

Université de Montréal

L'implémentation des relatives dans un réalisateur profond

Par

Charlotte Portenseigne

Département de linguistique et de traduction

Faculté des arts et des sciences

Mémoire présenté en vue de l'obtention du grade de maîtrise ès arts (M.A.)

en linguistique

Août 2021

© Charlotte Portenseigne, 2021

Université de Montréal

Département de linguistique et de traduction, Faculté des arts et des sciences

Ce mémoire intitulé

L'implémentation des relatives dans un réalisateur profond

Présenté par

Charlotte Portenseigne

A été évalué par un jury composé des personnes suivantes

Antoine Venant

Président-rapporteur

François Lareau

Directeur de recherche

Lyne Da Sylva

Membre du jury

Résumé

Ce mémoire porte sur l'implémentation des propositions relatives en français dans le réalisateur profond multilingue GenDR. Les réalisateurs de surface (SimpleNLG, JSReal ou RealPro) génèrent des propositions relatives, mais dans les réalisateurs profonds (MARQUIS, Forge ou GenDR) cette génération reste rudimentaire. Dans un corpus français de 21 461 phrases, 4505 contiennent une relative, soit environ une phrase sur cinq. Il s'agit donc d'un phénomène linguistique important que GenDR devrait couvrir.

Notre cadre théorique est la théorie Sens-Texte. Les propositions relatives se situent au niveau de l'interface sémantique-syntaxe. Nous présentons une typologie des propositions relatives. Nous définissons la relative et elle est divisée en deux grandes catégories : directe et indirecte. La définition des pronoms relatifs se base sur Riegel et al. (2018).

Nous avons utilisé GREW, afin d'analyser un corpus du français en SUD. Il y a plus de relatives directes ($\approx 78\%$) que d'indirectes ($\approx 22\%$). Les pronoms les plus fréquents sont *qui* (58,8 %), *que* (13,8 %), *dont* (10,2 %) et *où* (10 %), enfin viennent préposition suivie de *lequel* (5,7 %), préposition suivie de *qui* (0,7 %), *lequel* (0,4 %), préposition suivie de *quoi* (0,1 %). Le rôle syntaxique le plus fréquent du nom modifié est objet direct.

Puis, nous avons implémenté dans GenDR les règles pour la relative directe, la relative indirecte, et les pronoms relatifs *qui*, *que*, *dont*, préposition suivie de *qui* et préposition suivie de *lequel*. Notre implémentation couvre les types de relatives les plus communs en français. Les phénomènes qui nous résistent sont la génération des pronoms *lequel*, préposition suivie de *quoi*, *où* et *qui* objet, le traitement des verbes modaux et la génération des phrases avec un verbe à l'infinitif après un verbe modal, le traitement des verbes supports et autres collocatifs. Notre implémentation traite le français, mais peut être facilement adaptée à d'autres langues.

Mots clefs : proposition relative ; génération automatique de texte ; réalisateur de texte ; Théorie Sens-Texte

Abstract

This Master's thesis is about the implementation of French relative clauses in the multilingual deep realizer GenDR. Surface realizers (SimpleNLG, JSReal or RealPro) generate relative clauses, but in deep realizers (MARQUIS, Forge or GenDR) their handling remains rudimentary. In a French corpus of 21,461 sentences, 4,505 contain a relative, i.e. about one in five sentences. Thus, it is a core linguistic phenomenon that should be handled by GenDR.

Our theoretical framework is the Meaning-Text theory. Relative clause is relevant in the semantics-syntax interface. We offer a typology of relative clauses. The relative clause is defined, and it is divided into two main categories: direct and indirect. Our definition of relative pronouns is based on Riegel et al. (2018).

We used GREW to analyze a French corpus in SUD. There are more direct ($\approx 78\%$) than indirect ($\approx 22\%$) relatives. The most frequent pronouns are *qui* (58.8%), *que* (13.8%), *dont* (10.2%) and *où* (10%), then a preposition followed by *lequel* (5.7%), a preposition followed by *qui* (0.7%), *lequel* (0.4%), and a preposition followed by *quoi* (0.1%). The most frequent function of the modified noun is direct object.

We implemented in GenDR the rules for direct relative, indirect relative, and relative pronouns *qui*, *que*, *dont*, a preposition followed by *qui*, and a preposition followed by *lequel*. Our implementation covers the most common types of relatives. The phenomena that are not well handled by our rules are the generation of the pronouns *lequel*, a preposition followed by *quoi*, *où* and object *qui*, the treatment of modal verbs and the generation of sentences with an infinitive verb after a modal verb, the treatment of support verbs and other collocations. Our implementation is for French, but it can be easily adapted to other languages.

Keywords: relative clause; natural language generation; text realizer; Meaning-Text theory

Table des matières

Abstract	4
Table des matières	5
Liste des tableaux	12
Liste des figures	13
Liste des sigles et abréviations	18
Remerciements	19
Chapitre 1 — Introduction.....	20
1.1. Problématique	20
1.2. Objectifs du mémoire	22
1.3 Organisation du mémoire.....	22
Chapitre 2 – La théorie Sens-Texte et la génération de texte	24
2.1. La théorie Sens-Texte	24
2.1.1. Les principes Sens-Texte	24
2.1.2. La stratification du langage.....	25
2.1.3. L’interface sémantique-syntaxe	26
2.1.3.1. La représentation sémantique.....	26
2.1.3.2. La représentation syntaxique profonde	28
2.1.3.3. La représentation syntaxique de surface.....	30
2.2 La génération automatique de texte	31
2.2.1 Les réalisateurs linguistiques	32
2.2.2. Les propositions relatives dans les différents réalisateurs linguistiques	33
2.2.2.1. Le réalisateur SimpleNLG	34

2.2.2.2.	Le réalisateur JSreal	34
2.2.2.3.	Le réalisateur RealPro	35
2.2.2.4.	Le réalisateur MARQUIS	35
2.2.2.5.	Le réalisateur FORGe	35
2.3	Le projet GenDR.....	36
2.3.1	Au niveau sémantique	37
2.3.2	Au niveau de la syntaxe profonde	39
2.3.3	Au niveau de la syntaxe de surface	40
2.4	Synthèse	41
Chapitre 3 – Les propositions relatives		42
3.1	Point de vue traditionnel	42
3.1.1.	La relative sans antécédent	43
3.1.2.	La relative avec antécédent.....	44
3.1.2.1.	La relative restrictive	44
3.1.2.2.	La proposition relative explicative	45
3.2.	L’interface sémantique-syntaxe	46
3.2.1.	Au niveau sémantique	47
3.2.2.	Au niveau de la syntaxe profonde	50
3.2.3.	Au niveau de la syntaxe de surface	50
3.3.	Les pronoms relatifs	51
3.3.1.	Le pronom relatif <i>qui</i>	51
3.3.2.	Le pronom relatif <i>lequel</i>	52
3.3.3.	Le pronom relatif <i>que</i>	52
3.3.4.	Le pronom relatif <i>dont</i>	53

3.3.5.	Le pronom relatif <i>où</i>	54
3.3.6.	Une préposition suivie du pronom relatif <i>qui</i>	54
3.3.7.	Une préposition suivie du pronom relatif <i>quoi</i>	54
3.3.8.	Une préposition suivie du pronom relatif <i>lequel</i>	55
3.4.	La proposition relative.....	55
3.4.1.	Définition de la proposition relative	55
3.4.2.	La proposition relative directe.....	56
3.4.3.	La proposition relative indirecte.....	57
3.4.4.	Les spécificités de la proposition relative	58
3.4.5.	Exemples de propositions relatives directe et indirecte	59
3.5.	Synthèse	61
Chapitre 4 – Analyse de corpus		62
4.1.	Choix des corpus.....	62
4.1.1	SUD et les critiques de UD	63
4.1.2	Description des corpus	66
4.1.2.1	Le corpus GSD	66
4.1.2.2	Le corpus Sequoia	67
4.1.2.3	Le corpus ParTUT	68
4.1.2.4	Synthèse sur les corpus	68
4.2.	Collecte des données.....	69
4.2.1.	Patrons des pronoms.....	69
4.2.1.1.	Patron du pronom <i>qui</i> sujet.....	70
4.2.1.2.	Patron du pronom <i>lequel</i>	70
4.2.1.3.	Patron du pronom <i>que</i>	70

4.2.1.4.	Patron du pronom <i>dont</i>	71
4.2.1.5.	Patron du pronom <i>où</i>	71
4.2.1.6.	Patron d'une préposition suivie du pronom <i>qui</i>	71
4.2.1.7.	Patron d'une préposition suivie du pronom <i>quoi</i>	71
4.2.1.8.	Patron d'une préposition suivie du pronom <i>lequel</i>	72
4.2.2.	Patrons des propositions relatives	72
4.2.2.1.	Patron de la relative directe	72
4.2.2.2.	Patron des relatives indirectes	73
4.2.3.	Organisation des données	74
4.2	Analyse des données	74
4.2.1	Les parties du discours.....	75
4.2.1.1	Patron nom-verbe.....	75
4.2.1.2	Patron adverbe — verbe.....	76
4.2.1.3	Patron nom-nom	77
4.2.2	Analyse des pronoms.....	78
4.2.2.1	Les pronoms, fonctions et corpus.....	78
4.2.2.2	Les fonctions	79
4.2.2.3	Les pronoms sujets	80
4.2.2.4.	Les pronoms objets.....	81
4.2.2.5.	Les pronoms compléments prépositionnels.....	81
4.2.2.6.	Les pronoms par fonction	82
4.2.3.	Analyse du rôle syntaxique	83
4.2.3.1.	Le rôle syntaxique des pronoms dans le corpus	83
4.2.3.2.	Le rôle syntaxique du nom modifié par la proposition relative	84

4.2.4.	Analyse du type de proposition relative	85
4.2.4.1.	La fréquence des propositions relatives directes et indirectes	85
4.2.4.2.	Les pronoms relatifs en fonction du type de proposition relative.....	85
4.2.4.3.	Le rôle syntaxique du nom modifié par rapport au type de relative	86
4.2.4.4.	Synthèse de l'analyse de données du corpus	87
4.3	Synthèse	89
Chapitre 5 — Implémentation.....		91
5.1	Interface sémantique-syntaxe profonde	91
5.1.1	La règle de Dubinskaite.....	91
5.1.2.	La règle relative_direct	94
5.1.3.	La règle relative_indirect	96
5.1.4.	Modification de la règle des patrons de régime actant_gp.....	100
5.2	Interface syntaxe profonde-syntaxe de surface	102
5.2.1	La règle du pronom relatif <i>qui</i>	102
5.2.2	La règle du pronom relatif <i>que</i>	104
5.2.3	La règle du pronom relatif <i>dont</i>	107
5.2.4	La règle de la préposition suivie du pronom relatif <i>qui</i>	111
5.2.5	La règle d'une préposition suivie du pronom relatif <i>lequel</i>	115
5.3.	Implémentation du trait animacité	118
5.3.1.	Le dictionnaire Lexicon	118
5.3.2.	Au niveau des règles	119
5.3.3.	Les exceptions.....	120
5.4.	Synthèse	120
Chapitre 6 – Évaluation		121

6.1	Couverture.....	121
6.2.	Précision	121
6.2.1.	Avec un pronom relatif <i>qui</i> sujet	122
6.2.2.	Avec un pronom relatif <i>que</i> objet.....	124
6.2.3.	Avec un pronom relatif <i>dont</i> en complément du nom	125
6.2.4.	Avec un pronom relatif <i>dont</i> en oblique.....	127
6.2.5.	Avec une préposition suivie du pronom <i>qui</i>	128
6.2.6.	Avec une préposition suivie du pronom <i>lequel</i>	130
6.2.7.	Avec un pronom relatif <i>lequel</i>	131
6.2.8.	Avec un pronom relatif <i>où</i>	133
6.2.9.	Avec une préposition suivie du pronom <i>quoi</i>	134
6.2.10.	Avec un verbe modal	136
6.2.11.	Avec une fonction lexicale	138
6.2.12.	Synthèse	139
6.3.	Discussion	140
6.3.1.	Les pronoms relatifs	140
6.3.1.1.	Le pronom <i>lequel</i>	140
6.3.1.2.	Une préposition suivie du pronom <i>quoi</i>	141
6.3.1.3.	Le pronom <i>où</i>	141
6.3.2.	Les verbes modaux	142
6.3.3.	Les verbes supports	142
6.3.4.	Les fonctions lexicales	142
6.3.5.	Projectivité.....	143
6.4.	Synthèse	144

Chapitre 7 — Conclusion	145
Références bibliographiques	148

Liste des tableaux

Tableau 1	Tableau descriptif des représentations par niveau en TST.....	26
Tableau 2	Tableau des modules et étape par Reiter et Dale (2000, pp. 49, Figure 3.1)	32
Tableau 3	Les pronoms relatifs (Riegel, Pellat, & Rioul, 2018, p. 385)	51
Tableau 4	Nombre de pronoms relatifs par fonction et par corpus.....	74
Tableau 5	Nombre et pourcentages des pronoms par fonction dans le corpus	88
Tableau 6	Tableau récapitulatif des relatives générées	139
Tableau 7	Nombres et pourcentages de génération des propositions relatives.....	140

Liste des figures

Figure 1	Schéma des strates linguistiques de Mel'čuk (1997, p. 8)	25
Figure 2	RSém de la phrase (3)	28
Figure 3	RSyntP de la phrase (3)	30
Figure 4	RSyntS de la phrase (3)	30
Figure 5	RSém de la phrase (7) dans GenDR (forme textuelle)	39
Figure 6	RSém de la phrase (7) dans GenDR (forme graphique)	39
Figure 7	RSyntP de la phrase (7) dans GenDR	39
Figure 8	RSyntS de la phrase (7) dans GenDR.....	40
Figure 9	Focus sémantique (Mel'čuk, 2001, p. 116)	48
Figure 10	Ex. de thème et de rhème secondaires (Mel'čuk, 2001, p. 119)	48
Figure 11	Exemple d'arrière-plan (Mel'čuk, 2001, p. 202)	49
Figure 12	Exemple focalisé (Mel'čuk, 2001, p. 193)	49
Figure 13	RSém de la phrase (39)	57
Figure 14	RSyntP de la phrase (39)	57
Figure 15	RSém de la phrase (40)	58
Figure 16	RSyntP de la phrase (40)	58
Figure 17	RSyntS de la phrase (39)	59
Figure 18	RSyntS de la phrase (40)	60
Figure 19	Annotation de la phrase (41) en UD venant de UD	64
Figure 20	Schéma de la phrase (41) en UD.....	64
Figure 21	Annotation de la phrase (41) en SUD venant de SUD.....	65
Figure 22	schéma de la phrase (41) en SUD	65
Figure 23	Annotation de la phrase (42) en SUD venant de SUD.....	75
Figure 24	Schéma de la phrase (42) en SUD	75
Figure 25	Annotation de la phrase (43) en SUD venant de SUD.....	76
Figure 26	Schéma de la phrase (43) en SUD	77
Figure 27	Annotation de la phrase (44) en SUD venant de SUD.....	78

Figure 28	Schéma de la phrase (44) en SUD	78
Figure 29	Occurrence des pronoms par rôle syntaxique et corpus	79
Figure 30	Occurrence des rôles syntaxiques des pronoms	80
Figure 31	Occurrence des pronoms sujets	80
Figure 32	Occurrence des pronoms compléments prépositionnels	82
Figure 33	Occurrence des pronoms par rôle syntaxique	82
Figure 34	Occurrence des rôles syntaxiques des pronoms	83
Figure 35	Occurrence du rôle syntaxique du nom modifié par la relative	84
Figure 36	Proportion des relatives directes et indirectes	85
Figure 37	Occurrence des pronoms en fonction de la relative	86
Figure 38	Relation des noms en fonction des relatives directs et indirects	87
Figure 39	Annotation de l'exemple (45)	89
Figure 40	La règle des relatives de Dubinskaite	92
Figure 41	Schéma de la règle des relatives de Dubinskaite	92
Figure 42	RSém de la phrase (46)	93
Figure 43	RSyntP de la phrase (46) avec la règle de Dubinskaite	93
Figure 44	RSyntS de la phrase (46) avec la règle de Dubinskaite	93
Figure 45	Règle <code>relative_direct</code>	95
Figure 46	Schéma de la règle <code>relative_direct</code>	95
Figure 47	RSyntP de la phrase (46) avec la règle <code>relative_direct</code>	96
Figure 48	RSém de la phrase (47)	97
Figure 49	RSyntP de la phrase (47) avec la règle de Dubinskaite	97
Figure 50	Règle <code>relative_indirect</code>	98
Figure 51	Schéma de la règle <code>relative_indirect</code>	98
Figure 52	RSyntP de la phrase (47) avec la règle <code>relative_indirect</code>	99
Figure 53	Schéma de la règle de Dubinskaite	99
Figure 54	Schémas des règles relatives directes (à gauche) et indirectes (à droite)	100
Figure 55	Règle <code>actant_gp v.1.0</code>	101
Figure 56	Règle <code>actant_gp v.2.0</code>	101

Figure 57	Règle du pronom <i>qui</i>	103
Figure 58	Schéma de la règle du pronom <i>qui</i>	103
Figure 59	RSyntS de la phrase (46) avec le pronom <i>qui</i>	104
Figure 60	RSyntS de la phrase (48) avec la règle des relatives de Dubinskaite	105
Figure 61	Règle du pronom <i>que</i>	105
Figure 62	Schéma de la règle du pronom <i>que</i>	106
Figure 63	RSyntP de la phrase (48) avec le pronom <i>que</i>	106
Figure 64	RSyntS de la phrase (48) avec le pronom <i>que</i>	107
Figure 65	RSyntP de la phrase (49) avec la règle de Dubinskaite	107
Figure 66	RSyntS de la phrase (49) avec la règle de Dubinskaite	108
Figure 67	RSyntP de la phrase (50) avec la règle de Dubinskaite	108
Figure 68	RSyntS de la phrase (50) avec la règle de Dubinskaite	108
Figure 69	Règle du pronom relatif <i>dont</i>	109
Figure 70	Schéma de la règle du pronom relatif <i>dont</i>	109
Figure 71	RSyntP de la phrase (49) avec le <i>dont</i> oblique	110
Figure 72	RSyntP de la phrase (50) avec le <i>dont</i> complément du nom	110
Figure 73	RSyntS de la phrase (49) avec le <i>dont</i> oblique	110
Figure 74	RSyntS de la phrase (50) avec <i>dont</i> complément du nom	111
Figure 75	RSyntP de la phrase (51) avec la règle de Dubinskaite	112
Figure 76	RSyntS de la phrase (51) avec la règle de Dubinskaite	112
Figure 77	Règle de la préposition suivie du pronom relatif <i>qui</i>	113
Figure 78	Schéma de la règle de la préposition suivie du pronom relatif <i>qui</i>	113
Figure 79	RSyntP de la phrase (51)	114
Figure 80	RSyntS de la phrase (51)	114
Figure 81	RSyntP de la phrase (52) avec la règle de Dubinskaite	115
Figure 82	RSyntP de la phrase (52) avec la règle de Dubinskaite	115
Figure 83	Règle de la préposition suivie du pronom relatif <i>lequel</i>	116
Figure 84	Schéma de la règle de la préposition suivie du pronom <i>lequel</i>	116
Figure 85	RSyntP de la phrase (52) préposition + <i>lequel</i>	117

Figure 86	RSyntS de la phrase (52) avec préposition + <i>lequel</i>	117
Figure 87	Input de la phrase (58) dans GenDR pour le RSém.....	123
Figure 88	RSém de la phrase (58)	123
Figure 89	RSyntP de la phrase (58)	123
Figure 90	RSyntS de la phrase (58)	123
Figure 91	Input de la phrase dans GenDR pour la RSém	124
Figure 92	RSém de la phrase (59)	125
Figure 93	RSyntP de la phrase (59)	125
Figure 94	RSyntS de la phrase (59)	125
Figure 95	Input de la phrase (60) dans GenDR pour la RSém.....	126
Figure 96	RSém de la phrase (60)	126
Figure 97	RSyntP de la phrase (60)	126
Figure 98	RSyntS de la phrase (60)	127
Figure 99	Input de la phrase (61) dans GenDR pour la RSém.....	127
Figure 100	RSém de la phrase (61)	128
Figure 101	RSyntP de la phrase(61)	128
Figure 102	RSyntS de la phrase(61)	128
Figure 103	Input de la phrase (62) dans GenDR pour la RSém.....	129
Figure 104	RSém de la phrase (62)	129
Figure 105	RSyntP de la phrase(62)	129
Figure 106	RSyntS de la phrase (62)	129
Figure 107	Input de la phrase dans GenDR pour la RSém	130
Figure 108	RSém de la phrase (63)	131
Figure 109	RSyntP de la phrase (63)	131
Figure 110	RSyntS de la phrase(63)	131
Figure 111	Input de la phrase dans GenDR dans la RSém	132
Figure 112	Sém de la phrase (64)	132
Figure 113	SyntP de la phrase (64)	132
Figure 114	Synt de la phrase (64)	132

Figure 115	Input de la phrase dans GenDR dans la RSém	133
Figure 116	Sém de la phrase (65)	133
Figure 117	SyntP de la phrase (65)	134
Figure 118	SyntS de la phrase (65)	134
Figure 119	Input de la phrase dans GenDR dans la RSém	135
Figure 120	Sem de la phrase (66)	135
Figure 121	SyntP de la phrase (66)	136
Figure 122	SyntS de la phrase (66)	136
Figure 123	Input de la phrase dans GenDR dans la RSém	137
Figure 124	Sém de la phrase (67)	137
Figure 125	SyntP de la phrase (67)	137
Figure 126	SyntS de la phrase (67)	137
Figure 127	Input de la phrase dans la GenDR pour la RSém	138
Figure 128	Sém de la phrase (68)	138
Figure 129	SyntP de la phrase (68)	139
Figure 130	SyntS de la phrase (68)	139

Liste des sigles et abréviations

GAT : génération automatique de texte

RhSém : rhème sémantique

RSém : représentation sémantique

RSyntP : représentation syntaxique profonde

RSyntS : représentation syntaxique de surface

SSém : structure sémantique

SSém-Com : structure sémantico-communicative

SSyntP : structure syntaxique profonde

SSyntS : structure syntaxique de surface

SUD : Surface-Syntactic Universal Dependencies

ThSém : thème sémantique

TST : Théorie Sens-Texte

UD : Universal Dependencies

Remerciements

Je voudrais remercier mon directeur de recherche François Lareau.

J'aimerais aussi remercier ma mère qui m'a permis d'étudier à l'étranger.

Je souhaiterais remercier mes camarades en maîtrise de linguiste, les personnes de l'OLST, Linna He et Nadjat.

Pour finir, merci à mes amis de France Garance Richard-Badier, Margot Caffin-Malboeuf, Mohamed Tebah et d'ailleurs Justin, Artemis et Geun-Seok Lim.

Chapitre 1 — Introduction

1.1. Problématique

Ce mémoire s’inscrit en linguistique computationnelle, et plus spécifiquement en génération automatique de texte (GAT). En GAT, il y a deux grandes catégories de réalisateur linguistique : les réalisateurs de surface (par exemple, JREAL, SimpleNLG et RealPro) qui sont les plus connus, et les réalisateurs profonds (MARQUIS, FORGe, GenDR), qui traitent de l’interface sémantique – syntaxe. Le sujet de ce mémoire est l’implémentation des propositions relatives dans le réalisateur profond GenDR (Lareau et al., 2018). Notre cadre théorique linguistique est la Théorie Sens-Texte (TST) (Mel'čuk, 1997; Polguère, 1998; Milićević, 2006).

GenDR est un réalisateur profond qui couvre des phénomènes linguistiques comme la lexicalisation simple et complexe (Lambrey, 2017; Dubinskaite, 2017; Zhao, 2018), la complémentation par exemple les régimes verbaux à partir de VerbNet en anglais (Garreta-Piquette, 2018) et en mandarin (He, 2020), ou la modification, dans plusieurs langues telles que le français, l’anglais et le mandarin, dans l’interface sémantique-syntaxe. Le cadre théorique de ce réalisateur est la TST.

La proposition relative est une proposition qui modifie un nom¹. Elle est introduite par un pronom relatif. Ce pronom relatif est coréférent avec le nom modifié par la proposition relative. Traditionnellement, il y a deux types de propositions relatives : la proposition relative restrictive et la proposition relative explicative. Voici deux exemples de propositions relatives.

(1) *Le voisin joue avec la fille qui porte une robe verte.*

Qui porte une robe verte est une proposition relative explicative qui modifie le nom *fille*. Le pronom relatif qui introduit la proposition relative est *qui* et ce pronom est coréférent au nom *fille*.

¹ Nom est employé dans un sens générique il peut s’agir de nom, de pronom, mais aussi d’adverbe

(2) *Les chats qui sont noirs portent malheur.*

Qui sont noirs est une proposition relative restrictive qui modifie le nom *chats*. Le pronom relatif qui introduit la proposition relative est *qui* et ce pronom est coréférent au nom *chats*. Le chapitre 3 présente plus en détail les différentes structures des propositions relatives présentées.

Le générateur GenDR n'est pas encore suffisamment implémenter pour générer des propositions relatives avec une seule règle. Les propositions relatives sont classées comme directes ou indirectes dans ce mémoire pour des raisons techniques d'implémentation de GenDR et du logiciel de SUD, GREW. Cette distinction est basée sur le nombre d'éléments entre le coréférent du nom modifié dans la proposition relative et le verbe à la tête de la proposition relative. La proposition relative directe n'est qu'un cas particulier de la proposition relative indirect où n_0 est le nombre d'éléments entre le pronom relatif et le verbe à la tête de la proposition relative. Dans les deux exemples précédents (1) et (2) sont des relatives directes, car il n'y a aucun élément entre le coréférent du nom modifié, en syntaxe de surface c'est le pronom *qui* ici, et le verbe à la tête de la proposition relative *porter* pour l'exemple (1) et *être* pour l'exemple (2). Cette classification est purement technique et permet de créer plusieurs règles. GenDR peut ainsi générer les propositions relatives avec un ou plusieurs éléments entre le coréférent du nom modifié et le verbe de la proposition principale, et GREW peut détecter ainsi que spécifier le type de relatives avec le pronom relatif et le nom modifié. Cette classification est développée dans le chapitre 2.

Dubinskaite a fait une règle assez rudimentaire au niveau de la syntaxe profonde sur les propositions relatives. Cette règle se base sur le même principe qu'un nom a une relation de modification avec un adjectif et indique juste qu'un nom peut avoir une relation de modification avec un verbe. La proposition relative n'est pas représentée comme une proposition. Il n'y a pas de lien de coréférence entre le coréférent du nom modifié et le nom modifié dans la proposition principale. Dans l'implémentation de GenDR, le nom ne peut pas avoir une relation de modification. Dans le meilleur des cas, la proposition principale et la proposition relative sont vues comme deux phrases indépendantes et le nom coréférent du nom modifié ne se génère pas. De plus, des règles pour la génération du bon pronom relatif ont aussi été développées. Nous vous

présenterons des exemples illustrant les problèmes et l'efficacité de la règle de Dubiskaite2017 dans le chapitre 5.

Dans un corpus du français de 21 461 phrases, nous trouvons 4 505 propositions relatives, ce qui fait environ une phrase sur cinq contenant une proposition relative. La règle sur les propositions relatives se doit d'être améliorée. Les propositions relatives sont donc un phénomène linguistique important dans les langues, si nous partons du principe que le français n'est pas une langue particulière.

1.2. Objectifs du mémoire

Ce projet vise à définir la proposition relative, faire une typologie des propositions relatives, implémenter des règles sémantiques et syntaxiques profondes pour traiter les propositions relatives en français dans un réalisateur profond. Ainsi nos trois objectifs principaux sont :

- 1) Définir et faire une typologie des propositions relatives à partir des travaux des linguistes.
- 2) Créer de nouvelles règles dans l'interface sémantique-syntaxe profonde de GenDR qui rattachent la proposition principale à la proposition relative et qui prennent en compte la fonction de modification de la proposition relative.
- 3) Créer des règles dans l'interface syntaxe profonde-syntaxe de surface qui génèrent le bon pronom relatif en français.

1.3 Organisation du mémoire

Ce mémoire est composé de deux parties principales : une partie théorique et une partie pratique. D'une part, dans la partie théorique, il y a la mise en contexte avec la présentation du cadre théorique, de ce qui a été fait dans différents réalisateurs de texte et des différentes parties de GenDR (chapitre 2). Puis, il y a la typologie et la définition des propositions relatives (chapitre 3). D'autre part, la partie pratique est aussi divisée en plusieurs chapitres. Tout d'abord,

il y a une analyse des corpus afin de faire ressortir les types de relatives et les pronoms relatifs les plus fréquents (chapitre 4). Ensuite, l'implémentation des différentes règles de grammaire dans l'interface sémantique-syntaxe profonde et dans l'interface syntaxe profonde-syntaxe de surface (chapitre 5). Le dernier chapitre fait état de la couverture et des limites de notre implémentation dans GenDR (chapitre 6). Enfin, la conclusion résume ce mémoire (chapitre 7).

Chapitre 2 – La théorie Sens-Texte et la génération de texte

Ce mémoire s'inscrit dans le cadre théorique linguistique de la Théorie Sens-Texte (TST) et dans le domaine de la génération automatique de texte (GAT). Dans ce chapitre, la TST, la GAT et le projet GenDR seront présentés au regard de l'intérêt apporté pour les propositions relatives.

2.1. La théorie Sens-Texte

Le modèle Sens-Texte ou TST ou encore Meaning-Text Theory en anglais est une théorie linguistique formelle et descriptive qui peut aussi être qualifiée de fonctionnelle. Le modèle Sens-Texte a été développé par Igor Mel'čuk, Alexandre Žolkovskij et de Juri Apresjian dans les années 60 à Moscou, puis s'est développé en Russie, en Europe et au Canada (Mel'čuk I., 1973, 1988, 1997, 2001, 2012, 2015 ; Polguère, 1998 ; Milićević, 2006, 2007). C'est un modèle transformationnel et pas génératif. Cette théorie repose sur plusieurs principes. Le premier est que pour un sens donné, il y a plusieurs phrases possibles pour l'exprimer. Le deuxième est la stratification en sept niveaux de langue. De plus, la TST est utilisée pour le traitement automatique de la langue, la GAT, la traduction automatique, l'apprentissage de langues secondes ou maternelles, et la création de dictionnaires.

2.1.1. Les principes Sens-Texte

Dans un premier temps, le premier principe de la TST (Polguère, 1998 ; Milićević, 2006 ; Mel'čuk, 2001, 2013 ; Mel'čuk & Milićević, 2014, pp. 69-72) est que pour un sens, il y a plusieurs phrases possibles en fonction du contexte. Dans le Modèle Sens-Texte (Mel'čuk, 1988, pp. 43-91, 1997, 2012), il est expliqué que même s'il y a une transition du sens vers le texte et du texte vers le sens, il faut préférer la transition du sens vers le texte. Le choix d'une phrase plutôt qu'une autre se fait en fonction de la situation de communication.

2.1.2. La stratification du langage

Dans un deuxième temps, la stratification est un pilier important de la TST. En effet, il y a 4 strates linguistiques majeures dans la langue : sémantique, syntaxique, morphologique, phonologique. Chaque niveau est subdivisé en profond et surface, excepté le niveau sémantique. Au total, il y a sept niveaux : sémantique, syntaxe profonde, syntaxe de surface, morphologie profonde, morphologie de surface, phonologie de surface, phonologie de surface (Mel'čuk, 1988, p. 50, 1997, Polguère, 1998 ; Milićević, 2006 ; Mel'čuk & Milićević, 2014, pp. 73-77).

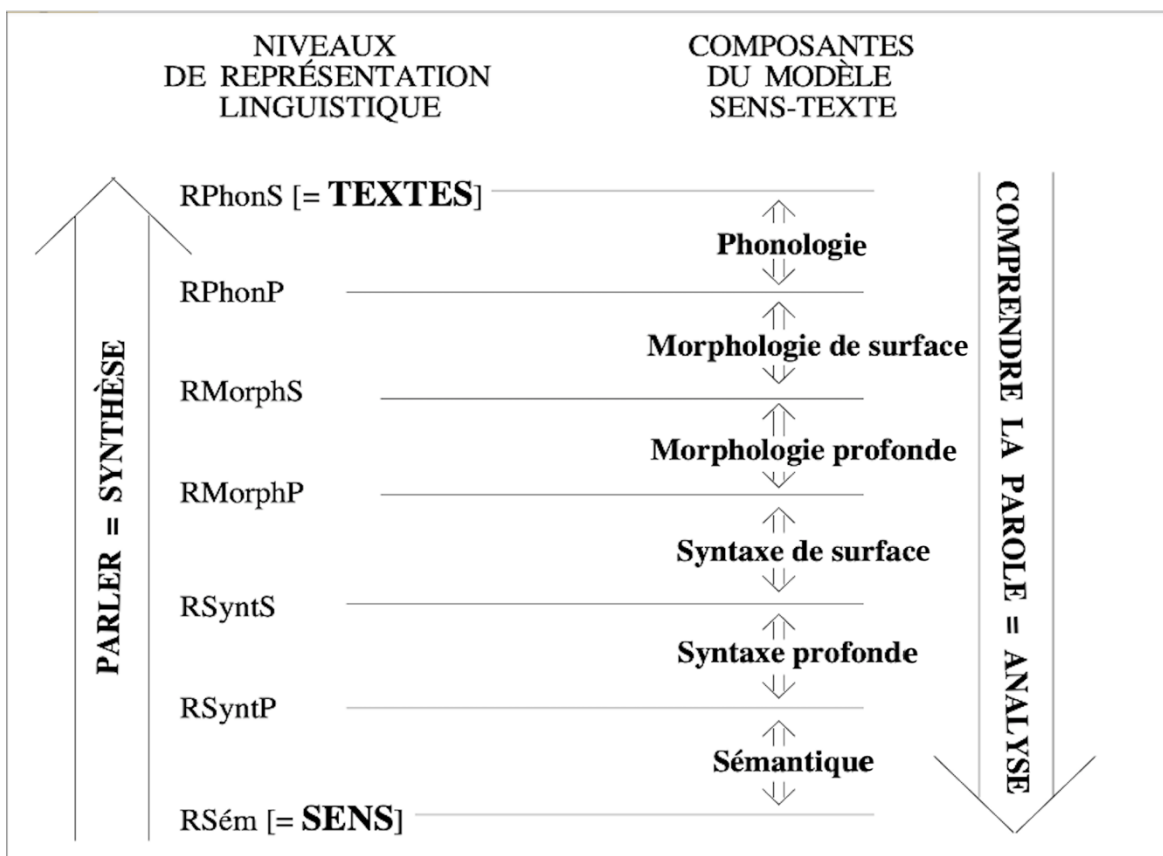


Figure 1 Schéma des strates linguistiques de Mel'čuk (1997, p. 8)

Pour chaque niveau, une représentation de ce niveau est faite pour le modéliser. Voici un tableau descriptif des représentations par niveau.

Niveau	Représentation	Références
Sémantique	Représentation sémantique (RSém)	(Mel'čuk I. , 1973, 1988, pp. 52-59, 1997, 2001, 2012, 2013, 2015; Polguère, 1998; Milićević, 2006; Mel'čuk & Milićević, 2014)
Syntaxe profonde	Représentation syntaxique profonde (RSyntP)	(Mel'čuk, 1973, 1988, pp. 59-67, 1997, 2013, 2015; Polguère, 1998; Mel'čuk & Milićević, 2014)
Syntaxe de surface	Représentation syntaxique de surface (RSynS)	(Mel'čuk, 1973, 1988, pp. 67-69, 1997, 2015; Polguère, 1998; Mel'čuk & Milićević, 2014)
Morphologie profonde	Représentation morphologique profonde	(Mel'čuk, 1973, 1988, pp. 23, 69-71, 1993, 1994, 1996, 1997, 2000, 2015; Polguère, 1998; Mel'čuk & Milićević, 2014)
Morphologie de surface	Représentation morphologique de surface	(Mel'čuk, 1973, 1988, pp. 71-72, 1993, 1994, 1996, 1997, 2000; Mel'čuk & Milićević, 2014)
Phonologie profonde	Représentation phonologique profonde	(Mel'čuk, 1997; Mel'čuk & Milićević, 2014)
Phonologie de surface	Représentation phonologique de surface	(Mel'čuk, 1997; Mel'čuk & Milićević, 2014)

Tableau 1 Tableau descriptif des représentations par niveau en TST

De plus, chaque représentation est constituée de plusieurs structures. Ce mémoire porte sur les propositions relatives et se situe dans l'interface sémantique-syntaxe ; seules les structures pertinentes pour notre travail seront expliquées plus en détail dans la suite de ce chapitre et dans le chapitre suivant sur les propositions relatives.

2.1.3. L'interface sémantique-syntaxe

Seule l'interface sémantique-syntaxe nous intéresse dans ce projet de recherche, autrement dit ce mémoire portera seulement sur les niveaux sémantique, syntaxique profond et syntaxique de surface.

2.1.3.1. La représentation sémantique

Dans un premier temps, il y a la représentation sémantique. D'après Mel'čuk (1997, pp. 9-11, 1988, pp. 52-59, 2012, 2013, 2015), Polguère (1998, pp. 7-9) et Milićević (2006, 2014) dans la

représentation sémantique, il y a quatre structures. La structure de base ou structure sémantique (SSém) est la structure principale de la représentation. La structure de base est le point de vue objectif de l'énoncé et la structure linguistique centrale. SSém est un réseau constitué de nœuds qui représentent des sémantèmes et d'arcs numérotés en fonction des relations actanciennes de chaque nœud. « L'unité sémantique de base, appelée sémantème est un sens lexical, c'est-à-dire le signifié d'une unité lexicale (=lexie de la langue) » (Mel'čuk & Milićević, 2014, p. 109). Il y a deux types de sémantèmes : les foncteurs, qui regroupent les prédicats, les quantificateurs et les connecteurs logiques, et les noms de classes d'objets.

