

Analyse exploratoire des propriétés sémantiques et syntaxiques des verbes « psychologiques » (classe P) dans *Les Verbes Français*

Nicolas Mazziotta

Universität Stuttgart – nicolas.mazziotta@ulg.ac.be

Abstract

The electronic dictionary *Les verbes français* (henceforth « LVF ») provides an wide-range syntactic and semantic set of features for French verbs. Such a resource can be analyzed using exploratory statistic methods that have generally been used to investigate corpora. To evaluate the efficiency of this approach, statistical results can be compared to the outcomes of theoretical linguistics. This paper studies the behaviour of psychological verbs in the LVF. It appears clearly that both approaches converge and combine very well. Statistical methods offer additional evidence to the theoretical classification and hint at hidden clusters. On the other hand, linguistic theory underlines lacking information in the database that should be formalized to achieve a proper description.

Résumé

Le dictionnaire électronique *Les verbes français* (ou « LVF ») procure une description syntaxique et sémantique des verbes du français qui est exploitable par les méthodes de l'analyse de données textuelles. Pour évaluer l'apport de cette approche, les résultats de l'analyse de données peuvent être comparés à ceux de la description linguistique théorique appliquée aux verbes classés comme « psychologiques » dans le LVF. Il apparaît clairement que les deux approches se complètent de différentes manières. Les méthodes statistiques soutiennent le classement théorique et attirent le chercheur sur des regroupements ignorés, alors que l'analyse théorique pointe les manques inhérents à l'encodage et invitent à formaliser d'autres aspects de la description linguistique.

Mots-clés : verbes français, sémantique, syntaxe, ACM

1. Introduction

À notre connaissance, les méthodes statistiques exploratoires ne sont pratiquement jamais exploitées pour traiter des questions relatives aux fonctionnements sémantique et syntaxique du lexique verbal français. Pourtant, au moins deux bases de données de taille conséquente sont accessibles : le *Dicovalence* et le dictionnaire *Les verbes français* (voir respectivement (Dubois et Dubois-Charlier, 1997) et (van den Eynde et Mertens, 2010)). Les deux dictionnaires fournissent des informations d'une richesse si importante qu'on aurait peine à en dégager intuitivement des tendances générales sans l'aide de méthode outillées. Poursuivant nos recherches exploitant cette méthodologie (Mazziotta et Martin, 2013), nous voudrions comparer le traitement non outillé d'une question par un linguiste faisant usage de la base de données *Les verbes français* (désormais « LVF ») aux résultats d'une analyse statistique. En l'occurrence, nous analyserons le classement des verbes dits « psychologiques » dans le LVF et nous confronterons nos résultats à ceux proposés par Rolf Kailuwet dans son article de 2007 intitulé « La classe P dans Les Verbes français et les verbes de sentiment », où il classe les verbes du LVF en fonction de ses propres critères.

Il s'agit donc ici de comparer deux *reclassements* de la classe P : celui qui émerge du traitement statistiques des informations du LVF et celui que Kailuwet élabore par observation non outillée.

1.1. Données du LVF

Le vaste dictionnaire électronique LVF décrit les propriétés sémantiques et syntaxiques des verbes du français ((Dubois et Dubois-Charlier 1997) ; forme XML publique (Hadouche et Lapalme 2010)). Il compte 12 308 articles, subdivisés en entrées spécifiques aux différentes acceptions et constructions de chaque verbe (pour un total de 25 610 entrées). Lorsqu'un verbe a plusieurs acceptions, son lemme est accompagné d'un indice numérique spécifique à chacune d'elles ; p. ex. : *abasourdir 01* « étourdir, assourdir » (sens physique) vs *abasourdir 02* « stupéfier » (sens psychologique). Nous ne décrivons pas intégralement la structure des articles de l'ouvrage (voir section 2 pour les éléments exploités, et (François et al., 2007) pour une présentation détaillée). Pour l'essentiel, le LVF fournit une description sémantique de chaque entrée sous la forme d'une définition partiellement formalisée, ainsi qu'une description syntaxique qui distingue les constructions transitives directes, transitives indirectes (p. ex. : *parler à quelqu'un*), intransitives et pronominales et en exemplifie une partie. Une analyse morphologique sommaire est également fournie (dérivations et classes de conjugaison).

L'ouvrage trie les verbes dans 246 sous-classes en fonction de leurs propriétés sémantiques et syntaxiques (*abasourdir 01* fait ainsi partie de la classe H « états physiques et comportements », alors qu'*abasourdir 02* fait partie de la classe P « verbe psychologique »). Selon les termes du LVF (pp. 285 sv.), les 2074 individus de la classe P sont les verbes exprimant : « avoir tel sentiment, telle activité consciente » (P1, 10 sous-classes), « faire avoir tel sentiment à qn » (P2, 3 sous-classes), et « manifester telle connaissance, telle analyse » (P3, 3 sous-classes). Ces trois groupes sont divisés en un total de 16 sous-classes (voir section 2).

Comme chacun des verbes de ces classes est associé à une série de constructions syntaxiques, il est possible d'étudier la répartition des verbes transitifs, transitifs indirects, intransitifs et pronominaux. L'intitulé des classes laisse transparaître la dimension syntaxique du classement, mais elle n'est pas explicitée *a priori* et mérite une vérification.

