



HAL
open science

Formalisation de la langue ukrainienne avec NooJ : préparation du module ukrainien

Olena Saint-Joanis

► **To cite this version:**

Olena Saint-Joanis. Formalisation de la langue ukrainienne avec NooJ: préparation du module ukrainien. Linguistique. Université Bourgogne Franche-Comté, 2024. Français. NNT: 2024UBFCC005 . tel-04690884

HAL Id: tel-04690884

<https://theses.hal.science/tel-04690884v1>

Submitted on 6 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNIVERSITÉ DE BOURGOGNE FRANCHE-COMTÉ
ÉCOLE DOCTORALE « LETTRES, COMMUNICATIONS, LANGUES, ARTS »

Thèse en vue de l'obtention du titre de docteur en
SCIENCES DU LANGAGE

présentée et soutenue publiquement par

Olena Saint-Joanis

Le 29 mars 2024

**FORMALISATION DE LA LANGUE UKRAINIENNE AVEC NOOJ :
PRÉPARATION DU MODULE UKRAINIEN**

Sous la direction de M. le Professeur **Max SILBERZTEIN**

et

co-direction de M. le Maître de Conférences HDR **Oleg CHINKAROUK**

Membres du jury :

M. Vincent BÉNET, Maître de Conférences émérite à l'Institut National des Langues et Civilisations Orientales, examinateur

M. Christian BOITET, Professeur émérite à l'Université Grenoble Alpes, président du jury

M. Thierry POIBEAU, Directeur de Recherche CNRS (LATTICE), rapporteur

M. Ivan ŠMILAUER, Maître de Conférences à l'Institut National des Langues et Civilisations Orientales, examinateur

Remerciements

Je tiens tout d'abord à remercier mes directeurs de thèse, le Professeur Max Silberztein et le Maître de Conférences HDR Oleg Chinkarouk. Je vous témoigne toute ma gratitude pour vos enseignements et conseils, pour le temps que vous m'avez consacré, l'intérêt que vous avez porté à mon travail, ainsi que pour votre bienveillance et votre patience.

J'ai beaucoup appris à vos côtés, autant sur les outils de TAL, leurs possibilités, que sur la grammaire de la langue ukrainienne.

La découverte de NooJ a été un réel plaisir et je souhaite poursuivre mes recherches avec NooJ, continuer à explorer ses possibilités.

Je remercie également les collègues Noojiens avec lesquels j'ai pu échanger lors des conférences ou sur le forum NooJ et qui m'ont suggéré des pistes à suivre.

Je tiens aussi à remercier ma famille, mes amis et collègues de l'INALCO qui ont été avec moi à toutes les étapes d'avancement de mon projet : votre soutien et vos encouragements m'ont été très précieux.

À ma famille
À mon Ukraine

Signes conventionnels utilisés

1. Pour retranscrire les noms, les toponymes et les quelques mots ukrainiens en français, nous avons décidé d'utiliser uniquement la transcription française¹. Nous prendrons en comptes les recommandations de la CNT pour l'écriture de certaines lettres dans des positions spécifiques. Cela permettra au grand public, non initié à la slavistique, de lire facilement les mots retranscrits.

Table de correspondance : les lettres ukrainiennes retranscrites avec des lettres françaises

А а (А а)	a	П п (П п)	p
Б б (Б б)	b	Р р (Р р)	r
В в (В в)	v	С с (С с)	s, ss (entre deux voyelles)
Г г (Г г)	h	Т т (Т т)	t
Ґ ґ (Ґ ґ)	g, gu (devant e et i)	У у (У у)	ou
Д д (Д д)	d	Ф ф (Ф ф)	f
Е е (Е е)	e	Х х (Х х)	kh
Є є (Є є)	ie, ie après a, o, e	Ц ц (Ц ц)	ts
Ж ж (Ж ж)	j	Ч ч (Ч ч)	tch
З з (З з)	z	Ш ш (Ш ш)	ch
И и (И и)	y	Щ щ (Щ щ)	chtch
І і (І і)	i	Ю ю (Ю ю)	iou, iou après a, o, e
Ї ї (Ї ї)	ï	Я я (Я я)	ia, ia après a, o, e
Й й (Й й)	y ²	Ь ь (Ь ь)	« ' » ; et ьо = io
К к (К к)	k	'	l'apostrophe, qui est un signe, n'est pas retranscrite
Л л (Л л)	l		
М м (М м)	m		
Н н (Н н)	n		
О о (О о)	o		

Pour retranscrire les textes, nous n'utiliserons pas la translittération³ scientifique ISO/R 9⁴, adoptée par les slavistes, car elle est utile surtout lorsqu'on donne uniquement le texte translittéré,

¹ La transcription française, ou romanisation de la langue ukrainienne en français, a été décrite par Jean Girodet dans le *Dictionnaire de la langue française* (Paris : Bordas, 1976). Elle permet de reproduire le mot phonétiquement. Elle est reprise également par la Commission nationale des toponymes (CNT) avec quelques recommandations pratiques. Cette transcription est référencée aussi dans tous les outils en ligne permettant de convertir les textes ukrainiens.