Puis, il y a la structure sémantico-communicative (= SSém-Comm) (Mel'čuk, 2001, 2012), qui montre le point de vue communicatif et subjectif de l'énoncé. Cette structure permet à l'énonciateur de montrer son intention et sur quoi ou sur qu'il veut mettre l'accent dans son énoncé. De plus, « La SSém-Comm est une division de la SSém en sous-réseaux communicatifs, avec 1) l'indication, dans chaque sous-réseau, d'un nœud dominant et 2) un étiquetage de ces sous-réseaux par des valeurs d'oppositions communicatives, telles que thème vs rhème, donnée vs nouveau, etc. » (Mel'čuk & Milićević, 2014, p. 131). Il y a huit oppositions communicatives (Mel'čuk, 2001; Milićević, 2006): la thémativité (thème, rhème, spécificateur), Givenness (donné, nouveau), la focalisation (focalisé, non focalisé), la perspective (arrière-plan, avant-plan, neutre), l'emphase (accentué, neutre), l'assertivité (affirmé, présupposé), l'unicité (unitaire, articulé) et la locution (communiqué, signalé, réalisé).

Ensuite, il y a la structure rhétorique, qui contient les figures de style et les effets littéraires ou artistiques du locuteur tel que le niveau de langue différent, le pathétique ou l'ironie.

Enfin il y a la structure référentielle, qui « est un ensemble de pointeurs allant des configurations sémantiques vers leurs référents, entités et faits du monde réel. » (Mel'čuk & Milićević, 2014, p. 131).

Voici un exemple d'une phrase et de sa RSém ci-dessous. Dans les exemples de ce mémoire, les règles des propositions relatives que nous avons implémentés dans GenDR ont été utilisé pour illustrer les schémas de la RSém, de la RSyntP et de la RSyntP. La SSém-Comm, la structure

rhétorique et la structure référentielle ne sont pas implémentées dans GenDR et ne sont pas utilisées par la suite.

(3) *La commission suit la ligne de conduite qu'il a définie.* [Sequoia, Europar.550_00473]

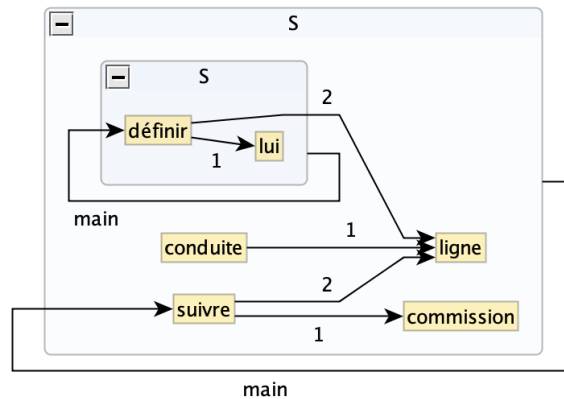


Figure 2 RSém de la phrase (3)

Les bulles S représentent les propositions et la relation main pointe vers le nœud dominant de chaque proposition. « Dans chaque sous-réseau communicatif, un sémantème est sélectionné par le locuteur en tant que communicativement dominant. C'est le sémantème auquel ce sous-réseau peut se réduire : c'est, pour ainsi dire, la paraphrase minimale du sous-réseau. » (Mel'čuk & Milićević, 2014, p. 142). Notre intérêt se porte sur la SSém et la SSém-Comm. Dans la partie 3.2.1, le rôle de la SSém-Comm sera abordé plus en détail pour la différenciation des propositions relatives restrictives et explicatives, ainsi que des propositions relatives et des propositions focalisées.

2.1.3.2. La représentation syntaxique profonde

Dans un deuxième temps, il y a la représentation syntaxique profonde. D'après Mel'čuk (1988, pp. 56-67, 1997, p. 11, 2013, 2015) Milićević (2014, p. 178), Polguère (1998, pp. 9-13), la RSyntP, contrairement à la RSém, représente la structure de la phrase. Les choix de lexicalisation et d'arborisation qui déterminent la phrase sont faits à ce niveau. Dans la RSyntP, il y a quatre

structures : la structure syntaxique profonde, la structure syntactico-communicative profonde, la structure syntactico-anaphorique profonde, la structure syntactico-prosodique profonde.

Tout d'abord, la structure syntaxique profonde (SSyntP) est la structure de base de la RSyntP. C'est dans cette structure que la structure de phrase se construit avec les lexies profondes, les grammèmes profonds et les relations universelles de dépendance syntaxique. « Une lexie profonde est soit une lexie véritable (sémantiquement) pleine, soit le nom d'une fonction lexicale, soit encore une lexie fictive. » (Mel'čuk & Milićević, 2014, p. 181). D'après Mel'čuk et Milićević (2014, p. 181), « une lexie pleine a un correspondant direct dans la structure sémantique : un sémantème ou une configuration de sémantème. Une telle lexie est soit un lexème soit une locution. » Et « Les grammèmes sémantiques, appelés aussi pleins, ou profonds, expriment les configurations de sémantèmes de la SSém initiale. En français, ce sont les grammèmes du nombre et de la détermination pour les noms et les grammèmes de la voix, du mode et du temps pour les verbes. » (Mel'čuk & Milićević, 2014, p. 185). La SSyntP est représentée par un arbre de dépendance. Les lexies profondes et les grammèmes profonds sont les nœuds et les relations universelles de dépendance syntaxique sont les branches.

Puis, la structure syntactico-communicative profonde est le reflet de la structure communicative du niveau sémantique.

Ensuite, la structure syntactico-anaphorique profonde est la structure où les coréférences sont indiquées, c'est-à-dire que les ellipses lexicales et la pronominalisation sont planifiées dans cette structure pour être effectuées dans le niveau de syntaxe de surface.

Enfin, la structure syntactico-prosodique profonde est la structure où il y a la prosodie induite sémantiquement, comme la prosodie neutre, emphatique, ironique, déclarative, interrogative, exclamative.

Voici la RSyntP pour la phrase (3).

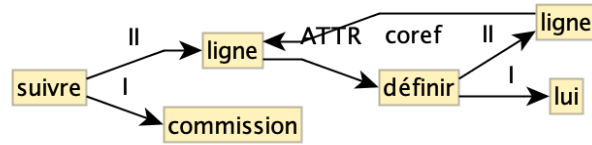


Figure 3 RSyntP de la phrase (3)

Nous avons les actants (I et II) des verbes, une relation de modification (ATTR) et une relation de coréférent (coref). *De conduite* n'apparaît pas dans la RSyntP car l'implémentation des verbes à l'infinitif comme actant n'est pas implémentée dans GenDR. Donc, la structure syntaxique profonde et la structure syntactico-anaphorique profonde nous intéressent pour les propositions relatives et leur implémentation dans un réalisateur profond.

2.1.3.3. La représentation syntaxique de surface

Finalement, d'après Mel'čuk (1988, pp. 65-66, 1997, pp. 12-13, 2015), Mel'čuk et Milićević (2014, pp. 65-66) et Polguère (1998, pp. 13-14), il y a quatre structures dans la représentation syntaxique de surface (RSyntS). Ces quatre structures sont les pendents des structures syntaxiques profondes. Donc, il y a la structure syntaxique de surface (SSyntS), la structure syntactico-communicative de surface, la structure syntactico-prosodique de surface et la structure syntactico-anaphorique de surface. La SSyntS, est comme la SSyntP, un arbre de dépendance, mais toutes les lexies de la phrase y sont présentes. C'est dans la structure syntactico-anaphorique qu'il y a la relation de coréférence entre le nom modifié et le pronom relatif par exemple.

Voici la RSyntS de la phrase (3).

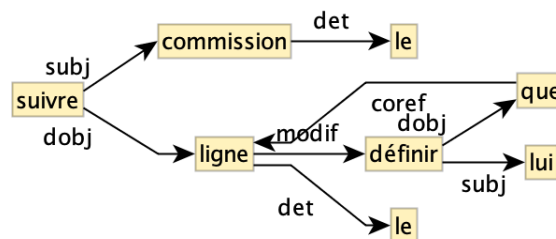


Figure 4 RSyntS de la phrase (3)

Nous avons une relation de modification (*modif*), une relation de coréférent (*coref*), des relations sujets (*subj*), objet direct (*dobj*) et déterminatif (*det*). La structure de base syntaxique de surface et la structure syntactico-anaphorique de surface, avec la relation de coréférence (*coref*), sont intéressantes pour les propositions relatives.

Pour conclure cette section, ce mémoire se situe dans le cadre théorique de la TST. Ce projet de recherche sur les propositions relatives se situe dans l'interface sémantique-syntaxe, et plus particulièrement, la structure de base sémantique, la structure sémantico-communicative, la structure de base syntaxique profonde, la structure syntactico-anaphorique profonde, la structure de base syntaxique de surface et la structure syntactico-anaphorique de surface.

2.2 La génération automatique de texte

La GAT est intéressante pour la linguistique parce qu'elle permet de tester les théories linguistiques de façon pratique (Danlos, 1983 ; Lareau, 2002). Les générateurs sont généralement composés de deux parties. D'une part il y a le *quoi-dire* (Danlos, 1983) ou *early processus* (Gatt & Krahmer, 2018), qui décide du contenu et de sa structure, évalue le contenu, décide de ce qu'il y aura en sortie et dans quel ordre. D'autre part, il y a le *comment-le-dire* (Danlos, 1983) ou *late process* (Gatt & Krahmer, 2018), qui est pour la décision du *quoi-dire*. Reiter et Dale (2000, p. 49) ont défini sept étapes des différents systèmes de GAT.

- 1) La sélection du contenu : est la sélection et la planification des informations pour le document de sortie.
- 2) La structuration du document : est l'assemblage et la structuration des différentes parties du contenu en termes rhétoriques.
- 3) La lexicalisation : est la sélection de la lexie pour exprimer le contenu sélectionné ultérieurement.
- 4) La génération des expressions de références : est la sélection des expressions pour se référer aux entités.

- 5) L'agrégation : est la planification de l'organisation des structures et de leur correspondance avec les structures linguistiques.
- 6) La réalisation linguistique : est la conversion des représentations abstraites des phrases en texte, en réalisation de surface.
- 7) La réalisation structurelle : est la conversion des structures abstraites comme des paragraphes, des sections, et des symboles de balisage en composant de présentation et de structuration du texte, en réalisation de surface.

Reiter et Dale (2000, pp. 49, Figure 3.1) ont fait un tableau récapitulatif des sept étapes des différents systèmes de GAT présentés ci-dessus.

Module	Étape pour le contenu	Étape de structuration
Planification du document	La sélection du contenu	La structuration du contenu
Micro planning	La lexicalisation	agrégation
	La génération des expressions de références	
Réalisation	Réalisation linguistique	Réalisation structurelle

Tableau 2 Tableau des modules et étape par Reiter et Dale (2000, pp. 49, Figure 3.1)

Dans la section suivante, différents réalisateurs linguistiques basés sur une grammaire et un dictionnaire seront présentés, ainsi que leur intérêt pour les propositions relatives.

2.2.1 Les réalisateurs linguistiques

Dans la génération de texte, il y a deux sortes de réalisateurs linguistiques : le réalisateur de surface et le réalisateur profond.

D'une part, le réalisateur de surface est plus proche du texte. Le réalisateur de surface est qualifié de surface, car il se situe sur les niveaux de syntaxe de surface, de morphologie profonde et de morphologie de surface au niveau de la TST. Autrement dit, le réalisateur de surface s'occupe principalement de la conjugaison, de l'accord sujet-verbe, et de la morphologie. Les réalisateurs

de surface peuvent s'appuyer sur d'autres modèles et théories. Les sept parties des différents systèmes de GAT décrit par Reiter et Dale (2000) sont au niveau de la réalisation de surface.

D'autre part, le réalisateur profond prend en compte le sens. Le réalisateur profond se place à un tout autre niveau, il se situe sur les niveaux sémantique, syntaxique profond et syntaxique de surface. Par exemple, des sémantèmes et un réseau sémantique sont implémentés dans un réalisateur profond et celui-ci génère une ou plusieurs phrases au niveau syntaxique profond puis au niveau de la syntaxe de surface. Plusieurs phrases peuvent être proposées en fonction du dictionnaire. Par exemple pour *médecin*, *médecin*, *docteur* et *toubib* ont le même sens et peuvent se générer. Les dictionnaires des réalisateurs profonds contiennent plus d'information en général pour traiter des niveaux sémantiques, syntaxique profond et syntaxique de surface que les réalisateurs de surface.

Pour conclure, les réalisateurs de surfaces et profonds sont complémentaires par rapport aux niveaux de la langue et leurs spécialisations sont pour des buts différents. Pour les réalisateurs de surfaces, la structure de la proposition relative doit être annotée avant comme relative afin d'être traité et généré par le réalisateur. Pour le réalisateur profond : le réalisateur doit identifier la structure puis la générer. La structure des propositions relatives sera traitée dans le chapitre 3. Dans la partie suivante, le traitement des propositions relatives en français par des réalisateurs linguistiques sera présenté.

2.2.2. Les propositions relatives dans les différents réalisateurs linguistiques

Dans cette partie, trois réalisateurs de surface (JREAL, SimpleNLG et RealPro) et deux réalisateurs profonds (MARQUIS, FORGe) seront présentés, ainsi que leur gestion des propositions relatives.

2.2.2.1. Le réalisateur SimpleNLG

SimpleNLG² (Gatt & Reiter, 2009) est un réalisateur de surface programmé en Java. Sa première version était en anglais, il existe maintenant dans plusieurs langues germaniques et romanes comme l'allemand (Bollmann, 2011), le français (Vaudry & Lapalme, 2013), le portugais brésiliens (De Oliveira & Sripada, 2014), l'italien (Mazzei, Battaglini, & Bosco, 2016) et l'espagnol (Ramos-Sato et al., 2017). L'entrée de ce réalisateur est une structure lexicalisée en XML. C'est un réalisateur basé sur des règles syntaxiques, des règles morphologiques et un dictionnaire, ce dernier indiquant les propriétés syntaxiques et morphologiques de chaque lexème.

Le réalisateur de surface SimpleNLG-EnFr (Vaudry & Lapalme, 2013) a un module pour traiter spécifiquement des propositions relatives en français qui n'a pas d'équivalent direct en anglais (Vaudry & Lapalme, 2013, p. 185). Une partie de la proposition relative est annotée de telle façon qu'elle est remplacée par le pronom relatif qui convient. S'il n'y a pas d'élément pour être remplacé par un pronom relatif, le système reconnaît la structure de la proposition relative et génère un pronom relatif. La fonction grammaticale, la préposition s'il y en a une et la partie de la proposition relative à remplacer par un pronom relatif sont les paramètres qui déterminent le choix du pronom relatif.

2.2.2.2. Le réalisateur JSreal

JSreal (JavaScript REALizer) (Daoust, 2013 ; Daoust & Lapalme, 2014) est un réalisateur de surface, c'est-à-dire qu'il s'occupe de la morphologie, des accords sujet-verbe et de la conjugaison. Ce réalisateur est fait pour le développement web. JSreal peut générer des expressions et des phrases en français et les mettre dans des documents HTML. Ce réalisateur peut aussi s'appliquer en JavaScript par le module node.js. JSreal génère des phrases à partir de règles syntaxiques, de règles morphologiques et d'un dictionnaire adapté SimpleNLG-EnFr. JSrealB (Molins & Lapalme, 2015; Lapalme, 2021) est une version bilingue anglais-français.

² <https://github.com/simplenlg/simplenlg>

Ce réalisateur peut aussi générer des propositions subordonnées, notamment des propositions relatives.

2.2.2.3. Le réalisateur RealPro

RealPro (Lavoie & Rambow, 1997) est un réalisateur de surface programmé en C++ et dans le cadre de la TST. L'entrée est une structure syntaxique profonde, c'est-à-dire que les lexèmes sont les nœuds et les relations universelles étiquettent les arcs. Chacune des transformations est traitée en module séparé. Il y a un dictionnaire qui encode les lexèmes sur les plans syntaxique et morphologique afin de faire fonctionner les différents modules. Il y a deux modules de règles grammaticales pour deux niveaux différents : de la syntaxe profonde à la syntaxe de surface et de la syntaxe de surface à la morphologie profonde. Le réalisateur de surface RealPro peut traiter des propositions subordonnées en anglais.

2.2.2.4. Le réalisateur MARQUIS

MARQUIS (Lareau & Wanner, 2007; Wanner et al., 2010) est un générateur de texte profond multilingue et le parent des réalisateurs profonds GenDR et FORGe. C'est un générateur de bulletins environnementaux. C'est un système de génération automatique de texte complet, mais seule la partie du réalisateur profond est mentionnée ici. Le modèle théorique de ce réalisateur est la TST, donc les niveaux sont stratifiés comme en TST (cf. Tableau 1).

Les articles ne font pas mention de génération des propositions relatives en français.

2.2.2.5. Le réalisateur FORGe

Forge (Mille & Wanner, 2017 ; Mille et al., 2017) est un réalisateur profond héritier de MARQUIS. Ce réalisateur est basé sur l'anglais et l'adaptation à d'autres langues comme l'allemand, le français, l'espagnol et le polonais est en cours. Comme pour le réalisateur MARQUIS, FORGe sont aussi dans le cadre de la TST et s'occupe de l'interface sémantique-syntaxe, c'est-à-dire que ce réalisateur génère d'abord RSèm à RSyntP, puis de RSyntP à RSyntS, enfin RSyntS au texte. Dans

un premier temps, l'algorithme traite de l'entrée lexicale et le réseau sémantique, puis des mots fonctionnels, enfin les structures sont linéarisées et les lexèmes sont générés par les règles syntaxiques et morphologiques afin d'obtenir le texte final.

FORGe a une couverture plus spécifique et donc moins flexible que GenDR. Il y a beaucoup de règles spécifiques. Il y a plusieurs règles pour générer les propositions relatives comme une règle générale sur les propositions relatives, une règle pour les propositions relatives restrictives, car la SSém-Comm est présente dans ce réalisateur, une règle générale pour introduire le pronom, puis une règle peut introduire les verbes supports et une autre règle pour générer un verbe support par défaut.

Pour conclure cette section, le générateur de surface SimpleNLG, JSReal et RealPro semblent très bien traiter les propositions relatives avec un module spécifique pour les propositions relatives, et le générateur FORGe peut générer des propositions relatives et il y a une différenciation entre les propositions relatives restrictives et explicatives. Il y a un manque de données pour savoir si le réalisateur profond comme MARQUIS, est capable de générer des propositions relatives en français.

2.3 Le projet GenDR

Dans cette partie, les différentes parties utilisées pour les propositions relatives seront expliquées à travers un exemple. GenDR (Lareau et al., 2018) est l'un des héritiers de MARQUIS. C'est un réalisateur profond multilingue avec un système qui traite l'anglais, le français, l'espagnol, le lituanien et le mandarin et basé sur la TST. Le transducteur de graphe est hérité de MARQUIS et s'appelle MATE (Bohnet, Langjahr, & Wanner, 2000; Bohnet & Wanner, 2010). GenDR a plusieurs branches dont celle pour les fonctions lexicales et les collocations (Lambrey & Lareau, 2015; Lambrey, 2017; Dubinskaite, 2017; Zhao, 2018) ou la branche de VerbNet en anglais (Garreta-Piquette, 2018) ou en mandarin (He, 2020), un mémoire est également en préparation sur le traitement des locutions (Dubé, à paraître). MATE est la plateforme informatique et GenDR est la grammaire et des dictionnaires.

Enfin, les différentes parties utilisées de GenDR seront présentées au fur et à mesure avec un exemple des étapes pour une phrase.

2.3.1 Au niveau sémantique

Dans un premier temps, nous entrons la RSém d'une phrase dans GenDR. Il est nécessaire chaque sémantème soit déjà dans le dictionnaire. Ce réalisateur fonctionne sur la combinaison d'un dictionnaire lexical, d'un dictionnaire sémantique et de règles de grammaire (Sem-SyntP et SyntP-SyntS). Les dictionnaires sont propres à chaque langue, mais pas la grammaire (bien que des règles puissent y être marquées comme spécifiques à une langue donnée).

Le dictionnaire lexical précise pour chaque lexie la partie du discours profonde, la partie du discours de surface, les collocations, les patrons de régime, la diathèse, les contraintes sur les actants et l'animacité, des lexies. Voici des exemples d'entrée du dictionnaire lexical. La partie du discours profonde est la partie du discours traditionnellement utilisé comme verbe, nom ou déterminant. La partie du discours de surface complète la partie du discours profond et donne plus d'information sur la lexie, par exemple verbe transitif ou nom commun. L'animacité est le trait animé ou inanimé de la lexie, ce qui est utile pour certains pronoms comme vue dans le chapitre 3.

(4) *La lexie combat*

```
combat : noun {  
  lf = {name=Func12 value=opposer}  
  lf = {name=Oper12 value=livrer}  
  lf = {name=Oper12 value=disputer}  
}
```

(5) *La lexie combattre*

```
combattre : verb_dt
```

(6) *La lexie combien*

```
combien : adverb
```

Nous avons la lexie *combat*, dont la partie du discours est nom, et les trois fonctions lexicales précisées sont pour des verbes supports : *opposer*, *livrer* et *disputer*. Puis, il y a la lexie *combattre*, dont la partie du discours est verbe. Enfin il y a la lexie *combien*, dont la partie du discours est adverbe. Pour chaque entrée (nom, verbe, adverbe), nous avons un renvoi vers une autre entrée faisant référence à cette partie du discours avec des informations paramétrées dans le dictionnaire lexical.

Le dictionnaire sémantique indique les lexicalisations possibles pour chaque sémantème, ce qui est une source de paraphrasage lexical. Par exemple, pour le sémantème '*combattre*', il y a plusieurs lexicalisations possibles, comme *combat*, *combattre*, *affrontement*, *affronter*.

Enfin, il y a un dictionnaire des fonctions lexicales (Lambrey & Lareau, 2015; Lambrey, 2017), qui ne sera pas utilisé dans ce projet de recherche.

Voici un exemple de RSém dans GenDR pour la phrase (7), sous forme textuelle et sous forme graphique.

(7) *Les antithrombotiques sont des médicaments qui empêchent la formation de caillot de sang.* [Sequoia, emea-fr-dev_00475³]

```
structure Sem s1 {
  S:1{
    sent_id="emea-fr-dev_00475" #numéro d'identification de la phrase
    text="Les antithrombotiques sont des médicaments qui empêchent la formation de caillots sanguins." #phrase écrite
    antithrombotique:1{number=PL definiteness=DEF}
    être:1{
      tense=PRES #verbe principal de la phrase
      #présent
      1-> antithrombotique:1 #actant 1 du verbe être
      2-> médicament:1 #actant 2 du verbe être
    }
    médicament:1{number=PL definiteness=INDEF}
    S:2{ # proposition relative
      empêcher:1{
        tense=PRES #verbe à la tête de la relative
        #présent
        1-> médicament:1 #actant 1 du verbe empêcher
        2-> "se former":1 #actant 2 du verbe empêcher
      }
      "se former":1{
        tense=PRES number=SG definiteness=DEF
        1-> caillot:1
      }
    }
    caillot:1{
      number=PL definiteness=INDEF
      1-> sang:1
    }
    sang:1{number=SG definiteness=NO}
    main-> empêcher:1 #nœud communicatif dominant de la relative
  }
}
```

³ Sequoia est le nom d'un des corpus utilisé et emea-fr-dev_00475 est le numéro d'identification de la phrase.

```

}
main-> être:1          #noeud communicatif dominant de la principale
}
}

```

Figure 5 RSém de la phrase (7) dans GenDR (forme textuelle)

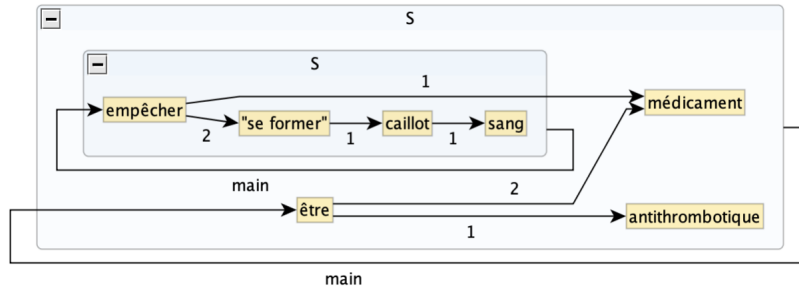


Figure 6 RSém de la phrase (7) dans GenDR (forme graphique)

Les bulles S représentent les propositions, la relation main marque le nœud dominant d’une proposition.

2.3.2 Au niveau de la syntaxe profonde

Dans un deuxième temps, c’est au niveau de l’interface sémantique-syntaxe profonde que le module de règles sémantiques-syntaxiques profondes (Sem-DSynt) est réalisé. Le passage de la structure de base sémantique à la structure de base syntaxique profonde se fait dans cette interface. Les règles permettent de transformer le réseau sémantique en arbre de dépendances syntaxiques profondes, autrement dit l’arborisation (Lareau et al., 2018; Galarreta-Piquette, 2018). Les dictionnaires et les règles correspondantes permettent de choisir les bonnes lexies pour exprimer les sémantèmes, c’est-à-dire d’opérer la lexicalisation profonde (Wanner & Bateman, 1990; Wanner, 1992 Lareau et al., 2018, Galaretta-Piquette, 2018). Voici un exemple d’un arbre de dépendance syntaxique profond dans GenDR pour la phrase (7)

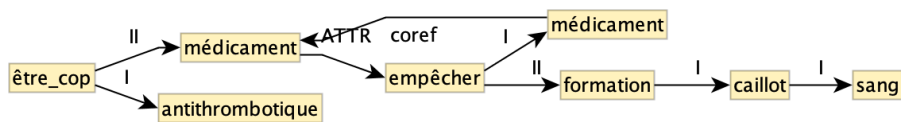


Figure 7 RSyntP de la phrase (7) dans GenDR

Le graphe ci-dessus représente un arbre de dépendance en syntaxe profonde dans GenDR avec une relation de modification (ATTR) et une coréférence (coref).

Puis, une fois que le graphe de la phrase au niveau de la syntaxe profonde est généré, il est possible de sélectionner différentes structures de phrase nous pouvons générer le niveau du réalisateur profond, c'est-à-dire le niveau de la syntaxe de surface.

2.3.3 Au niveau de la syntaxe de surface

Le module de règles de syntaxe profonde-syntaxe de surface (DSynt-SSynt) permet la transition au niveau de la syntaxe de surface. Il y a des règles communes aux langues et des règles propres à une langue (par exemple, pour les mots-outils). Les mots fonctionnels sont ajoutés à l'arbre de dépendance syntaxique pour la lexicalisation de surface. C'est à ce niveau qu'apparaissent les règles de pronominalisation. Voici un exemple d'un arbre de dépendance syntaxique de surface dans GenDR pour la phrase (7).

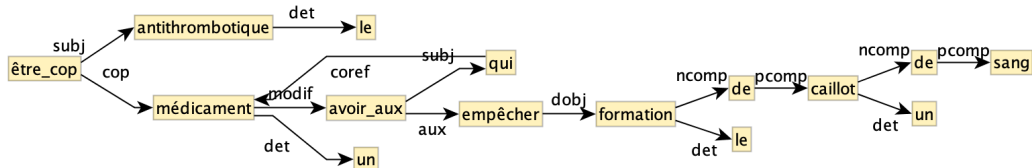


Figure 8 RSyntS de la phrase (7) dans GenDR⁴

Le graphe ci-dessus représente un arbre de dépendance de la syntaxe de surface dans GenDR, avec une relation de modification (modif), une relation de coréférence (coref) et l'apparition de pronom relatif *qui*.

Pour conclure cette sous-partie, GenDR est un réalisateur profond multilingue basée sur la TST. Les parties utilisées sont le dictionnaire de lexies en français, le dictionnaire de sémantèmes en français, et les dictionnaires de règles grammaticales au niveau sémantique-syntaxe profonde et

⁴ *Le* et *un* sont lexicalisés dans ce schéma. Dans le dictionnaire de GenDR, *le* a le trait défini et *un* le trait indéfini, mais nous ne le voyons pas sur le graphique.

au niveau syntaxe profonde-syntaxe de surface, notamment avec les règles grammaticales propres au français.

2.4 Synthèse

Dans ce chapitre, la TST est brièvement présentée comme le cadre théorique de ce mémoire. Ce projet de recherche sur les propositions relatives se situe dans l'interface sémantique-syntaxe, et plus particulièrement, la structure de base sémantique, la structure sémantico-communicative, la structure syntaxe profonde, la structure syntactico-anaphorique profonde, la structure de base syntaxe de surface et la structure syntactico-anaphorique de surface. Les générateurs de surfaces comme SimpleNLG, JSReal ou RealPro peuvent générer des propositions relatives, mais demandent d'avoir en entrée une structure syntaxique déjà fixée. La génération des propositions relatives dans les réalisateurs profonds reste un défi. Nous avons brièvement présenté le réalisateur profond GenDR, dans lequel nous implémenterons le traitement des propositions relatives. Dans le chapitre suivant, nous allons étudier plus en détail les propositions relatives d'un point de vue linguistique, avant de présenter notre implémentation au chapitre 4.

Chapitre 3 – Les propositions relatives

Ce projet de recherche porte sur les propositions relatives. Dans le chapitre précédent, le cadre théorique a été présenté, ainsi que l'interface sémantique-syntaxe où les propositions relatives se mettent en place. Dans ce chapitre, nous étudierons les propositions relatives sur le plan théorique. Nous aborderons d'abord les propositions relatives en adoptant un point de vue traditionnel. Ensuite, nous analyserons les structures et les relations de dépendance syntaxique dans l'interface sémantique-syntaxe. Pour finir, nous présenterons les différents pronoms relatifs en français, la définition de la proposition relative et ses spécificités.

3.1 Point de vue traditionnel

En français, les propositions relatives sont traditionnellement décrites comme des propositions subordonnées introduites par un pronom relatif. La proposition relative est dépendante de la proposition principale.

Il y a plusieurs définitions des propositions relatives qui se rejoignent. Riegel et al. définissent les propositions relatives comme « des propositions subordonnées qui sont introduites par un terme relatif, pronom ou, plus rarement, déterminant » (2018, p. 794). Ou encore, il y a la définition de Grevisse et Goose (2016) : « La proposition relative est une proposition commençant par un pronom relatif (*qui, que, quoi, dont, où, lequel, quiconque*) ou par un syntagme contenant le pronom relatif ou, parfois, dans la langue écrite surtout, notamment juridique, par un nom accompagné du déterminant relatif. » De plus, il y a d'autres auteurs qui définissent ces propositions relatives par leur fonction comme Leeman (2002, p. 55) : « la subordonnée relative complète un nom, avec lequel elle forme un GN (groupe nominal); elle est introduite par un pronom (le pronom relatif). »

En général, les pronoms relatifs sont : *qui, que, quoi, dont, où, lequel, auquel, duquel*. Il y a deux types de pronoms relatifs : les pronoms relatifs simples : *qui, que, quoi, dont, où*, et les pronoms relatifs composés comme *lequel, auquel* et *duquel* qui s'accordent (Riegel, Pellat, & Rioul, 2018,

p. 794). Les deux principaux types de propositions relatives sont : les relatives sans antécédent et celles avec antécédent (Riegel, Pellat, & Rioul, 2018, pp. 795-822).

3.1.1. La relative sans antécédent

Les propositions relatives sans antécédent, ou substantives, sont appelées ainsi, car le pronom relatif n'a pas d'antécédent nominal ou pronominal et se comporte plus comme un pronom indéfini. Les propositions relatives sont argument du verbe de la proposition principale. En français (Riegel, Pellat, & Rioul, 2018, p. 387; Delaveau, 1987; Grevisse & Goose, 2016). Grevisse et Goose (2016) décrivent trois types de relatives sans antécédent : le premier « *quiconque, celui, quelqu'il soit, qui* (il représente toujours des personnes) est sujet de propositions relatives qui, elles-mêmes, sont sujets ou compléments », le deuxième est les « propositions introduites par *qui* nominal représentant des personnes » et le troisième sont les autres cas, généralement pour les expressions figées, « avec *qui* nominale comme sujet, avec *que* nominale dans des expressions figées, avec *quoi* précédé d'une préposition, avec *où*. » Ces propositions relatives sont moins fréquentes que les relatives avec antécédent. Voici des exemples de propositions relatives sans antécédents.

(8) *Je suis qui je suis.*

(9) *Qui va à la chasse perd sa place*

(10) *J'irai où tu iras.*

Les propositions relatives sans antécédent ne sont pas le sujet de ce projet de recherche, parce qu'elles ne sont pas assez fréquentes et la plupart sont des expressions figées ou des locutions. De plus, pour des raisons techniques la génération des propositions relatives nécessite un verbe à la tête de la proposition relative pour GenDR, pour que les règles de GenDR fonctionnent et que les phrases soient générées.

3.1.2. La relative avec antécédent

L'antécédent d'une proposition relative (Grevisse & Goose, 2016) peut être un nom, un pronom, un adjectif, un participe épithète détaché, un adverbe de lieu, un adverbe de temps, ou une phrase. Grevisse et Goose distinguent plusieurs fonctions aux propositions relatives avec un antécédent comme épithète ou attribut. Dans les propositions relatives avec un antécédent, nous distinguons traditionnellement les propositions relatives restrictives et les propositions relatives explicatives. D'après Riegel et al. (2018), dans les propositions relatives à antécédent, il y a plusieurs sous-types, comme les relatives adjectives, les relatives de second degré, les relatives périphrastiques, les relatives substantives indéfinies, les relatives prädicatives et les relatives constituantes d'une expression concessive. Mais nous n'allons pas rentrer dans les détails, car nous n'en avons pas besoin dans ce projet de recherche puisqu'elles se traitent toutes de la même façon dans notre implémentation. Voici un exemple de proposition relative avec antécédent.

(11) *Les antithrombotiques sont des médicaments qui empêchent la formation de caillot sanguin*. [Sequoia, sequoia_emea-fr-dev_00475]

Dans cette phrase, *qui empêchent la formation de caillot sanguin* est une proposition relative qui contient un pronom relatif, qui, dont l'antécédent est *médicaments*.

3.1.2.1. La relative restrictive

Dans la proposition relative restrictive, l'antécédent est généralement un nom ou un pronom. D'après Grevisse et Goose (2016), la proposition relative a la fonction d'une épithète : « La relative déterminative ou restrictive, qui restreint l'extension du terme qu'elle accompagne (la suppression de la relative modifierait profondément le message) : [...] Il n'y a ni pause dans l'oral ni virgule dans l'écrit entre l'antécédent et la proposition ». Voici des exemples ci-dessous.

(12) *Le loup qui est le plus fort devient l'alpha de la meute*.

(13) *Les enfants qui sont fatigués ont fait la sieste⁵*. [termiumplus.qc.ca]

⁵https://www.btb.termiumplus.gc.ca/tpv2guides/guides/clefsfp/index-eng.html?lang=eng&lettr=indx_catlog_v&page=9jHyuL4-36fl.html

Parmi tous les loups, seul le loup le plus fort *devient l'alpha de la meute* et parmi tous les enfants seuls les enfants fatigués *ont fait la sieste*, donc *qui est le plus fort* et *qui sont fatigués* sont des propositions relatives restrictives.

3.1.2.2. La proposition relative explicative

La proposition relative explicative, ou appositive, ou descriptive, ou parenthétique, peut être difficile à distinguer de la proposition restrictive s'il n'y a pas de virgules qui l'encadrent.

D'après Grevisse et Goose (2016), « la relative non déterminative, souvent appelée explicative (ou descriptive, parfois appositive), qui ne restreint pas l'extension du terme qu'elle accompagne (la suppression de la relative ne modifierait pas vraiment le message) ; on peut la rapprocher de l'épithète détachée. La relative non déterminative est souvent séparée de l'antécédent par une pause dans l'oral et une virgule dans l'écrit. »

Voici une autre définition de Riegel et al. (2018, p. 804) : « La relative est explicative (ou appositive) lorsqu'elle ne joue aucun rôle dans l'identification référentielle de l'antécédent. Elle peut alors, sans dommage pour cette identification, être supprimée. Elle peut aussi être remplacée par une proposition coordonnée. » Voici un exemple avec la phrase ci-dessous.

(14) *Les loups qui sont des animaux sauvages vivent en meute ou seuls.*

(15) *Les enfants, qui sont fatigués, ont fait la sieste⁶.* [termiumplus.qc.ca]

Qui sont des animaux sauvages et *qui sont fatigués* sont des propositions relatives explicatives. La proposition relative *qui sont des animaux sauvages* peut être encadrée de virgules comme pour la deuxième phrase.

⁶https://www.btb.termiumplus.gc.ca/tpv2guides/guides/clefsfp/index-eng.html?lang=eng&lettr=indx_catlog_v&page=9jHyuL4-36fl.html

3.2. L'interface sémantique-syntaxe

Dans le chapitre précédent, la TST est brièvement présentée comme le cadre théorique de ce mémoire et ce projet de recherche se situe dans l'interface sémantique-syntaxe. Dans l'interface sémantique-syntaxe, il y a plusieurs structures qui sont intéressantes pour les propositions relatives. La structure de base sémantique, la structure sémantico-communicative, la structure de base syntaxe profonde, la structure syntactico-anaphorique profonde, la structure de base syntaxe de surface et la structure syntactico-anaphorique de surface sont les structures qui vont être étudiées dans cette section.

De plus, c'est dans cette interface que les relations⁷ de dépendances syntaxiques jouent un rôle. D'après Mel'čuk et Milićević (2014, p. 115), « Dans un énoncé, la dépendance linguistique est une relation syntagmatique hiérarchique entre deux unités lexicales, ou lexies, l'une appelée gouverneur et l'autre dépendant ». « Une lexie est soit un lexème soit une locution » (Mel'čuk & Milićević, 2014, p. 251). Il y a trois types de dépendances : la dépendance morphologique, la dépendance syntaxique et la dépendance sémantique.

