire un régissant peut commander plusieurs subordonnés [13] » et est visualisée par l'ensemble des traits de connexion que Tesnière nomme « stemma ».

8. *Ibid.*, p. 109.
9. *Ibid.*, p. 103. Il y a, nous semble-t-il, souvent confusion entre une structure discursive de type « thème » – « rhème » et une structure syntaxique.
10. *Ibid.*. p. 104 et 105.
11. Kita Tschenkeli : *Einführung in die georgische Sprache*, Amirani Verlag, Zurich, 1958, vol. 1, p. 154.
12. L. Tesnière, *Éléments de syntaxe structurale*, p. 14.
13. *Id.* Ce faisant Tesnière définit implicitement un arbre acyclique.

Tesnière note bien que ce stemma, qui préfigure un arbre de dépendances, représente l'ordre structural des mots qui est à plusieurs dimensions alors que le stemma est ramené nécessairement à deux dimensions. Il est également pleinement conscient de l'antinomie entre l'ordre structural à plusieurs dimensions et l'ordre linéaire qui n'en a qu'une.

Il en conclut, chapitre 6, paragraphe 1 : « Toute la syntaxe structurale repose sur les rapports qui existent entre l'ordre structural et l'ordre linéaire. »

Mais étrangement, il ne tient pas compte de l'ordre linéaire dans les stemmas, dont ce sera, à notre avis, leur faiblesse avec leur complexité due aux translations. Au plan graphique qui doit permettre de visualiser sa pensée, il établit les traits de connexion suivant un principe strictement vertical « puisqu'il symbolise le lien entre un terme supérieur et un terme inférieur ». De plus, rompant de manière encore plus flagrante avec l'ordre linéaire, il décide, en suivant une idée de formalisation (ou au moins de classement) d'un autre type, de positionner les actants sur la gauche et les circonstants sur la droite.

Tesnière aborde une autre question fondamentale, disons une règle fondamentale, qui n'est pas assez souvent respectée. Il nous semble primordial de ne jamais mélanger les niveaux d'analyse. Chaque niveau, phonologie, morphologie, syntaxe, sémantique, doit être examiné à l'intérieur de ses limites et avec les outils qui lui appartiennent. Ce n'est qu'après avoir mené à terme une analyse que l'on peut en rechercher une interprétation à un autre niveau. Le traitement automatique nous a appris que les deux démarches doivent être séparées. Dans le cas contraire, c'est une complication qui masque le système et rend la compréhension de la langue et son apprentissage nettement plus difficiles, voire incertains...

Tesnière consacre un chapitre entier à la distinction de la structure et du sens, autrement dit de la syntaxe et de la sémantique. Après avoir rappelé que « chapitre 20 : § 1. – Si la syntaxe est distincte de la morphologie, elle ne l'est pas moins de la sémantique. Autre chose est la structure d'une phrase, autre chose l'idée qu'elle exprime et qui en constitue le sens. Il y a donc lieu de distinguer entre le plan structural et le plan sémantique », c'est-à-dire souligner l'indépendance de la morphologie, de la syntaxe et de la sémantique, c'est-à-dire aussi élucider le premier niveau, puis l'interpréter. Il met en avant, comme nous venons de le faire, la nécessité de la compréhension du système :

> « § 4. – La distinction entre le plan structural et le plan sémantique est d'importance capitale pour la bonne compréhension de la syntaxe structurale. »

En substituant à chaque mot d'une phrase l'entrée qui suit ce mot dans le dictionnaire, Tesnière donne une illustration magistrale de l'indépendance de la syntaxe et de la sémantique. Cet exemple est resté dans les annales et est souvent cité *ex nihilo* :

> « § 17. – Le plan structural et le plan sémantique sont donc théoriquement entièrement indépendants l'un de l'autre. La meilleure preuve en est

> qu'une phrase peut être sémantiquement absurde tout en étant structuralement parfaitement correcte ("le signal vert indique la voie libre" *vs* "le silence vertébral indispose la voile licite"). »

En plus des connexions, des translations, des stemmas, Tesnière a apporté une réflexion sur la diathèse qu'il définit comme diathèse active, diathèse passive, diathèse réfléchie et diathèse réciproque.

Tesnière s'intéresse aussi à ce que Lazard a nommé, plus tard, les instruments de l'actance :

> « Chapitre 52 : § 1. – Il est de toute nécessité, pour qu'une phrase soit compréhensible, que les différents actants soient pourvus de signes distinctifs suffisants pour qu'on puisse les différencier.
>
> « § 2. – Ces signes distinctifs sont ou bien des indices plus ou moins agglutinés (prépositions et postpositions, préfixes, suffixes ou désinences), ou bien la position des actants sur la chaîne parlée [14]. »

Tesnière a explicitement défini trois actants, ceux des verbes de « dire » et de « don », mais il perçoit aussi que leur limite n'est peut-être pas aussi tranchée qu'il n'y paraît et en fait la remarque :

> « Chapitre 57 : § 1. – À première vue la limite entre actants et circonstants est nette. Mais, à y regarder de près, on s'aperçoit qu'elle est délicate à fixer avec précision. [...] le tiers actant présente déjà quelques caractéristiques de circonstant. Inversement, certains circonstants présentent avec les actants quelques analogies qui invitent à considérer attentivement les criteriums susceptibles de permettre un départ entre les actants et les circonstants [15]. »

Il aura fallu attendre longtemps en France pour que l'œuvre de Lucien Tesnière jouisse d'une certaine réputation en dehors du germaniste Jean Fourquet, son collègue et ami, et de quelques slavisants tels que Paul Garde (russe et serbo-croate) ou Claude Vincenot (russe et slovène). Pourtant ses travaux et notamment les questions de représentations des structures syntaxiques, la diathèse, les instruments de l'actance ou une définition plus précise des éléments constituant le cadre verbal demeurent des recherches d'actualité largement débattues.

C'est ainsi qu'une œuvre majeure, l'*Actance* revient sur ces questions sous la plume de Gilbert Lazard, iranologue, linguiste et traducteur. Fonctionnaliste comme la grammaire de Tesnière, l'*Actance* de Lazard se concentre uniquement sur les rapports entre le prédicat verbal et les éléments nominaux (actants) qui en dépendent. Il n'y a pas ici la construction d'une syntaxe structurale débouchant sur un mode de représentation formalisable, mais plutôt une vaste étude de ces relations actancielles et de leurs

14. *Ibid.*, p. 111-115. Tesnière alimente ses réflexions par sa connaissance personnelle de langues, mais aussi par une connaissance très précise des travaux d'André Basset sur le berbère.

15. *Ibid.*, p. 127.

modes de représentation basée sur une typologie de langues nettement supérieure à celle de Tesnière grâce à la participation aux séminaires de recherche de Lazard de linguistes confirmés, spécialistes de divers horizons.

L'*Actance* aborde des sujets qui n'étaient pas ou peu traités chez Tesnière, une attention importante étant donnée aux relateurs, aux indices actanciels intégrés ou associés au verbe, à l'ordre des termes ou encore aux cas de coalescence. Comme Tesnière, Lazard développe les questions de diathèse dans de nouvelles directions dues, entre autres, à l'étude de langues telles que le basque, les langues bantoues ou les langues du Caucase : diathèse accusative, ergative ou duale. Lazard propose également un classement des verbes sur la base de leur valence. On y retrouve aussi un certain nombre de points de vue qui renforcent les thèses de Tesnière et sur lesquelles nous souhaitons revenir ici.

La question de la distinction de la structure et du sens constitue un chapitre entier de la grammaire de Tesnière. Ce point est également d'importance stratégique pour Lazard qui écrit :

> « Lorsqu'elle [la grammaire traditionnelle] parle, par exemple, de "complément d'attribution" ou de "complément de manière", on ne sait pas sur quel plan se situent ces notions, celui du contenu sémantique ou celui de la structure morphosyntaxique... Nous procéderons... en suivant la voie sémasiologique. Il s'agit, dans cette perspective, de définir les fonctions actancielles sur des critères exclusivement morphosyntaxiques... de chercher à voir... comment ces systèmes syntaxiques plus ou moins différents sont utilisés pour exprimer des relations sémantiques présumées universelles.
>
> « Pour éviter le risque, toujours présent et menaçant de confusion entre le plan sémantique et le plan morphosyntaxique, nous employons des termes différents pour désigner les unités de l'un et de l'autre. Au plan morphosyntaxique appartiennent le prédicat verbal ou, plus simplement, le verbe et les actants (ainsi que les circonstants). Leur correspondent respectivement sur le plan sémantique le procès (ce terme est pris dans un sens très général et couvre aussi bien les états et les changements d'état que les actions et les évènements) et les participants (ainsi que les circonstances), ce que l'on résume :
>
> | | | | |
> |---|---|---|---|
> | « morphosyntaxe : | prédicat verbal / verbe | actants | circonstants |
> | « sémantique : | procès | participants | circonstances. » |

Tesnière avait soulevé le problème de la limite entre actants et circonstants. Cette question est actuellement posée par tous ceux qui travaillent sur la valence. Lazard y répond par une typologie de l'actant, véritable dégradé de valeurs :

> « De l'actant "obligatoire" (généralement unique) aux termes "requis et régis" [le verbe impose à la fois la présence d'un actant et la forme qu'il prend], puis aux termes simplement "régis" [les actants de ce genre sont facultatifs, mais leur forme n'est pas libre] et enfin aux "adjets" [ces termes

> ne sont pas régis, mais ils sont requis et en outre cumulables] (Feuillet, 1980 : 26), la contrainte se fait de moins en moins forte. Reste le cas de ceux dont la présence et la forme sont également libres. Ils se rangent normalement parmi les circonstants. »

L'*Actance* de Lazard aborde aussi la visée communicative qui a été introduite par les études tchèques, notamment l'école de Prague. La partie 4 du chapitre 5 (p. 209-215) lui est consacrée :

> « La visée communicative, c'est-à-dire la répartition au sein de la phrase des parties thématiques et rhématiques, joue dans les variations d'actance un rôle important qui peut se manifester notamment dans le choix de la diathèse, le marquage de l'objet, le jeu de l'accord verbal. »

Vladimír Šmilauer (1895-1983), professeur à l'Université de Prague, est le contemporain de Lucien Tesnière. Spécialiste des langues tchèque et slovaque (ainsi que du hongrois), spécialiste d'onomastique, son nom demeure lié avant tout aux études syntaxiques.

La publication en 1947 de la *Novočeská skladba* (*La syntaxe du tchèque moderne*), remaniée et republiée en 1966, est un évènement majeur dans le monde de la linguistique tchèque. À cette époque où sont conçues d'autres grammaires d'importance telles que le second volume de la grammaire de Trávníček consacré à la syntaxe (1949) ou la grammaire de Havránek et Jedlička (1950), on ne dispose encore que de la syntaxe de Gebauer et Ertl publiée en 1914.

La publication de Šmilauer ouvre une ère nouvelle en proposant une description syntaxique homogène du tchèque. Elle est organisée en structure de dépendances où seuls les mots pleins sémantiquement sont représentés. Deux traits importants distinguent la position de Šmilauer et celle de Tesnière [16]. Le premier serait, d'après nous, à l'avantage de Tesnière : Šmilauer représente en effet le sujet et le prédicat verbal au même niveau, car il considère qu'aucun d'eux ne domine l'autre, ce qui va à l'encontre de la théorie tesnièrienne et, d'une manière générale, à l'encontre des grammaires de dépendances [17]. Le second est d'une importance considérable et totalement à l'avantage de Šmilauer : ses structures de dépendances représentent simultanément l'ordre structural et l'ordre linéaire par une projection orthogonale sur un axe vectoriel horizontal.