1.2. Analyse de Kailuwet 2007

Dans son article, Rolf Kailuwet compare son analyse des verbes dits « de sentiment » en français et en italien (2005) à celle du LVF. Son classement se fonde sur des critères syntaxiques et des critères sémantiques. Du point de vue syntaxique, il distingue (reprenant la terminologie de (Tesnière, 1959)) : les verbes *monovalents* (ex. : *angoisser*), qui ne sont accompagnés que d'un sujet ; les verbes *divalents*, accompagnés d'un sujet et d'un complément direct (*transitifs directs* ; ex. : *déranger X*) ou indirect (*transitifs indirects* ; ex. : *craindre pour X*) et les verbes *trivalents*, accompagnés d'un sujet et de deux autres compléments (ex. : *dégoûter X de Y*). Du point de vue sémantique, Kailuwet insiste sur la nécessité de distinguer quatre rôles : EXPERIENT (personne affectée par le sentiment), CAUSEUR (entité causant le sentiment), CORRELAT d'intentionnalité (entité qui est l'objet du sentiment) et POINT DE REFERENCE (entité support de l'objet du sentiment, qui reste implicite). Par exemple (repris et adapté de (Kailuwet, 2007)) :

Paul (EXPERIENT) *crain*t son père (CORRELAT)

Paul (CAUSEUR) *terrifie* son père (EXPERIENT)

Paul craint pour son père (POINT DE REFERENCE)

Il propose ainsi un classement en six types qu'on peut résumer comme suit (où l'on voit que c'est essentiellement la notion d'EXPERIENT qui est mobilisée ; les autres rôles sont utilisés pour affiner le classement) :

1. Verbes bivalents transitifs directs à EXPERIENT objet = classe P2a
2. Verbes bivalents transitifs directs à EXPERIENT sujet = classes P1i et P1j, considérées comme indistinctes¹
3. Verbes bivalents transitifs indirects à EXPERIENT objet = classe P2b
4. Verbes bivalents transitifs indirects à EXPERIENT sujet = classe P1c
5. Verbes trivalents (aucune classe spécifique associée explicitement)
6. Verbes monovalents = portion réduite de la classe P1a (*angoisser, baliser, bicher, bisquer, complexer, culpabiliser, (ne pas) décolérer, déprimer, désespérer, flipper, paniquer, pétocher, stresser*)

La classe P2c « augmenter, diminuer, remplir tel sentiment chez qn » est disqualifiée par Kailuwet : n'associant pas le rôle d'EXPERIENT à un de leurs arguments (le sujet exerce en effet le rôle de causeur sur un objet non humain, ces verbes ne peuvent être qualifiés de verbes « de sentiment »). En dehors de P2c, les autres classes non associées à l'un des types de Kailuwet ne sont pas même mentionnées par l'auteur, mais on devine que nombre d'entre elles ont été écartées parce qu'elles ne correspondaient pas à la définition qu'il sélectionne pour identifier les verbes « de sentiment »². Notre étude se servira surtout de la notion d'EXPERIENT, qui n'est pas présente telle quelle dans le LVF, mais qu'il y a moyen de reconstituer de manière compositionnelle (voir section 2).

1.3. Plan de l'exposé

Dans un premier temps, nous présenterons la préparation des données (section 2). En nous servant de l'ACM et de classifications ascendantes³, nous allons déterminer la structure interne des données du LVF (section 3), puis nous concluons en comparant les résultats obtenus au classement de Kailuwet 2007 (section 4).

2. Préparation des données et sélection des individus

L'étude portant sur les verbes de la classe P, nous avons limité l'échantillon aux 2069 lexèmes verbaux non défectifs de cette classe. Chacun de ces verbes est défini par cinq variables.

La variable *Sc* comporte 16 modalités nommées en fonction de la sous-classe du lexème (cf. section 1.1). Les modalités sont (le nombre d'individus concernés figure entre parenthèses après la glose) :

P1a « être dans tel état psychologique ; avoir tel sentiment » (211) ; *P1b* « manifester physiquement [un] sentiment de joie, de douleur » (34) ; *P1c* « avoir telle

¹ « La distinction entre un objet humain et non-humain ne me paraît pas pertinente, étant donné que le CORRELAT est toujours un état de chose. On n'aime, ne craint, ne déteste, etc. pas un objet ou une personne en soi, mais en tant que porteur de certaines propriétés. » (Kailuwet, 2007 : 37).

² Il doit être clair que nous ne cherchons pas ici à apporter une définition de la classe de verbes étudiée.