D'après Mel'čuk et Milićević (2014, p. 116), « Une dépendance syntaxique est une dépendance entre deux lexies de la phrase L1 et L2, telle que l'une le gouverneur, détermine la distribution, c'est-à-dire les types de liens syntaxiques externes, de la configuration lexicale L1 — L2 entière ». Il existe trois sous-types de relations de dépendances syntaxiques, celle de complémentation, celle de modification et celle de coordination au niveau de la syntaxe de profonde. Plusieurs de ces sous-types de relations peuvent être appliqués à une même phrase. La proposition relative agit comme une modification de son antécédent et peut contenir une relation de coordination ou de complémentation à l'intérieur de la proposition relative.

⁷ « La notion de dépendance est une notion fondamentale dans l'approche Sens-Texte, qu'on pourrait caractériser comme relationnelle car elle considère les relations – en particulier les relations de dépendance – comme facteur essentiel de l'organisation linguistique » (Mel'čuk & Milićević, 2014, p. 115)

3.2.1. Au niveau sémantique

Dans la structure de base sémantique, il y a les sémantèmes et les relations sémantiques entre eux, comme dans la Figure 6, p. 38.

Les différences sémantiques entre proposition relative et proposition explicative sont connues depuis longtemps. Voici une comparaison des relatives restrictives et explicatives modifiée de Port-Royal : « La relative restrictive (ou déterminative) est alors définie comme étant une relative qui restreint l'extension du concept exprimé par l'antécédent (N ou SN) qu'elle modifie, et la relative appositive (ou explicative) comme une relative qui laisse cette extension inchangée. ». « Une relative appositive n'a aucune action sur l'extension du terme antécédent. » (Kleiber, 1987, p. 15)

Dans la structure sémantico-communicative, nous proposons une différence entre proposition relative restrictive et proposition relative explicative au niveau de la SSém basée sur Mel'čuk (2001) et des exemples qu'il propose. Cette proposition n'est pas testée, car GenDR n'est pas implémenté pour générer la SSém-Comm. GenDR ne peut que proposer un choix pour le nœud communicatif dominant dans une phrase. La proposition relative restrictive ferait partie du rhème de la phrase, et plus particulièrement, du focus sémantique. La définition de Mel'čuk du focus sémantique est « le focus sémantique dans la SemS (la structure sémantique) S est la partie du Sem-Rhème (rhème sémantique) qui est destiné par le locuteur à fournir la valeur de l'élément inconnu Q dans la question sous-jacente de la phrase S » (Mel'čuk, 2001, p. 116). Autrement dit, la proposition relative restrictive apporte une information au rhème. La proposition relative restrictive faisant partie du rhème répond à une question qui permet de définir plus précisément l'ensemble que comprend le rhème, c'est-à-dire qu'il fait partie du focus sémantique. Par exemple, les phrases (12) et (13) répondraient aux questions : parmi les loups lequel est devenu l'alpha, et parmi les enfants lesquels ont fait la sieste ? Voici un autre exemple de focus sémantique.

(16) *The system was used by four programmers.*

Le système a été utilisé par quatre programmeurs. (Mel'čuk, 2001, p. 116)

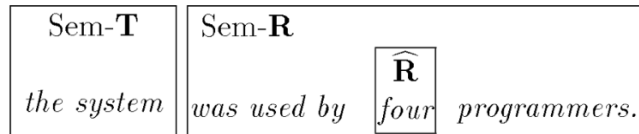


Figure 9 Focus sémantique (Mel'čuk, 2001, p. 116)

Nous proposons que pour la relative explicative, la proposition relative fasse partie du rhème et que dans le rhème il y ait une deuxième division thème-rhème où le pronom relatif est le sous-thème et le reste de la proposition relative forme le sous-rhème. En effet, « une représentation sémantique peut correspondre à une phrase complexe, qui contient une proposition subordonnée. Dans une telle représentation sémantique, il peut y avoir une division R-T (rhème-thème) supplémentaire en sous-domaines communicatifs à l'intérieur du RhSém (rhème sémantique) ou du ThSém (thème sémantique) primaire ; ces sous-domaines peuvent à leur tour être subdivisés. » (Mel'čuk, 2001, p. 119) Dans les exemples (17), (18) et la Figure 10, nous pouvons observer que la phrase a un sous-thème et un sous-rhème dans le rhème.

(17) *Alain realized that some strangers had removed his computer.*

Alain a réalisé que des étrangers ont enlevé son ordinateur. (Mel'čuk, 2001, p. 119)

(18) *Alain realized that his computer had been removed by some strangers.*

Alain a réalisé que son ordinateur a été enlevé par des étrangers. (Mel'čuk, 2001, p. 119)

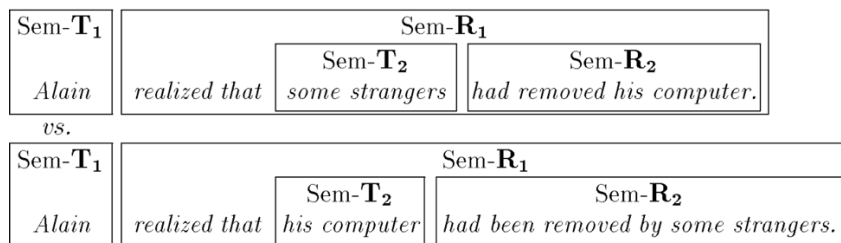


Figure 10 Ex. de thème et de rhème secondaires (Mel'čuk, 2001, p. 119)

Pour la proposition relative explicative encadrée par des virgules ou comportant une intonation particulière à l'oral par rapport au reste de la phrase, elle se situe au niveau de l'opposition communicative de perspective, et signale une mise à l'arrière-plan. Dans ce cas, la proposition relative explicative est une information parenthétique. (Mel'čuk, 2001, p. 202) Par exemple les phrases (15) et (19) et la Figure 11.

(19) *John, who is a very good carpenter, built this cabinet.*

John, qui est très bon charpentier, a construit ce meuble. (Mel'čuk, 2001, p. 202)

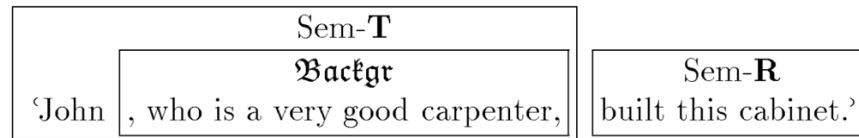


Figure 11 Exemple d'arrière-plan (Mel'čuk, 2001, p. 202)

De plus, la différence entre une proposition relative et une focalisée se trouve aussi à ce niveau. En effet, les focalisées peuvent avoir des structures syntaxiques proches des propositions relatives. Par exemple, les phrases qui ont la structure *c'est X qui/que* ou *ce que X est*, ne sont pas des propositions relatives, mais des focalisées. Les phrases *c'est X qui/que* ou *ce que X est* sont souvent analysées comme une relative dans le corpus. En effet, « pour exprimer un Sem-Th (thème sémantique) focalisé. Elle présente le sens qui correspond au Sem-Th focalisé sous la forme d'une structure de type relative. » (Mel'čuk, 2001, p. 192) Par exemple :

(20) *What John is reading is a whodunit. Ce que John lit est un roman policier. (Mel'čuk, 2001, p. 193)*

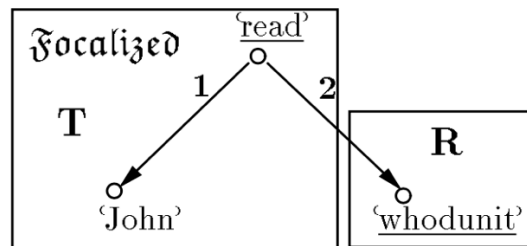


Figure 12 Exemple focalisé (Mel'čuk, 2001, p. 193)

(21) *C'est la fille qui porte une robe verte.*

(22) *Aria est avec la fille qui porte une robe verte.*

L'exemple (20) est une focalisé avec la Figure 12 qui représente sa structure communicative. L'exemple (21) est la focalisé et l'exemple (22) est la relative. Les deux exemples ont une proposition avec un verbe fini, son introduit par le pronom *qui* et ce pronom est coréférent à un nom dans la proposition principale. Les exemples (21) et (22) montrent à quel point la focalisé et

la proposition relative peuvent avoir des structures proches. Pour l'exemple (21) le focus est sur le nom *filles* et celle *qui porte une robe verte*, alors que dans l'exemple (22) le nom *filles* et *qui porte une robe verte* modifie le nom *filles*.

3.2.2. Au niveau de la syntaxe profonde

La structure syntaxique profonde est l'équivalent de la structure sémantique. Au niveau de la syntaxe profonde et de la syntaxe de surface, il y a trois sous-types de dépendances syntaxiques. Les dépendances syntaxiques de complémentation et de modification sont des types de dépendance de la subordination (Mel'čuk, 2009, pp. 51, figure 5). Dans la structure des propositions relatives, il y a la dépendance syntaxique de modification qui a pour définition : « un mot-forme *w2* est un modificateur, ou un attribut syntaxique, du mot-forme *w1* et que *w1* est le gouverneur syntaxique de *w2* » (Mel'čuk, 2009, p. 49). Il y a la relation ATTR qui est une relation de modification entre le nom modifié par la proposition relative et le verbe au sommet de la proposition relative. Par exemple, la Figure 3 est la RSyntP de la phrase (3).

Dans la structure syntactico-communicative profonde, il y a les différences de situation de communication entre la proposition relative restrictive et la proposition relative explicative et entre proposition relative et focalisée qui sont héritées de la structure sémantico-communicative au niveau sémantique.

3.2.3. Au niveau de la syntaxe de surface

Dans la structure de base syntaxique de surface, la relation ATTR devient la relation de modification. Par exemple, la Figure 4 est la RSyntS de la phrase (3).

Pour conclure cette partie, les structures de base sémantique, syntaxique profonde et syntaxique de surface sont importantes pour la relation de modification de la proposition relative. Les structures sémantico-communicative, syntactico-anaphorique profonde et syntactico-anaphorique de surface sont importantes pour distinguer une proposition relative restrictive d'une explicative, et une proposition relative d'une proposition focalisée. En effet, la proposition

relative et la focalisée sont très semblables au niveau de la syntaxe de surface. Le pronom relatif apparaît en syntaxe de surface.

3.3. Les pronoms relatifs

Il y a deux types de pronoms comme vu dans la partie 3.1. Le tableau ci-dessous présente les différents pronoms selon leur fonction et l’animacité de leur antécédent.

FONCTION	ANTÉCÉDENT		
	Animé	Non-animé	Animé ou non animé
Sujet	<i>qui</i>		<i>lequel, laquelle, etc.</i>
Objet direct, attribut	<i>qui/que</i>		
Complément prépositionnel	préposition + <i>qui</i>	préposition + <i>quoi</i> <i>où</i>	préposition + <i>lequel, etc.</i> <i>auquel</i> <i>duquel</i>
	<i>dont</i>		
	Simple		Composée
	FORME		

Tableau 3 Les pronoms relatifs (Riegel, Pellat, & Rioul, 2018, p. 385)

Dans la suite de ce mémoire, nous nous baserons sur ce tableau.

3.3.1. Le pronom relatif *qui*

Le pronom relatif *qui* occupe le rôle syntaxique de sujet. L’antécédent peut être animé ou non animé. « Le relatif a la forme *qui* et, sans porter lui-même aucune marque de nombre, de genre, ou de personne, il commande dans la relative, le cas échéant, les mêmes accords. » (Riegel, Pellat, & Rioul, 2018, p. 798). En linguistique, les pronoms relatifs *qui* et *que* sont traditionnellement considérés comme deux mots-formes du même lexème QUI, le mot-forme *qui* est pour le rôle syntaxique de sujet et le mot-forme *que* est pour le rôle syntaxique d’objet direct. « L’analyse traditionnelle dit que *qui* est le complémenteur sujet » et « *que* est le complémenteur objet. » (Kahane, 2009, p. 129) . Voici un exemple.

(23) *Des thromboses sont des caillots sanguins qui peuvent entraîner des complications sévères comme une crise cardiaque.* [Sequoia, emea-fr-fr-dev_00540]

Cette phrase a une proposition relative introduite par le pronom relatif *qui*.

Dans le corpus, il y a sept phrases détectées avec un pronom *qui* objet sur 4 505 phrases. Ces phrases sont des locutions ou des expressions figées. Nous avons décidé de ne pas considérer le pronom *qui* comme pouvant être objet par la suite, mais voici un exemple ci-dessous avec un *qui* objet.

(24) *Ils ont interdit qui que ce soit de s'approcher de ma chambre.* [GSD, fr-ud-train_01660]

3.3.2. Le pronom relatif *lequel*

Le pronom relatif *lequel* occupe le rôle syntaxique de sujet. C'est un lexème d'un déterminant amalgamé avec le pronom *quel*. Ce pronom relatif s'accorde en genre et en nombre avec son antécédent. L'antécédent peut être animé ou non. « L'emploi de *lequel* (pronom ou déterminant), archaïsant et rare, est réservé aux énoncés juridiques où il évite l'ambiguïté qui serait produite par *qui* dans le cas où il pourrait se rapporter à plus d'un antécédent. » (Riegel, Pellat, & Rioul, 2018, p. 798) Voici un exemple ci-dessous..

(25) *Nous discutons de l'actualité urgente, laquelle traite des prisonniers.* [Sequoia, Europar.550_00093]

3.3.3. Le pronom relatif *que*

Le pronom relatif *que* occupe le rôle syntaxique d'objet direct ou d'attribut. L'antécédent peut être animé ou non, ou encore un adverbe de temps comme *aujourd'hui*, *maintenant*, *à présent*, etc. (Grevisse & Goose, 2016). Riegel et al. (2018, p. 799) définissent le pronom relatif *que* comme « Il s'agit surtout de la relativisation du complément d'objet direct du verbe (exemples (27) et (28)), mais aussi d'autres syntagmes qui partagent avec lui la position postverbale et la

pronominalisation définie de la forme *le, la, les*, sujets logiques des verbes impersonnels (exemple (29)) ou compléments des présentatifs (exemple (30)). » Voici un exemple avec le pronom relatif *que*.

(26) *Maintenant que tout est prêt, nous pouvons commencer à cuisiner.*

(27) *Je regrette chacun des baisers que je donne. (G. Appollinaire) (Je donne des baisers)
(Riegel, Pellat, & Rioul, 2018, p. 799)*

(28) *Je lis un livre que ma sœur m'a offert. (Ma sœur m'a offert un livre) (Riegel, Pellat,
& Rioul, 2018, p. 799)*

(29) *C'est vraie solution qu'il faut à ce problème. (Il faut une vraie solution) (Riegel,
Pellat, & Rioul, 2018, p. 799)*

(30) *Je connais le garçon que voilà. (Voilà le garçon (Riegel, Pellat, & Rioul, 2018, p.
799)*

(31) *J'ai lu le livre que Luc pense que Marie aime.*

3.3.4. Le pronom relatif *dont*

Le pronom relatif *dont* occupe le rôle syntaxique de complément prépositionnel. L'antécédent peut être animé ou non. Le pronom relatif *dont* vient étymologiquement du latin *de unde* (Hadermann, 1987) et peut être : un *dont* oblique ou un *dont* complément du nom. Le premier est déduit de l'analyse des phrases du corpus où nous avons observé que *dont* pouvait avoir le rôle syntaxique de complément oblique. Le deuxième pronom relatif *dont* est complément du nom. « Elle équivaut à un pronom relatif précédé de la préposition *de*, quelle que soit la fonction du groupe prépositionnel ainsi formé, pourvu que le relatif ne soit pas complément d'un nom lui-même précédé d'une préposition. » (Riegel, Pellat, & Rioul, 2018, p. 386) L'exemple (32) contient le pronom *dont* complément du nom et l'exemple (33), le *dont* oblique.

(32) *L'arbre dont je vois la cime indique la direction.*

(33) *J'ai lu le livre dont Luc parle.*

3.3.5. Le pronom relatif *où*

Le pronom relatif *où* occupe la fonction de complément prépositionnelle et a un sens locatif. Avec le pronom relatif *où* (Grevisse & Goose, 2016), l'antécédent du pronom relatif peut être un nom ou un adverbe de lieu comme *ici, là, partout, quelque part*, etc. De plus, nous avons observé que ce pronom n'apparaît que s'il s'agit d'un lieu ou d'un temps. Le régime des verbes après ce pronom peut être particulier. Il s'agit de verbe, comme *se trouver, se situer, s'éteindre, mourir, aller*, pour lesquels il y a une indication spatiale ou temporelle. Voici un exemple.

(34) *Guy a consacré sa vie professionnelle à Besnier (lieu) où il comptait de nombreux amis.* [Sequoia, annodis.er_00074]

(35) *Aline se trouve à la bibliothèque où elle fait ses devoirs.*

3.3.6. Une préposition suivie du pronom relatif *qui*

Une préposition suivie du pronom relatif *qui* occupe le rôle syntaxique de complément prépositionnel. L'antécédent précédant la préposition et le pronom relatif *qui* est animé. Voici un exemple.

(36) *Chris offre des fleurs à Aria à qui j'ai prêté un livre.*

3.3.7. Une préposition suivie du pronom relatif *quoi*

Une préposition suivie du pronom relative *quoi* occupe le rôle syntaxique de complément prépositionnel. Le pronom relatif *quoi* est complémentaire au pronom relatif *qui* et est coréférent d'un antécédent qui n'est pas animé. Voici un exemple.

(37) *Jean parle du livre sur quoi j'ai fait un travail.*

3.3.8. Une préposition suivie du pronom relatif *lequel*

Une préposition suivie du pronom relatif *lequel* a le rôle syntaxique de complément prépositionnel. C'est un lexème formé d'un déterminant amalgamé avec le pronom *quel*. Ce pronom relatif s'accorde en genre et en nombre avec son antécédent. L'antécédent peut être animé ou non. Une préposition suivie du pronom relatif *lequel* est la forme complexe d'une préposition suivie du pronom relatif *qui* ou *quoi*. Voici un exemple.

(38) *La maladie de Piaget est une maladie dans laquelle la croissance des os est modifiée.* [Sequoia, emea-fr-test-00010]

Pour résumer, ce mémoire se base sur le Tableau 3 pour les pronoms relatifs. Les pronoms relatifs *qui* et *lequel* sont sujets, le pronom relatif *que* est objet, les pronoms relatifs *où*, *dont* et une préposition suivie du pronom *qui* ou *quoi* ou *lequel* sont complément prépositionnel.

3.4. La proposition relative

Dans un premier temps, les propositions relatives seront définies dans cette partie pour le reste de ce mémoire. Puis, deux catégories majeures de propositions relatives seront présentées, les propositions relatives directes et les propositions relatives indirectes. Enfin, une liste de caractéristiques définitoires des propositions relatives sera faite.

3.4.1. Définition de la proposition relative

Voici un exemple de proposition relative et son explication ci-dessous.

(39) *Vous suivez attentivement toutes les instructions que vous a données votre médecin.* [Sequoia_emea-test_00421]

La phrase (39) a une proposition relative *que vous a donnée votre médecin* qui modifie le nom *instructions* dans la proposition principale. Cette proposition relative est introduite par le pronom relatif *que*, qui est actant du verbe au sommet de la proposition relative *donner* et coréfèrent

d'*instructions* dans la proposition principale. Cette définition regroupe les propositions relatives explicatives et restrictives et permet d'accentuer les relations de dépendance syntaxique propre à la proposition relative. La compréhension de ces relations de dépendance syntaxique est nécessaire pour la bonne implémentation des propositions relatives dans un réalisateur profond. Dans les deux sous-parties suivantes, seulement deux catégories de propositions relatives seront définies et regroupent les relatives avec antécédent (§ 3.1.2).

3.4.2. La proposition relative directe

Les propositions relatives directes rassemblent tous les types de propositions relatives avec un antécédent vu précédemment. Une proposition relative est directe quand un actant du verbe au sommet de la proposition relative, c'est-à-dire le verbe communicatif dominant de la proposition relative, est aussi coréférent au nom modifié par cette proposition relative et qu'il n'y a pas de prédicat comme un verbe à l'infinitif ou un nom. La proposition relative directe est un cas particulier de la proposition relative indirecte, où le nombre de prédicats entre le verbe à la tête de la relative et le coréférent du nom modifié est zéro. La classification des propositions relatives en directes et indirectes est une classification technique, car avec GenDR et le logiciel GREW de SUD ne peuvent pas traiter en une règle les propositions relatives et ne détectent pas les propositions relatives avec des prédicats intermédiaires entre le verbe au sommet de la relative et le coréférent au nom modifié.

La phrase (39) est un exemple de proposition relative directe.

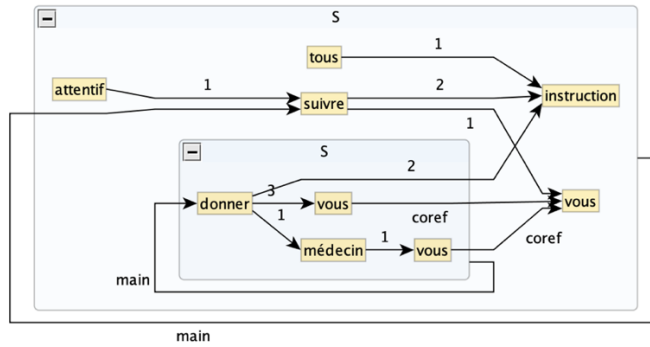


Figure 13 RSém de la phrase (39)⁸

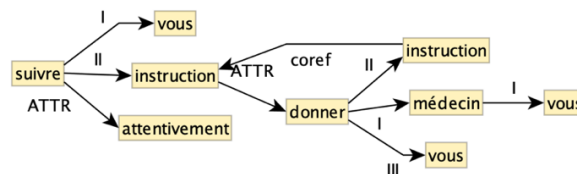


Figure 14 RSyntP de la phrase (39)

La proposition relative *que vous a donnée votre médecin* modifie le nom *instruction*. En effet, le nom *instructions* dans la proposition relative, sera remplacé par le pronom relatif *que* en syntaxe de surface, est l'actant du verbe au sommet de la proposition relative *donner* et coréférent au nom modifié *instructions* dans la proposition principale. Il n'y a pas de prédicat intermédiaire entre le nom coréférent du nom modifié *instruction* et le verbe au sommet de la proposition relative *donner*, donc le lien est direct.

3.4.3. La proposition relative indirecte

Les propositions relatives indirectes rassemblent aussi tous les types de propositions relatives avec un antécédent vu précédemment comme pour la relative directe. La proposition relative indirecte a un ou plusieurs prédicats entre l'actant du verbe au sommet de la proposition relative et le coréférent au nom modifié par cette proposition relative qui peut un verbe ou un nom ou

⁸ Il y a le nœud *vous* dupliqué en plusieurs fois, car GenDR n'est pas assez développé pour générer la pronominalisation correctement. C'est pourquoi le pronom est dupliqué de façon rudimentaire.

de n'importe quelle classe grammaticale. La phrase suivante (40) est un exemple de proposition relative indirecte.

(40) *J'ai lu le livre que Luc pense que Marie aime.*

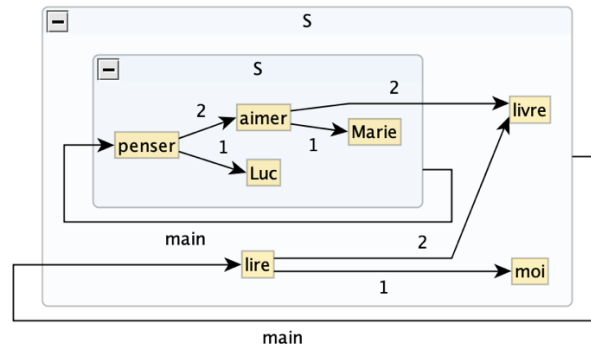


Figure 15 RSém de la phrase (40)

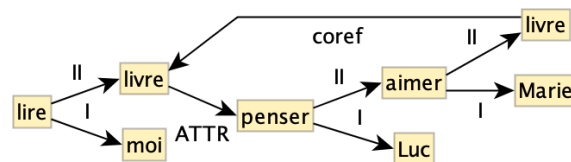


Figure 16 RSyntP de la phrase (40)

La proposition relative *que Luc pense que Marie aime* modifie le nom *livre* de la proposition principale *j'ai lu le livre*. *Luc* et *aimer* sont les deux actants du verbe au sommet de la proposition relative *penser*. Puis, *Marie* et *livre* de la proposition relative sont tous les deux actants du verbe *aimer*. *Livre* de la proposition relative est coréférent à *livre* de la proposition principale et sera remplacé par le pronom relatif *que* en syntaxe de surface. Donc *aimer* est entre le verbe au sommet de la proposition relative *penser* et *livre* de la proposition relative. C'est pourquoi *que Luc pense que Marie aime* est une proposition relative indirecte.

3.4.4. Les spécificités de la proposition relative

Dans cette partie, nous proposons une liste de spécificité propre à la proposition relative.

- Il doit y avoir au moins deux verbes finis, pour faire une proposition principale et une proposition relative.

- Un des verbes est le verbe principal et l'autre est le verbe au sommet de la proposition relative.
- Il y a une relation de modification entre la proposition relative et une lexie de la proposition principale.
- La proposition relative doit contenir un pronom relatif.
- La proposition relative ne doit pas être objet de la proposition principale
- Il y a une coréférence entre le pronom de la proposition relative et le coréférent de la proposition principale qui est modifiée par la proposition relative.

Dans la section suivante, un exemple de proposition relative directe et un exemple de proposition relative indirecte illustreront ces caractéristiques des propositions relatives.

3.4.5. Exemples de propositions relatives directe et indirecte

Voici un exemple illustré par la phrase (39) qui est de type direct, avec le pronom relatif *que* avec sa RSém dans la Figure 13, RSyntP dans la Figure 14 et RSyntS dans la Figure 17.

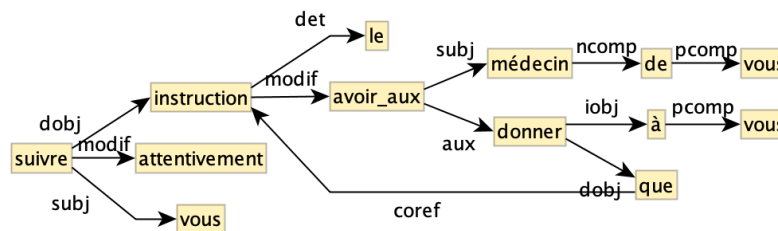


Figure 17 RSyntS de la phrase (39)

Pour la phrase (39) et sa RSyntS en Figure 17, nous reconnaissons toutes les caractéristiques des propositions relatives.

- Il y a deux propositions dans la phrase, *vous suivez attentivement toutes les instructions* et *que vous avez données votre médecin*.
- Il y a deux verbes finis *suivre* et *donner*.
- Il y a un pronom relatif *que* dans la RSyntS.

- La deuxième proposition *que vous a données votre médecin* modifie la première *vous suivez attentivement toutes les instructions*.
- Un actant *instruction* du verbe au sommet de la proposition relative *donner* est coréférent au nom *instruction* de la proposition principale *vous suivez attentivement toutes les instructions*.
- La deuxième proposition *que vous a données votre médecin* n'est pas objet de la proposition principale *vous suivez attentivement toutes les instructions*.
- Le verbe *donner* modifie le nom *instructions* et la relation entre le verbe au sommet de la proposition relative et l'actant du verbe *instructions* est coréférent au nom modifié *instruction* est direct.

La phrase (40) est un exemple de proposition relative indirecte avec un autre pronom, sa RSém est la Figure 15, sa RSyntP est la Figure 16 et sa RSyntS est la Figure 18.

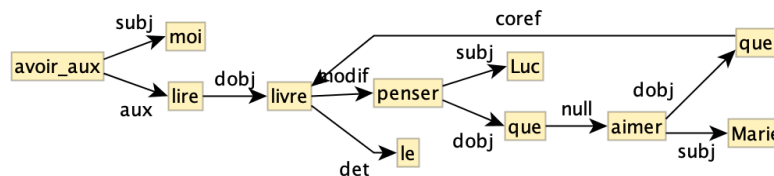


Figure 18 RSyntS de la phrase (40)

Les spécificités de la proposition relative sont respectées.

- Il y a deux propositions *j'ai lu le livre* et *que Luc pense que Marie aime*.
- Il y a trois verbes finis *lire*, *penser* et *aimer*.
- Il y a un pronom relatif *que* dans la RSyntS.
- La proposition *que Luc pense que Marie aime* modifie la proposition *j'ai lu le livre*.
- Il y a une coréférence entre *que* de la deuxième proposition *que Luc pense que Marie aime* et *livre* de la proposition principale *j'ai lu le livre*.
- La deuxième proposition *que Luc pense que Marie aime* n'est pas objet de la proposition principale *j'ai lu le livre*.

- Il y a un prédicat *aimer* entre le verbe au sommet de la proposition relative *penser* et le coréférent au nom modifié *livre*. Donc cette phrase est une proposition relative indirecte avec le pronom relatif *que*.

Dans la suite de ce mémoire, nous utiliserons la définition de la proposition relative discutée ici. Nous avons également listé les spécificités des propositions relatives et illustré par un exemple de proposition relative directe et indirecte.

3.5. Synthèse

Pour résumer ce chapitre, une typologie des propositions relatives a été faite dans la partie 3.1. Il y a deux typologies de propositions relatives : les relatives restrictives et les relatives descriptives. Puis, dans l'interface sémantique-syntaxe (§3.2), la structure de base sémantique, syntaxique profonde et syntaxique de surface sont importantes pour la relation de modification de la proposition relative. Les structures sémantico-communicative, syntactico-anaphorique profonde et syntactico-anaphorique de surface sont importantes. La structure sémantique et la structure SSém-Comm permettent de distinguer la proposition relative restrictive de l'explicative et de différencier une proposition relative d'une proposition focalisée. Ensuite, la catégorisation des pronoms relatifs (§ 3.3) suit le Tableau 3 de Riegel et al. (2018). Enfin, dans la partie 3.4 la proposition relative a été définie comme une proposition qui a une relation de modification et dont le pronom relatif a une relation de coréférence avec une lexie de la proposition principale. Il y a deux catégories de propositions relatives : la proposition relative directe et la proposition relative indirecte. La proposition relative directe a un lien sémantique direct entre le pronom relatif et le verbe au sommet de la proposition relative, tandis que la proposition relative indirecte a un lien sémantique indirect avec un ou des prédicats intermédiaires entre le verbe au sommet de la proposition relative et le coréférent au nom modifié. Les spécificités des propositions relatives ont aussi été listées : il doit y avoir au moins deux verbes finis, une relation de modification, un pronom relatif, une coréférence entre le pronom relatif et un lexème de la proposition principale.

Chapitre 4 – Analyse de corpus

Dans ce chapitre, nous observons les propositions relatives et les pronoms relatifs dans trois corpus français annotés en dépendances. Cette analyse des données permet de voir le comportement des propositions relatives et des pronoms relatifs dans l’usage. Nous pourrions ainsi conclure quel est le type de proposition relative le plus fréquent, quels sont les pronoms les plus fréquents et quel est le patron de proposition relative le plus fréquent. À la suite de cette analyse des propositions relatives et des pronoms dans le corpus, nous saurons ce que nous devons implémenter en premier et à quoi nous devons prêter attention pour la création des règles grammaticales sur les propositions relatives.

4.1. Choix des corpus

Les corpus GSD, Sequoia et ParTUT sont trois corpus en français. Ces corpus sont gratuits et téléchargeables gratuitement sur le site de Surface-Syntactic Universal Dependencies (SUD), Universal Dependencies (UD) et GitHub⁹. Ils ont été annotés automatiquement dans le format CoNLL-U, en suivant les schémas d’annotation de UD et de SUD. Ces corpus sont annotés avec les mêmes étiquettes, ont été lemmatisés et les relations sont également étiquetées. Il suffit de chercher la relation `acl:relcl` pour la version UD du corpus et `mod@relcl` pour la version SUD du corpus, pour trouver les phrases avec des propositions relatives.

Nous pouvons utiliser les corpus en UD et SUD sans modification majeure. L’utilisation de ces corpus permet d’économiser du temps et de l’énergie à collecter, nettoyer et annoter des phrases automatiquement. De plus, les problèmes d’autorisations pour l’utilisation des corpus, ainsi que le nettoyage et leur annotation sont évités, ce qui permet leur utilisation en traitement automatique de la langue.

⁹ <https://surfacesyntacticud.github.io/data/>; <https://github.com/surfacesyntacticud>

4.1.1 SUD et les critiques de UD

SUD (Osborne & Gerdes, 2019; Gerdes et al., 2019) est un schéma d’annotation traditionnel qui privilégie systématiquement la syntaxe plutôt que la sémantique. Notamment, ce schéma d’annotation positionne les verbes auxiliaires en tant que tête par rapport aux verbes de contenu, la copule en tant que tête par rapport aux éléments prédicatifs et les adpositions et subordinants en tant que tête par rapport aux noms et aux verbes. SUD vient de UD¹⁰ (De Marneffe & Manning, 2008; Zeman, 2008; Petrov, Das, & McDonald, 2011; SyntaxFest, 2019; Nivre, et al., 2020).

UD a été critiqué (Gerdes et al., 2019; Osborne & Gerdes, 2019) pour certains choix discutables du point de vue linguistique. Par exemple, pour certaines phrases des corpus en UD, le nom qui est modifié par la proposition relative est annoté par la relation `acl:relcl` en plus du verbe à la tête de la relative. Cela fait une double annotation de la relation relative pour la même proposition relative. Cela empêche de détecter correctement les propositions relatives et de savoir quelle est la relation du nom modifié. La phrase (41) ci-dessous illustre comment les choix de UD posent un problème.

(41) *Monsieur le Président, la Commission voudrait avant tout saluer la **déclaration** très claire du représentant du Conseil, selon laquelle le système Eurodac est un système qui n'est conçu que pour mettre en œuvre la convention de Dublin.* [Sequoia, Europar.550_195]

Voici ci-dessous l’annotation de la phrase en format de UD¹¹.

1	Monsieur	monsieur	NOUN	7	vocative	—	—
2	le	le	DET	3	det	—	—
3	Président	président	NOUN	1	nmod	—	SpaceAfter=No
4	,	,	PUNCT	7	punct	—	—
5	la	le	DET	6	det	—	—
6	Commission	commission	NOUN	7	nsubj	—	—
7	voudrait	vouloir	VERB	0	root	—	—
8	avant	avant	ADP	9	case	—	—
9	tout	tout	ADV	7	obl:mod	—	—
10	saluer	saluer	VERB	7	xcomp	—	—
11	la	le	DET	12	det	—	—

¹⁰ (<http://universaldependencies.org/>),

¹¹ Nous avons enlevé la colonne des traits morphologiques pour plus de visibilité et nous avons seulement gardé le trait `PronType =Rel` pour le pronom relatif. Nous avons mis le nom modifié, la relation relative et le trait du pronom relatif en gras pour plus de visibilité dans ce mémoire.

12	déclaration	déclaration	NOUN	10	obj	—	—
13	très	très	ADV	14	advmod	—	—
14	claire	clair	ADJ	12	amod	—	—
15-16	du	—	—	—	—	—	—
15	de	de	ADP	17	case	—	—
16	le	le	DET	17	det	—	—
17	représentant	représentant	NOUN	12	nmod	—	—
18-19	du	—	—	—	—	—	—
18	de	de	ADP	20	case	—	—
19	le	le	DET	20	det	—	—
20	Conseil	conseil	NOUN	17	nmod	—	SpaceAfter=No
21	,	PUNCT		12	punct	—	—
22	selon	selon	ADP	23	case	—	—
23	laquelle	lequel	PRON	PronType=Rel	29	obl:mod	—
24	le	le	DET	25	det	—	—
25	système	système	NOUN	29	nsubj	—	—
26	Eurodac	Eurodac	PROPN	25	nmod	—	—
27	est	être	AUX	29	cop	—	—
28	un	un	DET	29	det	—	—
29	système	système	NOUN	12	acl:relcl	—	—
30	qui	qui	PRON	PronType=Rel	33	nsubj:pass	—
31	n'	ne	ADV	33	advmod	—	SpaceAfter=No
32	est	être	AUX	33	aux:pass	—	—
33	conçu	concevoir	VERB	29	acl:relcl	—	—
34	que	que	ADV	36	advmod	—	—
35	pour	pour	ADP	36	mark	—	—
36	mettre	mettre	VERB	33	xcomp	—	—
37	en	en	ADP	38	case	—	—
38	oeuvre	oeuvre	NOUN	36	obl:arg	—	—
39	la	le	DET	40	det	—	—
40	convention	convention	NOUN	36	obj	—	—
41	de	de	ADP	42	case	—	—
42	Dublin	Dublin	PROPN	40	nmod	—	SpaceAfter=No
43	.	.	PUNCT	7	punct	—	—

Figure 19 Annotation de la phrase (41) en UD venant de UD

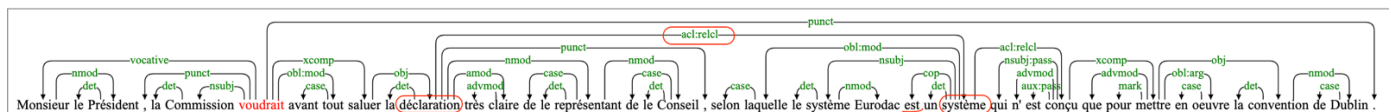


Figure 20 Schéma de la phrase (41) en UD¹²

La proposition relative dans cette phrase est *qui n'est conçu que pour mettre en œuvre la convention de Dublin*. *Concevoir* (33) qui est le verbe dans la proposition relative porte la relation `acl:relcl`, mais le nom modifié par cette proposition relative *système* (29) a également la relation `acl:relcl` propre aux propositions relatives. *Système* (29) est relié au nom *déclaration* (12). Les normes de UD sont surprenantes.

¹² Schéma réalisé par le Universal Dependencies graph/trees display de Guy Lapalme, <http://www-labs.iro.umontreal.ca/~lapalme/ShowUD/>

Au cours de notre recherche les corpus utilisés étaient en format UD, puis nous avons changé les corpus pour prendre les mêmes mais en format SUD (Osborne & Gerdes, 2019). Voici l'annotation de la phrase en format SUD.