L'amélioration et la transformation des stemmas en un arbre de dépendances par Šmilauer est liée à la connaissance qu'il avait des travaux de Tesnière antérieurs à la publication *post-mortem* des *Éléments de syntaxe structurale*, notamment de la *Petite grammaire russe* (1934) et, vraisemblablement, de « Comment construire une syntaxe » (*Bulletin de la Faculté des lettres de Strasbourg*, 12^{e} année, n° 7, mai-juin 1934, p. 219-229). Panevová (2016 [18]) cite à ce propos Šmejkalová (2015).

16. Il est une autre chose qui, curieusement, distingue Tesnière et Šmilauer : le premier était membre du Cercle linguistique de Prague et pas le second !

17. Nous ressentons ce fait comme une hésitation à se libérer totalement de l'opposition sujet – prédicat.

18. D'où nous tirons les informations relatives à l'oeuvre de Šmilauer.

Les structures de dépendances sont nées avec tous les avantages qu'elles comportent en caractère linguistiquement naturel, en simplicité, en héritage immédiat des propriétés. En liant ces deux ordres – structural et linéaire –, Šmilauer ouvre la voie de la calculabilité de la syntaxe tchèque une bonne vingtaine d'années avant que celle-ci ne soit mise en pratique. Cette représentation d'une structure arborescente de dépendances organisée autour du cadre verbal devient en quelque sorte l'emblème de l'école de Prague. Elle constitue l'héritage de Šmilauer. Elle s'étend à d'autres pays slaves et à la RDA (voir p. ex. les travaux de Jürgen Kunze [19]).

En dehors de l'intérêt pour le traitement automatique de la syntaxe, nous retiendrons, comme le faisait Šmilauer, la clarté de l'exposé didactique correspondant, simple à retenir et qui apporte un pouvoir explicatif important.

La *Description générative fonctionnelle* (DGF ; en tchèque : FGP : *Funkční Generativní Popis*), publiée pour la première fois en 1963 [20], s'attache à produire une description formelle explicite de la langue. Elle trouve ses racines dans une double tradition, la tradition linguistique européenne, d'une part, dont Petr Sgall souligne la richesse et la diversité de pensée tant dans les études diachroniques que dans les études synchroniques et la tradition de l'école de Prague, d'autre part.

Sgall fait ainsi remonter le concept de « division actuelle » à Henri Weil [21] et à Vilém Mathesius, l'un des principaux représentants du Cercle Linguistique de Prague. Il émet l'hypothèse que l'adoption des constituants immédiats par les pays anglo-saxons et les structures de dépendances liées au concept de valence par les pays slaves pour décrire la syntaxe est due en grande partie à la nature des langues concernées et au fait que l'Europe continentale, de par sa variété de langues, a recherché des descriptions plus générales.

La DGF est organisée autour de quatre principes fondamentaux : un principe génératiste, un principe de dépendances, un principe fonctionnel et un principe stratificationnel.

Elle s'est très rapidement constituée en un système basé sur les représentations de dépendances dont les avantages ont été immédiatement soulignés, notamment pour la description de la théorie des valences puisqu'elle est étroitement liée à la représentation du verbe comme centre de la proposition. Elle est ainsi adéquate au traitement des langues avec un ordre des mots dit libre.

La représentation d'une phrase au niveau de la signification linguistique est assurée par un arbre de dépendances dont le sommet est le prédicat (verbe) de la proposition principale. La forme d'un mot dans la phrase apparaît comme un noeud avec un étiquetage complexe.

L'arbre de dépendances permet de définir deux arrangements :

---

19. Jürgen Kunze : *Abhängigkeitsgrammatik für das Deutsche* (*Grammaire de dépendances pour l'allemand)*. Berlin, Akademie Verlag, 1975.

20. Sgall Petr : *Generativní popis jazyka a česká deklinace* (*Description générative de la langue et flexion tchèque*). Prague : Academia, 1967.

21. Henri Weil : *De l'ordre des mots dans les langues anciennes comparées aux langues modernes*, thèse principale, Paris, 1844 ; 3e éd., 1879. Réimpr. de la 3e éd., Paris, Didier érudition, 1991, IX-101 p. (ISBN 2-86460-166-4).

- structural : rapport de suzeraineté – vassalité des noeuds ;
- linéaire : répondant à la division actuelle et à l'arrangement des éléments dans la phrase suivant le dynamisme communicatif (*cf.* Mathesius, Firbas, Sgall, Hajičová, Buráňová, Daneš...).

La fonctionnalité dans la *Description générative fonctionnelle* est comprise comme un système d'oppositions, suivant en cela Saussure et Jakobson, qui s'exprime dans les rapports forme – fonction ou fonction – forme à toutes les strates de description, au niveau morphématique, voire morpho-phonologique et à deux niveaux syntaxiques (superficiel et profond, lié au sens).

La *Description générative fonctionnelle* a servi de cadre théorique à nombre de travaux, notamment au développement d'une théorie de la valence par des disciples de Sgall, essentiellement par Jarmila Panevová.

La valence est comprise comme un attribut lexico-syntaxique d'une « lexie », jonction d'une unité lexicale et de l'une de ses significations. Toute unité lexicale « autosémantique », c'est-à-dire porteuse de sens, peut recevoir un cadre valenciel. Cependant, le cadre valenciel est lié majoritairement au verbe, au point que nous parlerons, par raccourci, de cadre verbal.

Les travaux théoriques sur la valence de J. Panevová (de 1974 jusqu'en 2014) ont rendu possible la réalisation de vastes applications telles que Vallex et PDT-Vallex [22].

Le meilleur moyen de présenter la conception de la valence au sein de la DGF sera de le faire au travers de l'application Vallex. Dans le cadre de la DGF, la valence se situe au niveau tecto-grammatical, c'est-à-dire au niveau d'une syntaxe profonde, déjà un niveau de signification, ce qui explique la liaison de la valence aux seuls mots auto-sémantiques. C'est un point sur lequel nous reviendrons en conclusion de ce chapitre.

La première version de ce dictionnaire de valences (Vallex 1.0) est parue en 2003 et offrait de l'ordre de 2 500 cadres verbaux pour un ensemble de 1 500 lexèmes verbaux. L'équipe rassemblée autour de Markéta Lopatková, avec les conseils de J. Panevová, a travaillé à l'élaboration de plus en plus complète de ce dictionnaire : Vallex 1.0 en 2003, Vallex 2.0 en 2006, Vallex 2.5 en 2007 (parution du dictionnaire « papier » correspondant en 2008), Vallex 2.7, le premier à avoir un lien vers le PDT-Vallex (issu du corpus PDT), a bénéficié du support du projet « Delving Deeper : Lexicographic Description of Syntactic and Semantic Properties of Czech Verbs » (2012-2016) pour arriver à Vallex 3.0, fin 2016 à l'extinction du contrat précédent. Ce dernier dictionnaire [23] contient plus de 4 500 verbes et de l'ordre de 10 800 cadres verbaux.

Vallex a évolué, petit à petit, de manière pragmatique, sur les questions de limite et de nature entre les actants et les circonstants. La première division est traditionnelle : elle oppose les actants aux « extensions libres [24] », dites de nature « adverbiale » et correspondant nettement aux circonstants de Tesnière. La distinction entre ces deux

22. Le PDT (Prague Dependency Treebank) se réfère également au cadre théorique de la DGF.

23. M. Lopatková, V. Kettnerová, E. Bejček, A. Vernerová, Z. Žabokrtský : *Valenční slovník českých sloves (Vallex)*, Univerzita Karlova, Nakladatelství Karolinum, Prague, 2016.

24. En tchèque, « volná doplnění ».

catégories se fait à l'aide d'un test élaboré par J. Panevová (1974-75 et 1980) : si un complément ne peut apparaître qu'une seule fois avec un verbe donné (en dehors des questions de coordination et d'apposition), il s'agit d'un actant, la caractéristique même d'une extension libre étant de pouvoir apparaître un nombre quelconque de fois avec n'importe quel verbe.

Contrairement aux trois actants de Tesnière, ce sont cinq actants qui sont reconnus dans la DGF et donc dans Vallex. Ils sont habituellement présentés et énumérés selon l'ordre systémique. Il s'agit dans cet ordre de :

Un verbe tel que *předělávat*^{impf}, *předělat*^{pf} (transformer, changer en) peut présenter une occurrence de ces 5 actants :

otec předělal dětem loutku z Kašpárka na čerta.
[ACT] [ADDR][PAT] [ORIG] [EFF]

m.à.m. :

le père a transformé pour ses enfants la marionnette de guignol en diable.
[ACT] [ADDR] [PAT] [ORIG] [EFF]

En définitive, si l'on se reporte à l'*Actance* de Lazard, nous retombons ici aussi sur des problèmes de limite entre actants et circonstants illustrés par la création d'une catégorie intermédiaire de « quasi-actants ».

Il s'agit de DIFF (différence), INTT (intention) et OBST (obstacle). Nous trouvons des exemples tels que :

kluk zakopl o kořen (le garçon a buté contre une racine).
[OBST] [OBST]

Ces types de compléments ont des caractéristiques partagées entre celles des actants et celles des extensions libres.

Les « extensions libres » sont au nombre de 24. Elles représentent les circonstances de temps, de lieu, mais aussi toutes les autres : accompagnement, but, cause, manière, moyen...

Parmi les dictionnaires de valences des langues slaves, Vallex possède un avantage essentiel. Celui d'être doté d'un formalisme facile à comprendre et facile à utiliser.

Pour introduire ce formalisme, nous allons présenter un exemple très simple [fig. 1].

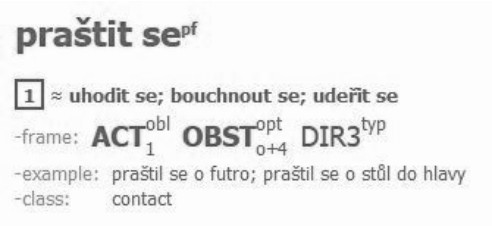


Figure 1 – Vallex

Le verbe est présenté par ses différents lemmes aspectuels à l'infinitif (jusqu'au nombre de six !). Le verbe « *praštit se* » (se cogner) est seul. Son aspect est perfectif (abréviation « pf »), ce qui est noté en exposant pour chaque lemme. Il est à noter que Vallex retient quatre indications aspectuelles : imperfectif ($^{\mathbf{impf}}$), perfectif ($^{\mathbf{pf}}$), bi-aspectuel ($^{\mathbf{biasp}}$) et itératif ($^{\mathbf{iter}}$), ordre suivant lequel les différents lemmes sont toujours ordonnés.

Les différentes significations du verbe sont dénommées « unités lexicales » et sont numérotées. Nous n'avons ici qu'une seule unité lexicale, donc de rang '1', dont le sens est donné par trois synonymes (*uhodit se ; bouchnout se ; udeřit se*) signifiant tous « se cogner, se heurter ».