³ Les calculs ont été effectués à l'aide du logiciel libre R (R core development team, 2012). Les calculs spécifiques à l'analyse factorielle et les graphiques ont été réalisés à l'aide du paquet FactoMineR (Lê et al., 2008) légèrement modifié (cf. note 4).

manifestation physique due à tel sentiment ; avoir tel sentiment pour/contre » (212) ; *P1d* « avoir tel sentiment portant sur qc ou poussant à qc » (70) ; *P1e* « avoir telle activité mentale » (178) ; *P1f* « avoir telle activité mentale dirigée sur tel objet » (71) ; *P1g* « prendre qc comme objet d'activité mentale » (89) ; *P1h* « avoir l'esprit tendu vers » (97) ; *P1i* « avoir tel sentiment portant sur qc » (102) ; *P1j* « avoir tel sentiment à l'égard de qn » (91) ; *P2a* « induire tel sentiment chez qn » (438) ; *P2b* « donner tel sentiment à qn » (18) ; *P2c* « augmenter, diminuer, remplir tel sentiment chez qn » (80) ; *P3a* « diriger son esprit, son activité réflexive sur qn » (71) ; *P3b* « diriger son esprit, son activité réflexive sur qc » (135) ; *P3c* « avoir ou ne pas avoir la connaissance de qc, de qn » (172).

Chaque lexème est également défini par ses propriétés syntaxiques. Si l'on s'en tient aux données strictement disponibles dans le LVF, on peut déterminer facilement pour chaque lexème s'il autorise la construction transitive directe (ici, variable **Td**), transitive indirecte (**Ti**), strictement intransitive (**I**) et pronominale (**P**). La difficulté pour comparer l'analyse du linguiste aux données du LVF est que ce dernier ne contient aucune information explicite à propos des quatre rôles de Kailuwet. Il est toutefois possible de *reconstruire* certains d'entre eux de manière compositionnelle sur la base des résultats d'un traitement automatique. L'analyse automatique de la définition semi-formalisée par (Bédaride, 2012) vérifie ainsi la correspondance entre les rôles sémantiques et les fonctions syntaxiques. Cela permet d'observer si le sujet de la construction transitive et celui de la construction pronominale sont agents ou patients. On départit ainsi les sujets CAUSEURS des sujets EXPERIENTS et PATIENTS (après un examen qualitatif pour confirmer les observations). Pour ce petit nombre de verbes (moins de 10 % de l'ouvrage), nous avons pu distinguer :

Td (construction transitive directe, 3 modalités) : \emptyset « construction absente » ; 1 « fonction de sujet alignée sur le premier rôle de la construction biactancielle (généralement AGENT en français) » ; 2 « sujet aligné sur le second actant (généralement PATIENT ou EXPERIENT en français) ».

P (construction pronominale, 3 modalités) : id.

Pour les deux autres variables syntaxiques aucune alternance fiable n'est dégagée par le traitement automatique. Les modalités sont donc limitées à :

Ti (construction transitive indirecte, 2 modalités) : \emptyset « construction absente » ; 1 « construction présente ».

I (construction intransitive, 2 modalités) : id.

Ainsi, chacun des 2069 individus sélectionnés est défini par une ligne comportant les valeurs de chaque variable ; par exemple :

Lemme		Td	P	I	Ti	Sc
abandonner 12(s)		\emptyset	1	\emptyset	\emptyset	P1h
abasourdir 02		1	\emptyset	\emptyset	\emptyset	P2a

3. Analyse des données

Après un examen des effectifs liés aux classes du LVF (section 3.1), nous soumettons les données à l'ACM (section 3.2) et à une classification hiérarchique ascendante (section 3.3).

3.1. Premier examen des effectifs

Selon les modalités mobilisées, les 2069 verbes se répartissent théoriquement en 576 groupes différents ($16 \times 3 \times 3 \times 2 \times 2$ modalités), dont seulement 55 sont attestés. On peut se poser d'emblée la question de l'homogénéité des groupes relevés par Kailuwet (section 1.2) et évaluer la systématisme des associations entre les modalités de **Sc** et celles des variables syntaxiques. Le tableau suivant reprend les 55 groupes (tri selon **Sc**) :

Td	P	I	Ti	Sc	Freq	Td	P	I	Ti	Sc	Freq	Td	P	I	Ti	Sc	Freq	Td	P	I	Ti	Sc	Freq
2	1	1	∅	P1a	9	∅	∅	∅	1	P1d	45	1	2	∅	∅	P1g	12	∅	1	1	∅	P2b	1
∅	1	1	∅	P1a	3	2	1	∅	∅	P1d	5	1	∅	∅	∅	P1g	56	∅	∅	∅	1	P2c	2
2	∅	1	∅	P1a	26	∅	1	∅	∅	P1d	17	∅	∅	∅	1	P1h	14	1	2	∅	∅	P2c	63
∅	∅	1	∅	P1a	62	2	∅	1	∅	P1e	64	1	1	∅	∅	P1h	7	1	∅	∅	∅	P2c	15
2	1	∅	∅	P1a	83	∅	∅	1	∅	P1e	63	∅	1	∅	∅	P1h	51	1	1	∅	∅	P3a	3
∅	1	∅	∅	P1a	28	2	1	∅	∅	P1e	10	1	∅	∅	∅	P1h	25	∅	1	∅	∅	P3a	2
∅	∅	1	∅	P1b	16	∅	1	∅	∅	P1e	41	1	∅	∅	∅	P1i	102	1	∅	∅	∅	P3a	66
∅	1	∅	∅	P1b	18	∅	∅	1	1	P1f	8	1	1	∅	∅	P1j	10	1	1	∅	∅	P3b	1
2	∅	∅	1	P1c	2	∅	∅	∅	1	P1f	27	∅	1	∅	∅	P1j	3	1	2	∅	∅	P3b	98
∅	∅	∅	1	P1c	17	∅	∅	1	∅	P1f	2	1	∅	∅	∅	P1j	78	1	∅	∅	∅	P3b	36
∅	∅	1	∅	P1c	47	2	1	∅	∅	P1f	3	1	1	∅	∅	P2a	41	1	1	∅	∅	P3c	4
2	1	∅	∅	P1c	75	∅	1	∅	∅	P1f	28	1	∅	∅	∅	P2a	397	1	2	∅	∅	P3c	76
∅	1	∅	∅	P1c	71	1	∅	∅	∅	P1f	3	∅	1	∅	1	P2b	4	1	∅	∅	∅	P3c	92
1	∅	∅	1	P1d	3	∅	1	∅	∅	P1g	21	∅	∅	∅	1	P2b	13						