1	Monsieur	monsieur	NOUN	7	vocative	-	-		
2	le	le	DET	3	det	-	-		
3	Président	président	NOUN	1	mod	-	-	SpaceAfter=No	
4	,	,	PUNCT	7	punct	-	-		
5	la	le	DET	6	det	-	-		
6	Commission	commission	NOUN	7	subj	-	-		
7	voudrait	vouloir	VERB	0	root	-	-		
8	avant	avant	ADP	7	mod	-	-		
9	tout	tout	ADV	8	comp:obj	-	-		
10	saluer	saluer	VERB	7	comp:obj@x	-	-		
11	la	le	DET	12	det	-	-		
12	déclaration	déclaration	NOUN	10	comp:obj	-	-		
13	très	très	ADV	14	mod	-	-		
14	claire	clair	ADJ	12	mod	-	-		
15-16	du	-	-	-	-	-	-		
15	de	de	ADP	12	udep	-	-		
16	le	le	DET	17	det	-	-		
17	représentant	représentant	NOUN	15	comp:obj	-	-		
18-19	du	-	-	-	-	-	-		
18	de	de	ADP	17	udep	-	-		
19	le	le	DET	20	det	-	-		
20	Conseil	conseil	NOUN	18	comp:obj	-	-	SpaceAfter=No	
21	,	,	PUNCT	12	punct	-	-		
22	selon	selon	ADP	27	mod	-	-		
23	laquelle	lequel	PRON	PronType=Rel 22	comp:obj	-	-		
24	le	le	DET	25	det	-	-		
25	système	système	NOUN	27	subj	-	-		
26	Eurodac	Eurodac	PROPN	25	mod	-	-		
27	est	être	AUX	12	mod@relcl	-	-		
28	un	un	DET	29	det	-	-		
29	système	système	NOUN	27	comp:pred	-	-		
30	qui	qui	PRON	PronType=Rel 32	subj@pass	-	-		
31	n'	ne	ADV	32	mod	-	-	SpaceAfter=No	
32	est	être	AUX	29	mod@relcl	-	-		
33	conçu	concevoir	VERB	32	comp:aux@pass	-	-		
34	que	que	ADV	35	mod	-	-		
35	pour	pour	ADP	33	comp:obl@x	-	-		
36	mettre	mettre	VERB	35	comp:obj	-	-		
37	en	en	ADP	36	comp:obl	-	-		
38	oeuvre	oeuvre	NOUN	37	comp:obj	-	-		
39	la	le	DET	40	det	-	-		
40	convention	convention	NOUN	36	comp:obj	-	-		
41	de	de	ADP	40	udep	-	-		
42	Dublin	Dublin	PROPN	41	comp:obj	-	-	SpaceAfter=No	
43	.	.	PUNCT	7	punct	-	-		

Figure 21 Annotation de la phrase (41) en SUD venant de SUD

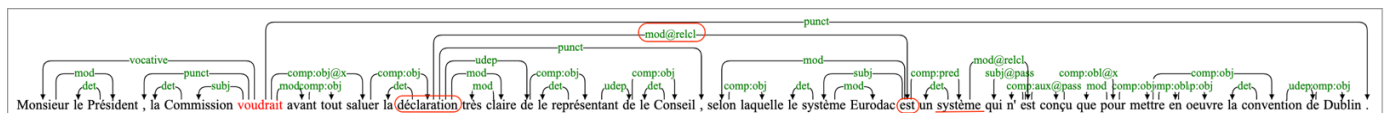


Figure 22 schéma de la phrase (41) en SUD¹³

¹³ Schéma réalisé par le Universal Dependencies graph/trees display de Guy Lapalme, <http://www-labs.iro.umontreal.ca/~lapalme/ShowUD/>

Nous pouvons constater que le problème en UD est réglé en SUD. Dans la phrase (41), la proposition relative est *qui n'est conçu que pour mettre en œuvre la convention de Dublin*, le nom modifié par cette proposition relative *système* (29) n'est plus marqué avec la relation proposition relative. La relation proposition relative `mod@re1` est portée par l'auxiliaire dans la proposition relative *être* (32).

4.1.2 Description des corpus

Les trois corpus utilisés, GSD, Sequoia et ParTUT, sont en français. Ces corpus ont été créés manuellement dans divers schémas d'annotation, puis ils ont été convertis en format UD, puis SUD. Ils ont des licences d'utilisation permettant de les télécharger et de les modifier. Le plus grand corpus est GSD, suivi de Sequoia, et le plus petit est ParTUT. Les trois corpus sont annotés, lemmatisés et les relations sont étiquetées en SUD.

4.1.2.1 Le corpus GSD

Le corpus français GSD¹⁴ vient de blogues, de journaux, de compte-rendu et de Wikipédia. Ce corpus peut être téléchargé sur GitHub¹⁵. Il a été conçu par Marie-Catherine de Marneffe, Bruno Guillaume, Ryan McDonald, Alane Suhr, Joakim Nivre, Matias Gioni, Carly Dickerson et Guy Perrier (Guillaume, de Marneffe, & Perrier, 2019). Il a été publié en 2015 et les dernières modifications ont été faites en novembre 2020, d'après le site GitHub¹⁶. Ce corpus est distribué avec une licence CC BY-SA 4.0¹⁷, c'est-à-dire que nous pouvons télécharger, modifier et partager le corpus avec les nouvelles modifications en mettant la même licence que le corpus original ou une licence compatible.

Le corpus GSD est le corpus le plus grand avec 16 341 phrases, comprenant 3 601 propositions relatives, et il est divisé en trois sections : fr-sud-train avec 14 449 phrases ou 354 662 tokens, fr-sud-dev avec 1 476 phrases ou 35 718 tokens et fr-sud-test avec 476 phrases ou 10 019 tokens.

¹⁴ https://universaldependencies.org/treebanks/fr_gsd/index.html

¹⁵ https://github.com/surfacesyntacticud/SUD_French-GSD

¹⁶ https://github.com/surfacesyntacticud/SUD_French-GSD

¹⁷ <https://creativecommons.org/licenses/by-sa/4.0/>

Les lemmes ont été étiquetés automatiquement avec quelques corrections manuelles. Quant à l'étiquetage des relations, cela a été fait manuellement, puis le corpus a subi quelques modifications manuelles lors de sa conversion en format UD et SUD. Enfin, deux catégories portent l'étiquette Re1 pour relative, les adverbes (ADV) et les pronoms (PRON). Ces adverbes sont *où* et *là*, et ces pronoms sont *qui*, *dont*, *qu'*, *que*, *lequel*, *laquelle*, *lesquelles* et *quoi*.

4.1.2.2 Le corpus Sequoia

Le corpus Sequoia¹⁸ a été conçu à l'origine par Marie Candito, Djamé Seddah (Candito & Djamé, 2012) Il a été ensuite converti en UD par Marie Candito, Djamé Seddah, Guy Perrier et Bruno Guillaume et en SUD par Bruno Guillaume et Guy Perrier. Ce corpus a été publié sur GitHub en 2017 et les dernières modifications ont été faites en novembre 2020. Il contient des extraits de discours en français du parlement européen (originellement en français ou traduits manuellement par des traducteurs professionnels du parlement européen), du Wikipédia français, de la presse régionale *Est républicain*, de l'agence de médecine européenne et d'essais non fictifs. Ce corpus est distribué sous licence LGPL-LR (*Lesser General Public License For Linguistic*)¹⁹, ce qui permet d'utiliser ce corpus, de le partager et le modifier.

Le corpus Sequoia est le deuxième plus grand avec 3 099 phrases, comprenant 602 propositions relatives, ou 68 615 tokens ou 70 567 mots syntaxiques. Ce corpus est réparti en trois sections : *fr_sequoia-sud-test* avec 10 048 tokens ou 456 phrases, *fr_sequoia-sud-dev* avec 10 002 tokens ou 412 phrases et *fr_sequoia-sud-train* avec 50 517 tokens ou 2 231 phrases. Les lemmes et les relations ont été étiquetés manuellement, puis automatiquement du format original en UD et SUD. Enfin, dans les types de pronoms (PronType), il y a les relatifs (Re1), soit *qui*, *que*, *dont*, *où*, *qu'*, *lequel*, *laquelle*, *lesquels*, *lesquelles* et *quoi*.

¹⁸ https://universaldependencies.org/treebanks/fr_sequoia/index.html

¹⁹ <https://spdx.org/licenses/LGPLLR.html>

4.1.2.3 Le corpus ParTUT

Ce corpus²⁰ a été conçu par Cristina Bosco et Manuela Sanguinetti à partir du corpus italien de l’université de Turin (Bosco & Sanguinetti, 2012). Il est basé sur les mêmes dépendances et le même contenu. Les phrases viennent de différents genres de textes, notamment des discours, des textes légaux, des articles Wikipédia. Le corpus est téléchargeable sur le site de GitHub²¹. Ce corpus est sous licence CC BY-NY-SA²².

Le corpus ParTUT est le corpus le plus petit avec 1 020 phrases, 27 658 tokens ou 28 595 mots syntaxiques, avec 301 propositions relatives. Les lemmes et les relations ont été étiquetés manuellement dans le corpus d’origine, puis automatiquement convertis en format UD, puis SUD. Enfin, dans les types de pronoms (PronType), il y a les relatifs (Rel), et à l’intérieur les pronoms sont *qui, que, qu’, dont, où, laquelle, quoi, lequel, auxquelles* et *lesquels*.

4.1.2.4 Synthèse sur les corpus

Ces trois corpus regroupent des genres de textes différents comme des articles de blogues, des récits non fictifs, des journaux, la presse régionale *Est Républicain*, des articles du Wikipédia français, des comptes rendus, des discours, des débats du parlement européen en français et traduits en français par des traducteurs du parlement européen, des textes légaux, des rapports médicaux et des rapports traduits en français de l’agence de médecine européenne. Les licences de ces trois corpus permettent de les télécharger gratuitement, notamment sur le site de GitHub, de les partager, de les modifier et de partager les corpus modifiés gratuitement.

Les trois corpus forment ensemble un corpus de 21 461 phrases, dont 4 505 contiennent une proposition relative en SUD, soit une phrase sur cinq environ. Les lemmes et les relations ont été étiquetés manuellement, puis automatiquement convertis en UD, puis SUD, sauf pour le corpus GSD où les lemmes ont été étiquetés automatiquement. Enfin, les types de pronoms relatifs présents sont *qui, que, qu’, quoi, dont, où, lequel, laquelle, lesquels, lesquelles* et *auxquelles*. Dans

²⁰ https://universaldependencies.org/treebanks/fr_partut/index.html

²¹ https://github.com/UniversalDependencies/UD_French-ParTUT

²² <https://creativecommons.org/licenses/by-nc-sa/4.0/>

le reste de ce mémoire, afin d'alléger le texte, nous utiliserons l'expression « corpus » pour référer à l'union de ces trois corpus.

Quelques traitements des données ont dû être apportés pour exploiter ces corpus dans le cadre de ce projet de recherche, que nous allons détailler ci-dessous.

4.2. Collecte des données

Pour chaque corpus téléchargé, les corpus sont en 3 sections : *train*, *dev* et *test*. Nous avons concaténé toutes les parties de tous les corpus en un seul fichier. Pour l'analyse des pronoms, de leur fonction et de leur fréquence, nous avons un fichier des propositions relatives par pronom venant de GREW²³. Puis nous avons organisé les données sous forme d'un tableau.

GREW est un outil de réécriture de graphes dédié aux applications dans le traitement des langues naturelles. Il peut manipuler de nombreux types de représentation linguistique. Il a été utilisé sur des séquences étiquetées, des parties du discours, la syntaxe de dépendance de surface, la syntaxe de dépendance profonde, la représentation sémantique, mais il peut être utilisé pour représenter n'importe quelle structure à base de graphe. Cet outil est téléchargeable gratuitement permet de traduire des corpus UD en SUD (et vice-versa), ainsi que de rechercher des éléments particuliers dans le corpus comme une relation ou une partie du discours.

4.2.1. Patrons des pronoms

Les patrons en GREW des pronoms et des types de relatives permettent de trouver les bons pronoms ou types de relatives, les phrases, les tokens des mots-formes qui nous intéressent. Les pronoms sont classés en fonction du Tableau 3 des pronoms relatifs en partie 3.3. Ainsi, il y a *qui* et *lequel* en fonction sujet, *que* en fonction objet, et une préposition suivie du pronom *qui*, une préposition suivie du pronom *quoi* et une préposition suivie du pronom *lequel* en fonction de complément prépositionnel. Enfin, des patrons pour GREW ont été faits afin d'avoir la fonction

²³ <https://grew.fr>

du pronom, le pronom, le rôle syntaxique et le type de proposition relative. Voici les patrons GREW des pronoms relatifs.

4.2.1.1. Patron du pronom *qui* sujet

Le patron pour le pronom *qui* sujet en format GREW est présenté ci-dessous.

```
pattern { P [lemma= "qui", PronType=Rel] }  
without { X-[comp:obj]-> P }
```

Le lemme sélectionné est *qui* (*lemma="qui"*) et le type du pronom est *Rel* (*PronType=Rel*), pour *relative*. Pour *qui* sujet, les relations d'objet direct (*comp:obj*) sont exclues ce qui permet de sélectionner toutes les relations sujets (sujet, sujet d'un causatif, sujet d'un passif, sujet d'un auxiliaire).

4.2.1.2. Patron du pronom *lequel*

Le patron pour le pronom *lequel* en format GREW est présenté ci-dessous.

```
pattern { P [lemma="lequel", PronType=Rel] }  
without { X-[comp:obj]-> P }
```

Le lemme sélectionné est *lequel* (*lemma="lequel"*) et le type du pronom est *Rel* (*PronType=Rel*), pour *relative*. Pour *lequel* sujet, les relations d'objet direct (*comp:obj*) sont exclues ce qui permet de sélectionner toutes les relations sujets (sujet, sujet d'un causatif, sujet d'un passif, sujet d'un auxiliaire).

4.2.1.3. Patron du pronom *que*

Le patron du pronom *que* en format GREW est présenté ci-dessous.

```
pattern { P [lemma="que", PronType=Rel] }
```

Pour le patron du pronom *que*, il n'est pas nécessaire de préciser la relation puisque ce pronom n'a qu'une fonction possible.

4.2.1.4. Patron du pronom *dont*

Le patron du pronom *dont* en format GREW est présenté ci-dessous.

```
pattern { P [lemma="dont", PronType=Rel] }
```

Pour le patron du pronom *P dont*, comme pour le patron de *que*, il n'y a pas d'ambiguïté et il n'y a qu'une seule fonction au pronom *dont*, que ce soit pour le *dont* oblique ou le *dont* complément du nom.

4.2.1.5. Patron du pronom *où*

Le patron du pronom *où* en format GREW est présenté ci-dessous.

```
pattern { P [lemma="où", PronType=Rel] }
```

Il en est de même pour le patron du pronom *où*. Il n'y a pas d'ambiguïté, car il n'y a qu'une seule fonction au pronom *où*.

4.2.1.6. Patron d'une préposition suivie du pronom *qui*

Le patron d'une préposition suivie du pronom *qui* en format GREW est présenté ci-dessous.

```
pattern { P [lemma="qui", PronType=Rel] }  
without { V-[subj]-> P }  
without { V-[subj@caus]-> P }  
without { V-[subj@pass]-> P }
```

Les relations sujets (*subj*, *subj@pass* (sujet d'un auxiliaire), *subj@caus* (sujet d'un causatif), entre un verbe *V* et le pronom *P* sont exclues pour n'avoir que le pronom *qui* portant la fonction complément prépositionnel.

4.2.1.7. Patron d'une préposition suivie du pronom *quoi*

Le patron d'une préposition suivie du pronom *quoi* en format GREW est présenté ci-dessous.

```
pattern { P [lemma="quoi", PronType=Rel] }
```

Pour le patron d'une préposition suivie du pronom *quoi*, il n'est pas nécessaire de préciser la relation puisque ce pronom n'a qu'une fonction possible.

4.2.1.8. Patron d'une préposition suivie du pronom *lequel*

Le patron d'une préposition suivie du pronom *lequel* en format GREW est présenté ci-dessous.

```
pattern { P [lemma="lequel", PronType=Rel] }
without { V-[subj]-> P }
without { V-[subj@caus]-> P }
without { V-[subj@pass]-> P }
```

Comme pour le patron d'une préposition suivie du pronom *qui*, le patron P d'une préposition suivie du pronom *lequel*, toutes les relations sujets (*subj*, *subj@pass* (sujet d'un auxiliaire), *subj@caus* (sujet d'un causatif)) entre un verbe V et le pronom P sont exclues pour que les relations compléments puissent être détecter.

4.2.2. Patrons des propositions relatives

4.2.2.1. Patron de la relative directe

Le patron de la proposition relative directe en format GREW est présenté ci-dessous.

```
pattern {
N-[mod@relcl]-> V;
V->P;
P [PronType=Rel];
}
```

La partie du discours n'est pas spécialisée pour les nœuds N, V et P afin de détecter toutes les propositions relatives du corpus. En général, le nœud N est pour le nom modifié par la proposition relative, le nœud V est pour le verbe au sommet de la proposition relative et le nœud P est pour le pronom avec le trait relatif (*PronType=Rel*).

4.2.2.2. Patron des relatives indirectes

Puis, il y a les patrons des propositions relatives indirectes. À cause de contraintes formelles dans GREW, nous devons distinguer plusieurs sous-types selon le nombre de nœuds intermédiaires. Le patron le plus fréquent de proposition relative indirecte est celui avec un seul nœud (prédicat) intermédiaire. Dans le corpus, il n’y a pas de proposition relative indirecte à plus de quatre nœuds (prédicats) X intermédiaires. Voici le patron de proposition relative indirecte à un nœud intermédiaire en GREW.

```
pattern {  
N -[mod@relcl]-> V;  
V -[^mod@relcl]-> X1;  
X1 -[^mod@relcl]-> P;  
P [PronType=Rel];  
}
```

Le principe des patrons pour les relatives indirectes est le même pour tous les patrons, nous ajoutons une ligne pour chaque nœud intermédiaire et la relation ne doit pas être une proposition relative (`^mod@relcl`).

Ces patrons sont enregistrés dans des fichiers destinés à être lus par GREW. GREW permet d’appliquer ces patrons dans les corpus et de générer un fichier de résultat en format JSON²⁴ pour chaque patron. Pour les pronoms relatifs, ces fichiers JSON sont composés du numéro d’identification de chaque phrase, du numéro du token du pronom relatif, du nom modifié, du verbe au sommet de la proposition relative et des éléments X intermédiaires propres aux propositions relatives indirectes.

²⁴ JavaScript Object Notation <https://www.json.org/json-en.html>

4.2.3. Organisation des données

Puis, nous utilisons Python pour lire ces fichiers et compiler les données présentées. Le tableau ci-dessous donne le nombre de résultats retournés par GREW pour chaque patron de pronoms relatifs dans chaque corpus.

Pronoms	Fonction	GSD	Sequoia	ParTUT	Total
qui	subj	2115	340	199	2654
	obj ²⁵	2	5	0	7
	comp	23	7	2	32
que	obj	458	105	60	623
dont	comp	380	62	17	459
où	comp	391	44	17	452
quoi	comp	3	1	3	7
lequel	sujet	12	4	1	17
	comp	206	32	16	254
total		3590	600	315	4505

Tableau 4 Nombre de pronoms relatifs par fonction et par corpus

4.2 Analyse des données

Dans cette partie, la fréquence des pronoms relatifs, le rôle syntaxique des pronoms relatifs, les types de propositions relatives, les parties du discours du nom modifié et du verbe au sommet de la proposition relative sont analysées dans l'ensemble du corpus.

²⁵ Les 7 propositions relatives avec le pronom relatif *qui* objet concernent des locutions

4.2.1 Les parties du discours

Dans l'ensemble du corpus, il y a trois patrons généraux de propositions relatives en fonction des parties du discours de ce qui est modifié et de ce qui modifie. Ces patrons peuvent décrire des propositions relatives explicatives ou restrictives et directes et indirectes.

4.2.1.1 Patron nom-verbe

Le premier patron, qui est le patron canonique et de loin le plus fréquent, est celui où l'élément modifié de la proposition principale est un nom et ce qui modifie ce nom est le verbe au sommet de la proposition relative. Ce patron est présent dans environ 90 % des phrases relatives. Les pronoms relatifs comme *qui*, *que*, *dont*, *où*, *lequel*, *auquel*, *duquel* sont présents dans ce patron de proposition relative. Voici un exemple de phrase avec une proposition relative avec le pronom *qui* et un nom modifié par le verbe au sommet de la proposition relative.

(42) *Les antithrombotiques sont des médicaments qui empêchent la formation de caillots sanguins (thrombose).* [Sequoia, emea-fr-dev_00475]

1	Les	le	DET	2	det	-	-
2	antithrombotiques	antithrombotique	NOUN	3	subj	-	-
3	sont	être	AUX	0	root	-	-
4	des	un	DET	5	det	-	-
5	médicaments	médicament	NOUN	3	comp:pred	-	-
6	qui	qui	PRON	PronType=Rel	7	subj	--
7	empêchent	empêcher	VERB	5	mod@relcl	-	-
8	la	le	DET	9	det	-	-
9	formation	formation.	NOUN	7	comp:obj	-	-
10	de	de	ADP	9	udep	-	-
11	caillots	caillot	NOUN	10	comp:obj	-	-
12	sanguins	sanguin	ADJ	11	mod	-	-
13	((PUNCT	11	punct	-	SpaceAfter=No
14	thrombose	thrombose	NOUN	11	appos	-	SpaceAfter=No
15))	PUNCT	10	punct	-	SpaceAfter=No
16	.	.	PUNCT	3	punct	-	-

Figure 23 Annotation de la phrase (42) en SUD venant de SUD

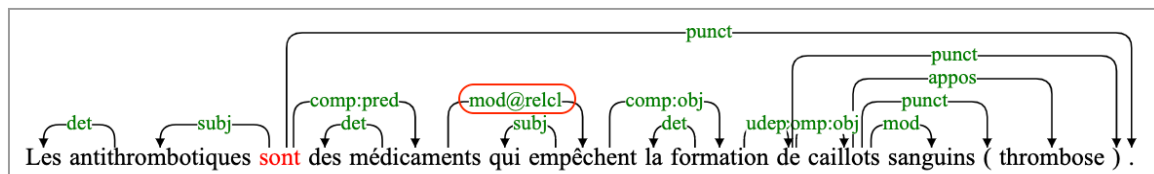


Figure 24 Schéma de la phrase (42) en SUD

Par exemple, dans la phrase (42), l'élément modifié par la proposition relative est le nom *médicaments* (5) et il est modifié par *empêcher* (7), le verbe au sommet de la proposition relative.

4.2.1.2 Patron adverbe — verbe

Puis, il y a un autre patron beaucoup moins fréquent où un adverbe est modifié par un verbe. Le pronom qui introduit la proposition relative pour ce patron est toujours *où*. Les adverbes présents pour ce patron sont des adverbes de temps ou de lieu comme *là*, *ici* ou *partout* et sont caractéristiques de ce patron. Voici un exemple.

(43) *Nous n'avons d'autre choix que d'aller là où ils se trouvent.* [Sequoia, Europar.550_00159]

Voici l'annotation de la phrase entière, mais pour simplifier la phrase seulement la partie de l'exemple (43) nous intéresse ici.

1	Je	il	PRON	2	subj	—	—
2	pense	penser	VERB	0	root	—	—
3	qu'	que	SCONJ	2	comp:obj	—	SpaceAfter=No
4	il	il	PRON	6	subj@expl	—	—
5	s'	se	PRON	6	unk	—	SpaceAfter=No
6	agit	agir	VERB	3	comp:obj	—	—
7	d'	de	ADP	6	comp:obl	—	SpaceAfter=No
8	un	un	DET	Art		9	det --
9	objectif	objectif	NOUN	7	comp:obj	—	—
10	important	important	ADJ	9	mod	—	—
11	parce	parce	SCONJ	6	mod	—	ExtPos=SCONJ
12	que	que	SCONJ	11	unk	—	—
13	si	si	SCONJ	22	mod	—	—
14	nous	il	PRON	15	subj	—	—
15	voulons	vouloir	VERB	13	comp:obj	—	—
16	toucher	toucher	VERB	15	comp:obj@x	—	—
17	les	le	DET	18	det	—	—
18	gens	gens	NOUN	16	comp:obj	—	SpaceAfter=No
19	,	,	PUNCT	22	punct	—	—
20	nous	il	PRON	22	subj	—	—
21	n'	ne	ADV	22	mod	—	SpaceAfter=No
22	avons	avoir	VERB	11	comp:obj	—	—
23	d'	un	DET	25	det	—	SpaceAfter=No
24	autre	autre	ADJ	25	mod	—	—
25	choix	choix	NOUN	22	comp:obj	—	—
26	que	que	SCONJ	25	udep	—	—
27	d'	de	ADP	26	comp:obj	—	SpaceAfter=No
28	aller	aller	VERB	27	comp:obj	—	—
29	là	là	ADV	28	comp:obl	—	—
30	où	où	PRON	PronType=Rel		33	comp:obl --
31	ils	il	PRON	33	subj	—	—
32	se	se	PRON	33	unk	—	—
33	trouvent	trouver	VERB	29	mod@relcl	—	SpaceAfter=No
34	.	.	PUNCT	2	punct	—	—

Figure 25 Annotation de la phrase (43) en SUD venant de SUD

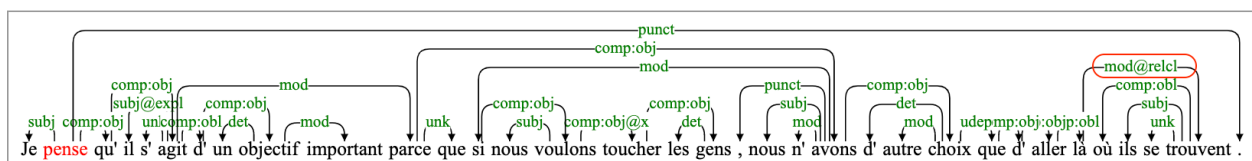


Figure 26 Schéma de la phrase (43) en SUD

La relation de modification est entre l'adverbe *là* (28), qui est l'élément modifié, et le verbe *trouvent* (33), qui est le verbe au sommet de la relative.

4.2.1.3 Patron nom-nom

Enfin, il y a le patron où un nom de la proposition principale est modifié par un nom enchâssé à l'intérieur de la proposition relative. Le pronom utilisé pour ce patron est toujours *dont* et les propositions relatives sont toujours restrictives. Voici un exemple qui illustre ce patron.

(44) *Plusieurs centaines de milliers d'Irakiens, dont quelque 500 000 enfants, sont morts prématurément en raison de l'embargo.* [Sequoia, Europar.550_00240]

Voici l'annotation de la phrase entière, mais pour simplifier la phrase seulement la partie de l'exemple (44) est pertinente ici.

1	Selon	selon	ADP	21	mod		
2	les	le	DET	3	det		
3	agences	agence	NOUN	1	comp:obj		
4	humanitaires	humanitaire	ADJ	3	mod		
5	de	de	ADP	3	udep		
6	l'	le	DET	7	det		SpaceAfter=No
7	ONU	O.N.U.	PROPN	5	comp:obj		SpaceAfter=No
8	,	,	PUNCT	21	punct		
9	plusieurs	plusieurs	DET	10	det		
10	centaines	centaine	NOUN	21	subj		
11	de	de	ADP	10	udep		
12	milliers	millier	NOUN	11	comp:obj		
13	d'	de	ADP	12	udep		SpaceAfter=No
14	Irakiens	irakien	NOUN	13	comp:obj		SpaceAfter=No
15	,	,	PUNCT	10	punct		
16	dont	dont	PRON		PronType=Rel	19	mod
17	quelque	quelque	ADV	19	mod		--
18	500 000	500 000	DET	19	det		
19	enfants	enfant	NOUN	10	mod@relcl		SpaceAfter=No
20	,	,	PUNCT	21	punct		
21	sont	être	AUX	0	root		
22	morts	mourir	VERB	21	comp:aux@tense		
23	prématurément	prématurément	ADV			21	mod
24	en	en	ADP	21	mod		--
25	raison	raison	NOUN	24	comp:obj		
26	de	de	ADP	25	udep		
27	l'	le	DET	28	det		SpaceAfter=No
28	embargo	embargo	NOUN	26	comp:obj		SpaceAfter=No
29	,	,	PUNCT	21	punct		
30	faute	faute	ADV	21	mod		

31	de	de	ADP	30	comp:obl	—	—
32	produits	produit	NOUN	31	comp:obj	—	—
33	alimentaires	alimentaire	ADJ	32	mod	—	SpaceAfter=No
34	,	,	PUNCT	35	punct	—	—
35	de	de	ADP	31	conj	—	—
36	médicaments	médicament	NOUN	35	comp:obj	—	—
37	et	et	CCONJ	38	cc	—	—
38	de	de	ADP	35	conj	—	—
39	soins	soin	NOUN	38	comp:obj	—	—
40	adéquats	adéquat	ADJ	39	mod	—	SpaceAfter=No
41	.	.	PUNCT	21	punct	—	—

Figure 27 Annotation de la phrase (44) en SUD venant de SUD

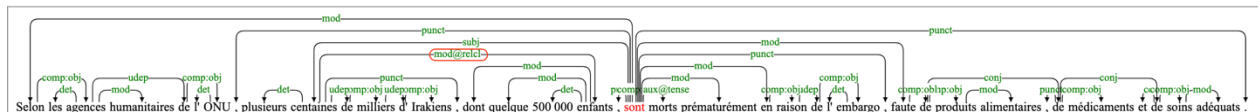


Figure 28 Schéma de la phrase (44) en SUD

Dans cette phrase, la modification est entre le nom *irakien* (14) qui est modifié par le nom *enfants* de la proposition relative. Ce patron est détecté par GREW mais il ne nous intéresse pas s'il n'y a pas de verbe fini dans la proposition relative. Pour implémenter les propositions relatives dans GenDR, pour le moment, nous avons besoin d'un verbe fini dans la proposition relative.

4.2.2 Analyse des pronoms

Nous allons analyser la distribution des pronoms relatifs et de leur rôle syntaxique dans les corpus. Les pronoms sont les pronoms relatifs et la fréquence est la fréquence absolue. Il n'y a pas de fréquence proportionnelle au nombre de pronoms, rôles syntaxiques, corpus ou types de relative dans les figures suivantes. Les graphes montrent leurs fréquences dans le corpus.

4.2.2.1 Les pronoms, fonctions et corpus

Dans un premier temps, voici un graphique qui montre la répartition des pronoms relatifs en fonction de leur rôle syntaxique par corpus.

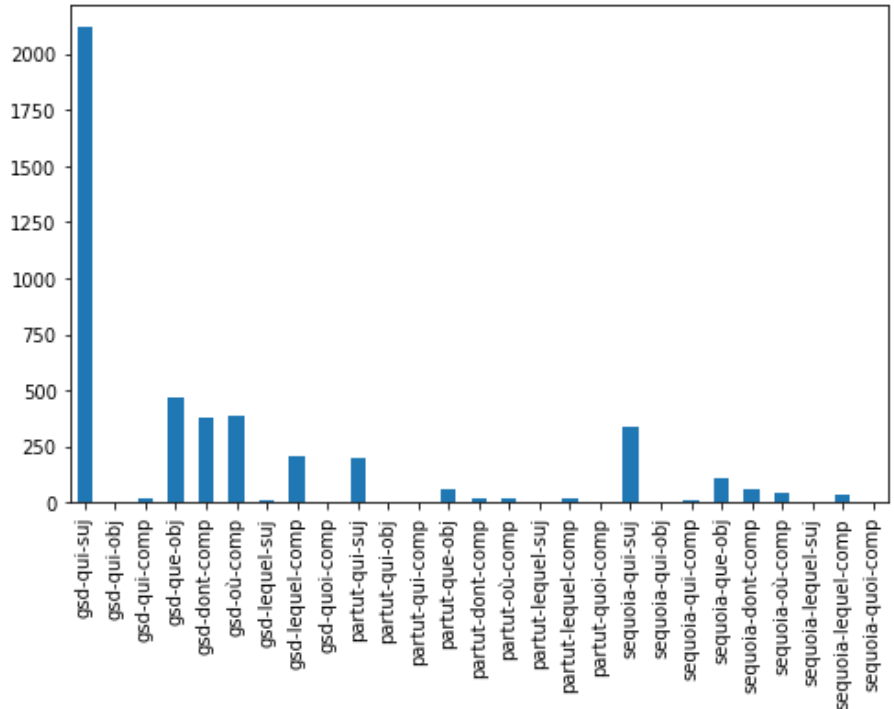


Figure 29 Occurrence des pronoms par rôle syntaxique et corpus

Ce corpus nous donne une vision d'ensemble des corpus, des pronoms et de leur rôle syntaxique. Dans ce graphique, le pronom *qui* en fonction sujet dans le corpus GSD est significativement plus élevé par rapport à tous les autres pronoms. Les pronoms les plus fréquents sont *qui* en fonction sujet (suj), *que* en fonction objet (obj), *dont*, *où* et *lequel* en fonction complément prépositionnel (comp) pour le corpus GSD. Puis, le pronom *qui* est le plus fréquent dans le corpus Sequoia. Donc, le pronom *qui* en fonction sujet est le plus fréquent.

4.2.2.2 Les fonctions

Dans la figure ci-dessous, les pronoms relatifs ont été regroupés par rôle syntaxique.

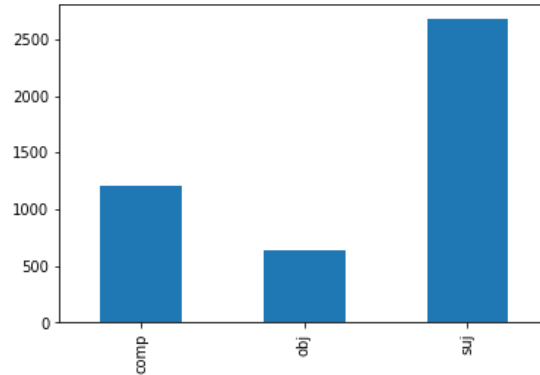


Figure 30 Occurrence des rôles syntaxiques des pronoms

Nous pouvons observer que la fonction la plus fréquente pour les pronoms relatifs est sujet (*suj*), puis vient complément prépositionnel (*comp*), enfin object (*obj*). Même s’il existe dans la langue beaucoup plus de pronoms qui peuvent être des compléments prépositionnels (*dont*, *où*, une préposition suivie de *qui*, une préposition suivie de *quoi*, une préposition suivie de *lequel*), les pronoms sujets sont significativement plus fréquents.

4.2.2.3 Les pronoms sujets

Nous avons vu dans les figures précédentes que les pronoms de rôle syntaxique sujet sont plus fréquents que les autres pronoms relatifs. Voici la figure suivante qui montre la fréquence des pronoms sujets *qui* et *lequel*.

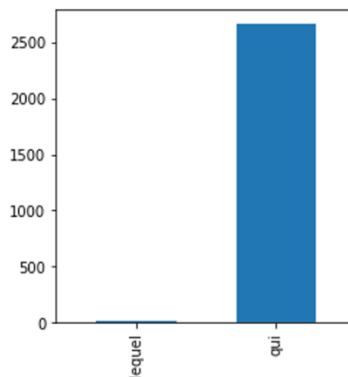


Figure 31 Occurrence des pronoms sujets

D'après la figure ci-dessus, le pronom relatif *qui* est significativement plus fréquent que le pronom relatif *lequel*. Donc avec les figures suivantes, nous pouvons déduire que l'implémentation du pronom *qui* de rôle syntaxique sujet est la plus importante à générer pour les propositions relatives.

4.2.2.4. Les pronoms objets

D'après le Tableau 3, des pronoms relatifs (§ 3.3) les pronoms relatifs de rôle syntaxique objet direct sont *que* et *qui*. Dans la Figure 29, nous pouvons constater que le pronom relatif *que* est parmi les pronoms relatifs les plus fréquents. Le pronom *que* est présent 623 fois sur 4 505 propositions relatives, soit 13,8 %. Il n'y a que sept propositions relatives avec le pronom relatif *qui* objet dans tout le corpus. Ces sept propositions sont des expressions figées ou des locutions, ce qui n'est pas traité dans ce sujet de recherche. Il y a quelques propositions relatives avec le pronom *que* qui sont marquées comme sujet, mais ce sont des inversions du sujet verbe qui ont été mal annotés par le logiciel d'annotation automatique. Après l'implémentation du pronom relatif *qui* sujet, l'implémentation du pronom relatif *que* objet est importante pour générer le plus de propositions relatives possible.

4.2.2.5. Les pronoms compléments prépositionnels

Les pronoms de rôle syntaxique complément prépositionnel sont les deuxièmes plus fréquents et sont les plus variés. Voici un graphique qui présente la fréquence des pronoms relatifs de rôle syntaxique complément prépositionnel.

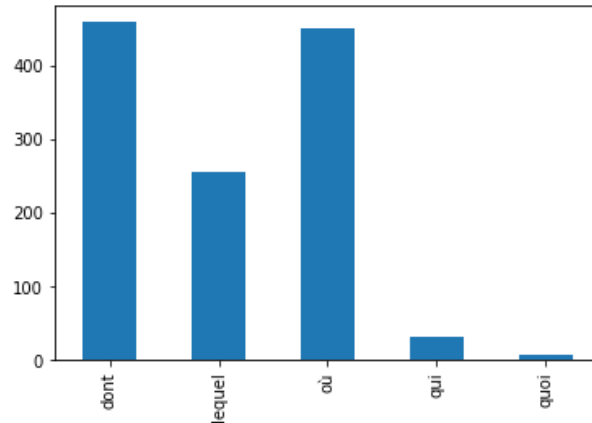


Figure 32 Occurrence des pronoms compléments prépositionnels

Le pronom *dont* est le plus fréquent suivie de près par *où*, puis dans une moindre mesure *lequel* (*préposition + lequel*), enfin, *qui* (*préposition + qui*) et *quoi* (*préposition + quoi*).

4.2.2.6. Les pronoms par fonction

Nous pouvons observer la proportion de chaque pronom par rôle syntaxique. Le tableau suivant présente les pronoms relatifs par rôle syntaxique.