Le cadre verbal (*frame*) est constitué de trois « foncteurs » représentant les groupes compléments du verbe, d'abord un acteur (un actant, en fait le prime-actant de Tesnière), ensuite un obstacle (OBST - un quasi-actant) et enfin DIR3, une extension libre exprimant la direction (par exemple, ici : *do hlavy* – m.à.m. : *dans la tête*). Les extensions libres sont écrites en caractères majuscules normaux tandis qu'actants et quasi-actants sont en caractères gras.

La manière dont les foncteurs sont présents au sein du cadre verbal est exprimée à l'aide d'exposants : les actants et quasi-actants peuvent être obligatoires ($^{\text{obl}}$) ou optionnels ($^{\text{opt}}$), les extensions libres qui appartiennent au cadre verbal sont dites « typiques » ($^{\text{typ}}$), c'est-à-dire qu'elles peuvent être présentes ou non. Dans l'exemple présenté, le prime-actant est obligatoire, le quasi-actant et l'extension libre peuvent être omis :

> *Jan se praštil* : Jean s'est cogné.
>
> *Jan se praštil do hlavy* : Jean s'est cogné à la tête.
>
> *Jan se praštil o futro* : Jean s'est cogné au chambranle de la porte.
>
> *Jan se praštil o futro do hlavy* : Jean s'est cogné la tête sur le chambranle de la porte.

Les valeurs syntaxiques sont liées aux foncteurs par un indice [25]. Le prime-actant porte un indice « 1 », ce qui signifie qu'il est au nominatif. Le quasi-actant porte l'indice

25. Rappelons ici que le tchèque a pré-formalisé l'usage de ses cas : on ne dit pas en tchèque « nominatif », mais « premier cas », « génitif », mais « deuxième cas », etc.

« o + 4 », c'est-à-dire la préposition « o » (marquant le contact) régissant l'accusatif, ce qui exprime le mouvement. Le foncteur DIR3 exprime la direction « vers où » avec mouvement à l'aide d'un ensemble de constructions prédéfinies, combinaisons préposition + cas ou adverbes [26].

Chaque unité lexicale est accompagnée d'exemples, des diathèses concernées (passive, déagentive ou possessive), des indications de réflexivité ou de réciprocité et, éventuellement d'une classe syntaxico-sémantique. Vallex en comprend vingt-deux telles que « action mentale », « verbe psychologique », « mouvement », « perception », « transport »,... mais aussi « contact » qui englobe 125 unités lexicales. Il va de soi que dans certaines occasions « contact » peut être un doux euphémisme !

La version 'Vallex 3' apporte plusieurs nouveautés dont une qui est particulièrement intéressante et nécessaire : les alternances lexicalisées.

> « *žnout trávu.*PAT *na louce.*LOC » (faucher l'herbe dans le pré)

face à :

> « *žnout louku.*PAT » (faucher le pré) est un exemple typique. Dans la construction syntaxique, le complément de lieu devient le patient.

'Vallex 3' offre ainsi une étude importante de ces phénomènes répartis en trois catégories :

a. Des conversions lexico-sémantiques concernant tant le sujet que l'objet :

> « *Rybáři prolomili díru do ledu.* » (Les pêcheurs ont creusé un trou dans la glace.)
> « *Rybáři prolomili led.* » (Les pêcheurs ont perforé la glace.)

b. Réalisation structurale différente d'un même participant (dans les verbes de mouvement, des verbes de positionnement, des verbes de jugement, de sentiments...) :

> « *Jan se v poslední době cítil jako stařík.* » (Ces derniers temps, Jean se sentait comme un vieux.)
> « *Jan se v poslední době cítil stař e.* » (Ces derniers temps, Jean se sentait vieux.)

26. L'exemplarité des solutions utilisées pour la représentation des compléments de lieu vaut la peine que l'on s'y arrête un instant. Nous avons en premier lieu un foncteur « LOC » qui représente un lieu sans mouvement. Le mouvement est exprimé par 3 foncteurs : « DIR1 » marque le lieu d'origine, « DIR2 » le lieu par lequel on passe et « DIR3 » le lieu vers lequel on se dirige. Certaines unités lexicales peuvent avoir le foncteur « DIR », c'est-à-dire l'ensemble des 3 foncteurs précédents.

Typiquement, des foncteurs tels que « LOC » et « DIR(X) » ne possèdent pas en indice d'indications syntaxiques, mais renvoient à un ensemble de possibilités telles que : • DIR1 : adverb, od+2, s+2, z+2, ze strany+2, zpod+2, zpoza+2, zpřed+2,... • DIR2 : 7, adverb, cestou+2, kolem+2, mezi+7, napříč+7, po+6, podél+2, přes+4, skrz+4, v+6,... • DIR3 : 7, adverb, do+2, do čela+2, k+3, kolem+2, mezi+4, mimo+4, na+4, na+6, nad+4, naproti+3, okolo+2, po+4, po+6, pod+4, proti+3, před+4, přes+4, v+4, vedle+2, za+4, za+7,...

c. Participant qui peut être réalisé sous deux formes valencielles :

« *Jan si pomyslel, že je Petra nafoukaná.* » (Jean pensait que Pierrette était prétentieuse.)

« *Jan si pomyslel o Petře, že je nafoukaná.* » (Jean pensait de Pierrette qu'elle était prétentieuse.)

Les recherches sur la valence avec des applications sous forme de dictionnaires, tels que le Vallex tchèque, sont présentes dans l'ensemble des langues slaves, certainement soutenues par des traits favorables tels que structure verbo-centrique et flexion externe. Nous explorerons ici, en plus des travaux sur le tchèque, choisis comme très représentatifs de ce que l'on peut faire, les travaux existants en slovaque comme support de nos projets ainsi que les travaux sur le haut-sorabe qui viennent compléter une vision de la valence et de ses traitements pour le groupe des langues slaves de l'Ouest.

Le *Valenčný slovník slovenských slovies na korpusovom základe*[27] (*Dictionnaire de valences des verbes slovaques sur la base de corpus*) qui nous servira de référence est l'héritier d'un courant de recherche sur la valence à l'Université de Prešov qui remonte aux années 1990. La présentation, en 4e de couverture, est faite par Markéta Lopatková. Les travaux pour la réalisation de ce dictionnaire ont bénéficié d'un contrat VEGA, mais contrairement à toute attente le dictionnaire qui en découle n'appartient pas au domaine public[28], ce qui nous semble contraire au but recherché, une telle publication étant incontournable pour la traduction vers le slovaque et la rédaction dans cette langue.

Le présent dictionnaire succède au *Valenčný slovník slovenských slovies* de J. Nižníková, M. Sokolová *et al.* publié en 1998 et au *Valenčný slovník slovenských slovies. 2. zväzok (na korpusovom základe)*, publié en 2005 par J. Nižníková. S'il conserve la matière des ouvrages précédents, celle-ci est modifiée par le recours systématique au Corpus national slovaque (*Slovenský Národný Korpus [SNK]*) de l'Académie des sciences.

Ce dictionnaire prend notamment en compte la classification sémantique des verbes slovaques de M. Sokolová (1993, 1995)[29], mais aussi l'expérience de dictionnaires de valences de langues slaves, en particulier du Vallex tchèque. Au plan théorique, une mention particulière est faite des travaux de M. Ivanová (2008), ainsi que des travaux tchèques de Daneš et Hlavsa (1981) et ceux de Panevová (1974, 1975, 1976, 1977, 1996, 2005, 2017), réalisés dans le cadre de la *Description générative fonctionnelle.*

Ce dictionnaire se situe dans l'héritage d'un ouvrage emblématique, *Štruktúra slovenského slovesa. Študia lexikálno-syntaktická* (*Structure du verbe slovaque. Étude lexico-syntaxique*) d'Eugen Pauliny, publié en 1943. Pauliny y élabore la théorie de l'intension verbale.

27. Ivanová, Sokolová, Kyseľová, Perovská : « Valenčný slovník slovenských slovies na korpusovom základe », Filozofická fakulta PU v Prešove, 2014. L'abréviation courante est : « VSSSKZ ».

28. L'exemplaire que nous avons pu consulter à la bibliothèque de la Faculté de mathématiques et physique de Prague (Malá Strana) porte le cachet : « Prešovská univerzita v Prešove. Filozofická fakulta. Autorské výtlačky – nepredajné » (*Université de Prešov. Faculté des lettres. Exemplaires auteur – interdits à la vente*). Il n'empêche que ce dictionnaire devrait appartenir à la bibliothèque de tout slavisant.

29. Miloslava Sokolová : « Sémantika slovesa a slovesný rod. » Bratislava, Veda, 1993.

L'intension est habituellement présentée comme la valence sémantique du verbe et, effectivement, les auteurs du présent dictionnaire l'expose en ces termes : « L'intension exprime les caractéristiques sémantiques du verbe en tant que prédicat, sur la base desquelles il est lié aux participants. » Le travail pionnier de Pauliny nous semble cependant relever davantage de la pragmatique du verbe que de sa sémantique.

Pour Pauliny, tout verbe exprime un procès (dans le sens d'action) qu'il s'agisse de verbes présentés traditionnellement en tant que verbes d'action ou verbes d'état. Par « intension du procès verbal », Pauliny entend le fait qu'un verbe en tant que prédicat requiert ou non l'expression d'un agent ou d'un patient de ce procès : « L'action en soi présuppose toujours d'une part un 'élément' [objet] qui réalise l'action et d'autre part un 'élément' [objet] qui subit l'action et en est le but. Au niveau sémantique, cela se manifeste par le fait qu'il existe un réalisateur de l'action, un agent (A), une action en soi (D) et un objet subissant l'action (Z). Ce sont ces attributs qui accompagnent toute action [30]. »

La typologie des verbes proposée par Pauliny repose sur des séries d'oppositions. La première dichotomie est créée par le type de l'agent, déterminé ou non. Sont ainsi définies deux premières classes : (I) agent déterminé et (II) agent indéterminé.

La première est subdivisée en (I A) agent exprimé de manière explicite et (I B) agent implicite. Ainsi, en (I A), nous trouverons la forme la plus courante a) « *otiec nesie drevo* » (mon/ton/notre/votre/le père porte du bois) où tous les éléments sont exprimés et correspondent au schéma (A – D – Z). La forme b) est représentée par le prototype « *brat ide* » (mon/ton/notre/votre/le frère vient/arrive) auquel correspond un schéma (AZ – D) dans la mesure où l'agent est en même temps le patient du procès (représentant quant au fond quelque chose qui pourrait se comparer à une formule du type « mon frère se ramène ici »). La forme c) a pour « vitrine » « *kostolník zvoní* » « le sacristain sonne » (sous-entendu, les cloches. En français, l'objet devrait être exprimé). Le schéma correspondant à la forme (A – DZ) justement dans la mesure où l'objet est compris dans le verbe. Pauliny y joint un 4^{e} type : « *brat valaší* » (mon frère est_berger) avec un schéma particulier (Az – Dz) où le verbe exprime l'état du sujet et, en même temps, que le sujet dans cet état est son propre agent !

La deuxième sous-classe (I B) offre deux types : (a) Il s'agit du modèle « *zabilo brata* » qui rappelle très fortement les structures impersonnelles du russe (*cf.* ci-avant) avec pour schéma (AD – Z) ([cela] a tué mon frère) et (b) « *hrmí* » ([il] tonne) avec le schéma (ADZ).