Dans certains cas, l'association entre les classes est effectivement transparente : la classe P2a comprend bien 438 tous transitifs ; seuls 3/193 verbes des classes P1i et P1j (très homogènes également) rejettent la construction transitive directe et seul 1/18 verbe de la classe P2b rejette la construction indirecte. Par contre, les verbes des autres classes identifiées n'ont pas la même homogénéité : pour P1c, 193/212 verbes rejettent la construction indirecte) et pour P1a, 118/211 verbes acceptent une construction transitive et 28/211 verbes sont strictement pronominaux. Le classement du LVF n'est donc pas entièrement déterminé par la syntaxe des arguments : il repose également sur la définition du prédicat, qui ne peut être réduit à sa valence. Ce premier examen montre en outre que les classes du LVF sont plus ou moins homogènes (P1i, P2a, P2b, P3a) ou hétérogènes (en particulier P1a, P1c, P1f) par rapport aux variables observées, mais que la tendance est tout de même à l'homogénéité de manière générale (moins de 10 % des combinaisons étant représentées).

3.2. Analyse factorielle des correspondances multiples

Les trois premiers facteurs de l'analyse factorielle des correspondances multiples sur la base des variables sélectionnées cumulent 94,1 % de l'inertie ajustée⁴. Nous nous limiterons donc à l'examen des deux premiers plans factoriels (dimensions 1 à 3). En suivant (Le Roux et Rouanet, 2010 : 52), les modalités nommées sur les graphes sont celles dont la contribution est supérieure ou égale à la contribution moyenne du facteur (soit 3.85 %) auquel elles contribuent (les autres modalités sont considérées comme étant dispensables sur le plan et sont représentées en filigrane).

3.2.1. Plan 1-2

Plan factoriel, dimensions 1 et 2

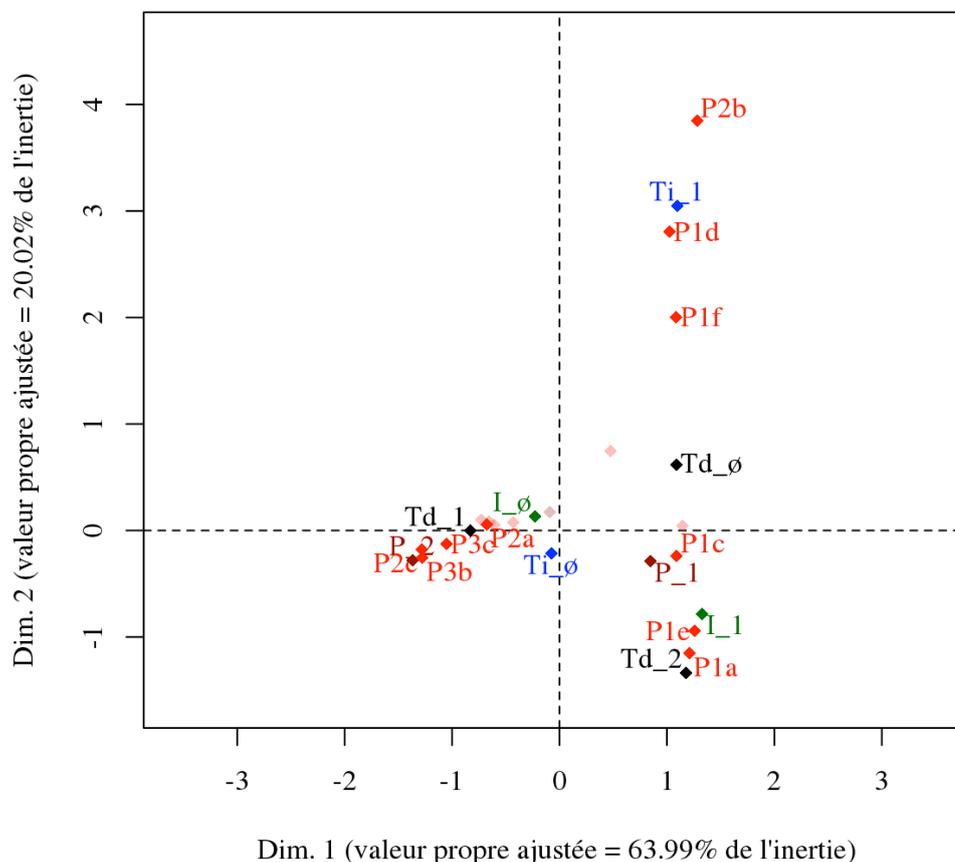


Figure 1

Le premier axe du plan croisant les deux premières dimensions (fig. 1) montre un groupe à gauche, qui comprend les verbes rejetant la construction intransitive (I_{\emptyset}), où le sujet de la construction transitive correspond au premier argument sémantique (Td_1) est généralement associé à une construction pronominale où le sujet correspond au deuxième argument (P_2) ;