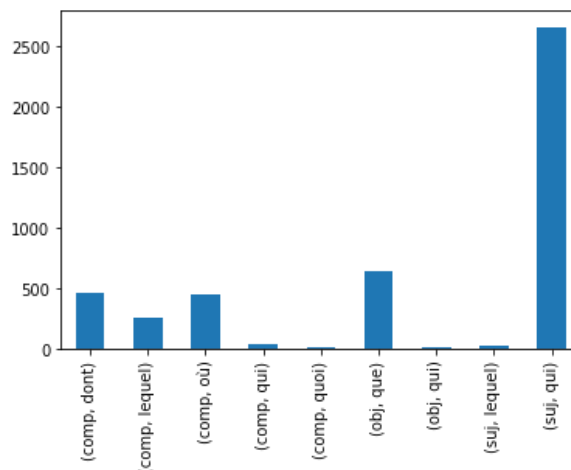


Figure 33 Occurrence des pronoms par rôle syntaxique

Dans ce graphique, le pronom *qui* sujet est significativement plus fréquent que les autres pronoms. Les pronoms *que* objet, *dont* complément prépositionnel, *où* complément prépositionnel et *lequel* complément prépositionnel n'ont pas une différence significative entre

eux. Les pronoms *lequel* sujet, *qui* objet, *qui* complément prépositionnel et *quoi* complément prépositionnel sont moins fréquents comparés aux autres pronoms. Il est donc important d'implémenter en premier le pronom *qui* sujet, mais aussi *que* objet, *dont* complément prépositionnel, *où* complément prépositionnel et *lequel* complément prépositionnel.

4.2.3. Analyse du rôle syntaxique

4.2.3.1. Le rôle syntaxique des pronoms dans le corpus

Nous avons observé les étiquettes des rôles syntaxiques des pronoms dans tout le corpus. Les relations sont plus spécifiques et variées que juste sujet, objet et complément prépositionnel. Voici un tableau qui présente la fréquence des rôles syntaxiques des pronoms relatifs.

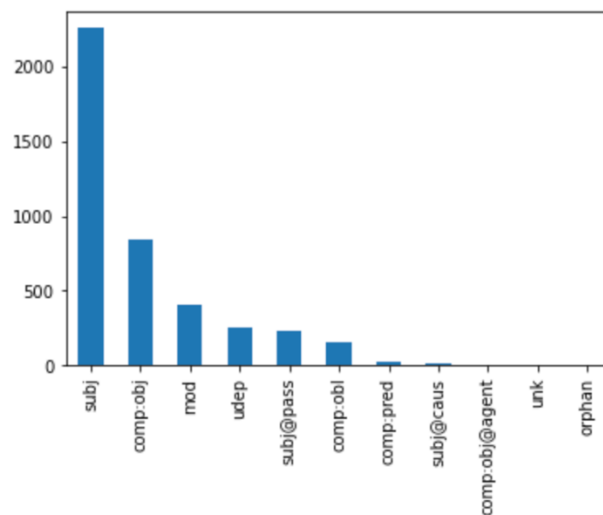


Figure 34 Occurrence des rôles syntaxiques des pronoms

L'étiquette sujet (`subj`, `subj@pass` [sujet d'un auxiliaire] et `subj@caus` [sujet d'un causatif]) est la plus fréquente. Les étiquettes `mod` sont pour les modificateurs de verbes, noms, adverbes, adjectives, auxiliaires, adpositions, appositions et conjonctions, `udep` pour les compléments, `unk` pour les relations inconnues, `orphan` pour orphelines. Toutes les relations `comp` : sont des relations spécifiques de l'étiquette `comp`, qui est utilisée pour les arguments en général. Les étiquettes `comp:obj` sont pour les objets directs, `comp:obl` pour les

arguments obliques, `comp:prep` pour les arguments prépositionnel, `comp:obj@agent` pour les objets agents. La relation qui est significativement plus fréquente est celle de sujet, qui correspond surtout au pronom *qui*. Puis, en deuxième il y a la relation `comp:obj` ou objet représenté par le pronom *que*. Enfin, les relations `mod`, `udep`, `subj@pass` et `comp:obl` sont aussi présents dans une moindre mesure. Quant aux relations `comp:prep`, `subj@caus`, `comp:obj@agent`, `unk` et `orphan`, elles sont presque inexistantes. Ces données viennent confirmer les résultats des figures précédentes où le pronom *qui* est le plus fréquent, puis *que*, *dont*, *où* et les autres pronoms relatifs.

4.2.3.2. Le rôle syntaxique du nom modifié par la proposition relative

Dans cette figure, le rôle syntaxique du nom modifié par la proposition relative est présenté ci-dessous. Nous pouvons observer le rôle syntaxique du nom modifié par la proposition relative pour déduire si cela a un impact sur le pronom relatif ou le type de proposition relative.

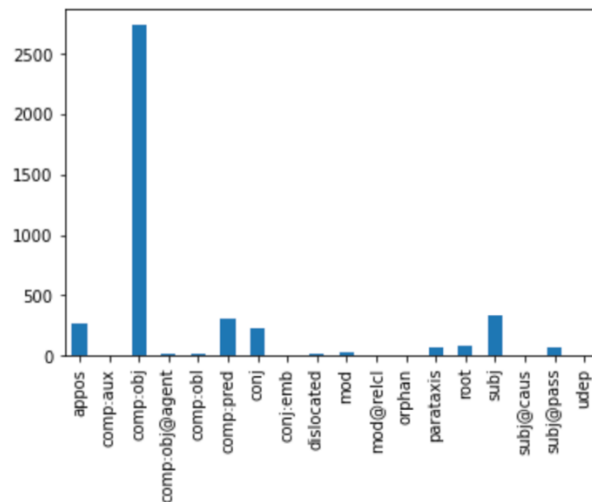


Figure 35 Occurrence du rôle syntaxique du nom modifié par la relative

Dans ce schéma, la relation des noms modifiés par une proposition relative significativement supérieure aux autres est `comp:obj`, c'est-à-dire que les noms sont des arguments objets. Cependant, même si c'est la relation de loin la plus fréquente, les noms modifiés peuvent être aussi `appos` (apposition), `subj` (sujet), `comp:pred` (les arguments prépositionnel), `conj` (conjonction), `parataxis` (parataxe), `root` (racine), `subj@pass` (sujet d'un

passif) et dans une nettement moindre mesure `comp:obj@agent` (complément d'agent), `comp:obl` (complément d'objet direct), `dislocated` (dislocation), `mod` (modificateur).

4.2.4. Analyse du type de proposition relative

4.2.4.1. La fréquence des propositions relatives directes et indirectes

Dans l'analyse de ce corpus, nous avons trouvé intéressant de mesurer la fréquence des propositions relatives directes et indirectes dans tout le corpus. Voici une figure qui représente la proportion entre les propositions relatives directes et indirectes.

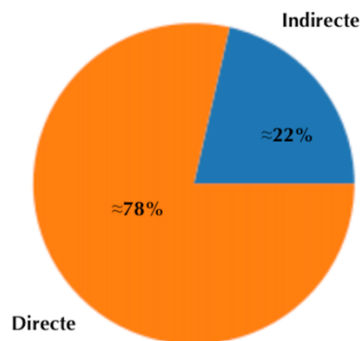


Figure 36 Proportion des relatives directes et indirectes

Les propositions relatives directes sont significativement plus fréquentes que les propositions relatives indirectes avec environ 78% de propositions relatives directes et environ 22% de toutes les propositions relatives indirectes (peu importe le nombre de nœuds intermédiaires) dans tout le corpus.

4.2.4.2. Les pronoms relatifs en fonction du type de proposition relative

Ce tableau de données illustre la fréquence des pronoms relatifs en fonction du caractère direct ou indirect de propositions relatives.

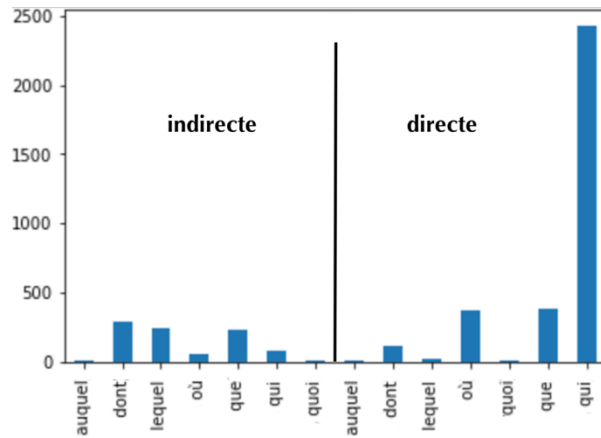


Figure 37 Occurrence des pronoms en fonction de la relative

Le pronom *qui* dans les propositions relatives directes est significativement plus élevé que tous les autres pronoms relatifs dans les deux types de propositions relatives. Puis, les pronoms *que* et *où* dans les propositions relatives directes, et les pronoms *dont*, *lequel*, *que* dans les propositions relatives indirectes qui sont les plus fréquents. Enfin, il y a les pronoms *auquel*, *où*, *qui*, *quoi* pour les propositions relatives indirectes et *auquel*, *dont*, *lequel*, *pourquoi* pour les propositions relatives directes qui sont presque inexistantes. De plus, ce graphique montre que les pronoms *qui* et *où* ont plus de chances de se trouver dans une proposition relative directe, tandis que *dont* et *lequel* ont plus de chance d’être dans une proposition relative indirecte. Cependant, les pronoms *que*, *auquel* et *quoi* ont la même fréquence dans une proposition relative directe qu’indirecte.

4.2.4.3. Le rôle syntaxique du nom modifié par rapport au type de relative

Pour compléter l’observation faite dans la Figure 35, le rôle syntaxique des noms modifiés peut être étudié en fonction des propositions relatives directes ou indirectes. La figure ci-dessous présente la fréquence des rôles syntaxiques pour le nom modifié selon le type de relative qui le modifie.

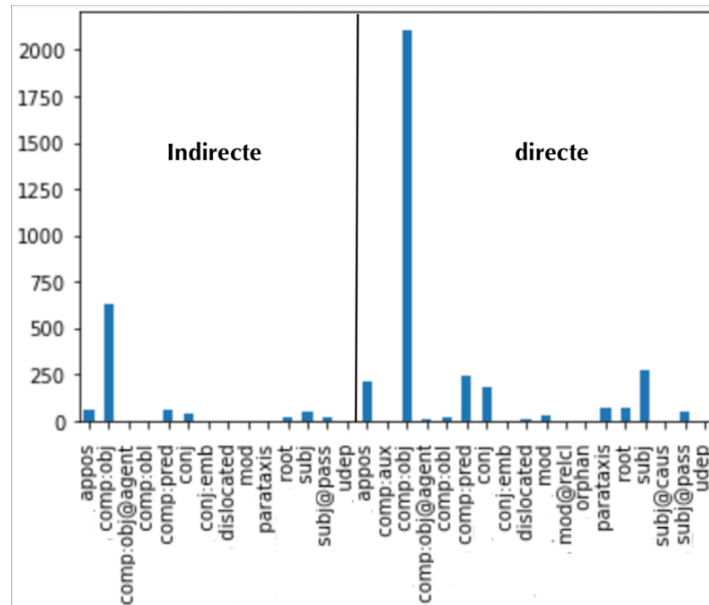


Figure 38 Relation des noms en fonction des relatives directs et indirects

Il y a plus de diversité dans les relations du nom modifié pour les propositions relatives directes que pour les propositions relatives indirectes, ce qui peut être expliqué par la fréquence plus élevée des propositions relatives directes. La relation significativement plus fréquente est `comp:obj` que les autres rôles syntaxiques comme pour la Figure 35 qui montre uniquement les relations des noms modifiés. Nous pouvons déduire que la relation du nom modifié n’influence pas le type de proposition relative qui suit ce nom modifié.

4.2.4.4. Synthèse de l’analyse de données du corpus

Pour résumer l’analyse sur le corpus, il y a significativement plus de propositions relatives directes environ 78 % que de propositions relatives indirectes environ 22 %. Puis, le rôle syntaxique du nom modifié le plus fréquent `comp:obj` ou objet. La relation du nom modifié n’influence pas le type de proposition relative qui suit. Voici un tableau récapitulatif du nombre de pronoms relatifs et de leurs pourcentages dans le corpus, à partir des données sur le Tableau 4.

Pronoms relatifs	Nombre dans le corpus	% dans le corpus
Qui sujet	2653	58,8%
Qui objet	13	0,3%

Qui comp	32	0,7%
Que objet	623	13,8%
Dont comp	459	10,2%
Où comp	452	10%
Lequel sujet	17	0,4%
Lequel comp	256	5,7%
Quoi comp	7	0,1%
Total	4 512	100%

Tableau 5 Nombre et pourcentages des pronoms par fonction dans le corpus

Donc en priorité, il faudra implémenter les propositions relatives directes, puis les pronoms relatifs *qui, que, dont, où*, une préposition suivie de *lequel*, une préposition suivie de *qui, lequel* et une préposition suivie de *quoi*.

Nous pouvons en déduire que les phrases avec des propositions relatives les plus représentatives sont des propositions relatives directes avec le pronom relatif *qui* sujet et la relation comp : obj pour le nom modifié par la proposition relative. De plus, les pronoms *qui* et *où* se retrouvent plus souvent dans une proposition relative directe, tandis que *dont* et *lequel* se retrouvent plus souvent dans une proposition relative indirecte, et les pronoms *que, auquel* et *quoi* sont aussi fréquents dans une proposition relative directe qu'indirecte. Autrement dit, les phrases les plus représentatives du corpus sont comme dans ces exemples (42) ou (45).

(45) *L'album de photos de police exposé au musée d'Histoire politique de la Russie de Saint-Pétersbourg révèle le visage de Raspoutine défoncé par des coups et son corps avec quatre points d'impacts de balles qui ont traversé le cœur, le cou et le cerveau.*

[GSD, fr-sud-dev_00025]

1	L'	le	DET	2	det	—	SpaceAfter=No wordform=1'
2	album	album	NOUN	19	subj	—	—
3	de	de	ADP	2	udep	—	—
4	photos	photo	NOUN	3	comp:obj	—	—
5	de	de	ADP	4	udep	—	—
6	police	police	NOUN	5	comp:obj	—	—
7	exposé	exposer	VERB	2	mod	—	—
8-9	au					—	—
8	à	à	ADP	7	comp:obl	—	—
9	le	le	DET	10	det	—	—
10	musée	musée	NOUN	8	comp:obj	—	—
11	d'	de	ADP	10	udep	—	SpaceAfter=No

12	Histoire	histoire	NOUN	11	comp:obj	—	wordform=histoire	
13	politique	politique	ADJ	12	mod	—	—	
14	de	de	ADP	12	udep	—	—	
15	la	le	DET	16	det	—	—	
16	Russie	Russie	PROPN	14	comp:obj	—	—	
17	de	de	ADP	16	udep	—	—	
18	Saint-Petersbourg	Saint-Petersbourg	PROPN	17	comp:obj	—	—	
19	révèle	révéler	VERB	0	root	—	—	
20	le	le	DET	21	det	—	—	
21	visage	visage	NOUN	19	comp:obj	—	—	
22	de	de	ADP	21	udep	—	—	
23	Raspoutine	Raspoutine	PROPN	22	comp:obj	—	—	
24	défoncé	défoncer	VERB	21	mod	—	—	
25	par	par	ADP	24	comp:obl@agent	—	—	
26	des	un	DET	27	det	—	—	
27	coups	coup	NOUN	25	comp:obj	—	—	
28	et	et	CCONJ	30	cc	—	—	
29	son	son	DET	30	det	—	—	
30	corps	corps	NOUN	21	conj	—	—	
31	avec	avec	ADP	30	udep	—	—	
32	quatre	quatre	NUM	33	det	—	—	
33	points	point	NOUN	31	comp:obj	—	—	
34	d'	de	ADP	33	udep	—	SpaceAfter=No	
35	impacts	impact	NOUN	34	comp:obj	—	—	
36	de	de	ADP	35	udep	—	—	
37	balles	balle	NOUN	36	comp:obj	—	—	
38	qui	qui	PRON	39	subj	—	—	
39	ont	avoir	AUX	37	mod@relcl	—	—	
40	traversé	traverser	VERB	39	comp:aux@tense	—	—	
41	le	le	DET	42	det	—	—	
42	cœur	cœur	NOUN	40	comp:obj	—	SpaceAfter=No	
43	,	,	PUNCT	45	punct	—	—	
44	le	le	DET	45	det	—	—	
45	cou	cou	NOUN	42	conj	—	—	
46	et	et	CCONJ	48	cc	—	—	
47	le	le	DET	48	det	—	—	
48	cerveau	cerveau	NOUN	45	conj	—	SpaceAfter=No	
49	.	.	PUNCT	19	punct	—	—	

Figure 39 Annotation de l'exemple (45)

Cet exemple a une proposition relative directe, avec le pronom relatif *qui* sujet (38) directement actant du verbe au sommet de la relative (39) et le nom modifié *balles* (37) portant la relation `comp:obj`.

4.3 Synthèse

La sélection des corpus s'est portée sur les trois corpus français GSD, Sequoia et ParTUT annotés en dépendances en SUD, car ils étaient déjà constitués, nettoyés, annotés et gratuits à télécharger, à modifier et à redistribuer.

Les informations de l'analyse des données des corpus aideront à guider l'implémentation des règles sur les propositions relatives. En effet, le patron nom-verbe est le plus courant pour une

proposition relative où un nom dans la proposition principale est modifié par un verbe de la proposition relative. Les propositions relatives directes sont les plus fréquentes (environ 72 %). Le rôle syntaxique du nom modifié le plus fréquent est `comp:obj`. Il faudra chercher à implémenter en premier le pronom relatif *qui* (58,8 %), puis, *que* (13,8 %), *dont* (10,2 %), *où* (10,0 %), une prépositionnel suivie de *lequel* (5,7 %), et les autres pronoms relatifs pour assurer la génération d'un plus grand nombre de propositions relatives.

Dans le chapitre suivant, nous présenterons notre implémentation.

Chapitre 5 — Implémentation

L'implémentation, ici, consiste à écrire des règles syntaxiques pour traiter les propositions relatives dans GenDR. Autrement dit, les théories linguistiques sur les propositions relatives seront traduites et écrites sous forme de règles formelles pour que GenDR puisse générer les bonnes structures. Comme il a été expliqué dans les parties précédentes 2.1.3 et 3.2, les propositions relatives s'articulent sur plusieurs niveaux en TST: la représentation sémantique, la représentation syntaxique profonde et la représentation syntaxique de surface. Ces différents niveaux se traduisent en deux interfaces dans GenDR, l'interface sémantique-syntaxe profonde et l'interface syntaxe profonde-syntaxe de surface. Suivant les conclusions de notre étude des corpus au chapitre 4, partie 4.3, notre objectif est de recréer le patron le plus fréquent, c'est-à-dire un nom modifié par un verbe, et de traiter les pronoms les plus fréquents comme *qui*, *que* et *dont*.

5.1 Interface sémantique-syntaxe profonde

Dans les chapitres 2 et 3, nous vous avons présenté la TST (§2.1), et l'interface sémantique-syntaxe. L'interface sémantique-syntaxe profonde est traitée en premier en commençant par la règle de Dubinskaite sur les propositions relatives.

5.1.1 La règle de Dubinskaite

Dans son mémoire, Dubinskaite (2017, pp. 89-91) a créé une règle pour les propositions relatives dans l'interface sémantique-syntaxe profonde. Cependant, cette règle ne pouvait générer que peu de phrases, et devait donc être améliorée afin de pouvoir s'appliquer à un plus grand nombre de propositions relatives. La règle de Dubinskaite est la suivante et son schéma est la Figure 41.

```

Sem<=>DSynt disabled relative_lex : relatives
leftside (v)
l:?Xl{ // main predicate of the relative
  l:?r-> ?Yl{ // noun to be modified
}
?L <- semanticon::(?Xl.sem).(lex)
rightside
rc:?Yr{ // noun to modify
rc:<=> ?Yl
ATTR-> ?Xr{ // main predicate of the relative
  <=> ?Xl
  dlex=?L
  dpos=lexicon::(?L).(dpos)
  dsynt=OK
}
}
conditions (≡)
?Yr.dsynt=OK;
//not ?Xr.dsynt=OK;
semanticon::(?Xl.sem).(lex);
lexicon::(?L).(dpos)=V;
not ?r=coref;

```

Figure 40 La règle des relatives de Dubinskaite

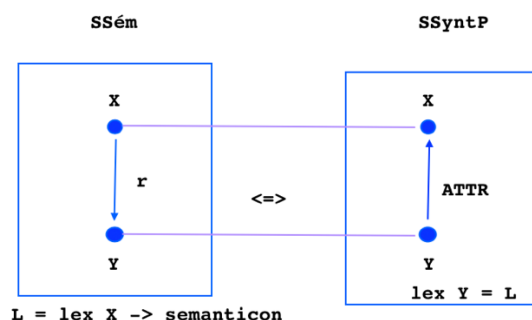


Figure 41 Schéma de la règle des relatives de Dubiskaite

Ce schéma représente la règle de Dubinskaite dans l'interface sémantique-syntaxe profonde du système de règles de GenDR. À droite, au niveau sémantique, nous avons un nœud X qui a une relation r avec un nœud Y. Le nœud X est lexicalisé à partir du dictionnaire sémantique. Cela est représenté à gauche par un nœud Y qui a une relation ATTR (de modification) avec un nœud X. Le nœud Y est aussi lexicalisé à partir du dictionnaire sémantique.

Par exemple, nous avons la phrase (46) avec sa RSém en la Figure 42. Ci-dessous, nous avons la RSyntP de la phrase (46) en Figure 43 et la RSyntS de la phrase (46) en Figure 44 générées par la règle de Dubinskaite pour les relatives.

(46) *Les antithrombotiques sont des médicaments qui empêchent la formation de caillot sanguin.* [Sequoia, sequoia_emea-fr-dev_00475]

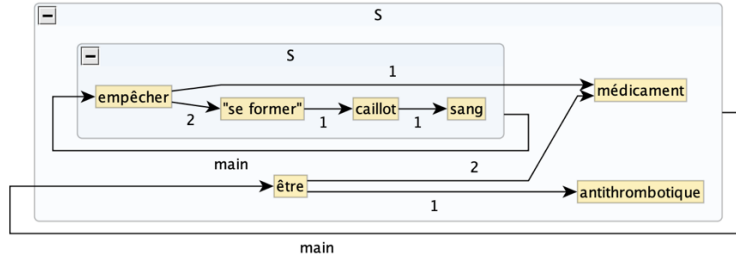


Figure 42 RSém de la phrase (46)

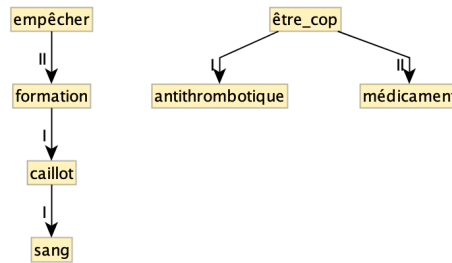


Figure 43 RSyntP de la phrase (46) avec la règle de Dubinskaite

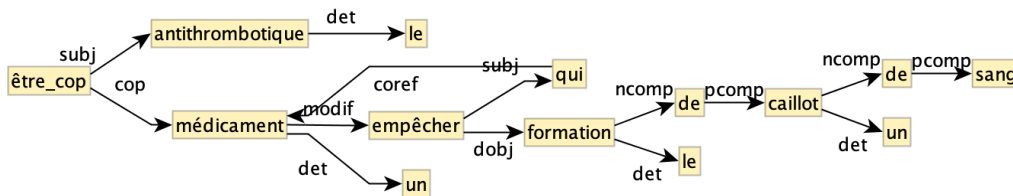


Figure 44 RSyntS de la phrase (46) avec la règle de Dubinskaite

Avec la règle de Dubinskaite sur les relatives, GenDR traite cette phrase comme deux phrases le niveau de syntaxe profonde et en fait de même pour le niveau de syntaxe de surface. GenDR produit deux arbres, un pour la proposition principale *les antithrombotiques sont des médicaments* et un autre pour la proposition relative *qui empêchent la formation d'un caillot sanguin*. À cause de l'implémentation des noms dans GenDR, le nom n'arrive pas à générer une relation ATTR avec le verbe. Il y a aucun pronom relatif, car les pronoms relatifs apparaissent en syntaxe de surface et il n'y a aucune règle dans le dictionnaire de règles grammaticales dans l'interface syntaxe profonde- syntaxe de surface qui permettrait de générer un pronom relatif. Cependant il n'y a pas de sujet pour le verbe au sommet de la proposition relative.

5.1.2. La règle `relative_direct`

Nous proposons deux nouvelles règles, `relative_directe` et `relative_indirecte`, qui s'appuient sur le principe qu'une proposition relative est une proposition dans une autre proposition. Cela se traduit par une bulle S pour la proposition relative, dans une autre bulle S qui est la proposition principale. Dans le chapitre 3, nous avons proposé une définition de la proposition relative (§3.4.1), ainsi que de la proposition relative directe (§3.4.2) et de la proposition relative indirecte (§3.4.3).

Donc quand le lien sémantique entre le nom modifié²⁶ de la proposition principale²⁷ et le verbe au sommet de la proposition relative est direct, la règle `relative_direct` s'applique. La règle des patrons de régime `actant_gp` a été légèrement modifiée comme il est expliqué dans la partie 5.1.4, afin d'avoir deux propositions (bulles S) et de permettre à la règle `relative_direct` de s'appliquer.

Nous avons dans la partie précédente que la phrase (46) n'est pas générée correctement, mais en deux phrases sans la relation ATTR avec la règle de Dubinskaite. Cette phrase (46) est un exemple de proposition relative directe et la Figure 42 est sa RSém. La proposition principale est *les antithrombotiques sont des médicaments* et la proposition relative est *qui empêchent la formation de caillot sanguin*. La proposition relative modifie le nom *médicament*. Le lien sémantique entre le nom modifié *médicament* et le verbe au sommet de la proposition relative *empêcher* est direct.

La règle des relatives directes et son schéma sont présentés ci-dessous.

²⁶ *Nom* est utilisé pour simplifier les choses, cela peut être le nom en général, donc il peut s'agir d'un nom, d'un pronom, d'un nom propre, ou d'autre chose comme un adverbe par exemple.

²⁷ *Proposition principale* est utilisée pour simplifier les choses, cela s'applique aussi aux propositions subordonnées

```

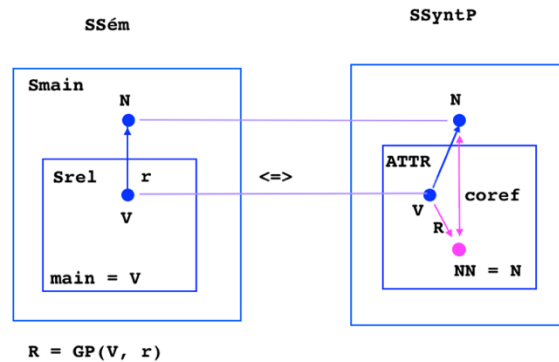
Sem<=>DSynt relative_direct : relatives
leftside (v)
?Smain{           // main clause
  ?Nl             // noun modified by relative
  ?Srel{         // subordinate clause
    ?Vl{        // main verb of the relative
      ?r-> ?Nl  // N is an argument of V
    }
    main-> ?Vl
  }
}
?R <- lexicon::(?Vr.dlex).(gp).(?r) // retrieve V's GP

rightside
rc:?Nr{
  rc:<=> ?Nl
  ATTR->rc:?Vr{
    rc:<=> ?Vl
    ?R-> ?NN{    // create a copy of Nr
      <-> rc:?Nr // coref with Nr
      dlex=?Nr.dlex // recopy all attributes from Nr
      dpos=?Nr.dpos
      det=?Nr.det
      definiteness=?Nr.definiteness
      number=?Nr.number
      case=?Nr.case
      relpro=yes
      animacy=?Nr.animacy
    }
  }
}

conditions (a)
?Smain.sem=S; // Bubble for the main clause
?Srel.sem=S; // Bubble for the relative clause
?Nr.dsyntax=OK; // Wait for full lexicalization
?Vr.dsyntax=OK; // Wait for full lexicalization

```

Figure 45 Règle relative_direct



$$R = GP(V, r)$$

Figure 46 Schéma de la règle relative_direct

Ce schéma représente la règle des propositions relatives directes dans GenDR dans l'interface sémantique (à gauche) syntaxe profonde (à droite). Le grand carré (Smain) illustre la proposition principale et le petit carré (Srel) illustre la proposition relative. À gauche, nous avons un nœud V dans la proposition relative qui a une relation r avec le nœud N dans la proposition principale. Cela est représenté à droite par le nœud V de la proposition relative, qui a une relation ATTR (de modification) avec le nœud N dans la proposition principale. Ce nœud V a une nouvelle relation R avec un nouveau nœud NN dans la proposition relative. Ce nœud NN est coréférent avec le nœud N de la proposition principale. La relation R est créée à partir du patron de régime du nœud V.

En d'autres termes, si nous reprenons la phrase (46) et la Figure 42 représentant sa RSém. À gauche de la règle, nous avons la proposition principale (Smain) *les antithrombotiques sont des médicaments*, avec *médicament* pour nœud N. Dans le petit carré représentant la proposition relative (Srel) *qui empêchent la formation de caillot sanguin*, nous avons le nœud V *empêcher*. Ce nœud V *empêcher* a une relation r avec le nœud N *médicament*.

Ensuite, si nous reprenons la phrase (46) et la Figure 47 représentant sa RSyntP.

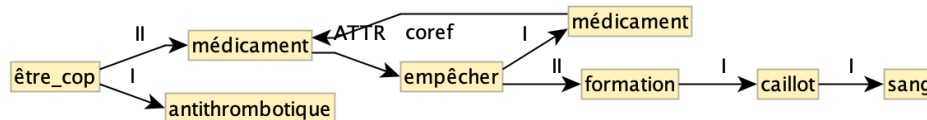


Figure 47 RSyntP de la phrase (46) avec la règle *relative_direct*

À droite de la règle, nous avons toujours la proposition principale *les antithrombotiques sont des médicaments* avec le nœud N *médicament*. Le nœud V *empêcher* de la proposition relative *qui empêchent la formation de caillot sanguin* a une relation ATTR (de modification) avec le nœud N *médicament*. Ce nœud V *empêcher* a une nouvelle relation R avec un nouveau nœud NN *médicament* coréférent et lexicalisé à partir du nœud N *médicament* de la proposition principale.

5.1.3. La règle *relative_indirect*

Quand il y a un ou plusieurs prédicat(s) entre le nom modifié et le verbe au sommet de la proposition relative, c'est un lien sémantique indirect et la règle *relative_indirecte* qui s'applique. La règle des patrons de régime *actant_gp* a été légèrement modifiée comme il est expliqué dans la partie 5.1.4, afin d'avoir deux propositions (bulles S) et de permettre à la règle *relative_indirecte* de s'appliquer.

Voici un exemple de relative indirecte avec la phrase (47) et sa RSém en Figure 48.

(47) *J'ai lu le livre que Luc pense que Marie aime.*

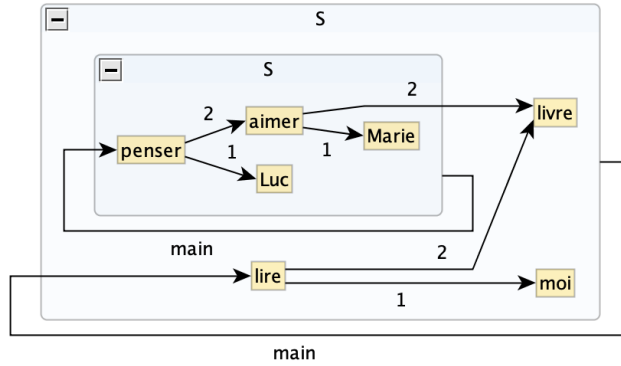


Figure 48 RSém de la phrase (47)

La proposition principale est *j'ai lu le livre*, et la proposition relative est *que Luc pense que Marie aime*. Le nom modifié par la proposition relative est *livre* et le verbe au sommet de la proposition relative est *penser*. Nous pouvons observer qu'entre le nom modifié de la proposition relative *livre* et le verbe au sommet de la proposition relative *penser*, il y a un prédicat *aimer* qui est un verbe ici.

Si nous essayons de générer cette phrase (47) avec la règle des relatives de Dubinskaite comme en Figure 49, nous obtenons un résultat similaire à la phrase avec la relative directe (46).

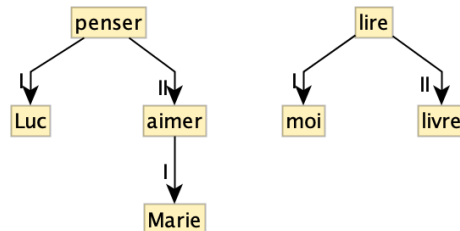


Figure 49 RSyntP de la phrase (47) avec la règle de Dubinskaite

La proposition principale, *j'ai lu le livre*, et la proposition relative, *que Luc pense que Marie aime*, sont générées comme deux phrases indépendantes et la relation ATTR ne s'est pas générée entre le nom *livre* et le verbe *penser*. La règle des relatives indirectes et son schéma sont décrits ci-dessous.

```

Sem<=>DSynt relative_indirect : relatives
leftside (v)
?Smain{          // main clause
  ?Nl            // modified noun
  ?Srel{        // subordinate clause
    ?Vl         // main verb of the relative
    ?Pl{       // predicate inside the relative
      ?r-> ?Nl // N is an argument of P
    }
  }
  main-> ?Vl
}
}
?R <- lexicon::(?Pr.dlex).(gp).(?) // retrieve V's GP

rightside
rc:?Nr{
  rc:<=> ?Nl
  ATTR-> rc:?Vr{<=> ?Vl}
}
rc:?Pr{
  rc:<=> ?Pl
  ?R-> ?NN{ // create a copy of Nr
    <-> rc:?Nr // coref with Nr
    dlex=?Nr.dlex // recopy all attributes from Nr
    dpos=?Nr.dpos
    det=?Nr.det
    definiteness=?Nr.definiteness
    number=?Nr.number
    case=?Nr.case
    relpro=yes
    animacy=?Nr.animacy
  }
}

conditions (a)
?Smain.sem=S; // Bubble for the main clause
?Srel.sem=S; // Bubble for the relative clause
?Nr.dsyntax=OK; // Wait for full lexicalization
?Vr.dsyntax=OK; // Wait for full lexicalization
?Pr.dsyntax=OK; // Wait for full lexicalization

```

Figure 50 Règle relative_indirect

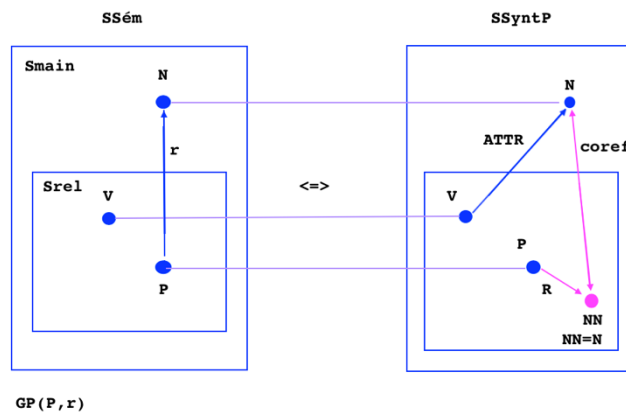


Figure 51 Schéma de la règle relative_indirect

Ce schéma représente la règle des propositions relatives indirectes dans GenDR dans l'interface sémantique (à gauche) syntaxe profonde (à droite). Le grand carré (Smain) illustre la proposition principale et le petit carré (Srel) illustre la proposition relative. À gauche, nous avons un nœud V et un nœud P dans la proposition relative. Ce nœud P a une relation r avec le nœud N dans la proposition principale. Cela est représenté à droite par le nœud V de la proposition relative, qui a une relation ATTR (de modification) avec le nœud N dans la proposition principale. Le nœud P a une nouvelle relation R avec un nouveau nœud NN dans la proposition relative. Ce nœud NN est coréférent avec le nœud N de la proposition principale. La relation R est créée à partir du patron de régime du nœud P.

Autrement dit, si nous reprenons l'exemple (47) et la Figure 48, sa RSém pour la partie gauche de la règle. *J'ai lu le livre* est la proposition principale dans le grand carré (Smain) avec le nœud N *livre*. *Que Luc pense que Marie aime* est la proposition relative dans le petit carré (Srel) avec le nœud V *penser* et le nœud P *aimer*. Le nœud P *aimer* a une relation *r* avec le nœud N *livre* de la proposition principale.

Ensuite, si nous reprenons avec la phrase (47) et sa RSyntP avec la Figure 52 pour illustrer la partie droite de la règle.

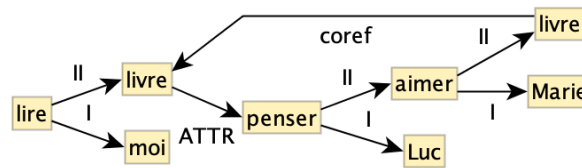


Figure 52 RSyntP de la phrase (47) avec la règle *relative_indirect*

Nous avons toujours la proposition principale *j'ai lu le livre* avec le nœud N *livre*. Dans la proposition relative *que Luc pense que Marie aime* Le nœud V *penser* de la proposition relative a une relation ATTR (de modification) avec le nœud N *livre*. Le nœud P *aimer* de la proposition relative a une nouvelle relation R avec un nouveau nœud NN *livre* coréférent et lexicalisé à partir du nœud N *livre* de la proposition principale.

Pour conclure sur les règles directes et indirectes, nous avons ci-dessous le schéma de la règle de Dubinskaite qui est le même que le schéma Figure 41, ainsi que les schémas des règles directes et indirectes, qui sont les mêmes qu'en Figure 46 et Figure 51.