Dans la deuxième classe, la non détermination de l'agent, indépendante de la volonté de l'énonciateur (II A) s'oppose à une non expression volontaire (II B). Nous trouvons deux types de verbes dans la première sous-classe (II A) : (a) « *otec starne* » (mon/ton/notre/votre/le père vieillit) avec un schéma en (D – Z) où le sujet est en

30. „Dejanie samo o sebe vždy predpokláda jednak predmet, ktorý dejanie uskutočňuje, jednak predmet, ktorý je dejaním zasahovaný a je jeho cieľom. Na pláne významovom sa to preto javí tak, že existuje uskutočňovateľ dejania, agens (A), samo dejanie (D) a predmet dejaním zasiahnutý (Z). Toto sú atribúty, ktoré sprevádzajú každé dejanie", cité par M. Dokulil dans son compte-rendu dans « Slovo a Slovesnost », 1949, n° 2.

même temps le patient du procès et (b) « *slepý nevidí* » (l'aveugle ne voit pas) avec un schéma (D) où le procès existe en tant que propriété.

La deuxième sous-classe (II B) offre quatre types marqués par la particule réfléchie « sa ». (a) « *dom sa stavia* » (la maison se construit) avec un schéma (D – Z) pour les mêmes raisons que pour « *otec starne* », l'énonciateur ne connaissant pas l'agent. Les trois types suivants sont (b) « *ide sa* » (on va / aller), (c) « *zvoní sa* » (on sonne) et (d) « *valaší sa* » (on est_berger).

Il est de tradition de se référer aux travaux de Pauliny par les exemples-types qui correspondent aux différentes situations.

Afin d'expliciter le rapport entre le sens du prédicat verbal et le cas du nom régi, Pauliny a également proposé une typologie des cas : au plan syntaxique selon que l'élément au cas donné exprime ou non une participation au procès et au plan lexical selon qu'il s'agit d'un être vivant ou d'une chose. Ainsi, une participation au procès est-elle au nominatif s'il s'agit d'un être vivant, à l'instrumental s'il s'agit d'une chose.

La théorie de l'intension de Pauliny n'a non seulement jamais été sérieusement critiquée (sauf la critique de F. Miko, restée sans écho [31]), mais a été le centre d'intérêt d'autres chercheurs, tels que J. Ružička (1966) ou J. Kačala (1989) qui ont continué à travailler dans le champ ouvert par Pauliny et à le développer. J. Ružička introduit ainsi le concept et le terme de « valeur intensionnelle du verbe » que J. Kačala dit proche du concept de valence [32].

Les auteurs du dictionnaire présentent la valence très exactement avec les termes de Tesnière, même s'ils donnent de ce dernier une image minorée [33]. La valence et l'intension ont des liens étroits même si les concepts diffèrent. La différence essentielle tient du fait que la valence répond, suivant la conception que l'on en a, du verbe et de ses actants ou du cadre valenciel global alors que l'intension ne considère dans une suite SVO qu'un « terme » gauche et un « terme » droit du verbe correspondant à un prime actant et à un second actant, si l'on reprend les termes de Tesnière.

Ces diverses structures (valence, intension, mais aussi structure sémantique) cohabitent judicieusement dans la description très éclairante qui est faite des verbes slovaques et dont nous présentons l'ordonnancement ci-dessous en détaillant la présentation des entrées du dictionnaire.

Ainsi, le verbe est d'abord donné avec son aspect, p. ex. *ROBIŤ ndk* (c'est-à-dire imperfectif). Il s'agit ici de ce que la tradition linguistique slovaque nomme un « imperfectif tantum », c'est-à-dire un verbe sans correspondant perfectif. De plus, les dépouillements ont mis plusieurs significations en évidence. Ce verbe est donc qualifié de « polysémantique ». Dans un tel cas, les différentes valeurs sont d'abord présentées groupées :

31. Citée par J. Kačala dans *Sloveso a sémantická štrutúra vety*. Bratislava, Veda, 1989, p. 70.

32. *Ibid.*, p. 68.

33. Ceci est peut-être dû au fait que l'ouvrage de référence en la matière semble être l'esquisse réalisée en 1934 et publiée dans le *Bulletin de la Faculté des lettres de Strasbourg* (13 pages), projet d'une grammaire à venir et non pas l'ouvrage lui-même, *Éléments de syntaxe structurale*, fort de 670 pages, publié *post-mortem*.

robiť 1 vykonávať činnosť (exercer / avoir une activité, ce qui est au demeurant le sens général de ce verbe) ;
robiť 2 zaoberať sa niečím (s'occuper de quelque chose) ;
robiť 3 prácou utvárať, pretvárať (former, transformer par le travail) ;
robiť 4 mať funkciu, zamestnanie (avoir une fonction, une profession) ;
robiť 5 byť niekde zamestnaný (être employé quelque part).

Les différentes significations du verbe sont ensuite traitées séparément, à la suite les unes des autres, présentées dans un cadre avec le verbe et son numéro, le tout en gras suivie immédiatement par une glose du sens traité, en gras :

robiť 1

vykonávať činnosť (exercer une activité).

Ce qui constitue l'indéniable particularité de ce dictionnaire est la présence contiguë et ordonnée de plusieurs cadres verbaux. Le premier ensemble étudié concerne la structure sémantique représentée par un ensemble de traits définitoires : action (± A), dynamisme (± D), étendue (± R), relation (± Re), transitivité (± Tr) et mutation (± M).

La structure sémantique de « robiť 1 » est donnée en +action, +dynamisme, +étendue, -relation, -transitivité et -mutation :

SŠ : +A, +D, +R, -Re, -Tr, -M

Vient ensuite :

SYN : pracovať (soit : synonyme : « travailler ») ;
ST : realization (soit : classe sémantique : « réalisation ») ;
VŠ : $\mathbf{S_n}$ – VF – $\mathrm{ADV_{temp}}$ – $\mathrm{ADV_{loc}}$ – $\mathrm{ADV_{mod}}$

Il s'agit de la « structure valencielle » (en slovaque : (VŠ) « *valenčná štruktúra* ») qui note en gras sans parenthèse les éléments obligatoires, en gras entre parenthèses les éléments potentiels et sans marque les éléments facultatifs (de type complémentations adverbiales) qui sont considérés ne pas appartenir à la structure valencielle. « VF » renvoie à une forme personnelle du verbe (par opposition à des formes impersonnelles, infinitif, gérondif ou participe).

IŠ : $\mathbf{AG_{rea}}$ – PRED
Sn : [HUM] AGrea

Cette structure « intensionnelle » représente les rôles sémantiques des participants (gauche et droit) et est liée à la classification sémantique des verbes (classes de verbes statiques, processuels ou d'action avec chacune leurs sous-classes). Ainsi, des verbes d'un type sémantique donné appartiennent-ils à des modèles intensionnels identiques.

La structure intensionnelle est basée sur le prédicat qui est central et est accompagné de participants et de compléments adverbiaux obligatoires. Il est représenté par le

symbole « PRED ». L'analyse de la structure intensionnelle se focalise ensuite sur l'étude du participant gauche et des participants droits.

Au titre du participant gauche (un verbe n'est accompagné que d'un seul participant gauche, mais éventuellement de plusieurs participants droits), on trouve prioritairement un agent noté « **AG** » pour « agens ». Celui-ci est subdivisé en 14 catégories différentes. Nous avons ici un sujet (Sn) représenté par un « AG_{rea} » qui est le symbole pour « *agens realizátor* » : il s'agit d'un être vivant [HUM] (év. d'une institution) qui réalise une action sur le participant droit. Ce cadre intensionnel correspond aux verbes d'action. Il existe donc deux autres symboles pour un participant gauche selon que le verbe est considéré comme exprimant un processus (nous avons le symbole « PROC » avec 17 sous-types) ou une valeur statique exprimée par le symbole « STAT », subdivisé en 13 sous-types.

Quant au participant droit, il est affecté de l'une de vingt-neuf catégories de rôles sémantiques tels que « destinataire », « donateur », « évaluateur », « phénomène », « porteur de modalité », « origine », « patient », « possesseur »,... Les terminologies afférentes et les abréviations associées sont d'origine gréco-latine, passées au filtre anglo-saxon, p. ex. « cognate (COG) ».

En fin de la description, nous trouvons une série d'exemples tels que :

Budeme spolu robiť : ja môžem pomôcť jemu a on mne.
(*Nous allons travailler ensemble : je peux l'aider et lui peut m'aider.*)

Aj v robote bol vždy prvý, a robil za troch.
(*Il était toujours premier même au travail et il travaillait pour trois.*)

Po nociach bdiem a robím do rána.
(*La nuit je reste éveillé et je travaille jusqu'au matin.*)
...

Comme dans Vallex, les articles afférents à une entrée se terminent avec les transformations possibles de la structure de base, suivant un modèle du type suivant :

TRANSF :
1b) $\mathbf{VF_{refl,\ imp}}$ – ADV_{loc} – ADV_{temp} – ADV_{mod}
V momente predaja v Elteku sa robilo na plné obrátky.
(*Au moment des ventes chez Eltek, on travaillait à plein régime.*)

1d) $\mathbf{VF_{refl,\ imp}}$ **– (Sa) –** $\mathbf{ADV_{mod}}$ – ADV_{loc} – ADV_{temp}
Na poliach sa dnes robí ľahšie ako v minulosti.
(*De nos jours le travail des champs est plus facile que par le passé.*)
(m.-à-m. : *On travaille dans les champs aujourd'hui plus facilement que dans le passé.*)

Il existe quatre types de transformations avec d'éventuels sous-types :

a. passif réfléchi ;
b. passif participial ;
c. construction résultative ;
d. participe.

Bien que langues minoritaires et isolats slaves en milieu germanique, les langues sorabes (haut-sorabe et bas-sorabe) bénéficient de l'intérêt soutenu de la part d'universitaires sorabophones (et de quelques autres) ainsi que d'une production scientifique importante, notamment dans le domaine linguistique. C'est ainsi que le haut-sorabe possède également un dictionnaire de référence en matière de valence verbale. Il s'agit de l'ouvrage en deux volumes *Aussagenstruktur im Sorbischen* (*Structure de l'énoncé en sorabe*) avec le sous-titre *Untersuchungen zur Syntax und Satzsemantik* (*Recherches sur la syntaxe et la sémantique de la phrase*). Il s'agit de la version publiée [34] de la thèse d'habilitation de Timo Meškank [35] soutenue à l'Université de Leipzig en 2008.

Meškank se réfère explicitement à Tesnière, mais le phénomène du double accusatif [36] en allemand l'amène à examiner de près les travaux de Fillmore (*The Case for Case*, 1968) et à adopter des cas profonds, rôles thématiques universels. À la suite de Fillmore (*The Case for Case Reopened*, 1977), il élabore la description d'une valence sémantique comparable à la valence syntaxique.

Par ailleurs, Meškank, familier de la Pologne et de la République tchèque, en connaît bien les travaux, notamment de l'école de Prague.

Mais les sources concernent également la langue. C'est ainsi que les recherches sur la fonction des cas conduites par Helmut Faßke et Siegfried Michalk en morphologie, présentées dans *Grammatik der obersorbischen Schriftsprache der Gegenwart. Morphologie* (*Grammaire de la langue littéraire contemporaine du haut-sorabe. Morphologie*) (1981) sont prises en compte ainsi que les travaux de Sonja Wölke, *Verbale Phraseme im Obersorbischen. Untersuchungen zur Valenz und Struktur* (*Phrasèmes verbaux en haut-sorabe. Études structurales sur la valence*) (1992).