⁴ Soit 63.99 + 20.02 + 9.88 %. L'inertie des axes d'une ACM doit être ajustée (Benzécri, 1979). Nous calculons de pseudo-valeurs propres $\tau(\lambda)$ sur la base de chaque valeur propre λ du tableau disjonctif et du nombre de variables actives Q selon la formule $\tau(\lambda) = [Q/(Q-1)]^2 (\lambda - 1/Q)^2$ (Lebart et al., 1997 : 368).

par exemple (fréquence des profils extrêmes sur le premier axe ; l'avant-dernière colonne indique la coordonnée) :

Td	P	I	Ti	Sc	COOX	Freq
1	2	∅	∅	P2c	-1.0330	126
1	2	∅	∅	P3b	-1.0320	196
1	2	∅	∅	P3c	-0.9704	152
1	2	∅	∅	P1g	-0.8004	24
1	∅	∅	∅	P2c	-0.6839	30

Dans ce cas, le pronominal est clairement passif. L'examen qualitatif (lecture des articles concernés) confirme qu'il s'agit de verbes où le rôle d'EXPERIENT n'est pas assigné à un argument. Ainsi, pour *alimenter 04* (P2c), le LVF livre les exemples *On alimente sa haine de tels propos* et *Sa haine s'alimente de tels propos*. Le terme *haine* n'est pas un EXPERIENT, mais un PATIENT, alors que le sujet de la construction transitive assume le rôle de CAUSEUR. Les classes sémantiques associées à ce groupe sont P2c, P3b et P3c. P2a semble se rapprocher de ces classes à cause de **I_∅**.

Ce premier groupe s'oppose par contre nettement à un ensemble caractérisé par les constructions intransitives (**I_1**), transitives indirectes (**Ti_1**) et celles où le sujet de la construction transitive ne correspond pas au premier argument sémantique, contrairement au sujet de la construction pronominale (**Td_2, P_1**) ; les profils les plus extrêmes sont :

Td	P	I	Ti	Sc	COOX	Freq
∅	∅	1	∅	P1b	0.9265	32
∅	∅	1	∅	P1a	0.9441	124
∅	∅	1	∅	P1e	0.9571	126
2	∅	1	∅	P1a	0.9698	52
2	∅	1	∅	P1e	0.9828	128
∅	1	∅	1	P2b	1.1160	8
∅	1	1	∅	P1a	1.2010	6
∅	1	1	∅	P2b	1.2200	2
2	1	1	∅	P1a	1.2270	18
∅	∅	1	1	P1f	1.2310	16

Cogiter 02, croire 06, ..., rêver 04 alternent les constructions intransitives et transitives indirectes, alors qu'*angoisser (s), berlurer 02(s)*, ainsi que d'autres verbes non reportés dans la liste alternent les constructions intransitives et les constructions pronominales où le sujet, comme on peut le déduire du contraste avec le sujet de la construction transitive, est EXPERIENT. Cela apparaît clairement dans les exemples que le LVF donne pour *cogiter 02* et pour *angoisser (s)* : *On cogite à la question* et *On angoisse, s'angoisse, est angoissé de le voir/Cette prédiction angoisse P*.

Au total, le groupe situé à droite de l'axe est associé à six sous-classes des verbes P : P1a, P1c, P1d, P1e, P1f, et P2b. Comme cela est visible dans le tableau ci-dessus (section 3.1), ces

classes peuvent présenter un EXPERIENT sujet (exprimé par **Td_2** ; P2b, qui ne compte par ailleurs que 18 individus, fait exception). Dans ce cas, les constructions pronominales ne sont pas exactement passives, mais correspondent plutôt à une sorte de moyen.

On oppose donc une forte présence de **P_2** à gauche à la possibilité de **Td_2** à droite, ce qui suggère que la première dimension correspond au critère de l'alignement des fonctions syntaxiques sur les rôles sémantiques.

La deuxième dimension oppose principalement les verbes qui ont une coordonnée positive sur le premier axe. Ainsi, dans ce groupe, ceux des verbes qui acceptent la construction transitive indirecte (**Ti_1**) au détriment de la construction transitive à ceux qui manifestent une tendance à l'intransitivité ou à l'expression de l'EXPERIENT en position de sujet de la construction pronominale. Les premiers cinq profils les plus extrêmes de part et d'autre de l'origine sont :

Td	P	I	Ti	Sc	COOY	Freq
2	1	1	∅	P1a	-1.2120	18
2	∅	1	∅	P1a	-1.0640	52
2	∅	1	∅	P1e	-0.9960	128
2	1	∅	∅	P1a	-0.9184	166
2	1	∅	∅	P1e	-0.8505	20
∅	∅	∅	1	P1f	1.919	54
1	∅	∅	1	P1d	1.978	6
∅	∅	∅	1	P1d	2.177	90
∅	1	∅	1	P2b	2.363	8
∅	∅	∅	1	P2b	2.511	26

Des verbes comme *angoisser (s)*, *berlurer 02(s)* alternent des constructions transitives, pronominales et proprement intransitives, alors que des verbes comme *agréer 02* sont accompagnés d'un complément indirect (cf. l'exemple : *Cette proposition agréée au ministre*).