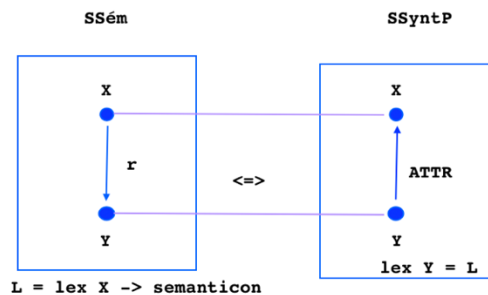


Figure 53 Schéma de la règle de Dubinskaite

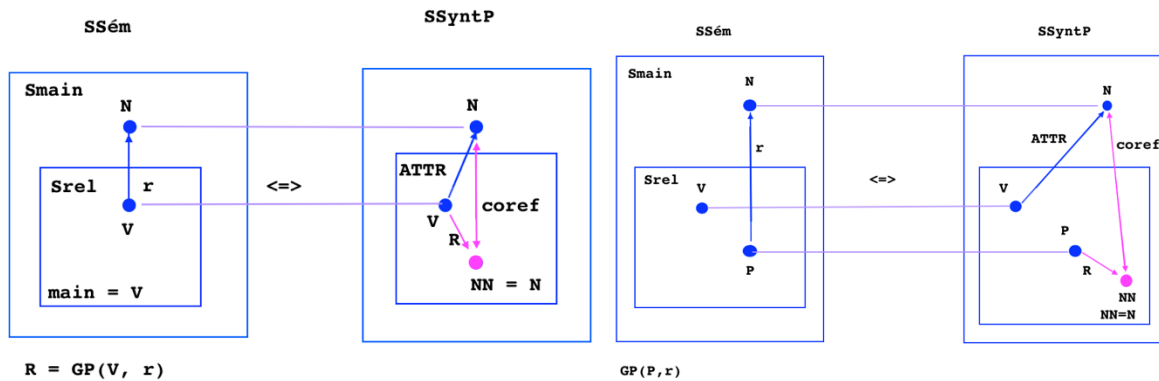


Figure 54 Schémas des règles relatives directes (à gauche) et indirectes (à droite)

Nous pouvons constater que la règle de Dubinskaite n’essaie que de générer la relation ATTR et qu’il n’y a pas de nœuds créés pour être le futur pronom relatif. Dans la règle directe et la règle indirecte, la structure des règles est plus complexe. Dans un premier temps, il y a une proposition dans une autre proposition, puis une relation ATTR qui sort de sa proposition, puis nous avons la création d’un nouveau nœud qui sera remplacé par un pronom relatif au niveau de la syntaxe de surface, enfin nous avons aussi une relation de coréférence. Ces règles essayaient de prendre en compte toutes les particularités des relatives afin de les détecter et générer au mieux.

5.1.4. Modification de la règle des patrons de régime actant_gp

La règle `actant_gp`, `gp` pour government pattern ou patron de régime, génère les actants d’un nœud en fonction de son entrée dans le dictionnaire de GenDR. Cette règle permet d’appliquer le régime des verbes, par exemple pour le verbe *lire*, quelqu’un lit quelque chose. Il y aura donc deux nœuds créés à partir du nœud *lire* pour les actants de *lire*. La règle `actant_gp`, a dû être légèrement modifiée pour prendre en compte les cas des règles des propositions relatives. Nous avons ajouté qu’une proposition peut contenir une autre proposition avec deux nœuds pour que la proposition principale et la proposition relative puissent être générées par GenDR dans une même phrase. La règle `actant_gp` avant d’être modifiée est présentée ci-dessous.

```

Sem<=>DSynt actant_gp : actants
leftside (v)
?Xl{
  l:?r-> ?Yl{}
}
?R <- lexicon::(?Xr.dlex).(gp).(?r)

rightside
rc:?Xr{
  rc:<=> ?Xl
  ?R-> ?Yr{
    <=> ?Yl
    dpos=lexicon::(?Xr.dlex).(gp).(?R).(dpos)
    finiteness=lexicon::(?Xr.dlex).(gp).(?R).(finiteness)
    mood=lexicon::(?Xr.dlex).(gp).(?R).(mood)
    subject=lexicon::(?Xr.dlex).(gp).(?R).(subject)
    det=lexicon::(?Xr.dlex).(gp).(?R).(det)
    definiteness=lexicon::(?Xr.dlex).(gp).(?R).(definiteness)
    case=lexicon::(?Xr.dlex).(gp).(?R).(case)
  }
}

conditions (a)
not ?Xr.lf; // Xr is not a LF
not ?R="---"; // the realisation of ?R must not be blocked
?Xr.dsynt=OK and not ?Xr.split=top; // not sure if always has to be the case.

```

Figure 55 Règle actant_gp v.1.0

Il y a un nœud X (?Xl) qui a une relation r (l : ?r->) avec un autre nœud Y (?Yl) et ce nœud X est lexicalisé à partir des informations du dictionnaire lexical (?R <- lexicon::(?Xr.dlex).(gp).(?r)). Puis, voici la règle actant_gp modifiée pour que les règles sur les relatives puissent fonctionner.

```

Sem<=>DSynt actant_gp : actants
leftside (v)
?Xl{
  l:?r-> ?Yl{}
}
?R <- lexicon::(?Xr.dlex).(gp).(?r)
?S{
  ?Xl{}
  ?Yl{}
}

rightside
rc:?Xr{
  rc:<=> ?Xl
  ?R-> ?Yr{
    <=> ?Yl
    dpos=lexicon::(?Xr.dlex).(gp).(?R).(dpos)
    finiteness=lexicon::(?Xr.dlex).(gp).(?R).(finiteness)
    mood=lexicon::(?Xr.dlex).(gp).(?R).(mood)
    subject=lexicon::(?Xr.dlex).(gp).(?R).(subject)
    det=lexicon::(?Xr.dlex).(gp).(?R).(det)
    definiteness=lexicon::(?Xr.dlex).(gp).(?R).(definiteness)
    case=lexicon::(?Xr.dlex).(gp).(?R).(case)
  }
}

conditions (a)
not ?Xr.lf; // Xr is not a LF
not ?R="---"; // the realisation of ?R must not be blocked
?Xr.dsynt=OK and not ?Xr.split=top; // not sure if always has to be the case.

```

Figure 56 Règle actant_gp v.2.0

La règle est la même, mais nous avons ajouté (dans le rectangle rouge) qu'une proposition S (?S) peut contenir deux autres nœuds X (?Xl) et Y (?Yl). Cela permettra d'avoir une proposition contenir une autre proposition.

C'est au niveau de la syntaxe de surface que la question des pronoms est posée. Effectivement, les règles au niveau de la syntaxe de surface sont spécialisées pour les pronoms relatifs en français dans ce projet de recherche.

5.2 Interface syntaxe profonde-syntaxe de surface

Dans le chapitre 3, nous avons vu dans la partie 3.2.3 que c'est au niveau de la syntaxe de surface que la pronominalisation se fait. La création de ces règles est inspirée de deux types de règles déjà existantes, celles de lexicalisation et celles de la gestion des actants en fonction du patron de régime du verbe au niveau de la syntaxe profonde et de la syntaxe de surface. Une modification de la règle de lexicalisation a dû être faite afin d'éliminer la compétition entre les règles des pronoms : pour éviter qu'il n'y ait une double lexicalisation ou que les deux propositions ne soient pas générées ensemble. Les règles des pronoms permettent de lexicaliser les pronoms *qui*, *que*, *dont*, une préposition suivie de *qui* et une préposition suivie de *lequel*.

5.2.1 La règle du pronom relatif *qui*

Le pronom relatif le plus présent dans le corpus est *qui* et il est traité dans cette section. Il a été présenté dans le chapitre 3, dans la partie 3.3.1, que le pronom *qui* est de rôle syntaxique sujet. Dans le chapitre 4, nous avons pu constater en regarder les sept phrases avec le pronom relatif *qui* objet que le pronom faisait partie d'une locution ou d'une expression figée, donc en nous basant sur le corpus nous considérons qu'il n'y a pas de *qui* objet. Nous avons aussi vu dans la partie, que le pronom relatif *qui* et la phrase (46) ne se génèrent pas correctement avec la règle sur les relatives de Dubinskaite en Figure 44.

Voici ci-dessous la règle du pronom relatif *qui* et son schéma, dans l'interface syntaxe profonde-syntaxe de surface.

```

DSynt<=>SSynt rel_qui : lex_relatives
leftside (v)                                rightside
?Xl{                                         rc:?Xr{
  l:?r-> l:?Yl{ //main verb of the relative  rc:<=> ?Xl
  <-> ?Zl{ //copy of N                      subj-> ?Yr{
  relpro=yes //noun modified by relative    <=> ?Yl
  }                                           <-> rc:?Zr {rc:<=> ?Zl} //Zr is not lexicalized
}                                           relpro=yes
?GP <- lexicon::(?Xr.slex).(gp).(r)        slex=qui //replace the noun by the pronoun
                                           dlex=qui
                                           dpos=lexicon::(qui).dpos
                                           spos=lexicon::(qui).spos
                                           }
                                           }
conditions (a)
?GP.rel=subj; // This rule handles subjects only
?Xl.finiteness=FIN; // Only finite verbs have a subject
not ?Xr.split=bottom; // ?Xr is not the bottom part of a split node
not ?Xr.auxdep; // Attach on the top Aux
?Xr.ssynt=OK; // Wait for full lexicalization

```

Figure 57 Règle du pronom *qui*

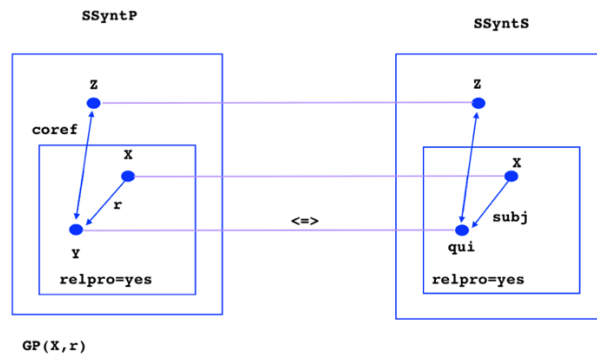


Figure 58 Schéma de la règle du pronom *qui*

Ce schéma représente la règle du pronom *qui* dans GenDR dans l'interface syntaxe profonde (à gauche) syntaxe de surface (à droite). Le grand carré illustre la proposition principale et le petit carré illustre la proposition relative. À gauche, nous avons un nœud X et un nœud Y dans la proposition relative. Ce nœud X a une relation r avec le nœud Y. Ce nœud Y a un trait relatif (relpro=yes) et est coréférent au nœud Z de la proposition principale. La relation R est créée par le patron de régime du nœud X. À droite, le nœud X de la proposition relative a une relation sujet avec le nœud Y de la proposition relative, qui est devenu le pronom relatif *qui*. Le nœud *qui* porte le trait relatif (relpro=yes) et il est coréférent au nœud Z de la proposition principale.

Nous reprenons la phrase (46) comme exemple de relative avec un pronom relatif *qui* et la Figure 47 représentant la RSyntP de cette phrase. À gauche de la règle, nous avons un nœud X empêcher

et un nœud Y *médicament* dans la proposition relative. Ce nœud X empêcher a une relation r (I) avec le nœud Y *médicament*. Ce nœud Y *médicament* a un trait relatif (relpro=yes) et est coréférent au nœud Z *médicament* dans la proposition principale. La relation R (I) est créée par le patron de régime du nœud X *empêcher*.

Ensuite, voici l'explication de la partie droite de la règle avec la phrase (46) et la Figure 59 est sa RSyntS.

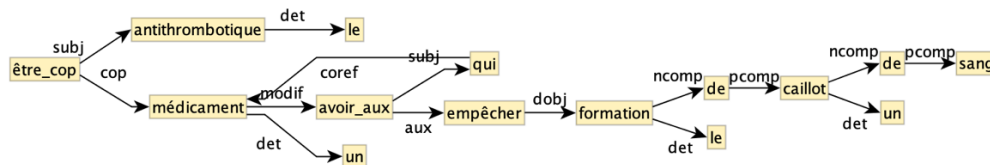


Figure 59 RSyntS de la phrase (46) avec le pronom *qui*

À droite de la règle, le nœud X *empêcher* de la proposition relative a une relation sujet avec le nœud Y de la proposition relative, *médicament* en syntaxe profonde, qui est devenu le pronom relatif *qui*. Le nœud Y *qui* porte le trait relatif (relpro=yes) et il est coréférent au nœud Z *médicament* dans la proposition principale.

5.2.2 La règle du pronom relatif *que*

La règle sur le pronom relatif *qui* est d'autant plus intéressante et efficace qu'elle est assez abstraite pour permettre à d'autres pronoms relatifs comme *que* et *dont* de se baser sur cette règle. Cette règle a le potentiel d'être réutilisable par d'autres langues en changeant la lexicalisation. En effet, les règles pour les pronoms *que* et *dont* ne diffèrent que par la lexicalisation du pronom relatif et la relation venant du patron de régime du verbe. Le pronom *que* est un pronom de rôle syntaxique objet direct ou attribut d'après le Tableau 3, comme il a été vu précédemment dans le chapitre 3, la partie 3.3.

Nous pouvons voir ci-dessous que le pronom relatif *que* et la phrase (48) et la **Erreur ! Source du r envoi introuvable**. Figure 60 qui est sa RSyntS ne se génèrent pas correctement au niveau de la syntaxe de surface avec la règle sur les relatives de Dubinskaite.

(48) *Vous suivez attentivement les instructions que vous a données votre médecin.*
 [Sequoia_emea-test_00421]

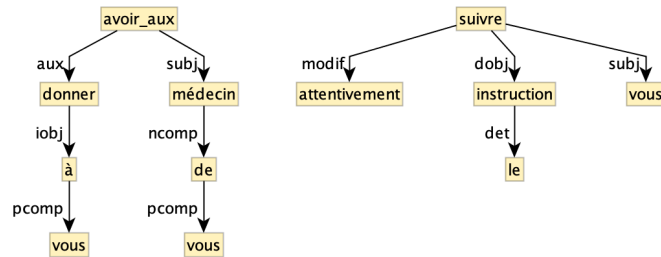


Figure 60 RSyntS de la phrase (48) avec la règle des relatives de Dubinskaite

La proposition principale *vous suivez attentivement les instructions* et la proposition relative *que vous a données votre médecin* sont générées séparément comme deux phrases par GenDR. Il n’y a pas de relation de modification entre le nom modifié *instruction* et le verbe au sommet de la proposition relative *suivre*, ni de pronom relatif *que*.

La règle du pronom relatif *que* est présenté ci-dessous.

```

DSynt<=>SSynt rel_que : lex_relatives
leftside (v)
?Xl{
  //main verb of the relative
  l:?r-> l:?Yl{ //copy of N
    <=> ?Zl //noun modified by relative
    relpro=yes //relative features
  }
}
?GP <- lexicon::(?Xr.slex).(gp).(?r)

rightside
rc:?Xr{
  rc:<=> ?Xl
  dobj-> ?Yr{
    <=> ?Yl
    <=> rc:?Zr {rc:<=> ?Zl} //Zr
    relpro=yes
    slex=que //replace the noun by the pronoun
    dlex=que
    dpos=lexicon::(que).dpos
    spos=lexicon::(que).spos
  }
}

conditions (a)
?GP.rel=dobj; // This rule handles subjects only
not ?Xr.split=bottom; // ?Xr is not the bottom part of a split node
//not ?Xr.auxdep; // Attach on the top Aux
?Xr.ssynt=OK; // Wait for full lexicalization

```

Figure 61 Règle du pronom *que*

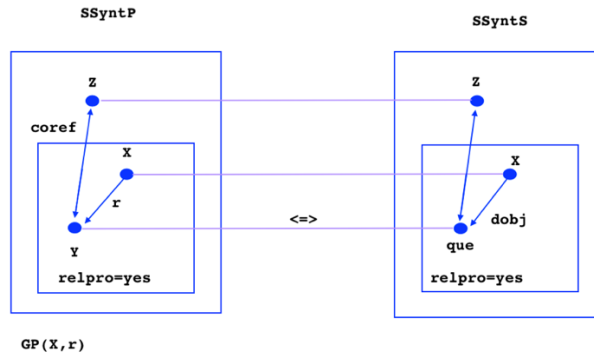


Figure 62 Schéma de la règle du pronom *que*

Ce schéma représente la règle du pronom *que* dans GenDR dans l'interface syntaxe profonde (à gauche) syntaxe de surface (à droite). Le grand carré illustre la proposition principale et le petit carré illustre la proposition relative. À gauche, nous avons un nœud X et un nœud Y dans la proposition relative. Ce nœud X a une relation *r* avec le nœud Y. Ce nœud Y a un trait relatif (*relpro=yes*) et est coréférent au nœud Z. La relation *r* est créée par le patron de régime du nœud X. À droite, le nœud X de la proposition relative a une relation objet direct (*dobj*) avec le nœud Y de la proposition relative, qui est devenu le pronom relatif *que*. Le nœud *que* porte le trait relatif (*relpro=yes*) et il est coréférent au nœud Z de la proposition principale.

Nous illustrons ici cette règle du pronom relatif *que* avec la phrase (48) et sa RSyntP en Figure 63 ci-dessous.

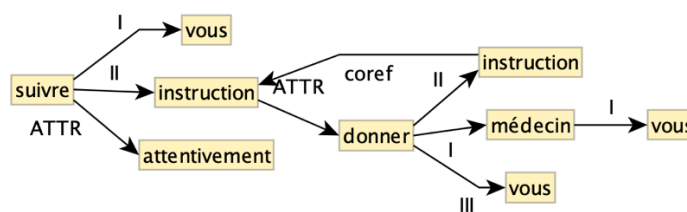


Figure 63 RSyntP de la phrase (48) avec le pronom *que*

À gauche de la règle, nous avons un nœud X *donner* et un nœud Y *instruction* dans la proposition relative. Ce nœud X *donner* a une relation *r* (II) avec le nœud Y *instruction*. Ce nœud Y *instruction* a un trait relatif (*relpro=yes*) et est coréférent au nœud Z *instruction* dans la proposition principale. La relation R (II) est créée par le patron de régime du nœud X *donner*.

Voici l'explication de la partie droite de la règle du pronom relatif *que* avec la phrase (48) et sa RSyntS en Figure 64.

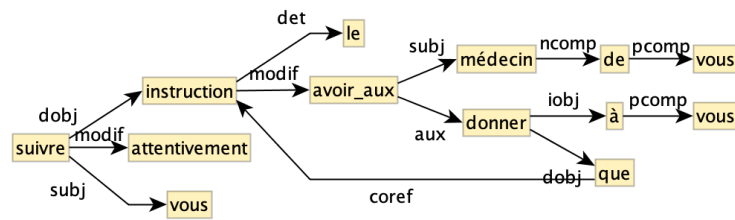


Figure 64 RSyntS de la phrase (48) avec le pronom *que*

À droite de la règle, le nœud X *donner* de la proposition relative a une relation objet direct avec le nœud Y de la proposition relative, *instruction* en syntaxe profonde qui est devenu le pronom relatif *que*. Le nœud Y *que* porte le trait relatif (*relpro=yes*) et il est coréférent au nœud Z *instruction* de la proposition principale.

5.2.3 La règle du pronom relatif *dont*

Il y a le pronom relatif *dont* qui est de rôle syntaxique complément prépositionnel d'après le Tableau 3 sur les pronoms relatifs au chapitre 3, partie 3.3. Nous pouvons constater ci-dessous que le pronom relatif *dont* oblique dans la phrase (49) et complément du nom dans la phrase (50) ne se génèrent pas correctement aux niveaux de la syntaxe profonde et de la syntaxe de surface avec la règle sur les relatives de Dubinskaite en Figure 65, Figure 66, Figure 67 et Figure 68.

(49) *J'ai lu le livre dont Luc parle.*

(50) *L'arbre dont je vois la cime indique la direction.*

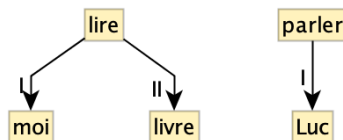


Figure 65 RSyntP de la phrase (49) avec la règle de Dubinskaite

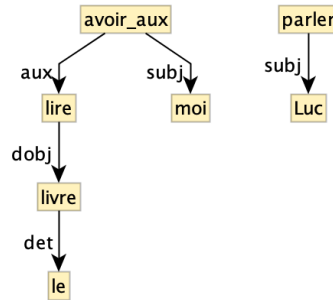


Figure 66 RSyntS de la phrase (49) avec la règle de Dubiskaite

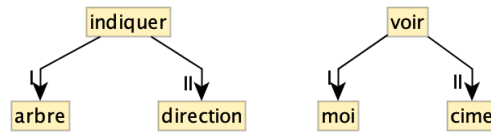


Figure 67 RSyntP de la phrase (50) avec la règle de Dubinskaite

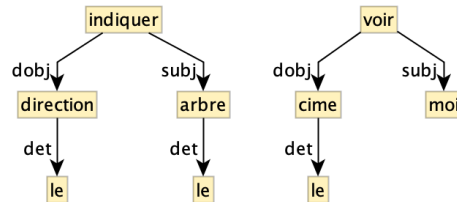


Figure 68 RSyntS de la phrase (50) avec la règle de Dubinskaite

Dans les deux phrases, les propositions principales et les propositions relatives sont générées par GenDR comme deux phrases différentes. Il n'y a pas la relation de modification entre le nom modifié et le verbe à la tête de la relative ni le pronom relatif *dont*.

La règle suivante traite du pronom *dont* comme un complément du nom introduit par la préposition *de* et comme oblique.

```

DSynt<=>SSynt rel_dont : lex_relatives
leftside (v)                                rightside
?Xl{                                         rc:?Xr{
  l:?r-> l:?Yl{ //main verb of the relative  rc:<=> ?Xl
    <-> ?Zl    //copy of N                  ?GP.rel-> ?Yr{
    relpro=yes //noun modified by relative  <=> ?Yl
  }                                           <-> rc:?Zr {rc:<=> ?Zl} //Zr
}                                           relpro=yes
?GP <- lexicon::(?Xr.slex).(gp).(?)        slex=dont //replace the noun by the pronoun
                                           dlex=dont
                                           dpos=lexicon::(dont).dpos
                                           spos=lexicon::(dont).spos
                                           }
                                           }
conditions (z)
?GP.prep=de; // This rule handles prep de
not ?Xr.split=bottom; // ?Xr is not the bottom part of a split node
//not ?Xr.auxdep; // Attach on the top Aux
?Xr.ssynt=OK; // Wait for full lexicalization

```

Figure 69 Règle du pronom relatif *dont*

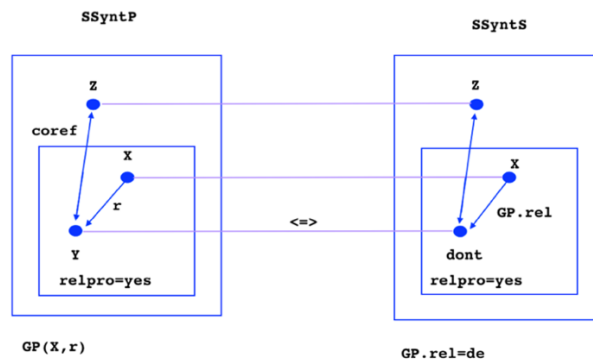


Figure 70 Schéma de la règle du pronom relatif *dont*

Ce schéma représente la règle du pronom *dont* dans GenDR dans l'interface syntaxe profonde (à gauche) syntaxe de surface (à droite). Le grand carré illustre la proposition principale et le petit carré illustre la proposition relative. À gauche, nous avons un nœud X et un nœud Y dans la proposition relative. Ce nœud X a une relation r avec le nœud Y. Ce nœud Y a un trait relatif (*relpro=yes*) et est coréférent au nœud Z de la proposition principale. La relation r est créée par le patron de régime du nœud X. À droite, le nœud X de la proposition relative a une relation créée par le patron de régime du nœud X avec le nœud Y de la proposition relative, qui est devenu le pronom relatif *dont*. Le nœud *dont* porte le trait relatif (*relpro=yes*) et il est coréférent au nœud Z dans la proposition principale. Le nœud Y est lexicalisé par le pronom relatif *dont* que si la relation du patron de régime du nœud X introduit une relation avec la préposition *de*.

Nous illustrons ici cette règle du pronom relatif *dont* oblique la phrase (49) et sa RSyntP en Figure 71 ci-dessous.

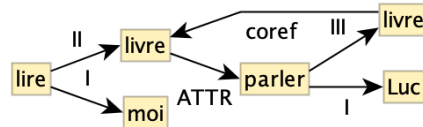


Figure 71 RSyntP de la phrase (49) avec le *dont* oblique

À gauche de la règle, nous avons un nœud X *parler* et un nœud Y *livre* dans la proposition relative. Ce nœud X *parler* a une relation r (III) avec le nœud Y *livre*. Ce nœud Y *livre* a un trait relatif (relpro=yes) et est coréférent au nœud Z *livre* de la proposition principale. La relation r est créée par le patron de régime du nœud X *parler*.

Voici un exemple du côté gauche de la règle avec le pronom relatif *dont* complément du nom avec la phrase (50) et sa RSyntP en Figure 72 ci-dessous.

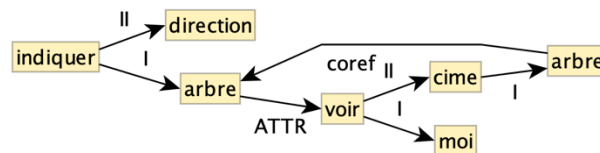


Figure 72 RSyntP de la phrase (50) avec le *dont* complément du nom

À gauche de la règle, nous avons un nœud X *cime* et un nœud Y *arbre* dans la proposition relative. Ce nœud X *cime* a une relation r (I) avec le nœud Y *arbre*. Ce nœud Y *arbre* a un trait relatif (relpro=yes) et est coréférent au nœud Z *arbre* de la proposition principale. La relation r est créée par le patron de régime du nœud X *cime*.

Voici l'explication de la partie droite avec la phrase (49) et sa RSyntS en Figure 73 pour le *dont* oblique.

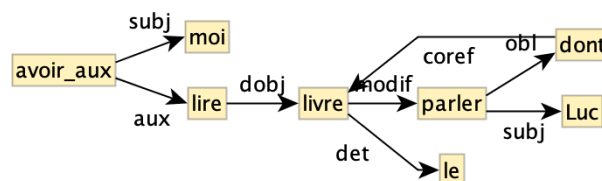


Figure 73 RSyntS de la phrase (49) avec le *dont* oblique

À droite, le nœud *X parler* de la proposition relative a une relation oblique (obl) créée par le patron de régime du nœud *X parler* avec le nœud Y de la proposition relative, qui est devenu le pronom relatif *dont*. Le nœud *dont* porte le trait relatif (relpro=yes) et il est coréférent au nœud Z *livre* dans la proposition principale. Le nœud Y est lexicalisé par le pronom relatif *dont* que si la relation du patron de régime du nœud *X parler* introduit une relation avec la préposition *de*, ici parler du livre.

Voici l'explication de la partie droite avec la phrase (50) et sa RSyntS en Figure 74 pour le *dont* complément du nom.

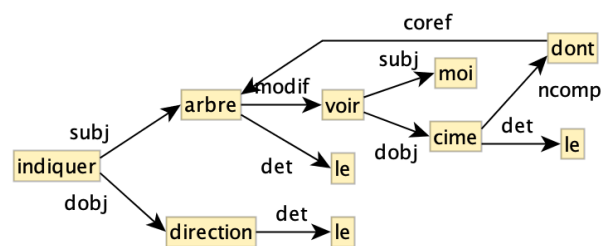


Figure 74 RSyntS de la phrase (50) avec *dont* complément du nom

À droite, le nœud *X cime* de la proposition relative a une relation complément du nom (ncomp) créée par le patron de régime du nœud *X cime* avec le nœud Y de la proposition relative, qui est devenu le pronom relatif *dont*. Le nœud *dont* porte le trait relatif (relpro=yes) et il est coréférent au nœud Z *arbre* dans la proposition principale. Le nœud Y est lexicalisé par le pronom relatif *dont* que si la relation du patron de régime du nœud *X cime* introduit une relation avec la préposition *de*, ici la cime de l'arbre.

5.2.4 La règle de la préposition suivie du pronom relatif *qui*

Nous pouvons constater ci-dessous que la préposition suivie du pronom relatif *qui* dans la phrase (51) ne se génère pas correctement aux niveaux de la syntaxe profonde et de la syntaxe de surface avec la règle sur les relatives de Dubinskaite en Figure 75 et Figure 76.

(51) *Chris offre une fleur à Aria à qui j'ai prêté un livre*

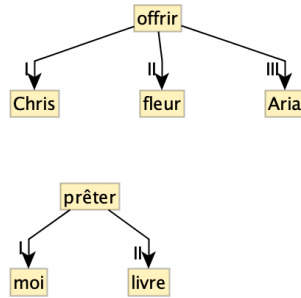


Figure 75 RSyntP de la phrase (51) avec la règle de Dubinskaite

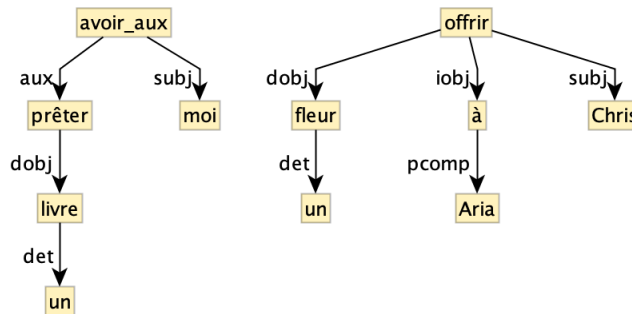


Figure 76 RSyntS de la phrase (51) avec la règle de Dubinskaite

Nous pouvons constater dans la et la que la proposition principale et la proposition relative sont générées séparément comme deux phrases, qu'il n'y a pas de relation de modification entre le nom modifié et le verbe au sommet de la relative et que la préposition suivie du pronom relatif *qui* ne se sont pas générés.

Une préposition suivie du pronom *qui* ou *quoi* était à la limite de GenDR, car il n'y avait pas de trait d'animacité dans le système. Comme il a été présenté dans le chapitre 3, partie 3.3 et dans le Tableau 3 sur les pronoms relatifs, c'est un trait de différenciation fondamental pour ces deux pronoms. Donc, nous avons ajouté ce trait dans le dictionnaire lexical, et cette partie est développée dans la partie 5.3. Cette règle est assez générale et pourrait convenir à d'autres langues. Dans cette règle, la préposition générée par GenDR dépend du patron de régime du verbe. Les informations du patron de régime du verbe et la sélection de la bonne préposition se trouvent dans le dictionnaire lexical de GenDR. La règle d'une préposition suivie du pronom *qui* est présenté ci-dessous.


```

DSynt<=>SSynt rel_prep_qui : lex_relatives
leftside (v)
?Xl{
  //main verb of the relative
  l:?r-> l:?Yl{ //copy of N
    relpro=yes //relative features
    <-> ?Zl{ //noun modified by relative
      animacy=ANIM
    }
  }
}
?GP <- lexicon::(?Xr.slex).(gp).(?r)

rightside
rc:?Xr{
  rc:<=> ?Xl
  ?GP.rel-> ?Pr{
    slex=?GP.prep //replace the noun by the pronoun
    dlex=?GP.prep
    dpos=lexicon::(?GP.prep).dpos
    spos=lexicon::(?GP.prep).spos
    lexicon::(?GP.prep).gp.II.rel-> ?Yr{
      <=> ?Yl
      <-> rc:?Zr {rc:<=> ?Zl} //Zr
      relpro=yes
      slex=qui //replace the noun by the pronoun
      dlex=qui
      dpos=lexicon::(qui).dpos
      spos=lexicon::(qui).spos
      animacy=ANIM
    }
  }
}

conditions (a)
not ?GP.prep=de; // prep de is handled by rel_dont
?Xr.ssynt=OK; // Wait for full lexicalization
?GP.prep;

```

Figure 77 Règle de la préposition suivie du pronom relatif *qui*

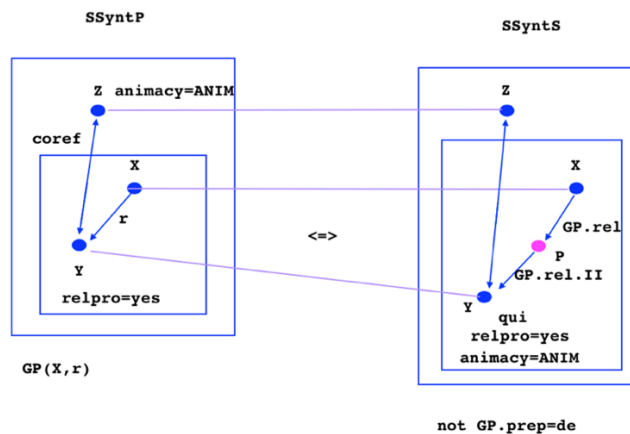


Figure 78 Schéma de la règle de la préposition suivie du pronom relatif *qui*

Ce schéma représente la règle de la préposition suivie du pronom relatif *qui* dans GenDR dans l'interface syntaxe profonde (à gauche) syntaxe de surface (à droite). Le grand carré illustre la proposition principale et le petit carré illustre la proposition relative. À gauche, nous avons un nœud X et un nœud Y dans la proposition relative. Ce nœud X a une relation r avec le nœud Y. Ce nœud Y a un trait relatif (relpro=yes) et est coréférent au nœud Z de la proposition principale qui a le trait animé (animacy=ANIM). La relation r est créée par le patron de régime du nœud X. À droite, le nœud X de la proposition relative a une relation créée par le patron de régime du nœud X avec le nouveau nœud P de la proposition relative. Le nouveau nœud P a une autre relation

crée par le patron de régime du nœud X avec le nœud Y de la proposition relative, qui est devenu le pronom relatif *qui*. Le nœud *qui* porte les traits relatif (relpro=yes) et animé (animacy=ANIM). Le nœud *qui* est aussi coréférent avec le nœud Z de la proposition principale. Le nouveau nœud P est lexicalisé par une préposition que si la relation du patron de régime du nœud X n'introduit pas une relation avec la préposition *de*.

Nous illustrons cette règle de la préposition suivie du pronom relatif *qui* avec la phrase (51) et sa RSyntP en Figure 79 ci-dessous.

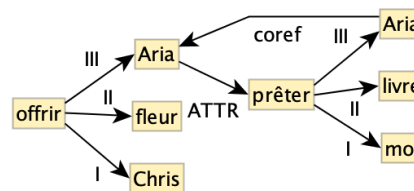


Figure 79 RSyntP de la phrase (51)

À gauche, nous avons un nœud X *prêter* et un nœud Y *Aria* dans la proposition relative. Ce nœud X *prêter* a une relation r (III) avec le nœud Y *Aria*. Ce nœud Y *Aria* a un trait relatif (relpro=yes) et est coréférent au nœud Z *Aria* de la proposition principale qui a le trait animé (animacy=ANIM). La relation r est créée par le patron de régime du nœud X *prêter*.

Voici l'explication de la partie droite avec la phrase (51) et sa RSyntS en Figure 80 pour la préposition suivie du pronom relatif *qui*.

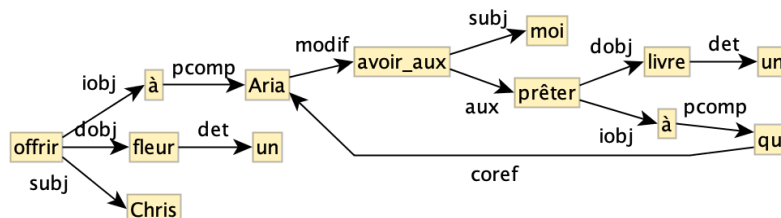


Figure 80 RSyntS de la phrase (51)

À droite, le nœud X *prêter* de la proposition relative a une relation créée par le patron de régime du nœud X *prêter* avec le nouveau nœud P *à* de la proposition relative. Le nouveau nœud P *à* a une autre relation créée par le patron de régime du nœud X *prêter* avec le nœud Y *Aria* de la proposition relative, qui est devenu le pronom relatif *qui*. Le nœud *qui* porte les traits relatif

(relpro=yes) et animé (animacy=ANIM). Le nœud *qui* est aussi coréférent avec le nœud Z *Aria* de la proposition principale. Le nouveau nœud P à est lexicalisé par une préposition que si la relation du patron de régime du nœud X *prêter* n'introduit pas une relation avec la préposition *de*, mais ici c'est *prêter à quelqu'un*.

5.2.5 La règle d'une préposition suivie du pronom relatif *lequel*

Nous pouvons constater ci-dessous que la préposition suivie du pronom relatif *lequel* dans la phrase (52) ne se génère pas correctement aux niveaux de la syntaxe profonde et de la syntaxe de surface avec la règle sur les relatives de Dubinskaite en Figure 81 et Figure 82.