Les études syntaxiques sont documentées grâce à une monographie polonaise, celle de Kazimierz Polański *Składnia zdania złożonego w języku górnołużyckim* (*Syntaxe de la phrase complexe en haut-sorabe*) publiée en 1967.

Nous obtenons ainsi la caractéristique principale des travaux de Timo Meškank : celle d'une étude systématique de la valence à trois niveaux : sémantique (rôles thématiques), syntaxique et morphologique. L'objet principal de ces recherches est l'étude des structures sémantiques porteuses et les formulations syntaxiques correspondantes. Chemin faisant, cela donne à Timo Meškank l'occasion de s'intéresser aux diathèses et de revoir les constructions infinitives.

34. Timo Meškank : « Aussagenstruktur im Sorbischen. » Verlag Dr Kovač, Hamburg, 2009.

35. Privatdozent à l'Institut de Sorabistik de l'Université de Leipzig, Timo Meškank est l'auteur d'autres dictionnaires. En premier lieu, le « *Retrogradny słownik hornjoserbskeje rěče* » (*Dictionnaire rétrograde du haut-sorabe*) (2001), mais aussi un dictionnaire qu'il a réédité : « *Prawopisny słownik hornjoserbskeje rěče* » (*Dictionnaire orthographique du haut-sorabe*) (2014).

36. Par exemple : « *Wer hat dich (1) solche Streiche (2) gelehrt* » (Qui t'a appris de pareilles plaisanteries) ?

Il en résulte un dictionnaire de valences contenant 1 800 verbes haut-sorabes répartis en trois hyperclasses contenant globalement seize classes sémantiques.

1. ***Classe des verbes de comportement*** :
   Avec les classes sémantiques suivantes : 1.1 – existence, 1.2 – capacité, 1.3 – sensations et sentiments, 1.4 – intérêts et devoirs, 1.5 – mouvements de l'âme et du corps.

2. ***Classe des verbes sensoriels*** :
   2.1 – perception, 2.2 – réflexion et connaissance, 2.3 – communication.

3. ***Classe des verbes d'action*** :
   3.1 – besoins vitaux, 3.2 – mouvement, 3.3 – activité, 3.4 – influence, 3.5 – efficience, 3.6 – réciprocité, 3.7 – possession et échange, 3.8 – phase.

Nous allons donner en illustration quelques exemples succincts :

353 ᵃgrawać 1.5
(4)–(18)
a ⟨1⟩ EXPERIENCER / Individuum — O_{Akk} / Subst
⟨2⟩ *fak.* KAUSA / Ursache — O_{Gen} / Subst, Infinitiv, NS
O_{Instr} / před+Subst
b ›ekeln‹
c •*Aktiv*
Wokoło nohow šuskachu myše a kwičachu, a mje⟨1⟩ grawašče. (MŁYNKOWA 1994-1:97) Sydnych so přihódnišo. Ale přeco hišće mje⟨1⟩ grawaše, ⟨so lehnyć⟩⟨2⟩. (MŁYNKOWA 1994-1:97) „Eh, wostaj

FIGURE 2 – grawać

Dans le présent exemple, nous avons en entrée le verbe « *grawać* » qui porte le numéro d'ordre 353 dans l'ouvrage de référence. Il appartient à la classe sémantique 1.5, c'est-à-dire « mouvements de l'âme et du corps » de l'hyper-classe « comportements ».

La ligne située sous l'entrée est de la forme : (4) – (18), l'entrée suivante « *grawać so* » (présentée ci-dessous) qui n'est que la forme réfléchie de l'entrée actuelle possède l'indication (1) – (18). Timo Meškank a organisé la présentation du cadre verbal sous la forme prédicat – arguments qui sont des entités cognitives interprétées aux trois niveaux présentés plus-haut : à un niveau sémantique en tant que rôles thématiques, cas profonds issus des travaux de Fillmore, à un niveau syntaxique et à un niveau morphologique. Les chiffres donnés entre parenthèses renvoient au rôle thématique du premier argument (de 1 – Agens à 4 – Experiencer) et du second (de 5 – Koagens à 24 – Goal), tous deux en relation avec les classes sémantiques. Le prédicat « *grawać* » a donc pour valeurs « 4 – Experiencer » et « 18 – Kausa » alors que pour « *grawać so* » le premier argument est « 1 – Agens ».

La section ***a*** décrit en parallèle une valence sémantique et une valence syntaxique. La valence sémantique, à gauche, comprend un premier argument obligatoire « Experiencer », représenté par un « individu », le deuxième argument, facultatif, représente la cause.

La valence syntaxique correspondante contient un premier argument, substantif à l'accusatif et un second argument représenté potentiellement par deux structures. La

première est soit un substantif au génitif, un infinitif ou une subordonnée, la seconde est un syntagme nominal prépositionnel introduit par la préposition « před » (*devant*) suivie d'un substantif à l'instrumental.

La section ***b*** apporte le sens : « *ekeln* » (dégoûter, répugner).

La section ***c*** présente de nombreux exemples extraits de la littérature.

Une éventuelle section ***d*** énumère sous une forme abrégée les verbes préfixés afférents au verbe de l'entrée.

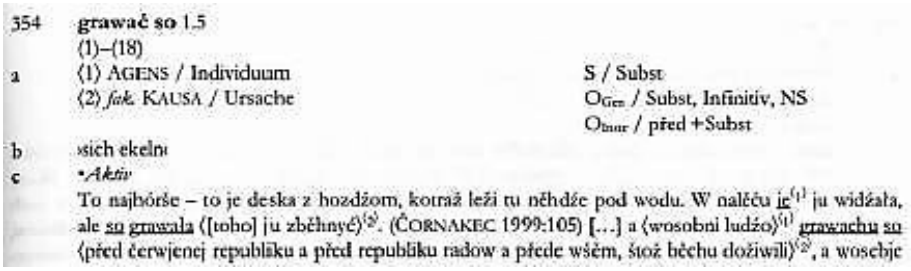

354 **grawać so** 1.5
(1)–(18)
a ⟨1⟩ AGENS / Individuum — S / Subst
⟨2⟩ *fak.* KAUSA / Ursache — O_{Gen} / Subst, Infinitiv, NS
O_{instr} / před +Subst
b ›sich ekeln‹
c •*Aktiv*
To najbórše – to je deska z hozdžom, kotraž leži tu něhdže pod wodu. W naléču je⟨1⟩ ju widźała, ale so grawała ⟨[toho] ju zběhnyć⟩⟨2⟩. (ČORNAKEC 1999:105) […] a ⟨wosobni ludźo⟩⟨1⟩ grawachu so ⟨před čerwjenej republiku a před republiku radow a přede wšěm, štož běchu doživili⟩⟨2⟩, a woschje

Figure 3 – grawać so

Nous donnerons rapidement un autre exemple, celui de « *chwatać* » (se presser, se dépêcher) (« eilen » en section ***b*** qui renvoie au dictionnaire de Helbig et Schenkel [37]), afférent à la classe sémantique 3.2 – mouvement.

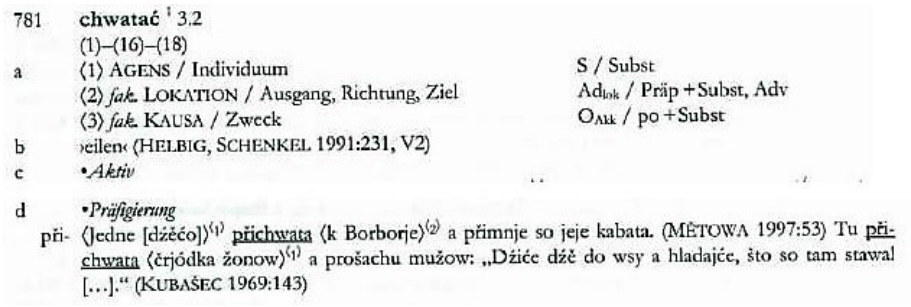

781 **chwatać** [1] 3.2
(1)–(16)–(18)
a ⟨1⟩ AGENS / Individuum — S / Subst
⟨2⟩ *fak.* LOKATION / Ausgang, Richtung, Ziel — Ad_{lok} / Präp +Subst, Adv
⟨3⟩ *fak.* KAUSA / Zweck — O_{Akk} / po +Subst
b ›eilen‹ (HELBIG, SCHENKEL 1991:231, V2)
c •*Aktiv*
d •*Präfigierung*
při- ⟨Jedne [džéćo]⟩⟨1⟩ přichwata ⟨k Borborje⟩⟨2⟩ a přimnje so jeje kabata. (MĚTOWA 1997:53) Tu přichwata ⟨črjódka žonow⟩⟨1⟩ a prošachu mužow: „Džiće dźě do wsy a hladajće, što so tam stawa […].“ (KUBAŠEC 1969:143)

Figure 4 – chwatać

C'est un verbe qui possède trois arguments :

- le premier est un « 1 – Agens / Individu », représenté en syntaxe par un sujet / substantif ;
- le second est une « 16 – Lokation » facultative, expression du lieu (origine, direction, but) avec au niveau syntaxique soit une préposition et un substantif, soit un adverbe. Ce rôle thématique correspond au « DIR » de Vallex, somme de « DIR1 », « DIR2 » et « DIR3 », donc sans préciser davantage le type de mouvement réalisé ;

37. Helbig, Schenkel : „*Wörterbuch zur Valenz und Distribution deutscher Verben*“. Niemeyer, 1991, 2011.

- le troisième argument, facultatif, est une « 18 – Kausa », la cause du déplacement, c'est-à-dire son but. Syntaxiquement, c'est un Objet à l'accusatif représenté par la préposition « po » suivie d'un substantif.

Une section ***d*** « préfixation » mentionne le verbe préfixé « *při-chwatać* » (arriver en courant) qui utilise le même cadre que le verbe simple. Nous supposons que l'économie de l'ensemble de l'appareil pour les verbes dérivés doit être susceptible de produire des descriptions imparfaites.

La mise en parallèle de la valence sémantique et de la valence syntaxique est certainement ce qui caractérise le mieux le travail de Timo Meškank. Cela constitue un outil de premier ordre pour l'étude du haut-sorabe, la rédaction en haut-sorabe et la traduction vers le haut-sorabe. C'est aussi, en puissance, un outil incontournable pour les traitements automatiques du haut-sorabe.

Mais, à notre connaissance, la partie dictionnaire de l'ouvrage n'existe pas en ligne avec tous les avantages de l'informatique qui permettrait différents types d'accès, ne serait-ce qu'un accès alphabétique des verbes. En effet, tel qu'il est conçu, le dictionnaire « papier » est classé par rôles thématiques, ce qui n'est que l'un des modes d'accès possibles. Enfin, un tel dictionnaire en ligne mériterait une description morphosyntaxique plus fine (*cf.* ci-dessus notre remarque sur « 16 – Lokation »).

Nous avons, à plusieurs reprises, souligné les qualités intrinsèques des langues slaves pour une description de la valence en s'appuyant sur les catégories casuelles d'un système linguistique à flexion externe. C'est certainement à ce niveau-là que le bât blesse en ce qui concerne les langues romanes, français en tête. Cela entraîne à notre avis certaines difficultés de représentation du cadre valenciel pour le français.