Comme cela est visible sur le premier plan, ces deux groupes sont clairement associés à des classes spécifiques : les classes P1d, P1f et P2b correspondent aux verbes transitifs indirects, alors que les classes P1a et P1e (et, dans une certaine mesure, P1c) sont davantage liés à une construction transitive permettant la présence d'un réfléchi avec des constructions inergatives. On remarque en outre que les constructions intransitives sont proches des constructions transitives du second type. En opposant de la sorte les constructions directes et intransitives aux constructions indirectes, la deuxième dimension est de nature proprement syntaxique. L'examen du plan dans son ensemble montre que les transitifs indirects sont les plus nettement distincts des pronominaux passifs.

3.2.2. Plan 2-3

Sur le deuxième axe du second plan (dimension 3, fig. 2), on voit clairement se dessiner une opposition entre les verbes permettant la construction pronominale (**P_1** et **P_2**, en haut) et ceux qui ne le permettent pas (**P_∅**, en bas). Comme le montrent les positions des modalités de **Td** sur cet axe et la faible contribution de **Td_1**, ce potentiel est plus fortement lié aux classes sémantiques concernées qu'à l'orientation de la transitivité. Toutefois, on observe

également une association entre l'intransitivité stricte et l'absence de construction pronominale.

T _d	P	I	Ti	Sc	COOY	Freq
∅	∅	1	∅	P1e	-0.6592	126
1	∅	∅	∅	P1i	-0.6405	204
2	∅	1	∅	P1e	-0.6243	128
1	∅	∅	∅	P3a	-0.5828	132
1	∅	∅	∅	P2a	-0.5716	2
1	∅	∅	∅	P2a	-0.5716	788
2	1	∅	∅	P1c	0.6655	150
1	1	∅	∅	P3b	0.7716	2
1	2	∅	∅	P3c	0.7747	152
1	2	∅	∅	P3b	1.1240	196
1	2	∅	∅	P2c	1.2050	126

Plan factoriel, dimensions 2 et 3

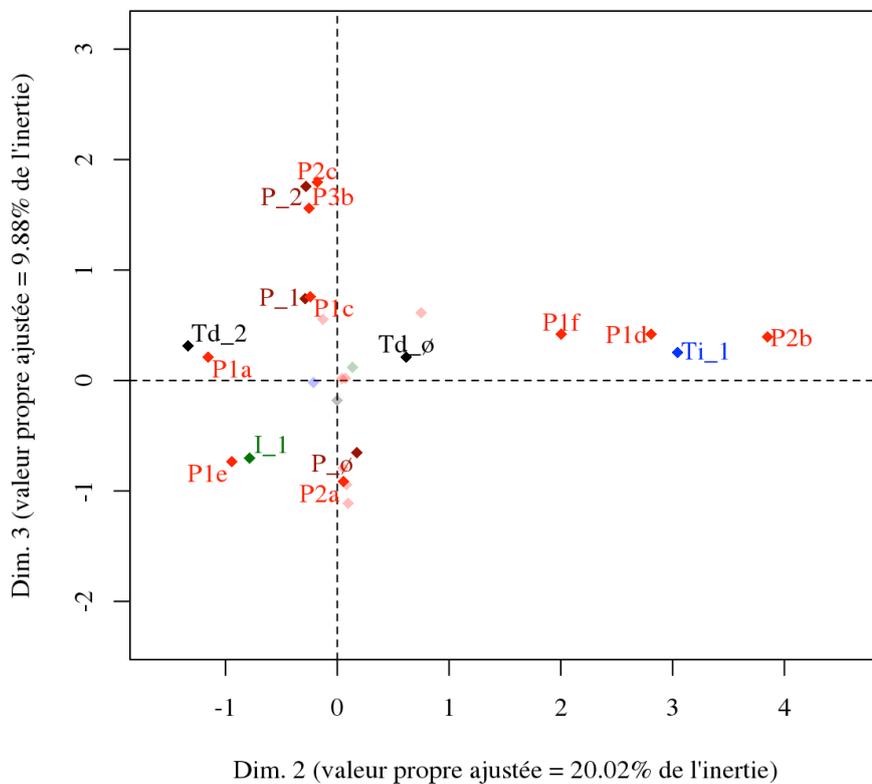


Figure 2

Comme on le voit, la classe P2c se détache de la classe P2a et se caractérise par l'absence d'EXPERIENT en tant qu'argument (association **Td_1**, **P_2**). Cela apparaît clairement dans les exemples du LVF : *Le journaliste anime un débat, On apaise une grève, une querelle/Le conflit s'apaise*. Dans ces phrases, ni la construction transitive, ni la construction pronominale n'actualisent un argument humain envisagé comme le siège d'un processus ou d'un état psychologique.