(52) *Aria a lu un livre avec lequel elle s'est endormie.*

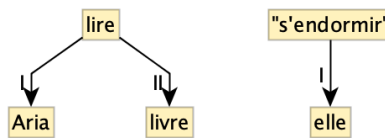


Figure 81 RSyntP de la phrase (52) avec la règle de Dubinskaite

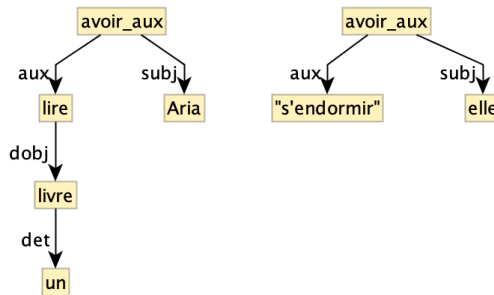


Figure 82 RSyntP de la phrase (52) avec la règle de Dubinskaite

Cette règle d'une préposition suivie du pronom *lequel* est exactement la même que celle d'une préposition suivie du pronom *qui*, mais avec le trait inanimé. La règle aurait dû être une préposition suivie du pronom *quoi*, mais il n'y a que quatre phrases avec le pronom *quoi*. Ces quatre phrases restent grammaticales si nous les remplaçons par le pronom *lequel*. Donc, le pronom a été changé pour *lequel*, qui peut être animé ou inanimé. Pour garder la distinction animée et inanimée, cette règle concerne seulement les noms inanimés. De plus, dans cette règle,

la préposition générée par GenDR dépend du patron de régime du verbe. Les informations du patron de régime du verbe et la sélection de la bonne préposition se trouvent dans le dictionnaire de GenDR. La règle d'une préposition suivie du pronom *lequel* est présentée ci-dessous.

```

DSynt<=>SSynt rel_prep_lequel : lex_relatives
leftside (v)
?Xl{
  l:?r-> l:?Yl{ //main verb of the relative
    relpro=yes //copy of N
    relpro=yes //relative features
    <-> ?Zl{ //noun modified by relative
      animacy=INANIM
    }
  }
}
?GP <- lexicon::(?Xr.slex).(gp).(?r)
rightside
rc:?Xr{
  rc:<=> ?Xl
  ?GP.rel-> ?Pr{
    slex=?GP.prep //replace the noun by the pronoun
    dlex=?GP.prep
    dpos=lexicon::(?GP.prep).dpos
    spos=lexicon::(?GP.prep).spos
    lexicon::(?GP.prep).gp.II.rel-> ?Yr{
      <=> ?Yl
      <-> rc:?Zr {rc:<=> ?Zl} //Zr
      relpro=yes
      slex=lequel //replace the noun by the pronoun
      dlex=lequel
      dpos=lexicon::(lequel).dpos
      spos=lexicon::(lequel).spos
      animacy=INANIM
    }
  }
}
conditions (a)
not ?GP.prep=de; // prep de is handled by rel_dont
?Xr.ssynt=OK; // Wait for full lexicalization
?GP.prep;

```

Figure 83 Règle de la préposition suivie du pronom relatif *lequel*

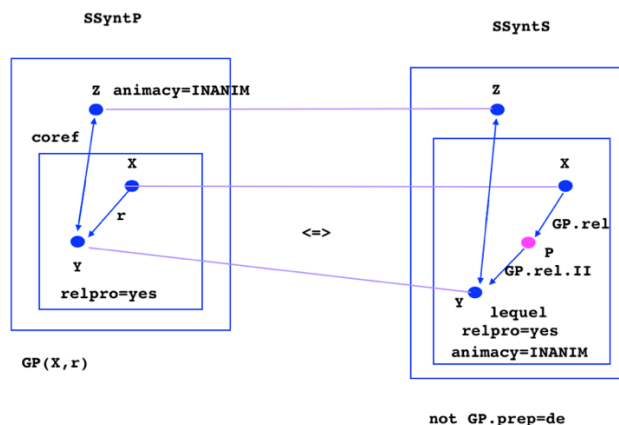


Figure 84 Schéma de la règle de la préposition suivie du pronom *lequel*

Ce schéma représente la règle de la préposition suivie du pronom relatif *lequel* dans GenDR dans l'interface syntaxe profonde (à gauche) syntaxe de surface (à droite). Le grand carré illustre la proposition principale et le petit carré illustre la proposition relative. À gauche, nous avons un nœud X et un nœud Y dans la proposition relative. Ce nœud X a une relation r avec le nœud Y. Ce nœud Y a un trait relatif (relpro=yes) et est coréférent au nœud Z de la proposition principale qui

a le trait inanimé (animacy=INANIM). La relation *r* est créée par le patron de régime du nœud *X*. À droite, le nœud *X* de la proposition relative a une relation créée par le patron de régime du nœud *X* avec le nouveau nœud *P* de la proposition relative. Le nouveau nœud *P* a une autre relation créée par le patron de régime du nœud *X* avec le nœud *Y* de la proposition relative, qui est devenu le pronom relatif *lequel*. Le nœud *lequel* porte les traits relatif (relpro=yes) et inanimé (animacy=INANIM). Le nœud *lequel* est aussi coréfèrent avec le nœud *Z* de la proposition principale. Le nouveau nœud *P* est lexicalisé par une préposition que si la relation du patron de régime du nœud *X* n'introduit pas une relation avec la préposition *de*.

Nous illustrons cette règle de la préposition suivie du pronom relatif *qui* avec la phrase (52) et sa RSyntP en Figure 85 ci-dessous.

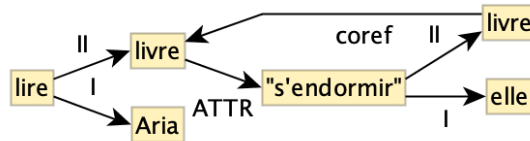


Figure 85 RSyntP de la phrase (52) préposition + *lequel*

À gauche, nous avons un nœud *X* *s'endormir* et un nœud *Y* *livre* dans la proposition relative. Ce nœud *X* *s'endormir* a une relation *r* (II) avec le nœud *Y* *livre*. Ce nœud *Y* *livre* a un trait relatif (relpro=yes) et est coréfèrent au nœud *Z* *livre* de la proposition principale qui a le trait inanimé (animacy=INANIM). La relation *r* est créée par le patron de régime du nœud *X* *s'endormir*.

Voici l'explication de la partie droite avec la phrase (52) et sa RSyntS en Figure 86 pour la préposition suivie du pronom relatif *lequel*.

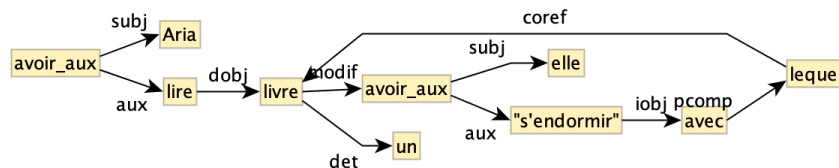


Figure 86 RSyntS de la phrase (52) avec préposition + *lequel*

À droite, le nœud *X* *s'endormir* de la proposition relative a une relation objet indirect (iobj) créée par le patron de régime du nœud *X* *s'endormir* avec le nouveau nœud *P* *avec* de la proposition

relative. Le nouveau nœud P *avec* a une autre relation complément prépositionnelle créée par le patron de régime du nœud X *s'endormir* avec le nœud Y *livre* de la proposition relative, qui est devenu le pronom relatif *lequel*. Le nœud *lequel* porte les traits relatif (relpro=yes) et inanimé (animacy=INANIM). Le nœud *lequel* est aussi coréférent avec le nœud Z *livre* de la proposition principale. Le nouveau nœud P *avec* est lexicalisé par une préposition que si la relation du patron de régime du nœud X *s'endormir* n'introduit pas une relation avec la préposition *de*, ici *s'endormir* avec quelque chose.

5.3. Implémentation du trait animacité

Pour les règles d'une préposition suivie du pronom *qui* et d'une préposition suivie du pronom *lequel*, le trait d'animacité ou animéité ou encore *animacy* en anglais est nécessaire. Nous avons implémenté ce trait d'animacité avec pour principes de modifier le moins possible la structure existante de GenDR et de ne pas avoir à entrer manuellement ce trait, ce qui nous ferait prendre des risques d'erreur manuelle. L'implémentation de ce trait se fait donc à plusieurs niveaux : au niveau du dictionnaire lexical, puis au niveau de l'interface sémantique-syntaxe profonde et enfin au niveau de l'interface syntaxe profonde-syntaxe de surface. Dans les sous-parties suivantes, les différentes modifications apportées sur les différentes parties de GenDR seront présentées.

5.3.1. Le dictionnaire Lexicon

L'implémentation de ce nouveau trait de base, *animacy*, a été comme pour le trait *countable*, c'est-à-dire que dans les paramètres par défaut de la partie du discours du nom, pour chaque partie du discours au niveau de la syntaxe de surface, le trait *animacy* a été rajouté. Par exemple, pour les noms communs (Commun nouns), les noms de ville (city) et les noms de masse (mass_noun) le trait est inanimé par défaut (animacy = INANIM) comme les exemples (53) et (54), tandis que pour les noms propres (proper_noun) et les pronoms (pronoun), le pronom le trait est animé (animacy = ANIM) par défaut comme dans les exemples (55) et (56). Voici des exemples qui pourraient être d'entrée dans le dictionnaire lexical pour le trait animacité.

(53) *Information* [animacy = INANIM]

(54) *Montréal* [animacy = INANIM]

(55) *Elle* [animacy = ANIM]

(56) *Aria* [animacy = ANIM]

Contrairement au trait de définitude (où le trait défini est lexicalisé par *le* et le trait indéfini par *un*), le trait animacy n'a pas de lexicalisation propre. C'est pourquoi le trait d'animacité n'est pas dans le dictionnaire sémantique.

5.3.2. Au niveau des règles

À l'interface sémantique-syntaxe profonde et à l'interface syntaxe profonde-syntaxe de surface, nous avons ajouté des modifications à deux règles de lexicalisation. La première est la règle de lexicalisation générale, qui lexicalise les lexies sauf les cas particuliers qui sont traités par la deuxième règle modifiée.

Pour la règle de lexicalisation générale `lex_lu`, nous voulons que l'animacité de la lexie soit prise en compte à partir des informations du dictionnaire lexical pendant la lexicalisation (`animacy=lexicon::(?L).(animacy)`).

Pour la règle de lexicalisation des cas particuliers `lex_class`, qui s'applique pour les noms propres comme *Aria*, nous voulons aussi que l'animacité de la lexie soit prise en compte à partir des informations du dictionnaire lexical pendant la lexicalisation (`animacy=lexicon::(?Xl.class).(animacy)`).

Donc, l'ajout de la prise en compte de l'animacité des lexies dans la lexicalisation des lexies à partir des informations du dictionnaire lexical permet d'avoir l'animacité pour les lexies comme nous pouvons le voir dans les exemples (53), (54), (55) et (56).

Cependant même si deux règles aux interfaces sémantique-syntaxe profonde et syntaxe profonde-syntaxe de surface ont été modifiées, et des paramètres par défaut pour la partie du discours du nom sont implémentés, il y a des cas où les paramètres par défaut ne sont pas corrects pour la lexie. Ces exceptions vont être traitées dans la sous-partie suivante.

5.3.3. Les exceptions

Lorsque les paramètres par défaut dans le dictionnaire lexical ne correspondent pas au bon trait d'animacité des lexies, il n'est pas nécessaire de corriger manuellement dans chaque phrase le trait d'animacité, il suffit d'ajouter le bon trait animé ou inanimé dans le dictionnaire lexical. Par exemple les noms de métiers désignent des êtres animés.

(57) *Infirmière* : noun {animacy=ANIM}

Nous avons vu que les paramètres par défauts les noms communs sont inanimés comme dans l'exemple (53), mais ici il s'agit d'une personne donc cette lexie se réfère à un être animé donc l'animacité a été spécifié comme animé.

Enfin, nous avons implémenté l'animacité en modifiant au minimum les règles existantes aux interfaces sémantique-syntaxe profonde et syntaxe profonde-syntaxe de surface et en ajoutant des paramètres dans le dictionnaire lexical.

5.4. Synthèse

Pour résumer ce chapitre, le point de départ est la règle de Dubinskaite (§5.1.1). Nous avons remplacé cette règle par deux nouvelles règles sur les propositions relatives, une pour la proposition relative directe (§5.1.2) et une pour la proposition relative indirecte (§5.1.3) dans l'interface sémantique-syntaxe profonde. Puis, nous avons créé les règles dans l'interface syntaxe profond-syntaxe de surface pour les pronoms *qui* (§5.2.1), *que* (§5.2.2), *dont* (§5.2.3), une préposition suivie de *qui* (§5.2.4) et une préposition suivie de *lequel* (§5.2.5). Le trait d'animacité a aussi été implémenté au niveau du dictionnaire lexical (§5.3.1), dans l'interface sémantique-syntaxe profonde et dans l'interface syntaxe profonde-syntaxe de surface (§5.3.2), ainsi que les exceptions (§5.3.3).

Dans le prochain chapitre, la couverture et les limites de toutes les nouvelles règles seront étudiées.

Chapitre 6 – Évaluation

Nous avons implémenté des règles syntaxiques dans le réalisateur profond GenDR dans l'interface sémantique-syntaxe profonde et dans l'interface syntaxe profonde-syntaxe de surface, cependant tester toutes les phrases du corpus dans le réalisateur prendrait des mois. Il faudrait décrire tout le vocabulaire des phrases, compléter de nombreuses règles sémantiques et syntaxiques pour couvrir toutes les phrases sur les deux interfaces et peut-être même régler des problèmes informatiques liés au réalisateur en lui-même. C'est pourquoi la couverture et de la précision des règles sont plus abstraites en ne montrant qu'une phrase par pronom et règle.

6.1 Couverture

Nous avons vu une analyse du corpus dans le chapitre 4 et l'implémentation des règles sur les pronoms relatifs. La proposition relative a été définie (§3.4) en deux grandes catégories : les propositions relatives directes (3.4.2) et les propositions relatives indirectes (§3.4.3). Les règles `relative_direct` (§5.1.2) et `relative_indirect` (§5.1.3) dans l'interface sémantique – syntaxe profonde s'occupe de ces deux catégories. Puis, en syntaxe de surface, nous avons les pronoms relatifs qui apparaissent (§3.2.3). Les pronoms relatifs (§3.3) sont : *qui*, *qui* objet, *que*, *dont*, *où*, *lequel*, une préposition suivie de *qui*, une préposition suivie de *quoi* et une préposition suivie de *lequel*. Sur les neuf pronoms relatifs, il y a cinq règles qui ont été faites sur les pronoms *qui* (§5.2.1), *que* (§5.2.2), *dont* (§5.2.3), une préposition suivie de *qui* (§5.2.4) et une préposition suivie de *lequel* (§5.2.5).

6.2. Précision

Dans le chapitre 4 sur l'analyse de corpus, nous avons regroupé les pronoms par pronom et par fonction dans l'ensemble du corpus dans le Tableau 5. Au niveau de l'interface sémantique-syntaxe, la règle pour les relatives directe fonctionne mieux, car il n'y a pas de prédicat

intermédiaire. Les propositions relatives indirectes fonctionnent, mais plus il y a de prédicats intermédiaires entre le verbe au sommet de la relative et le nom coréférent au nom modifié par la proposition relative, plus la proposition est sensible à d'éventuels bruits ou problèmes de génération.

De plus, nous avons essayé une phrase pour chaque pronom et fonction, ainsi que les phrases avec un verbe modal ou une fonction lexicale, au niveau de l'interface syntaxe profonde-syntaxe de surface, si les règles des propositions relatives directes et indirectes permettent de générer toutes les propositions relatives dans l'interface sémantique-syntaxe profonde. Nous allons partir du principe que si la phrase est générée dans l'interface sémantique-syntaxe profonde, peu importe le pronom, cela s'applique aussi aux autres phrases dans l'interface sémantique-syntaxe profonde.

6.2.1. Avec un pronom relatif *qui* sujet

Pour le pronom *qui*, nous avons générer la phrase (58), à partir de l'input Figure 87, la Figure 88 pour la RSém, la Figure 89 pour la RSyntP et la Figure 90 pour la RSyntS. La phrase (58) est la même que la (46), la Figure 88 est la même que la Figure 42, la Figure 89 est la même que la Figure 47 et la Figure 90 est la même que la Figure 59.

(58) *Les antithrombotiques sont des médicaments qui empêchent la formation de caillot sanguin.* [Sequoia, sequoia_emea-fr-dev_00475]

```

structure Sem s1 {
  S:1{
    sent_id="emea-fr-dev_00475"
    text="Les antithrombotiques sont des médicaments qui empêchent la formation de caillots sanguins."
    antithrombotique:1{number=PL definiteness=DEF}
    être:1{
      tense=PRES
      1-> antithrombotique:1
      2-> médicament:1
    }
    médicament:1{number=PL definiteness=INDEF}
    S:2{
      empêcher:1{
        tense=PRES
        1-> médicament:1
        2-> "se former":1
      }
      "se former":1{
        tense=PRES number=SG definiteness=DEF
        1-> caillot:1
      }
    }
  }
}

```

```

}
caillot:1{
  number=PL definiteness=INDEF
  1-> sang:1
}
sang:1{number=SG definiteness=NO}
main-> empêcher:1 #noeud communicatif dominant de la relative
}
main-> être:1 #noeud communicatif dominant de la principale
}
}

```

Figure 87 Input de la phrase (58) dans GenDR pour le RSém

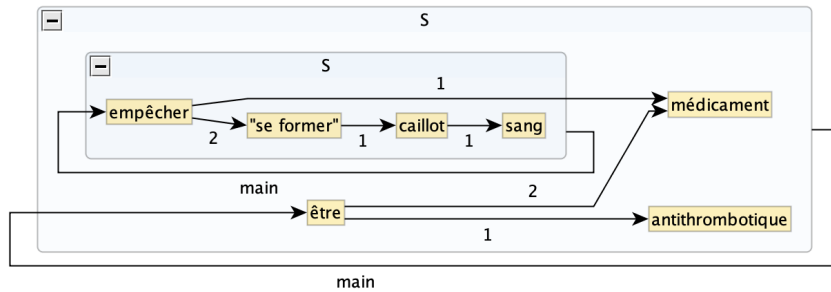


Figure 88 RSém de la phrase (58)

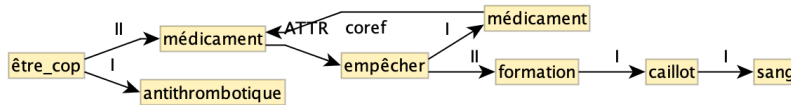


Figure 89 RSyntP de la phrase (58)

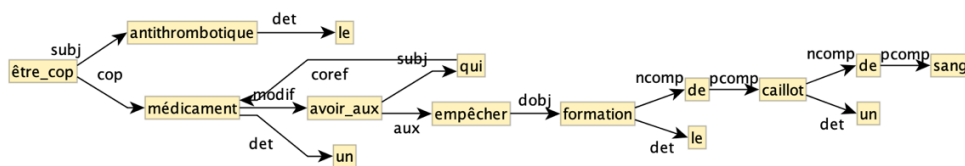


Figure 90 RSyntS de la phrase (58)

Nous pouvons observer qu'à l'interface sémantique-syntaxe profonde la règle *relative_direct* est utilisée et que nous n'avons pas de problème pour générer la phrase. À l'interface syntaxe profonde-syntaxe de surface, la règle sur le pronom *qui* est utilisé.

6.2.2. Avec un pronom relatif *que* objet

Pour le pronom *que*, nous avons généré la phrase (59) à partir de l'input Figure 91, la Figure 92 pour la RSém, la Figure 93 pour la RSyntP et la Figure 94 pour la RSyntS. La phrase (59) est la même que la (48), la Figure 93 est la même que la Figure 63 et la Figure 94 est la même que la Figure 64.

(59) *Vous suivez attentivement les instructions que vous a données votre médecin.*

[Sequoia_emea-test_00421]

```
structure Sem s1 {                                     #pour ouvrir une RSém dans GenDR
  S:1{                                                #proposition principale
    sent_id="emea-fr-test_00421"                    #numéro d'identification de la phrase
    text="Suivez attentivement toutes les instructions qui vous sont données par votre médecin."
    vous:1{}
    suivre:1{                                        #verbe principal de la phrase
      mood=PRES
      1-> vous:1
      2-> instruction:1
    }
    instruction:1{number=PL definiteness=DEF}
    attentif:1{
      1-> suivre:1
    }
    tous:1{
      1-> instruction:1                             #nom modifié
    }
  }
  S:2{                                                #proposition relative
    vous:2{
      <-> vous:1
    }
    donner:1{                                        #verbe au sommet de la relative
      tense=PAST
      1-> médecin:1
      2-> instruction:1
      3-> vous:2
    }
    médecin:1{
      number=SG definiteness=NO
      1-> vous:3
    }
    vous:3{
      <-> vous:1
    }
    main-> donner:1                                 #noeud communicatif dominantde la relative
  }
  main-> suivre:1                                   #noeud communicatif dominant de la principale
}
}
```

Figure 91 Input de la phrase dans GenDR pour la RSém

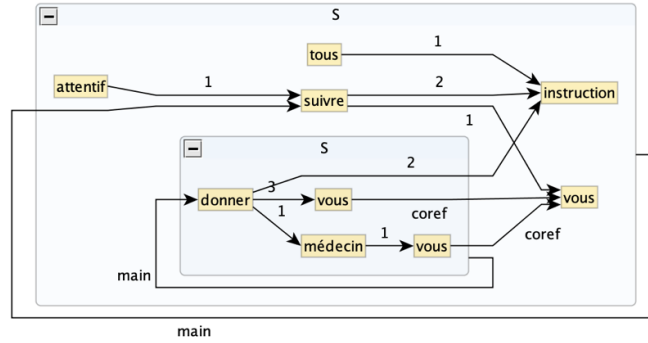


Figure 92 RSém de la phrase (59)

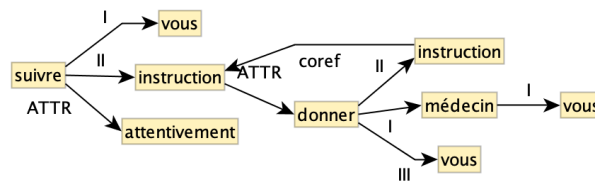


Figure 93 RSyntP de la phrase (59)

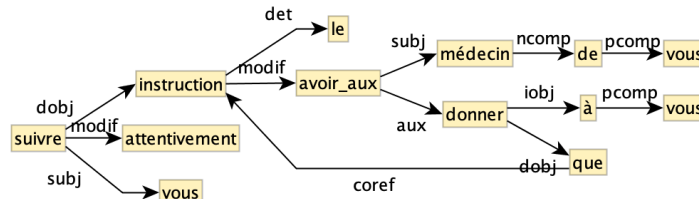


Figure 94 RSyntS de la phrase (59)

Nous pouvons observer qu'à l'interface sémantique-syntaxe profonde la règle *relative_direct* est utilisée et qu'à l'interface syntaxe profonde-syntaxe de surface la règle sur le pronom *que* est utilisée.

6.2.3. Avec un pronom relatif *dont* en complément du nom

Pour le pronom *dont* en complément du nom, nous avons générer la phrase (60) à partir de l'input Figure 95, la Figure 96 pour la RSém, la Figure 97 pour la RSyntP et la Figure 98 pour la RSyntS. La phrase (60) est la même que la (50), la Figure 97 est la même que la Figure 72 et la Figure 98 est la même que la Figure 74.

(60) *L'arbre dont je vois la cime indique la direction.*

```

structure Sem s1 {
  S:1{
    text="J'ai lu le livre dont Luc parle."
    arbre:1{
      S:2{
        moi:1{
          voir:1{
            tense=PRES
            1-> moi:1
            2-> cime:1
          }
          cime:1{
            number=SG definiteness=DEF
            <-> arbre:1
          }
          main-> voir:1
        }
        indiquer:1{
          tense=PRES
          1-> arbre:1
          2-> direction:1
        }
        direction:1{number=SG definiteness=DEF}
        main-> indiquer:1
      }
    }
  }
}

```

#pour ouvrir une RSém dans GenDR
 #la proposition principale
 #le nom modifié
 #la proposition relative
 #verbe au sommet de la relative
 #noeud communicatif dominant de la relative
 #verbe principal
 #noeud communicatif dominant de la principale

Figure 95 Input de la phrase (60) dans GenDR pour la RSém

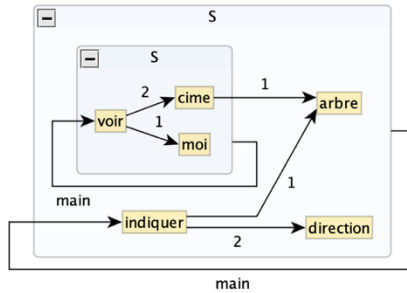


Figure 96 RSém de la phrase (60)

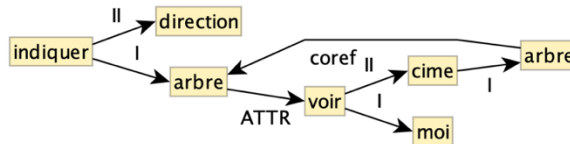


Figure 97 RSyntP de la phrase (60)

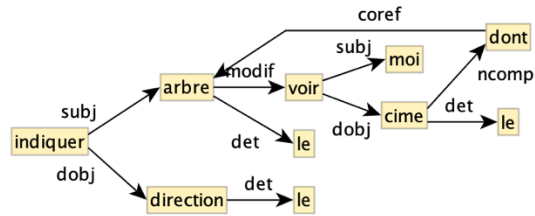


Figure 98 RSyntS de la phrase (60)

Nous pouvons observer qu'à l'interface sémantique-syntaxe profonde la règle `relative_indirect` est utilisée et qu'à l'interface syntaxe profonde-syntaxe de surface la règle sur le pronom *dont* est utilisée.

6.2.4. Avec un pronom relatif *dont* en oblique

Pour le pronom *dont* en oblique, nous avons généré la phrase (61) à partir de l'input Figure 99, la Figure 100 pour la RSém, la Figure 101 pour la RSyntP et la Figure 102 pour la RSyntS. La phrase (61) est la même que la (49), la Figure 101 est la même que la Figure 71 et la Figure 102 est la même que la Figure 73.

(61) *J'ai lu le livre dont Luc parle.*

```

structure Sem s1 {
  S:1{
    text="J'ai lu le livre dont Luc parle."
    moi:1{}
    lire:1{
      tense=PAST
      1-> moi:1
      2-> livre:1
    }
    livre:1{number=SG definiteness=DEF}
  }
  S:2{
    Luc:1{class=proper_noun}
    parler:1{
      tense=PRES
      1-> Luc:1
      3-> livre:1
    }
    main-> parler:1
  }
  main-> lire:1
}

```

#pour ouvrir une RSém dans GenDR
#proposition principale
#verbe principal
#nom modifié
#proposition relative
#verbe au sommet de la relative
#nœud communicatif dominant de la relative
#nœud communicatif dominant de la principale

Figure 99 Input de la phrase (61) dans GenDR pour la RSém

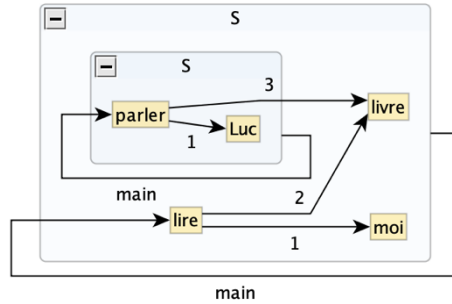


Figure 100 RSém de la phrase (61)

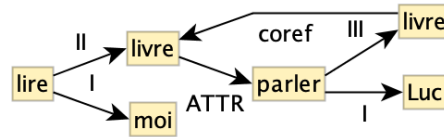


Figure 101 RSyntP de la phrase(61)

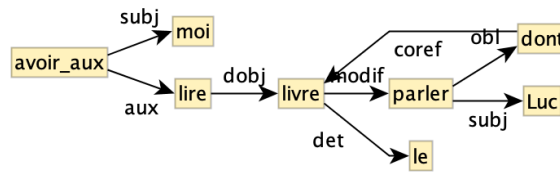


Figure 102 RSyntS de la phrase(61)

Nous pouvons observer qu'à l'interface sémantique-syntaxe profonde la règle `relative_direct` est utilisée et qu'à l'interface syntaxe profonde-syntaxe de surface la règle sur le pronom `dont` est utilisée.

6.2.5. Avec une préposition suivie du pronom *qui*

Pour une préposition suivie du pronom *qui*, nous avons généré la phrase (62) à partir de l'input Figure 103, la Figure 104 pour la RSém, la Figure 105 pour la RSyntP et la Figure 106 pour la RSyntS. La phrase (62) est la même que la (51), la Figure 105 est la même que la Figure 79 et la Figure 106 est la même que la Figure 80.

(62) *Chris offre une fleur à Aria à qui j'ai prêté un livre.*

```
structure Sem s1 {
  S:1{
    text="Chris offre des fleurs à Aria à qui j'ai prêté un livre."
  }
}
```



```

Chris:1{class=proper_noun}
offrir:1{
    tense=PRES
    1-> Chris:1
    2-> fleur:1
    3-> Aria:1
}
fleur:1{number=SG definiteness=INDEF}
Aria:1{class=proper_noun}
S:2{
    moi:1{
        prêter:1{
            tense=PAST
            1-> moi:1
            2-> livre:1
            3-> Aria:1
        }
        livre:1{number=SG definiteness=INDEF}
        main-> prêter:1
    }
    main-> offrir:1
}

```

Figure 103 Input de la phrase (62) dans GenDR pour la RSém

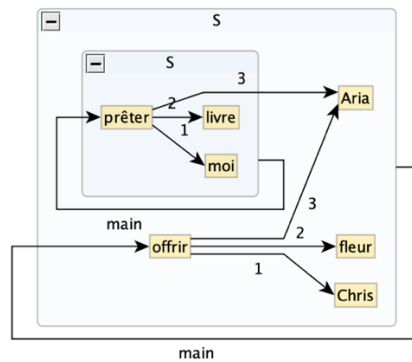


Figure 104 RSém de la phrase (62)

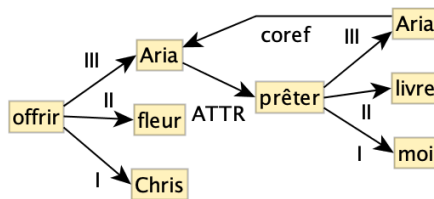


Figure 105 RSyntP de la phrase(62)

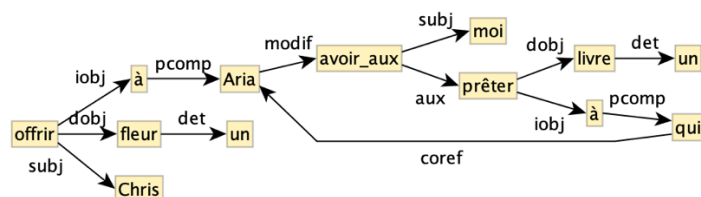


Figure 106 RSyntS de la phrase (62)

Nous pouvons observer qu'à l'interface sémantique-syntaxe profonde la règle `relative_direct` est utilisée et qu'à l'interface syntaxe profonde-syntaxe de surface la règle sur une préposition suivie du pronom *qui* est utilisé.

6.2.6. Avec une préposition suivie du pronom *lequel*

Pour une préposition suivie du pronom *lequel*, nous avons généré la phrase (63) à partir de l'input Figure 107, la Figure 108 pour la RSém, la Figure 109 pour la RSyntP et la Figure 110 pour la RSyntS. La phrase (63) est la même que la (52), Figure 109 est la même que la Figure 85 et la Figure 110 est la même que la Figure 86.

(63) *Aria a lu un livre avec lequel elle s'est endormie.*

```

structure Sem s1 {                                     #pour ouvrir une RSém dans GenDR
  S:1{                                                #proposition principale
    text="Aria a lu un livre avec quoi elle s'est endormi le soir"
    Aria:1{class=proper_noun}
    lire:1{                                          #verbe principal
      tense=PAST
      1-> Aria:1
      2-> livre:1
    }
    livre:1{number=SG definiteness=INDEF}          #nom modifié
  }
  S:2{                                                #proposition relative
    elle:1{
      <-> Aria:1
    }
    "s'endormir":1{                                  #verbe au sommet de la relative
      tense=PAST
      1-> elle:1
      2-> livre:1
    }
    soir:1{
      number=SG definiteness=DEF
      1-> "s'endormir":1
    }
    main-> "s'endormir":1                            #noeud communicatif dominant de la relative
  }
  main-> lire:1                                       #noeud communicatif dominant de la principale
}
}

```

Figure 107 Input de la phrase dans GenDR pour la RSém

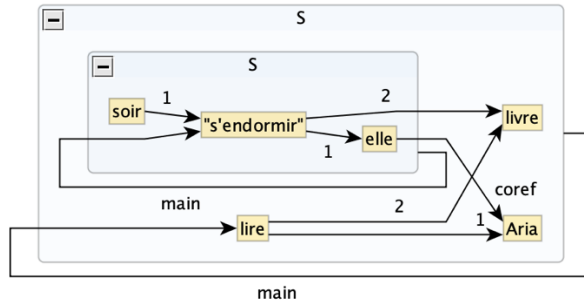


Figure 108 RSém de la phrase (63)

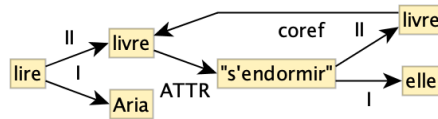


Figure 109 RSyntP de la phrase (63)

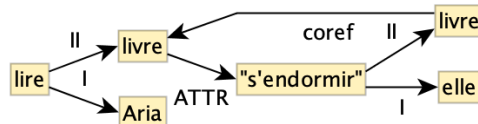


Figure 110 RSyntS de la phrase(63)

Nous pouvons observer qu'à l'interface sémantique-syntaxe profonde la règle *relative_direct* est utilisée et qu'à l'interface syntaxe profonde-syntaxe de surface la règle sur une préposition suivie du pronom *lequel* est utilisé.

6.2.7. Avec un pronom relatif *lequel*

Pour le pronom *lequel*, nous avons généré la phrase (64) à partir de l'input Figure 111, la Figure 112 pour la RSém, la Figure 113 pour la RSyntP et la Figure 114 pour la RSyntS.

(64) *Nous traitons de l'actualité urgente laquelle traite des prisonniers.*

[Europar.550_00093]

```
structure Sem s1 {
  S:1{
    sent_id="Europar.550_00093" text="Nous discutons de l'actualité urgente, laquelle traite des
    prisonniers."
    nous:1{
      discuter:1{
        tense=PRES
      }
    }
  }
}
#pour ouvrir une RSém dans GenDR
#proposition principale
#verbe principal
```

```

1-> nous:1
2-> actualité:1
}
actualité:1{number=SG definiteness=DEF}
urgent:1{
1-> actualité:1 #nom modifié
}
S:2{ #proposition relative
#verbe au sommet de la relative
traiter:1{
tense=PRES
1-> actualité:1
2-> prisonnier:1
}
prisonnier:1{number=PL definiteness=INDEF}
main-> traiter:1 #noeud communicatif dominant de la relative
}
main-> discuter:1 #noeud communicatif dominant de la principale
}
}

```

Figure 111 Input de la phrase dans GenDR dans la RSém

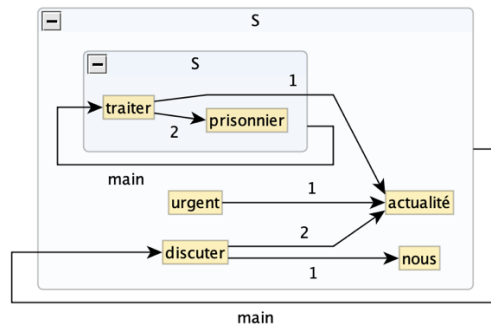


Figure 112 Sém de la phrase (64)

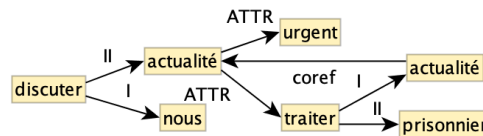


Figure 113 SyntP de la phrase (64)

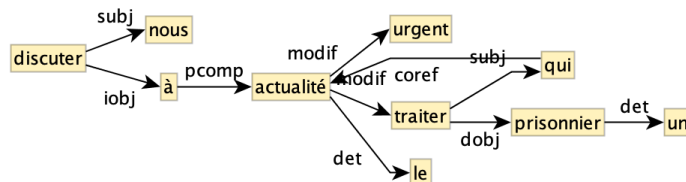


Figure 114 Synt de la phrase (64)

Nous pouvons observer qu'à l'interface sémantique-syntaxe profonde la règle `relative_direct` est utilisée et qu'à l'interface syntaxe profonde-syntaxe de surface la règle

du pronom *qui* est utilisé. C'est le mauvais pronom qui est généré à l'interface syntaxe profonde-syntaxe de surface, car il n'y a pas de règle du pronom *lequel* sujet dans GenDR.

6.2.8. Avec un pronom relatif où

Pour le pronom *où*, nous avons généré la phrase (65) à partir de l'input Figure 115, la Figure 116 pour la RSém, la Figure 117 pour la RSyntP et la Figure 118 pour la RSyntS.

(65) *Aria vit à Paris où Aline étudie.*

```

structure Sem s1 {
  S:1{
    text="Aria vit à Paris où Aline étudie"
    Aria:1{class=proper_noun}
    vivre:1{
      tense=PRES
      1-> Aria:1
      2-> Paris:1
    }
    Paris:1{
      class=city
      1-> étudier:1
    }
  }
  S:2{
    étudier:1{
      tense=PRES
      1-> Aline:1
    }
    Aline:1{class=proper_noun}
    LocIn:1{
      class=noun
      1-> étudier:1
      2-> Paris:2
    }
    Paris:2{class=city}
    main-> étudier:1
  }
  main-> vivre:1
}

```

#pour ouvrir une RSém dans GenDR
 #proposition principale
 #verbe principal
 #nom modifié
 #proposition relative
 #verbe au sommet de la relative
 #lieu
 #noeud communicatif dominant de la relative
 #noeud communicatif dominant de la principale

Figure 115 Input de la phrase dans GenDR dans la RSém

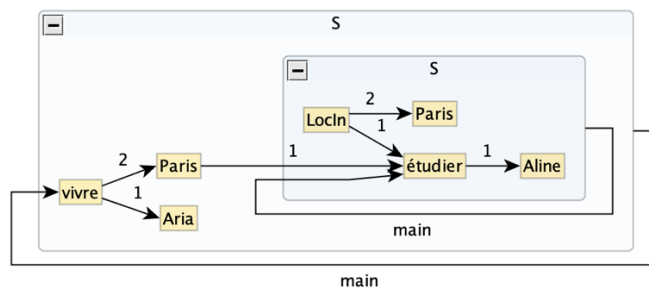


Figure 116 Sém de la phrase (65)

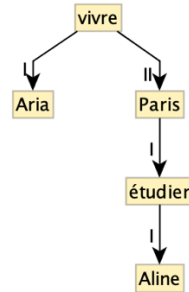


Figure 117 SyntP de la phrase (65)

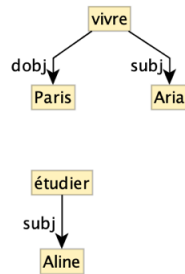


Figure 118 SyntS de la phrase (65)

Nous pouvons observer qu'aucune règle n'est utilisé ni à l'interface sémantique-syntaxe profonde avec les règles sur les relatives ni à l'interface syntaxe profonde-syntaxe de surface avec les règles sur les pronoms relatifs. Les propositions principale et relative sont séparées.