Pourtant de nombreuses réalisations pratiques, sous forme de bases de données, généralement accessibles en ligne, existent tant en Belgique qu'en France.

Il convient de citer ici les travaux pionniers de Dubois et Dubois-Charlier, rediffusés en 2013 par Paul Sabatier (LIF – Université d'Aix-Marseille) sous l'intitulé « LVF+1 » pour « LVF » (*Les verbes français*), nouvelle version. Sabatier distingue très clairement deux œuvres souvent confondues : d'une part, le « *Dictionnaire électronique des verbes* (français) (DEV) que Jean Dubois et Françoise Dubois-Charlier (D&D) ont réalisé en dBase et terminé en 1992 et qui a été diffusé à partir de septembre 2007 par MoDyCo dans un format Excel », et d'autre part « l'ouvrage *Les verbes français* (LVF) que D&D ont publié en 1997 chez Larousse. Cet ouvrage étant épuisé, une reproduction au format PDF est diffusée par le laboratoire MoDyCo depuis septembre 2007 ».

Paul Sabatier précise :

> « L'ouvrage LVF est une présentation de certaines données contenues dans le DEV. Dans LVF, les verbes sont présentés par classes et sous-classes sémantiques. Dans le DEV, les verbes sont présentés par ordre alphabétique. »

Sabatier a eu le mérite de mettre en clair tout ce qui était sous forme d'abréviation pour raison d'économie de place à une époque où les supports informatiques étaient limités. Cela permet une lecture nettement plus aisée et immédiate des données qui sont publiées dans l'optique d'exploitations automatiques ultérieures. Nous avons ainsi accès à un travail linguistique extrêmement précis (on le verra plus loin, L. Danlos considère que la précision va éventuellement trop loin - cet avis est, bien entendu, lié à l'usage que l'on souhaite faire de ces données et le point de vue est certainement différent suivant que l'on souhaite les utiliser pour une connaissance scientifique, des applications didactiques ou des traitements automatiques et dans chacune de ces situations peuvent intervenir des visées différentes : l'utilisation de ces données sera-t-elle la même pour l'enseignement à des écoliers, des adultes, des étrangers... ?). Cette masse de 25 609 entrées est structurée à l'aide de 14 classes génériques, 54 classes sémantico-syntaxiques, 248 sous-classes syntaxiques et 533 sous-types (Sabatier, 2013).

Nous donnons ci-dessous l'exemple du verbe « abattre » (11 entrées dont 2 pronominales). Étant donné le nombre de données liées à un verbe, nous présentons l'exemple sous forme de trois tableaux avec, en première ligne, l'intitulé réduit de la colonne. Si le 3^{e} tableau [fig7] concernant la dérivation se lit sans grande difficulté, il n'en va pas de même pour certaines colonnes. Il vaut certainement la peine de s'arrêter sur l'interprétation à donner à la colonne « classe ». Dans le code utilisé, le 1er caractère représente l'une des 14 classes génériques. Ainsi, dans l'exemple présenté, nous avons « E » : « entrée / sortie », « F » : « frapper / toucher », « P » : « verbes psychologiques » et « M » : « mouvement sur place ». Le 2^{e} caractère du code indique une sous-classe syntaxico-sémantique, le 3^{e} caractère distingue un type de construction syntaxique et le 4^{e} caractère est l'un des 533 sous-types.

D'autres éléments se lisent également avec difficulté (il faut avoir recours à la version PDF de l'ouvrage écrit pour obtenir la clé). Ainsi dans les opérateurs, « f » veut dire « avoir une activité » et « f.ire » « faire aller quelque part », « f.mvt » : « faire un mouvement » et « f.sent » : « prêter tel sentiment à quelqu'un ».

<ID>	MOT	DOMAINE	*DOMAINE en clair*	*N↑ LANGUE*	OPERATEUR	CLASSE
29	abattre 01	OBJ	objet		(#) [f.ire] qc en bas	E3f.3
30	abattre 02	MIL	militaire		(#) [ict] mauvais aéronef	F3c.1
31	abattre 03	CUI	cuisine, pâtisserie		(#) [ict] animal à mort	F1b.1
32	abattre 04	PATf	pathologie	familier	(#) [ict] qn à mort	F1b.1
33	abattre 05	PSY	psychologie		(abstrait) [f.sent] vif à qn	P2a.1
34	abattre 06	POL	politique		(#) [ict] total qn abstrait	F2a.1
35	abattre 07	NAV	navigation, bateau		(bateau) [ire] VERS	E3d.3
36	abattre 08	MON	monnaie, argent		(#) [f.mvt] moins valeur	M4b.1
37	abattre 09	LOC	locatif, lieu		(#) [ire] per lieu	E3e
38	abattre 10 (s')	LOC	locatif, lieu		(qc) [ire] SUR	E3d.3
39	abattre 11 (s')	SOCt	sociologie	littéraire	(abstrait) [ire] SUR abstrait	E4c.4

Figure 5 – 1. abattre

	PHRASE	CONJUGAISON	CONSTRUCTION
29	On a~ les arbres sur le sol. Le chêne est a~ par la tempête.	5eZ	T13g8
30	On a~ un avion avec un missile.	5eZ	T1308
31	On a~ les boeufs avec une masse dans un abattoir.	5eZ	T1208
32	Le tueur a~ P d'une balle à la sortie de l'hôtel.	5eZ	T1108
33	Ce coup du sort a~ P. La fièvre a~ P. On est a~.	5eZ	T3100
34	On a~ le gouvernement par un vote.	5eZ	T1108
35	Le bateau a~ vers la gauche.	5eZ	N3g
36	On a~ les impôts de dix mille francs.	5eZ	T1306 - P3006
37	On a~ beaucoup de kilomètres avec cette voiture à cette vitesse.	5eZ	T1300
38	L'avion s'a~ sur la maison. Le vent a~ l'arbre.	5eZ	P30g0 - T33g0
39	Un malheur s'a~ sur la famille.	5eZ	P30g0

Figure 6 – 2. abattre

	DERIVATION	A_able	A_é	A_ant	N_age	N_ment	N_ion	N_eur	N_oir	N_ure	NOM	lexique
29	1-1 1- --RA --	1	-	1	1	-	--	RA	-	-	-	1
30	--- -- ---- --	-	-	-	-	-	--	--	-	-	-	5
31	--- 1- ---- 1-	-	-	-	1	-	--	--	1	-	-	3
32	--- -- ---- --	-	-	-	-	-	--	--	-	-	-	3
33	-1- -1 ---- --	-	1	-	-	1	--	--	-	-	-	3
34	--- -- ---- --	-	-	-	-	-	--	--	-	-	-	5
35	--- 1- ---- --	-	-	-	1	-	--	--	-	-	-	5
36	--- -1 ---- --	-	-	-	-	1	--	--	-	-	-	5
37	--- -- ---- --	-	-	-	-	-	--	--	-	-	-	5
38	--- -- ---- --	-	-	-	-	-	--	--	-	-	3*	3
39	--- -- ---- --	-	-	-	-	-	--	--	-	-	-	3

Figure 7 – 3. abattre

D'une manière générale, il est recommandé de consulter les deux documents (la base de données et le livre) pour d'éventuels traitements automatiques ultérieurs.

Une autre référence incontournable est l'œuvre initiée et conduite pendant plusieurs décennies par Maurice Gross. Maurice Gross est l'un de ces polytechniciens devenu linguiste [38] qui a fortement influencé le domaine et qui a porté loin de nos frontières le drapeau de la linguistique française. Quelques années après sa sortie de polytechnique, Maurice Gross a la chance de pouvoir séjourner à Harvard et au MIT où il fera la connaissance de Chomsky, mais aussi celle de Marcel-Paul Schutzenberger, ce qui sera peut-être encore plus déterminant. Il fera ainsi un ouvrage en collaboration avec André Lentin [39], autre élève de Schutzenberger, « Notions sur les grammaires formelles » (1967) que l'on connaît sous l'appellation « Le Gross et Lentin ».

C'est aussi grâce à Schutzenberger que Maurice Gross pourra collaborer avec Zellig Harris à l'Université de Pennsylvanie (Philadelphie). Il en rapportera une formation

38. Il en va de même en ce qui concerne son cadet, Christian Boitet, spécialiste de TA/TAO, successeur de Bernard Vauquois.

39. Il y avait dans l'entourage de Schutzenberger d'autres personnalités connues telles que Maurice Nivat ou Pierre Rosenstiehl, ses élèves. Maurice Gross rencontre au Centre Henri-Poincaré des linguistes tels que Pottier ou Greimas (« Maurice Gross, un grand linguiste », hommage funéraire de Jean-Claude Chevalier dans Le Monde du 12/12/2001).

unique et le « cadre du lexique-grammaire » comme le rappelle justement Amr-Helmy Ibrahim [40] :

> « [. . .] la langue fonctionne comme un lexique-grammaire et c'est une triple erreur théorique, méthodologique et pratique d'avoir des niveaux d'analyse distincts, pour ne pas dire indépendants, du lexique, de la syntaxe et de la sémantique. Ainsi ne peut-on rien dire de consistant sur le lexique sans avoir exploré la grammaire de chaque mot et il n'existe pas de règles grammaticales totalement indépendantes des particularités lexicales des unités auxquelles elles s'appliquent. Le lexique-grammaire recouvre donc à la fois une conception de la langue et un programme pour l'analyser.
>
> « Le programme surtout ne se confond pas tout à fait – et pour certains pas du tout – avec le cadre plus général de la linguistique harrissienne dont il est, pour l'essentiel, issu. Il se distingue aussi, nettement, des différentes hypothèses lexicalistes liées ou non au courant générativiste. Enfin il est résolument incompatible tant dans ses pratiques concrètes de description que dans ses présupposés méthodologiques et théoriques avec une linguistique qui partirait d'un ensemble d'intuitions sémantiques fondées ou non sur des associations bi-univoques récurrentes entre formes et interprétations. »

Décrire l'œuvre de Maurice Gross et notamment le *Lexique-Grammaire* en quelques lignes est chose très difficile. Un trait dominant à mettre en exergue est certainement ce qu'a retenu Ibrahim dans l'un de ses articles, à savoir « une méthode raisonnée de classement des unités linguistiques » comparable à la démarche de Carl von Linné connu par sa taxinomie binomiale. On aurait pu, tout autant, parler du tableau périodique des éléments de Mendeleïev. Mais la description linguistique rappelle bien davantage les sciences biologiques dont les taxinomies sont en constante évolution, pour ne pas dire révolution avec par exemple la protistologie.

C'est dès son retour de Philadelphie que Maurice Gross met en place le lexique-grammaire (LG) tout en publiant abondamment en parallèle. Il crée le Laboratoire d'automatique documentaire et linguistique (LADL) et bientôt le réseau international « Relex ». Il publie *Syntaxe du verbe* (1968) qui sera suivie de *Syntaxe du nom* (1977) et de *Syntaxe de l'adverbe* (1986). Les *Méthodes en syntaxe* (1975) exposent ses positions théoriques.

Le LG, c'est, entre autres, la classification d'environ 5 000 verbes simples et des prédicats complexes (verbe support suivi d'un substantif « prédicatif »). On aura bientôt un lexique-grammaire comportant plus de 25 000 expressions verbales figées sans compter les expressions avec « être » et « avoir ».