3.3. Classifications hiérarchiques ascendantes

L'application de la classification hiérarchique (CH) sur les individus révèle cinq grandes partitions non monothétiques (effectifs : 301, 877, 397, 151 et 343 individus). La tabulation entre les différentes modalités de **Sc** et les partitions donne le tableau suivant.

P/Sc	P1a	P1b	P1c	P1d	P1e	P1f	P1g	P1h	P1i	P1j	P2a	P2b	P2c	P3a	P3b	P3c
1	-	-	-	-	-	-	12	-	-	-	-	-	<u>78</u>	-	<u>135</u>	76
2	-	-	-	-	-	3	56	25	<u>102</u>	<u>88</u>	<u>438</u>	-	-	<u>69</u>	-	96
3	111	18	148	5	-	31	21	58	-	3	-	-	-	2	-	-
4	-	-	17	65	-	35	-	14	-	-	-	18	2	-	-	-
5	100	16	47	-	178	2	-	-	-	-	-	-	-	-	-	-

On voit dans le tableau qui suit que les modalités de **Sc** n'ont pas toutes la même homogénéité (décomptes des classes dont les individus sont tous ou presque tous dans la même partition soulignés) : P1e, P1i, P2a, P2b et P3b sont intégralement dans la même partition, P1d, P1j, P2c et P3a ne souffrent que quelques exceptions et les 7 autres modalités (sur 16) sont moins concentrées. En comparant ce classement au tableau de la section 3.1, on constate que l'hétérogénéité apparente de la classe P1e s'efface lors de la classification, alors que cette dernière fait apparaître certaines disparités passées sous silence par Kailuwet (P1a et P1c). Inversement, les cinq partitions principales contiennent toutes des individus issus de différentes classes sémantiques. La CH montre donc que les groupements se font d'une manière dépendante de la variable **Sc**, mais que d'autres facteurs interviennent.

Si nous travaillons sur les modalités des variables. Les branches du dendrogramme⁵ (fig. 3) correspondent à l'analyse factorielle et renforcent les conclusions précédentes. On remarque ainsi cinq grandes partitions qui correspondent respectivement aux partitions de la CH sur les individus) :

1. Le groupe à l'extrême gauche met en évidence l'association assez rapide entre les membres de la classe P2c et les membres des classes P3b et P3c autour de la modalité **P_2**, qui représente les constructions pronominales passives.
2. Ce premier groupe est proche des verbes ayant tendance à être exclusivement transitifs directs. À l'intérieur de ce deuxième groupe, qui contient évidemment P2a, les classes P1i et P1j sont rapidement associées à la classe P3a.
3. Le troisième groupe est celui des transitifs indirects.
4. Le quatrième groupe est celui qui rejette le plus la transitivité directe (**Td_0**) et privilégie les constructions pronominales moyennes.

⁵ Construit avec la fonction R *hclust* (méthode de Ward) sur le tableau des distances (tab. Burt) au carré.

5. Le cinquième groupe, à l'extrême droite, associe les constructions intransitives aux constructions transitives à sujet EXPERIENT.