6.2.9. Avec une préposition suivie du pronom *quoi*

Pour une préposition suivie du pronom *quoi*, nous avons généré la phrase (66) à partir de la Figure 119, la Figure 120 pour la RSém, la Figure 121 pour la RSyntP et la Figure 122 pour la RSyntS.

(66) *Une infirmière donne la perfusion de bivalirudine au patient pendant une heure, après quoi l'infirmière retire la perfusion.* [Sequoia, emea-fr-dev_00110]

```

structure Sem s1 {
  S:1{
    sent_id="emea-fr-dev_00110"
    text="Une infirmière donne la perfusion de bivalirudine au patient pendant 1 heure, après
    quoi l'infirmière retire la perfusion."
    infirmière:1{number=SG definiteness=INDEF}
    perfusion:1{number=SG definiteness=DEF}
    bivalirudine:1{
      number=SG definiteness=INDEF
      1-> perfusion:1
    }
    donner:1{
  }
}
#pour ouvrir une RSém dans GenDR
#proposition principale
#numéro d'identification de la phrase
#verbe principal
  
```

```

tense=PRES
1-> infirmière:1
2-> perfusion:1
3-> patient:1
}
patient:1{number=SG definiteness=INDEF}
pendant:1{
  1-> donner:1
}
1:1{
  class=numeral
  1-> heure:1
}
heure:1{
  number=SG definiteness=INDEF
  1-> pendant:1
}
S:2{
  retirer:1{
    tense=PRES
    1-> infirmière:2
    2-> perfusion:2
    3-> patient:1
  }
  infirmière:2{
    number=SG definiteness=DEF
    <-> infirmière:1
  }
  perfusion:2{
    number=SG definiteness=DEF
    <-> perfusion:1
  }
  main-> retirer:1
}
main-> donner:1
}
}

```

#nom modifié

#proposition relative
#verbe au sommet de la relative

#noeud communicatif dominant de la relative

@noeud communicatif dominant de la principale

Figure 119 Input de la phrase dans GenDR dans la RSém

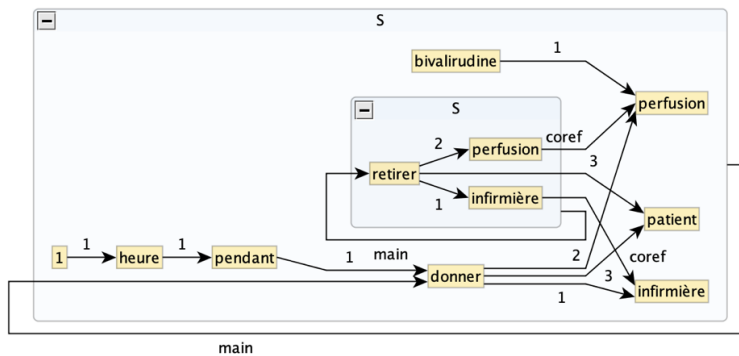


Figure 120 Sem de la phrase (66)

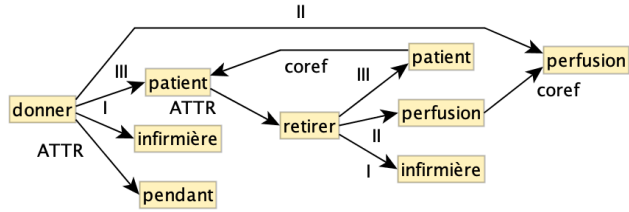


Figure 121 SyntP de la phrase (66)

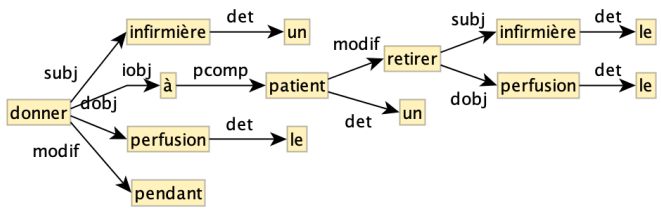


Figure 122 SyntS de la phrase (66)

Nous pouvons observer qu'à l'interface sémantique-syntaxe profonde aucune règle sur les relatives n'est utilisée. La mauvaise règle des pronoms est utilisée à l'interface syntaxe profonde-syntaxe de surface, nous avons la règle sur le pronom *qui* qui a été utilisé au lieu d'avoir la règle d'une préposition suivie de *lequel* ou rien.

6.2.10. Avec un verbe modal

Pour une phrase avec un verbe modal, nous avons généré la phrase (67) à partir de l'input Figure 123, la Figure 124 pour la RSém, la Figure 125 pour la RSyntP et la Figure 126 pour la RSyntS.

(67) *Leo offre des fleurs à Tina qui veut aller dans la forêt.*

```

structure Sem s1 {
  S:1{
    text="Leo offre des fleurs à Tina qui veut aller dans la forêt"
    Leo:1{class=proper_noun}
    offrir:1{
      tense=PRES
      1-> Leo:1
      2-> fleur:1
      3-> Tina:1
    }
    fleur:1{number=PL definiteness=INDEF}
    Tina:1{class=proper_noun}
    S:2{
      vouloir:1{
        tense=PRES
        1-> Tina:1
        2-> aller:1
      }
    }
  }
}

```



```

aller:1{
  tense=INF
  1-> forêt:1
}
forêt:1{number=SG definiteness=DEF}
main-> vouloir:1
}
main-> offrir:1
}
}

```

#noeud communicatif dominant de la relative
#noeud communicatif dominant de la principale

Figure 123 Input de la phrase dans GenDR dans la RSém

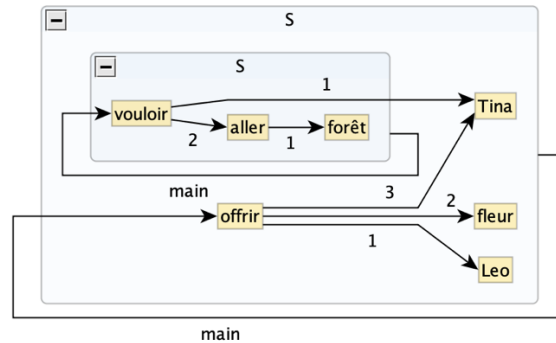


Figure 124 Sém de la phrase (67)

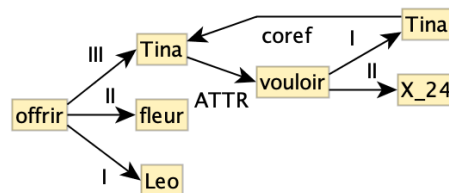


Figure 125 SyntP de la phrase (67)

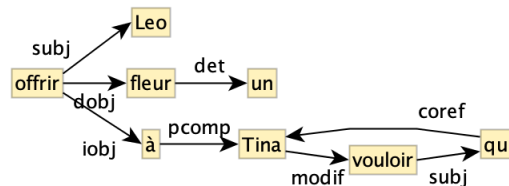


Figure 126 SyntS de la phrase (67)

Nous pouvons observer qu'à l'interface sémantique-syntaxe profonde la règle *relative_direct* est utilisée, mais le verbe non modal n'est pas généré et qu'à l'interface syntaxe profonde-syntaxe de surface la règle sur le pronom *qui* est utilisé, mais le verbe non modal n'est pas généré. Cependant, c'est un meilleur résultat qu'avec la règle de Dubinskaite en

Erreur ! Source du renvoi introuvable. au niveau SyntP et en **Erreur ! Source du renvoi introuvable.** au niveau SyntS .

6.2.11. Avec une fonction lexicale

Pour une phrase avec une fonction, nous avons la (68), la Figure 128 pour la RSém, la Figure 129 pour la RSyntP et la Figure 130 pour la RSyntS.

(68) *Anne admire l'affection que Bob à pour Carla.*

```

structure Sem s1 {
  S:1{
    text="Anne admire l'affection que Bob éprouve pour Carla."
    Anne:1{class=proper_noun}
    admirer:1{
      tense=PRES
      1-> Anne:1
      2-> affectionner:1
    }
  }
  S:2{
    affectionner:1{
      tense=PRES number=SG definiteness=DEF
      1-> Bob:1
      2-> Carla:1
    }
    Bob:1{class=proper_noun}
    Carla:1{class=proper_noun}
    main-> affectionner:1
  }
  main-> admirer:1
}

```

Figure 127 Input de la phrase dans la GenDR pour la RSém

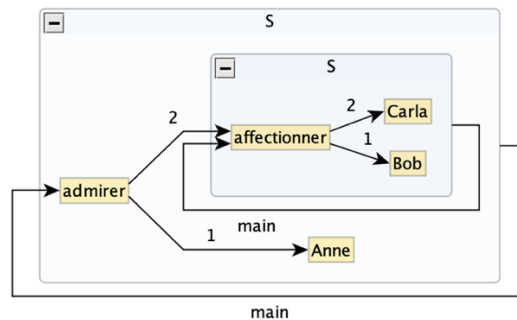


Figure 128 Sém de la phrase (68)

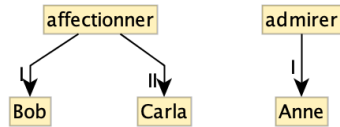


Figure 129 SyntP de la phrase (68)

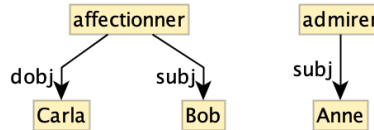


Figure 130 SyntS de la phrase (68)

Nous pouvons observer qu’aucune règle n’est utilisé ni à l’interface sémantique-syntaxe profonde avec les règles sur les relatives ni à l’interface syntaxe profonde-syntaxe de surface avec les règles sur les pronoms relatifs. Les propositions principale et relative sont séparées.

6.2.12. Synthèse

Voici un tableau récapitulatif des pronoms et de la génération ou non des phrases.

Pronoms	Sém-SyntP	SyntP-SyntS
Qui	Direct	Qui
Que	Direct	Que
Dont complément du nom	Indirect	Dont
Dont oblique	Direct	Dont
Où	Échec	Échec
Lequel	Direct	Qui, même fonction
Préposition + qui	Direct	Préposition + qui
Préposition + quoi	Direct	Pas de pronom généré
Préposition + lequel	Direct	Préposition + lequel
Verbe modal	Échec	Qui, bon pronom, le verbe qui suit échec
Fonction lexicale	Échec	Échec

Tableau 6 Tableau récapitulatif des relatives générées

Voici le tableau regroupant le nombre de pronoms par fonctions qui peuvent générer une phrase, les nombres viennent du Tableau 4.

Sur 4512 propositions relatives	Bon pronom	Phrase générée
Nombre	4023	4060
%	89,1%	90%

Tableau 7 Nombres et pourcentages de génération des propositions relatives

Nous pouvons voir dans ce tableau que 89,1 % des propositions peuvent être générées avec le bon pronom (*qui, que, dont*, une préposition suivie de *qui* et une préposition suivie de *lequel*) et jusqu'à 90 % des propositions relatives peuvent être générées avec ou sans le bon pronom relatif (*qui, lequel, que, dont*, une préposition suivie de *qui* et une préposition suivie de *lequel*).

6.3. Discussion

Des phrases avec proposition relative directe et indirecte dans l'interface sémantique-syntaxe profonde et les pronoms *qui, que, dont*, préposition suivie du pronom *qui* et préposition suivie de *lequel* et dans une moindre mesure le pronom *lequel* ont pu être générés par GenDR. Comme nous l'avons vu dans la partie précédente, ces règles ont des limites et les voici.

6.3.1. Les pronoms relatifs

La lexicalisation des pronoms se fait par des règles spécifiques sur les pronoms relatifs au niveau de la syntaxe profonde et de la syntaxe de surface, certains pronoms ne peuvent pas se générer dans les phrases dans GenDR.

6.3.1.1. Le pronom *lequel*

Comme nous l'avons vu dans la partie précédente, les phrases avec le pronom *lequel* peuvent se générer au niveau de l'interface sémantique-syntaxe profonde par la règle `relative_direct`

ou la règle `relative_indirect`. Cependant au niveau de l'interface syntaxique profonde-syntaxe de surface, il n'y a pas de règle pour le pronom *lequel* qui ferait de la concurrence au pronom *qui* de la même fonction. Donc les phrases se génèrent bien, mais avec le pronom *qui* au lieu de *lequel*, car nous n'avons pas trouvé un moyen de faire coexister des règles générales pour les pronoms d'une même fonction. Le choix s'est porté sur le pronom *qui* car il est le pronom le plus fréquent dans le corpus. Nous avons aussi une hypothèse, d'après la Figure 37 suivante, peut-être que le pronom *qui* est privilégié pour la proposition relative directe et le pronom *lequel* pour la proposition relative indirecte.

6.3.1.2. Une préposition suivie du pronom *quoi*

Comme nous l'avons vu dans la partie précédente, les phrases avec une préposition suivie du pronom *quoi* peuvent se générer au niveau de l'interface sémantique-syntaxe profonde par la règle `relative_direct` ou la règle `relative_indirect`. Cependant au niveau de l'interface syntaxique profonde-syntaxe de surface, il n'y a pas de règle pour une préposition suivie du pronom *quoi* qui ferait peut-être de la concurrence à une préposition suivie du pronom *lequel* de la même fonction, mais surtout parce qu'il n'y a que quatre phrases sur 4 505 propositions relatives dans le corpus. Donc, nous avons préféré implémenter la règle une préposition suivie du pronom *lequel* de la même fonction et pouvant être animé ou inanimé. Aucun pronom ne se génère à l'interface syntaxique profonde-syntaxe de surface.

6.3.1.3. Le pronom *où*

Les phrases avec le pronom *où* ne peuvent pas se générer au niveau de l'interface sémantique-syntaxe profonde par la règle `relative_direct` ou la règle `relative_indirect` ni à l'interface syntaxique profonde-syntaxe de surface, par une règle des pronoms relatif. Il n'y a pas de règle pour le pronom *où*. Le pronom *où* porte le trait lieu ou temps, ce qui n'est pas implémenté dans GenDR et il ne peut pas être remplacé par un autre pronom. Pour pouvoir implémenter une règle pour le pronom *où*, il faudrait compléter le système de règles sémantiques et syntaxiques de GenDR, ce qui n'est pas l'objet de ce mémoire.

6.3.2. Les verbes modaux

Les phrases avec un verbe modal qui sont dans leur sens modal dans la phrase ne peuvent pas être générées par les règles des propositions relatives, surtout au niveau de l'interface sémantique-syntaxique profonde par la règle `relative_direct` ou la règle `relative_indirect`. Il arrive qu'un pronom relatif puisse se générer au niveau de l'interface syntaxique profonde-syntaxe de surface s'il y a les règles des pronoms relatifs. Cependant, l'actant du verbe modal n'est pas généré au niveau de l'interface sémantique-syntaxe profonde ni au niveau de l'interface syntaxique profonde-syntaxe de surface dans tous les cas. GenDR n'a pas de module sur les verbes modaux et un verbe à l'infinitif, comme dans l'exemple (67), ne peut être considéré comme un actant d'un verbe modal. Ce problème dépasse le cadre de ce mémoire.

6.3.3. Les verbes supports

Sur le même principe que les verbes modaux en français, certains verbes ont un comportement différent de la plupart des verbes lexicaux. « Un verbe support est un collocatif verbal sémantiquement vide dans le contexte de la collocation, dont la fonction linguistique est de “verbaliser” une base nominale, c'est-à-dire de la faire fonctionner dans la phrase comme si elle était elle-même un verbe. » (Polguère, 2016, p. 205)

Autrement dit, les verbes supports tiennent plus des collocations ou une fonction lexicale. Les phrases avec des verbes supports sont générées comme des verbes sémantiquement pleins si le patron du régime du verbe n'est pas dans le dictionnaire Lexicon ou ne se génère pas comme pour les fonctions lexicales. Les systèmes des fonctions lexicales et des propositions relatives ne sont pas compatibles pour l'instant dans GenDR. Ce problème n'a rien avoir avec les propositions relatives.

6.3.4. Les fonctions lexicales

Le système des fonctions lexicales de GenDR ne fonctionne pas avec la partie sur les propositions relatives. Les phrases ne se génèrent pas entièrement comme dans l'exemple (68). Les fonctions

lexicales sont définies « Une fonction lexicale donnée f décrit une relation existante entre une lexie L – appelée l'argument de f – et un ensemble de lexies ou de syntagmes appelé la valeur de l'application de f à la lexie L . La fonction lexicale f est telle que :

1. l'expression $f(L)$ représente l'application de f à L ;

2. chaque élément de la valeur de $f(L)$ est lié à L (à peu près) de la même façon.

Il existe autant de fonctions lexicales qu'il existe de types de relations lexicales et chaque fonction lexicale est identifiée par un nom particulier : *Syn* (pour la synonymie), *Anti* (pour l'antonymie), etc. » (Polguère, 2016, p. 196)

Autrement dit, une fonction lexicale est relation entre un groupe de lexie qui a une valeur particulière et qui peut changer le sens littéral des mots. Par exemple, dans l'expression *peur bleu*, *bleu* ne signifie pas la couleur, mais la fonction lexicale *Magn*, donc « très » ou « intense ».

Le système des fonctions lexicales n'est pas compatible avec les propositions relatives et les phrases ne se génèrent pas comme dans l'exemple (68). Le système des fonctions lexicales est complexe et contient une centaine de règles. Ce problème de compatibilité dépasse le cadre de notre mémoire.

6.3.5. Projectivité

La notion de projectivité est liée à l'ordre des mots dans la phrase (Mel'čuk, 2015, 1988). Avec la structure des règles syntaxiques des propositions relatives au niveau de l'interface de la syntaxe profonde et de la syntaxe de surface avec les règles sur les pronoms relatifs, les phrases sont difficiles à linéariser et donc ne permet pas de faire une phrase projective pour le réalisateur GenDR. Néanmoins, les propositions relatives ont tendance à ne pas être des structures projectives en français, notamment avec le pronom *dont* (Mel'čuk, 2015, p.486). De plus, le problème de linéarisation des propositions relatives n'est pas un problème de proposition relative, mais de GenDR.

6.4. Synthèse

Nous avons implémenté les règles `relative_direct` (§5.1.2) et `relative_indirect` (§5.1.3) dans l'interface sémantique-syntaxe profonde, puis dans l'interface syntaxe profonde - syntaxe de surface nous avons implémenté les règles pour les pronoms relatifs suivant : *qui* (§5.2.1), *que* (§5.2.2), *dont* (§5.2.3), une préposition suivie de *qui* (§5.2.4) et une préposition suivie de *lequel* (§5.2.5). Autrement dit, nous avons couvert les propositions relatives directes et indirectes dans l'interface sémantique-syntaxe profonde et cinq des neuf pronoms relatifs dans l'interface syntaxe profonde-syntaxe de surface. Cela correspond à 89,1 % des propositions relatives qui peuvent être générées avec le bon pronom du corpus de proposition relative.

Les phrases avec les pronoms *lequel*, *où* et une préposition suivie de *qui* ne peuvent pas être générées dans l'interface syntaxe profonde-syntaxe de surface. Les verbes supports, les fonctions lexicales et les verbes modaux ne peuvent être générés ni dans l'interface sémantique-syntaxe profonde ni dans l'interface syntaxe profonde-syntaxe de surface.

Chapitre 7 — Conclusion

Ce projet cherche à implémenter des règles syntaxiques profondes et des règles syntaxiques de surface pour les propositions relatives en français dans un générateur de texte automatique profond GenDR. Ainsi les trois objectifs principaux de ce mémoire sont atteints.

- 1) Nous avons défini les propositions relatives et établi deux grandes catégories les propositions relatives directes et indirectes.
- 2) À partir de l'ancienne règle des propositions relatives de Dubinskaite, nous avons créé des nouvelles règles dans l'interface sémantique-syntaxe profonde, les règles `relative_direct` et `relative_indirect`. Ces règles rattachent la proposition principale à la proposition relative, prennent en compte la dimension de modification de la proposition relative et la coréférence dans la proposition relative.
- 3) Nous avons créé des règles dans l'interface syntaxe profonde-syntaxe de surface qui peuvent générer les pronoms relatifs *qui*, *que*, *dont*, préposition suivie de *qui* et préposition suivie de *lequel*.

Nous avons vu que le traitement des propositions relatives se fait dans l'interface sémantique-syntaxique (§ 2.1.3), et implique plus particulièrement, la structure sémantique, la structure sémantico-communicative, la structure syntaxique profonde, la structure syntactico-anaphorique profonde, la structure syntaxique de surface et la structure syntactico-anaphorique de surface. Puis, nous avons observé comment les réalisateurs linguistiques traitent les propositions relatives (§2.2). En GAT, il y a deux types de réalisateurs linguistiques, les réalisateurs de surface qui se situent sur les niveaux de syntaxe de surface, de morphologie profonde, de morphologie de surface, et les réalisateurs profonds qui se situent sur les niveaux sémantiques, syntaxe profonde, syntaxe de surface. Les réalisateurs de surfaces traitent assez bien les propositions relatives, mais exigent en entrée une structure syntaxique déjà largement déterminée. La génération des propositions relatives au niveau des réalisateurs profonds reste un défi, et nous avons proposé une implémentation des relatives dans GenDR, un réalisateur profond multilingue basé sur la TST (§2.3).

D'abord, une typologie des propositions relatives a été faite (§3.1). Seules les relatives avec antécédent nous intéressent, et les relatives restrictives et descriptives sont traitées de la même façon par notre implémentation. Enfin, la proposition relative a été définie comme une proposition qui a un rôle de modification et dont le pronom relatif a une relation de coréférence avec la lexie modifiée (§ 3.4). Puis, deux grands types de propositions relatives, direct et indirect, ont été définis. La proposition relative directe a un lien sémantique direct entre le nom modifié et le verbe à la tête de la proposition relative, tandis que la proposition relative indirecte a un lien sémantique indirect avec un ou des prédicats intermédiaires entre le verbe au sommet de la proposition relative et le nom modifié. Les spécificités des propositions relatives sont :

- Il doit y avoir deux verbes finis.
- Il doit y avoir une relation de modification.
- Une proposition doit être introduite par un pronom relatif.
- Il doit y avoir une coréférence entre la proposition relative et la proposition principale.

De plus, sur le plan pratique, un corpus a été analysé, formé de trois corpus, GSD, Sequoia et ParTUT en SUD (§ 4.1). Ces corpus annotés sont intéressants, car ils sont déjà constitués, nettoyés, annotés et gratuits à télécharger, à modifier et à redistribuer gratuitement. Des règles dans le logiciel GREW ont été faites pour trouver les pronoms relatifs et les propositions relatives directes et indirectes dans le corpus. Le patron nom-verbe est le plus courant pour une proposition relative où un nom dans la proposition principale est modifié par le verbe de la proposition relative (§ 4.2). Les propositions relatives directes sont les plus fréquentes à environ 78 %, alors qu'il n'y a qu'environ 22 % de propositions relatives indirectes dans tout le corpus. Le rôle syntaxique du nom modifié le plus fréquent est `comp:obj`. Le rôle syntaxique du nom modifié n'influence pas le type de proposition relative qui le modifie. Le pronom relatif le plus fréquent est *qui* à 58,8 %, puis, *que* à 13,8 %, *dont* 10,2 %, *où* 10 %, une préposition suivie de *lequel* 5,7 %, préposition suivie de *qui* 0,7 % et préposition suivie de *quoi* 0,1 %.

Ensuite, avec pour point de départ l'ancienne règle de 2017 sur les propositions relatives au niveau sémantique-syntaxe profonde (§5.1.1) nous avons créé de nouvelles règles syntaxiques pour les propositions relatives. En effet, nous avons implémenté deux nouvelles règles sur les

propositions relatives, la règle `relative_direct` (§5.1.2) et la règle `relative_indirect` (§5.1.3) dans l'interface sémantique-syntaxe profonde. Puis, en syntaxe de surface, nous avons implémenté cinq règles pour les pronoms *qui* (§5.2.1), *que* (§5.2.2), *dont* (§5.2.3), une préposition suivie de *qui* (§5.2.4) et une préposition suivie de *lequel* (§5.2.5), ainsi que la gestion du trait d'animacité, ainsi que les exceptions qui ne correspondent pas aux paramètres par défauts (§5.3).

Enfin, les règles créées peuvent en théorie (à condition d'avoir les ressources lexicales nécessaires) générer environ 90 % des phrases du corpus avec une proposition relative (§6.2). Les limites des règles syntaxiques des propositions relatives sont la génération des pronoms *lequel* sujet et *quoi* complément prépositionnel, la génération des phrases avec le pronom *où* et *qui* objet au niveau de la syntaxe profonde et de la syntaxe de surface, le traitement des verbes modaux et la génération des phrases avec un verbe à l'infinitif après un verbe modal, le traitement des verbes supports et des collocations (§6.3).

Pour conclure, les règles de l'interface sémantique-syntaxe pour les propositions relatives directes et indirectes sont assez générales pour convenir à d'autres langues que le français. De plus, puisque nos règles suivent les patrons GREW définis (§ 4.2), nous pouvons supposer qu'une réimplémentation de notre grammaire en GREW serait possible.

Références bibliographiques

- Bohnet, B., & Wanner, L. (2010). Open Source Graph Transducer Interpreter and Grammar Development Environment. *Proceedings of the International Conference on Language Resources and Evaluation*. Valette, Malte.
- Bohnet, B., Langjahr, A., & Wanner, L. (2000). A development Environment for an MTT-based Sentence Generator. *Proceedings of the First International Conference on Natural Language Generation*, (pp. 260-263).
- Bollmann, M. (2011). Adapting SimpleNLG to German. *Proceedings of the 13th European Workshop on Natural Language Generation*, (pp. 133-138).
- Bosco, C., & Sanguinetti, M. L. (2012). The Parallel-TUT: a Multilingual and Multiformat Treebank. *Proceedings of the Eight Conference on Language Resources and Evaluation*. Istanbul, Turquie.
- Candito, M., & Djammé, S. (2012). Le corpus Sequoia : annotation syntaxique et exploitation pour l'adaptation d'analyseur par pont lexical. *Proceedings of the Joint Conference JEP-TALN-RECITAL. 2 : TALN*, pp. 321-334. Grenoble, France: ATALA/AFCP.
- Danlos, L. (1983). Présentation d'un modèle de génération automatique. *Revue québécoise de linguistique*, 13(1), 203-228.
- Daoust, N. (2013). *JSreal : un réalisateur de texte pour la programmation web*. Mémoire de maîtrise: Université de Montréal.
- Daoust, N., & Lapalme, G. (2014). JSreal: A Text Realizer for Web Programming. Dans *Language Production, Cognition, and the Lexicon* (Vol. 48, pp. 363-378). Springer.
- De Marneffe, M.-C., & Manning, C. D. (2008). The Stanford Typed Dependencies Representation. *Proceedings of the workshop on Cross-Framework and Cross-Domain Parser Evaluation* (pp. 1-8). Manchester, UK: Coling 2008 Organizing Committee.

- De Oliveira, R., & Sripada, S. (2014). Adapting SimpleNLG for Brazilian Portuguese realisation. *Proceedings of the 8th International Natural Language Generation Conference*, (pp. 93-94).
- Delaveau, A. (1987). Pronom relatif sans antécédent et pronom interrogatif - La forme qui. *Linx*, 16, 5-27.
- Dubé, M. (à paraître). *Le traitement des locutions en génération automatique de texte*. Montréal: Université de Montréal [Mémoire de maîtrise].
- Dubinskaite, I. (2017). *Développement de ressources lituaniennes pour un générateur automatique de texte multilingue*. Montréal: Université de Montréal [Mémoire de maîtrise].
- Garreta-Piquette, D. (2018). *Intégration de VerbNet dans un réalisateur profond*. Montréal: Université de Montréal [Mémoire de maîtrise].
- Gatt, A., & Krahmer, E. (2018). Survey of the State Art in Natural Language Generation : core tasks, applications and evaluation. *Journal of Artificial Intelligence Research* 61, 65-170.
- Gatt, A., & Reiter, E. (2009). SimpleNLG: A realisation engine for practical applications. *Proceedings of European Workshop on Natural Language Generation*, (pp. 90-93).
- Gerdes, K., Guillaume, B., Kahane, S., & Perrier, G. (2019). Improving Surface-syntactic Universal Dependencies (SUD): MWEs and deep syntactic features. *Proceedings of the 18th International Workshop on Treebanks and Linguistic Theories* (pp. 126-132). Paris, France: Association for Computational Linguistics.
- Grevisse, M., & Goose, A. (2016). *Le bon usage : Grevisse langue française*. Paris: De Boeck supérieur.
- Guillaume, B., de Marneffe, M.-C., & Perrier, G. (2019). Conversion et améliorations de corpus du français annotés en Universal Dependencies. *Traitement Automatique des Langues*, (pp. 71-95).

- Hadermann, P. (1987). *Histoire d'un pronom relatif : "dont"*. Roumanie: Editura Academiei Republicii Socialiste Romania.
- He, L. (2020). *Un dictionnaire de régimes verbaux en mandarin*. Montréal: Université de Montréal [Mémoire de maîtrise].
- Kahane, S. (2009). On the Status of Phrases in Head-Driven Phrase Structure Grammar: Illustration by a Fully Lexical Treatment of Extraction. Dans I. Mel'čuk, & A. Polguère, *Dependency in Linguistic Description* (pp. 111-150). Amsterdam/Philadelphie: John Benjamins.
- Kahane, S., & Polguère, A. (2001). Formal foundation of lexical functions. *Collocation: Computational Extraction, Analysis and Exploitation*, (pp. 8-15). Toulouse.
- Kleiber, G. (1987). *Relatives restrictives et relatives appositives : une opposition "introuvable"?* Germany:Tübingen: Max Niemeyer Verlag Tübingen.
- Lambrey, F. (2017). *Implémentation des collocations pour la réalisation de texte multilingue*. Montréal: Université de Montréal [Mémoire de maîtrise].
- Lambrey, F., & Lareau, F. (2015). Le traitement des collocations en génération de texte multilingue. *Actes de la 22e conférences sur le traitement automatique des langues naturelles*, (pp. 263-269).
- Lapalme, G. (2021). *The jsRealB Text Realizer: Organization and Use Cases*.
- Lareau, F. (2002). *La synthèse automatique de paraphrases comme outil de vérification des dictionnaires et grammaires de type Sens-Texte*. Montréal: Université de Montréal [Mémoire de maîtrise].
- Lareau, F., & Wanner, L. (2007). Towards a generic multilingual dependency grammar for text generation. *Proceedings of the Gramma Engineerings Across Frameworks Workshop* (pp. 203-223). CSLI Publications.
- Lareau, F., Lambrey, F., Dubinskaite, I., Galarreta-Piquette, D., & Nejat, M. (2018). GenDR: A Generic Deep Realizer with Complex Lexicalization. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*. Miyazaki, Japon.

- Lavoie, B., & Rambow, O. (1997). A Fast and Portable Realizer for Text Generation Systems. *Fifth Conference on Applied Natural Language Processing*, (pp. 265-268).
- Leeman, D. (2002). *La phrase complexe*. Paris: Duculot.
- Mazzei, A., Battaglino, C., & Bosco, C. (2016). SimpleNLG-IT: adapting Simple NLG to Italian. *Proceedings of the 9th International Natural Language Generation Conference*, (pp. 184-192).
- Mel'čuk, I. (1973). Towards a Linguistic "Meaning \Leftrightarrow Text" Model. *Trends in Soviet Theoretical Linguistics*(Kiefer, Ferec), 33-57.
- Mel'čuk, I. (1988). *Dependency Syntax : Theory and Practice*. Albany : N. Y.: The SUNY Press.
- Mel'čuk, I. (1993). *Cours de morphologie général. Introduction + Le mot* (Vol. 1). Montréal/Paris: PUM/CNRS éditions.
- Mel'čuk, I. (1994). *Cours de morphologie général. Significations morphologiques* (Vol. 2). Montréal/Paris: PUM/CNRS éditions.
- Mel'čuk, I. (1996). *Cours de morphologie général. Syntactiques morphologiques* (Vol. 3). Montréal/Paris: PUM/CNRS éditions.
- Mel'čuk, I. (1997). *Cours de morphologie général. Signes morphologiques* (Vol. 4). Montréal/Paris: PUM/CNRS éditions.
- Mel'čuk, I. (1997). *Vers une linguistique Sens-Texte. Leçon inaugurale*. Paris: Collège de France.
- Mel'čuk, I. (2000). *Cours de morphologie général. Modèles morphologiques* (Vol. 5). Montréal/Paris: PUM/CNRS éditions.
- Mel'čuk, I. (2001). *Communication Organization in Natural Language: The Semantic-Communicative Structure of Sentences* (Vol. Studies in Language Companion Series, 57). John Benjamins Publishing.
- Mel'čuk, I. (2001). *Communicative Organization in Natural Language. The Semantic-Communicative Structure of Sentences*. Amsterdam/Philadelphie: John Benjamin.

- Mel'čuk, I. (2001). Semantics and the Lexicon in Modern Linguistics. *article non publié*, 17.
- Mel'čuk, I. (2009). Dependency in Natural Language. Dans I. Mel'čuk, & A. Polguère, *Dependency in Linguistic Description* (pp. 1-110). Amsterdam/Philadelphie: John Benjamins.
- Mel'čuk, I. (2012). *Semantics: From meaning to text* (Vol. 1). Amsterdam/Philadelphie: John Benjamins.
- Mel'čuk, I. (2013). *Semantics: From Meaning to Text* (Vol. 2). Amsterdam/Philadelphie: John Benjamins.
- Mel'čuk, I. (2015). *Semantics: From Meaning to Text* (Vol. 3). Amsterdam/Philadelphie: John Benjamins.
- Mel'čuk, I. (2015). Un modèle linguistique fonctionnel : Le modèle Sens-Texte. INALCO.
- Mel'čuk, I., & Milićević, J. (2014). *Introduction à la linguistique* (Vol. 2). Paris: Herman.
- Mel'čuk, I., & Milićević, J. (2014). *Introduction à la linguistique* (Vol. 1). Paris: Hermann.
- Mel'čuk, I., & Milićević, J. (2014). *Introduction à la linguistique* (Vol. 3). Paris: Hermann.
- Milićević, J. (2006). A short Guide to the Meaning-Text Linguistic Theory. *Journal of Koralex*, 8, 187-233.
- Milićević, J. (2007). *La paraphrase. Modélisation de la paraphrase langagière*. Peter Lang.
- Mille, S., & Wanner, L. (2017). a demo of FORGe: the Pompeu Fabra Open Rule-based Generator. *Proceedings of the 10th International Natural Language Generation conference* (pp. 245-246). Santiago de Compostela, Espagne: Association for Computational Linguistics.
- Mille, S., Carlini, R., Burga, A., & Wanner, L. (2017). FORGe at SemEval-2017 Task9: Deep sentence generation based on a sequence of graph transducers. *Proceedings of the 11th International Workshop on Semantic Evaluations* (pp. 920-923). Vancouver, Canada: Association for Computational Linguistics.
- Molins, P., & Lapalme, G. (2015). Approche systémique pour la réalisation multilingue de textes.

- Nivre, J., De Marneffe, M.-C., Ginter, F., Hajič, J. M., Pyysalo, S., Shuster, S., . . . Zeman, D. (2020). Universal Dependencies v2: An Evergrowing Multilingual Treebank Collect. *Proceedings of the 12th Language Resources and Evaluation Conference*, (pp. 4043-4043). Marseille, France.
- Osborne, T., & Gerdes, K. (2019). The status of function words in dependency grammar: A critique of Universal Dependencies (UD). *Glossa: a journal of general linguistics*, 4; 17, 1-28.
- Petrov, S., Das, D., & McDonald, R. (2011). A Universal Part-of-Speech Tagset. *Computing Research Repository - CORR*, 2089–2096.
- Polguère, A. (1998). La théorie Sens-Texte. *Dialangue*, 8-9, 9-30.
- Polguère, A. (2016). *Lexicologie et sémantique lexicale*. Montréal: PUM.
- Ramos-Sato, A., Janeiro-Gallardo, J., & Bugarin Diz, A. (2017). Adapting simple NLG to Spanish. *Proceedings of the 10th International Conference on Natural Language Generation*, (pp. 144-148).
- Reiter, E., & Robert, D. (2000). *Building Natural Language Generation System*. Cambridge: Cambridge University Press.
- Riegel, M., Pellat, J.-C., & Rioul, R. (2018). *Grammaire méthodique du français* (Vol. 7e éditions). Paris: PUF.
- SyntaxFest. (2019). Proceedings of the Third Workshop on Universal Dependencies. *Proceedings of the Third Workshop on Universal Dependencies*. Paris.
- Vaudry, P.-L., & Lapalme, G. (2013). Adapting SimpleNLG for Bilingual English-French Realisation. *Proceedings of the 14th European Workshop on Natural Language Generation* (pp. 183-187). Sofia, Bulgarie: Association for Computational Linguistics.
- Wanner, L. (1992). Lexical Choice and the Organization of Lexical Resources in Text Generation. *Proceedings of the 10th European Conference on Artificial Intelligence*, (pp. 495-499).

- Wanner, L. (1996). *Lexical Functions in Lexicography and Natural Processing*. Amsterdam/Philadelphie: John Benjamin.
- Wanner, L., & Bateman, J. A. (1990). A collocational based approach to salience-sensitive lexical selection. *Proceedings of the 5th International Workshop on Natural Language Generation*.
- Wanner, L., Bohnet, B., Bouayad-Agha, N., Lareau, F., & Niclaß, D. (2010). MARQUIS: Generation of User-Tailored Multilingual Air Quality Bulletins. *Applied Artificial Intelligence*, 914-952.
- Zeman, D. (2008). Reusable Tagset Conversion Using Tagset Drivers. *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2008*, (pp. 213-218). Marrakech, Maroc.
- Zhao, X. (2018). *Les collocations du champs sémantique des émotions en mandarin*. Montréal: Université de Montréal [Mémoire de maîtrise].