En 1990, Christian Leclère [41] indiquait que le LG contenait alors 31 000 entrées réparties en 81 tables : 20 tables de verbes composés (20 000 phrases figées), 18 tables de

40. Ibrahim, Amr Helmy, Le cadre du Lexique-grammaire, LINX 48, 2003. *Cf.* bibliographie.
41. Leclère Christian. Organisation du lexique-grammaire des verbes français. *Cf.* bibliographie.

verbes simples à construction complétive (3 000 verbes) et 43 tables de verbes simples à construction nominale (8 000 verbes).

En tant que scientifique, à une époque où la traduction automatique devient une question stratégique, Maurice Gross a pensé de tout temps (et nous en sommes témoin) au traitement automatique des langues. C'est avec cette finalité que sont construits les dictionnaires électroniques DELAS (dictionnaires des mots simples), DELAF (dictionnaire des formes fléchies), DELAC (dictionnaires des mots composés) et quelques autres.

Le LG est ainsi constitué d'un grand nombre de vastes matrices que les capacités limitées de mémoire des ordinateurs de l'époque ne permettaient pas encore de traiter. Il en va aujourd'hui autrement. Nous présentons ci-dessous un minuscule extrait, issu des travaux d'Elsa Tolone, pour illustrer le principe :

N0 =: Nhum	N0 =: N-hum	N0 =: Nnr	Ppv	Ppv =: se figé	Ppv =: en figé	Ppv =: les figé	Nég	<ENT>	N0 V	N0 être V-ant	N1 =: Nhum	N1 =: N-hum	N1 =: le fait Qu P	Ppv =: lui	Ppv =: y	N0hum V W sur ce point	[extrap]	<OPT>
+	-	-	<E>	-	-	-	-	renaître	+	+	-	+	-	-	+	-	-	Max renaît au bonheur de vivre
+	-	-	se	+	-	-	-	rendre	+	-	+	+	+	-	+	+	+	Max s'est rendu à mon opinion
+	-	-	se	+	-	-	-	rendre	+	-	+	-	-	-	-	-	-	Le caporal s'est rendu à l'ennemi
+	-	-	<E>	-	-	-	-	renoncer	-	-	+	+	-	-	+	-	-	Max renonce à son héritage

Figure 8 – Extrait de la classe 33 des verbes

Nous connaissons plusieurs tentatives de réutiliser en l'aménageant le lexique-grammaire pour créer des lexiques d'analyse syntaxique du français. C'est le cas du projet Synlex réalisé par Claire Gardent et plusieurs de ses collègues dans le cadre de laboratoires de l'Université de Nancy. Claire Gardent signale deux caractéristiques du LG importantes pour le traitement automatique : la sous-catégorisation et les alternances. Ce dernier point nous semble encore plus important que le premier. Il constitue une partie essentielle des derniers travaux pour le Vallex tchèque. Ce sont dans le LG des transformations telles que passif, réciproque et surtout telles que « alternance locative » : « Les fautes pullulent dans ce texte », face à : « Ce texte pullule de fautes. »

Un vaste projet de réutilisation du LG, plus récent, est également conduit par Elsa Tolone, sujet sur lequel elle a présenté sa thèse de doctorat intitulée *Analyse syntaxique à l'aide des tables du Lexique-Grammaire du français* à l'Université de Paris-Est en 2011. Ces travaux sont réalisés par les héritiers du LADL au sein du Laboratoire d'informatique Gaspard-Monge.

Claire Blanche-Benveniste est connue essentiellement pour ses travaux sur le français oral. Elle est bien connue aussi pour ses travaux, tout aussi pionniers que les

précédents, sur l'intercompréhension[42] romane qui ont débouché sur la réalisation de la méthode Eurom4[43]. Mais il y a une autre composante (ce n'est pas la seule au demeurant) qu'il convient de mettre en avant, ce sont ses travaux de recherche en syntaxe et tout particulièrement sur « l'approche pronominale ». L'approche pronominale est l'œuvre commune de Claire Blanche-Benvéniste et de Karel Van Den Eynde, ce qui donnera une remarquable application, *Dicovalence*, que nous verrons ci-dessous. L'approche pronominale est aussi le thème de la thèse d'État de Claire Blanche-Benveniste : *Recherche en vue d'une théorie de la grammaire française : essai d'application à la syntaxe des pronoms*, publiée en 1975. Elle publie abondamment avec Van den Eynde, notamment « Syntaxe et mécanismes descriptifs : présentation de l'approche pronominale » dans les *Cahiers de lexicologie* en 1978 et *Pronom et syntaxe. L'approche pronominale et son application au français* à la Selaf, en collaboration avec Deulofeu, Stéfanini et Van den Eynde en 1984.

Dans un article publié en 2002 dans *Tranel* (*Travaux neuchâtelois de linguistique*), « La complémentation verbale : petite introduction aux valences verbales », Claire Blanche-Benveniste démontre clairement et simplement les difficultés à déterminer les cadres verbaux. Elle prend comme point de départ les travaux de Dominique Willems, qui sur la base d'une étude englobant 3 000 verbes français, en déduit une soixantaine de classes au comportement syntaxique et sémantique parfaitement descriptible.

Les questions primordiales d'acceptibilité et de grammaticalité trouvent chez Claire Blanche-Benveniste et Karel Van Den Eynde une réponse grâce à l'approche pronominale, c'est-à-dire par le recours aux pronoms ou plus exactement à des « pro-formes ». Ainsi, les valences verbales sont-elles décrites à l'aide de la liste des pro-formes que les verbes acceptent. Au niveau des actants, ce sont les pronoms personnels aux formes fléchies adéquates qui sont utilisés. En ce qui concerne les locatifs et les temporels, ils sont représentés « par toutes les pro-formes disponibles, interrogatifs compris », la manière par des termes tels que « ainsi », « comme ça », « comment », la quantité par « tant », « autant », « combien », etc. Il est intéressant de constater que Claire Blanche-Benveniste considère l'usage de ces pro-formes comme bien plus naturel que l'usage d'étiquettes grammaticales : « Je le lui fais connaître », car c'est très exactement ce que fait la tradition slave pour énoncer les cas voulus lors d'une hésitation (en tchèque : « vidět koho, co » (*voir qui/quoi* et donc accusatif).

Il en découle la possibilité de déterminer le domaine du verbe, ce qui demeure un problème essentiel. Ainsi, Claire Blanche-Benvéniste écrit-elle :

> « Les éléments du domaine verbal… ont une propriété essentielle : ils sont liés au verbe par une relation qui ne dépend pas directement de leur lexique (c'est pourquoi on peut les représenter par des pronoms), ni de la modalité, affirmative, négative ou restrictive, sous laquelle ils apparaissent (c'est pourquoi on peut en changer). Les éléments qui ne font pas partie du domaine verbal n'ont pas cette propriété. »

42. En ce qui concerne l'intercompréhension, il faut noter aussi le caractère pionnier des travaux de Louise Dabène, à l'origine du pôle d'intercompréhension grenoblois.

43. Devenue par la suite Eurom5 sous la direction d'Elisabetta Bonvino, professeur à l'Université de Rome.

L'approche pronominale a été le ferment d'un vaste projet de dictionnaire de valences réalisé à l'Université de Louvain (dont Claire Blanche-Benveniste a reçu un *doctora honoris causa*) par Karel Van den Eynde et ses collègues.

C'est une entreprise qui dure depuis plusieurs décennies. Le premier projet, le dictionnaire de valences Proton, a été réalisé par Karel Van den Eynde et son équipe entre les années 1986 et 1992. Ce sont 3 700 verbes simples extraits du Petit Robert qui produisent 8 600 entrées syntaxico-sémantiques différentes. Les schémas valenciels ont été constitués à partir du *Petit Robert*, du *Larousse* et du dictionnaire de valences de Busse & Dubost déjà existant (1983). Une typologie des compléments est constituée dans la perspective de l'approche pronominale, plus riche que les descriptions grammaticales traditionnelles : P0 (sujet), P1 (objet), P2 (objet indirect), P3 (objet indirect), PQ (quantité), PL (locatif), PD (délocatif), PT (temps), PM (manière) et PP (autre complément à préposition fixe). Comme dans LG, un certain nombre de transformations (nommées ici « reformulations ») sont déjà présentes, p. ex. les reformulations passives. Proton traite aussi les infinitives simples ou précédées d'une préposition, les « que-phrases ».

Pour des raisons d'exploitation automatique, on a substitué aux pronoms un système de traits, ce qui a permis d'intégrer Proton dans le système de traduction automatique METAL. L'analyse en traits a permis de lancer des recherches sur l'intégration de Proton dans un analyseur en arbres de dépendances (Mertens, 2002).

En 2006, le dictionnaire est entièrement revu par Karel Van den Eynde et Piet Mertens et prend le nom de *Dicovalence*. En 2010, Piet Mertens réalise une version ultérieure *Dicovalence 2*. Chacune des nouvelles versions a apporté son lot de changements. La première introduit notamment une identification plus large des différentes classes de verbes : verbes simples, locutions verbales, verbes copules, verbes auxiliaires, verbes de dispositif, etc. Les apports les plus marquants de la seconde version sont :

- Le champ FRAME qui fournit le cadre de sous-catégorisation. Pour chaque place de rection il indique le caractère obligatoire ou facultatif, la fonction syntaxique, les réalisations syntagmatiques possibles (pronom, groupe nominal, complétive, infinitive, adverbe…), ainsi que les principales restrictions de sélection, formulées à l'aide des traits sémantiques « humain, concret non humain, abstrait »[44].
- Le champ AUX qui indique le ou les auxiliaires d'aspect employées pour l'entrée[44].

44. Extrait de : Le dictionnaire de valence *Dicovalence* : Manuel d'utilisation du dictionnaire (version du 25-6-2010).

À titre d'illustration, nous présentons le verbe « dégager », en premier, dans la version de Proton :

```
VAL$      dégager
PRED$     simple_predicator
NUM$      24750
EX$       r : les plantes dégagent du gaz carbonique
TR$       ontwikkelen, verspreiden, uitwasemen
P0$       que, ceci, elle, il, ils, celui-ci, ceux-ci, ça
P1$       la, le, les, en Q, que, celui-ci, ceux-ci, ça
RP$       passif être
NEWEX$    ses cheveux dégagent une odeur agréable
RESLEX$   [lexicalization of] p1, [restricted to] une odeur, ...
LC$       elle en dégage, ça se dégage de là
LINK$     to 24760
```

Figure 9 – Proton

Puis la version *Dicovalence 2* :

VAL$	dégager: P0 P1
VTYPE$	predicator simple
VERB$	DEGAGER/dégager
NUM$	24750
EG$	ses cheveux dégagent une odeur agréable
TR_DU$	ontwikkelen, verspreiden, uitwasemen
TR_EN$	develop, give off a smell; exhale
FRAME$	subj:pron|n:[nhum,?abs], obj:pron|n:[abs]
P0$	que, elle, il, ils, ça, ceci, celui-ci, ceux-ci
P1$	que, la, le, les, en Q, ça, celui-ci, ceux-ci
RP$	passif être
LC$	24750-24760 elle en dégage, ça se dégage de là
AUX$	avoir

Figure 10 – *Dicovalence 2*

On notera également dans cette dernière version la présence d'équivalents anglais en plus de ceux en néerlandais.