Classification hiérarchique ascendante

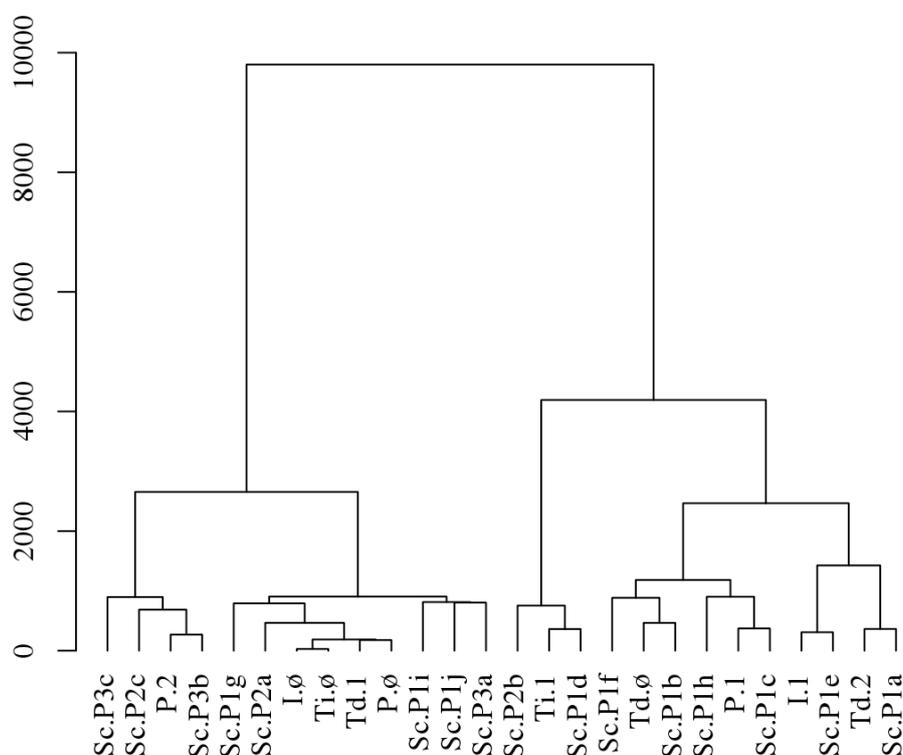


Figure 3

4. Conclusion

Les 16 classes initiales peuvent être regroupées comme suit. Les classes P3b, P3c et P2c représentent les verbes transitifs non proprement « de sentiment », caractérisés par un sujet CAUSEUR et un objet PATIENT (pas de réalisation de l'EXPERIENT). Kailuwet exclut la classe P2c et l'on observe qu'elle se distingue d'elle-même sur la base des données du LVF. Les classes P3b et P3c s'en rapprochent pour les mêmes raisons, mais leur exclusion n'est pas explicitée par le linguiste. Tout au plus signale-t-il : « Cependant, il faut se demander combien de verbes de la sous-classe P sont des verbes de sentiment au sens propre si on y applique des critères sémantiques plus rigides. » (Kailuwet, 2005 : 38). L'analyse des données fournit les arguments pour éliminer deux classes supplémentaires.

Les verbes transitifs indirects forment bel et bien deux groupes distincts suivant la fonction exprimant le rôle d'EXPERIENT : 1/ P1d et P2b si l'EXPERIENT est objet ; 2/ P1c si l'EXPERIENT est sujet. Cette distinction est également rendue possible par l'analyse de l'alignement des rôles sémantiques et des fonctions.

Les classes P1i et P1j sont regroupées *a priori* par Kaluwet et l'analyse des données révèle effectivement cette association. Il faut ajouter à cela : 1/ que la classe P2a n'en pas distinctement séparée par l'analyse, alors que le linguiste en fait une classe à part (la distinction repose sur l'opposition entre EXPERIENT objet ou sujet) ; 2/ la classe P3a est quant

à elle très proche de P1i et de P1j. Syntactiquement, donc, les sens « avoir un sentiment portant sur qn/qc » ou « voir telle connaissance portant sur qn/qc » se comportent similairement. On gagnerait à étudier mieux l'alignement de l'objet sur un rôle sémantique (LVF) et à proposer des critères distinguant sentiment et connaissance.

Ce dernier point est également visible dans l'association entre P1a et P1e. La proximité de ces classes correspond à la proximité syntaxique des constructions exprimant les sens « avoir une activité mentale » et « éprouver un sentiment ». Il faut affiner l'analyse de Kailuwet, qui se contente de dire que les verbes P1a sont intransitifs, alors qu'on voit que les classes sont aussi associées en fonction du sens de la construction pronominale.

Ainsi, certaines distinctions de Kaluweit émergent selon une approche plus empirique. Mais l'analyse de données montre que le classement est trop peu nuancé : les classes du LVF ne correspondent pas toujours à des patrons sémantico-syntaxiques stables. Cela invite à enrichir le LVF. En particulier, l'analyse des verbes trivalents est difficile à mener, et ni Kaluweit, ni nous-mêmes ne sommes actuellement en mesure de comparer le classement du LVF à celui du linguiste à cet égard.

Références

- Bédaride P. (2012). Raffinement du *Lexique des Verbes Français*. In *Actes de la conférence conjointe JEP-TALN-RECITAL 2012, volume 2 : TALN*. ATALA/AFCP : 155-168. [<http://www.aclweb.org/anthology/F/F12/F12-2012>]
- Benzécri J.-P. (1979). Sur le calcul des taux d'inertie dans l'analyse d'un questionnaire. *Cahiers d'analyse des données*, 3 : 377-378.
- Dubois J. et Dubois-Charlier F. (1997). *Les verbes français*. Larousse.
- François J., Le Pesant D. et Leeman D. (2007). Présentation de la classification des Verbes Français de Jean Dubois et Françoise Dubois-Charlier. *Langue française* 153 : 3-19.
- Lê S., Josse J. et Husson F. (2008). FactoMineR : An R Package for Multivariate Analysis. *Journal of Statistical Software* 25(1) : 1-18.
- Lebart L., Morineau A. et Piron M. (1998). *Statistique exploratoire multidimensionnelle*. Dunod.
- Le Roux B. et Rouanet H. (2010). *Multiple correspondence analysis*. Sage.
- Mazziotta N. et Martin F. (2013). An Exploratory Approach to Transitivity Morphemes in French. In Online Proceedings of the 5th Conference on Quantitative Investigations in Theoretical Linguistics : 54–57. [<http://www.ling.arts.kuleuven.be/QITL5/QITL5-proceedings.pdf>]
- Kailuwet R. (2005). *Linking : Syntax und Semantik französischer und italienischer Gefühlsverben*. Niemeyer.
- Kailuwet R. (2007). La classe P dans *Les Verbes français* et les verbes de sentiment . *Langue française* 153 : 33-39.
- Hadouche F. et Lapalme G. (2010). Une version électronique du LVF comparée avec d'autres ressources lexicales. *Langages* 179-180, 193-220.
- Tesnière L. (1959). *Éléments de syntaxe structurale*. Klincksieck.
- R Core Team (2012). *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.
- Van Den Eynde K. et Mertens P. (2010). *Le dictionnaire de valence DICOVALENCE: manuel d'utilisation (version 1.2)*. [http://bach.arts.kuleuven.be/dicovalence/manuel_100625.pdf]