Un autre projet de dictionnaire de valences n'est pas non plus sans rapport avec Claire Blanche-Benveniste, puisque André Valli, coauteur d'Eurom4 et cofondateur avec quelques autres collègues du Groupe aixois de recherches en syntaxe (GARS), joue un rôle décisif dans la construction de LEXVALF, le lexique électronique des valences verbales du français.

D'autres chercheurs ont également participé au projet LEXVALF parmi lesquels nous pouvons citer Morris Salkoff, Paul Sabatier et Laure Brieussel, chargée des aspects informatiques. Comme beaucoup d'autres projets de ce type, LEXVALF exploite les tables du lexique-grammaire de Maurice Gross. Mais les entrées de cette base de données regroupent beaucoup d'informations venues d'autres sources électroniques telles que le *Petit Robert* (PR), le *Grand Robert* (GR) et le *Trésor de la langue française* (TLF). L'un des buts affichés de cette base est d'offrir un inventaire des compléments prépositionnels, afin de « distinguer les arguments verbaux de simples

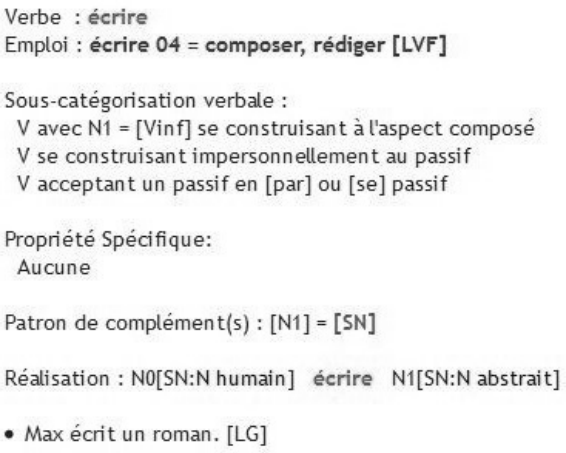
ajouts droits du verbe (les "circonstants" de la tradition grammaticale)[45] » et de sérier les problèmes de rattachement en traitement automatique[46]. Pour cela, les auteurs se sont appuyés sur les critères d'identification développés par Salkoff dans sa *Grammaire en chaîne du français* (1980). Les auteurs adoptent également un « typage des arguments » « à la Gaston Gross » à l'aide d'une liste de traits sémantiques.

Si nous reprenons l'exemple donné par André Valli, nous trouverons avec le verbe « écrire », un schéma syntaxique :

N0[SN :N humain] écrire N1[SN :N abstrait]

qui vaut aussi bien pour l'illustrissime Max lorsqu'il écrit un roman (exemple du lexique-grammaire) « Max écrit un roman » que pour le « musicien écrit cette symphonie » (exemple des *Verbes français*).

Nous présentons ci-dessous les deux réalisations extraites du site LEXVALF. Elles éclairent parfaitement la raison de l'usage de LVF (*Les verbes français*) qui permet de décrire la polysémie verbale de manière adéquate, ici la distinction entre production littéraire et production musicale. Le recours à LVF permet de classer les « différentes réalisations des compléments d'un verbe, soit sur la base de la structure, soit sur celle de l'emploi sémantique ». On ne peut cependant s'empêcher, notamment dans l'éventualité d'une utilisation dans le cadre de programmes de traitement automatique, de se demander s'il n'y a pas une certaine « surgénération » des nuances, par exemple ici des types de production.


Verbe : **écrire**
Emploi : **écrire 04 = composer, rédiger [LVF]**

Sous-catégorisation verbale :
V avec N1 = [Vinf] se construisant à l'aspect composé
V se construisant impersonnellement au passif
V acceptant un passif en [par] ou [se] passif

Propriété Spécifique:
Aucune

Patron de complément(s) : [N1] = [SN]

Réalisation : N0[SN:N humain] **écrire** N1[SN:N abstrait]

• Max écrit un roman. [LG]

Figure 11 – LEXVALF (1)

45. André Valli : « L'intérêt d'adopter l'interface sémantique de LVF pour un nouveau lexique syntaxique de la complémentation verbale en français », Langages 2010/3 (n° 179-180), pp 175-192.

46. Ce que j'appellerais volontiers le syndrome de « l'escalope de dinde normande », même si l'expression standard, non raccourcie devrait être « escalope de dinde *à la* normande ».

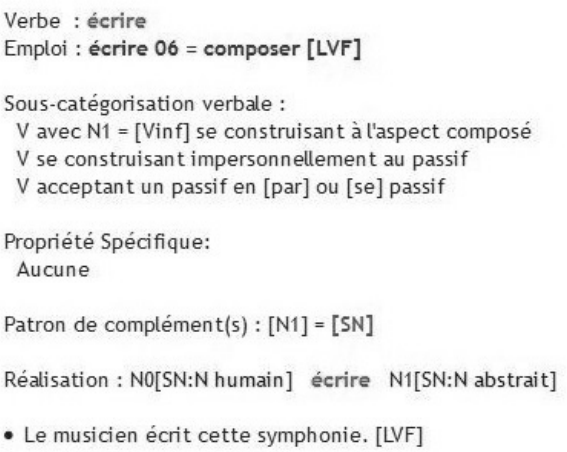


Figure 12 – LEXVALF (2)

La base de données, qui est l'expression concrète du projet LEXVALF, contient actuellement 1018 verbes qui donnent un ensemble important de 23 000 « réalisations ». Ses auteurs la présentent comme devant réunir, à terme, les entrées de plus de 6 000 verbes, ce qui serait équivalent au volume de *Dicovalence*.

Signalons enfin l'existence de *Lefff*, un dictionnaire morphologique du français destiné au traitement automatique. Ce projet est conduit depuis 2003 (jusqu'à la réalisation d'une version 3.4 en 2015 [47]) par Benoît Sagot et ses collaborateurs. Une composante syntaxique a été construite à partir d'informations complémentaires issues du Lexique-grammaire et de *Dicovalence* [48].

L'anglais, privé de la quasi-totalité des flexions germaniques, se comporte comme le français et c'est certainement grâce à cela que nous viennent des perspectives intéressantes pour le traitement de la valence. Concrètement, c'est Laurence Danlos [49], spécialiste de traitement automatique des langues, universellement connue pour ses travaux sur la génération automatique, qui nous les apporte sous la forme d'un « verbənet du français ».

Verbənet [50] est construit sur la base de VerbNet, thèse de doctorat de Kipper-Schuler en 2005 [51] à l'Université de Pennsylvanie. VerbNet est un lexique électronique (créé de manière manuelle) des verbes anglais avec leurs différents sens et pour chacun d'eux le cadre valenciel (nombre et forme des arguments auxquels sont associés les rôles thématiques correspondants), ce qui est vraisemblablement une nouveauté pour

47. Voir p. ex. : http ://alexina.gforge.inria.fr/
48. Voir https ://www.irit.fr/~Dominique.Longin/TALN2007/
49. L. Danlos, Q. Pradet, L. Barque, T. Nakamura, M. Constant : *Un verbənet du français*, TAL, vol. 57, n° 1-2016, p. 33-58.
50. Disponible sur https ://github.com/aymara/verbenet.
51. Kipper-Schuler K., *VerbNet : A broad-coverage, comprehensive verb lexicon*, PhD thesis, University of Pennsylvania, 2005.

le français. VerbNet est une reprise et une extension des travaux de Levin, organisé en 109 super-classes et 270 classes possédant elles-mêmes des sous-classes. VerbNet est une ressource en ligne, documentation comprise, gérée par l'Université de Boulder (Colorado) sous la direction de Martha Palmer.

Laurence Danlos explique que l'entreprise de constituer un Verbənet à partir de VerbNet provient du constat que « les ressources verbales existantes ne permettaient pas de développer facilement un lexique codant les rôles thématiques, soit parce que trop complexes pour être aisément enrichies (LVF et LG), soit parce que non structurées sémantiquement (*Dicovalence* et *Lefff*), soit parce qu'en cours de développement (ASFALDA et RL-Fr)[52] ».

Pour la réalisation de Verbənet, les 270 classes de Verbnet ainsi que les informations sémantiques sont conservées. Chaque classe reçoit (en partie manuellement) les classes du LVF et de LG qui correspondent. Même si la traduction du verbe anglais guide cette opération, ce n'est pas elle qui est déterminante, mais la correspondance avec la définition sémantique de Verbnet et l'invariance des rôles thématiques dans une position syntaxique déterminée.

Il peut y avoir, bien sûr, nombre de dissymétries entre les classes anglaises et les classes françaises : Danlos montre que la différence de transitivité entraîne des dimensions de classes très différentes.

Les opérations qui suivent sont entièrement réalisées à la main et concernent l'adaptation des *frames* de l'anglais, la redéfinition éventuelle de sous-classes, la vérification des rôles thématiques et celle de la réalisation syntaxique des compléments...

Un point important, souvent négligé tant dans les projets que leur évaluation, a été la fabrication d'outils spécifiques pour réaliser les tâches définies. Pour Verbənet, des interfaces réalisées à l'aide du *framework* Django et du SGBD PostGreSQL, tant pour l'exécution du projet que pour des consultations ultérieures, semble être une opération optimale.

Des éléments extrêmement importants tant au plan théorique que pratique pour la réalisation de ressources nécessaires aux analyses automatiques syntaxico-sémantiques ont été développés pendant l'élaboration de Verbənet. Il convient de citer, ici aussi, une attention particulière aux alternances : alternance passive (bien que non codée dans Verbənet), alternances locatives, symétriques, pronominales. Une forme pronominale « se V » peut être précisée à l'aide de restrictions <+middle>, <+neutre>, <+réciproque> ou <+réflexive> comme dans l'exemple suivant :

a. Luc a cassé la carafe = Agent V Patient.

b. La carafe a cassé = Patient V.

c. La carafe s'est cassée = Patient se V<+neutre>[53].

52. Danlos, Pradet, Barque, Nakamura, Constant : Un Verb@net du français, TAL. Volume 57 – n° 1/2016, p. 33-58.

53. *Ibid.*, p. 49.

Un autre point qui nous semble important est le traitement des clitiques, notamment pour la complémentation indirecte. Enfin, le recours à des restrictions sur certains rôles thématiques permet une gestion fine des compléments phrastiques et infinitivaux, notamment avec un « faux de » comme dans :

> Luc a projeté de partir aux US : Agent V Theme<+de VAgent-inf>

Alors que :

> Luc doute de pouvoir partir aux US : Agent V {de} Theme<+VAgent-inf>.

En effet, une infinitive introduite par la « vraie » préposition « de » se pronominalise en « de cela », ce qui n'est pas le cas ici de « projeter de » (projeter quoi/cela).

En conclusion, Verbənet, qui bénéficie de l'expérience de LVF et de LG, nous emmène beaucoup plus loin que ses prédécesseurs, notamment en ce qui concerne l'analyse automatique du français à un niveau syntaxico-sémantique.

En ce qui concerne un exposé grammatical de la valence en slovaque dans le cadre d'une grammaire contextualisée une connaissance approfondie et, si possible, une pratique des différents systèmes exposés seront très précieuses. Il nous semble cependant, et cela apparaît dans tous les systèmes à partir des années 1970, qu'une théorie de la valence doit être testée par une analyse automatique. Cela serait de nature à nous inciter à réaliser une analyse automatique du slovaque parallèlement à la rédaction de la grammaire.