² Dans la combinaison d'une voyelle et de la lettre « y », cette dernière retranscrit toujours la lettre « ь ».

³ La translittération permet de reproduire la graphie, mais ne donne pas d'indications sur la prononciation.

⁴ La translittération scientifique ISO/R 9 des langues slaves est basée sur le latin croate. Exemple : ж (ua) → j (fr) → ž (ISO/R 9); щ (ua) → chtch (fr) → šč (ISO/R 9), <http://translit.kh.ua/?lat&isor9#ling>.

parce qu'elle permet de retrouver le texte original avec précision. Or, nous allons donner le texte original en italique, puis sa traduction en italique et entre crochets [] :

повний [plein]

Dans quelques rares cas, nous serons amenée à donner la transcription des mots ukrainiens. Le mot et sa translittération seront alors notés en italique, entre guillemets, puis la transcription entre barres obliques // :

« *русинський* » /roussyns'kyu/

Nous avons noté que dans les études consacrées au traitement automatique de l'ukrainien, les translittérations ISO/R 9 et ISO 9 ne sont pas utilisées.

Les quelques mots russes et les noms propres russes seront retranscrits du russe à l'aide de la même transcription française.

Table de correspondance des lettres cyrilliques russes différentes retranscrites avec des lettres françaises

Г г (Г г)	g, gu (devant e et i)
Е е (Е е)	e, ie à l'initiale, après ъ et ь, ie après а, о et у
Ё ё (Ё ё)	e
И и (И и)	i
Ъ ъ (Ъ ъ)	(signe dur) n'est pas retranscrit
Ы ы (Ы ы)	y
Э э (Э э)	è

Les toponymes seront écrits selon la tradition française, par exemple *Галичина [Galicie]*. Par ailleurs, nous utiliserons l'endonyme⁵ *Київ* pour nommer la capitale de l'Ukraine, et *Дніпро* pour nommer le fleuve principal, et non Kiev et Dniepr, en usage en français, mais retranscrits du russe. Selon les normes ukrainiennes de 1995-1996⁶, révisées en 2010 et adoptées par les pays anglophones, tous les toponymes doivent être transcrits ou translittérés de la langue ukrainienne. La CNT a adopté cette pratique pour toutes les villes hormis la capitale et le fleuve principal⁷. En outre, la division francophone des Nations unies stipule :

« Un système de translittération national qui n'a pas reçu d'approbation de résolution par le GENUNG est pris en considération dès lors que son usage s'avère officiel. Exemple : système officiel ukrainien 1996, révisé 2010⁸. »

Pour les mêmes raisons, nous préférons utiliser le toponyme Bélarus et l'ethnonyme bélarussien⁹.

⁵ L'endonyme est un toponyme transcrit selon les normes actualisées d'usage du pays.

⁶ Protocole n° 1 du 14 octobre 1995, traduit en français : https://www.axl.cefan.ulaval.ca/europe/ukraine_loi-1995-toponymie.htm.

⁷ Communication de la CNT du 18 mars 2022 : https://cnig.gouv.fr/IMG/documents_wordpress/2022/04/CNT-CNIG_CR-projet_18-mars-2022_vuPJ.pdf

⁸ http://www.toponymiefrancophone.org/divfranco/Bougainville/Systemes_de_romanisation.aspx

⁹ <http://www.inalco.fr/sites/default/files/asset/document/belarusien-2022-2023.pdf>

2. Pour marquer une transformation, nous utiliserons le symbole « → » et nous mettrons les éléments qui se transforment en vert :

повний [plein] → повен [plein].

3. Pour mettre en évidence un élément de la phrase, nous le marquerons en gras :

*За зелен **садом** [derrière le **jardin** vert]. Ici, le nom **jardin** est à l'instrumental.*

4. Dans les références, nous indiquerons la page lorsque nous nous appuyerons sur un article, un chapitre ou un livre entier.

Dans la bibliographie en ukrainien, nous donnerons d'abord le nom de l'auteur et le titre en ukrainien, puis la transcription du nom et la traduction du titre en français entre crochets, l'édition et la ville seront données uniquement en ukrainien :

56. Русанівський, Віталій (2001). *Історія української літературної мови*. [Roussanivs'kyy. Histoire de la langue ukrainienne]. Артк, Київ.

Les références des articles encyclopédiques dans le texte seront notées en ukrainien :

(Encyclopédie 2022 : *Українська мова*).

Abréviations

ALU – unité linguistique atomique

f. – féminin

m. – masculin

n. – neutre

pl. – pluriel

an. – animé

inan. – inanimé

PF – perfectif

IPF – imperfectif

Nom. – nominatif

Acc. – accusatif

Gén. – génitif

Loc. – locatif

Dat. – datif

Inst. – instrumental

« j » marque le son « yod », qui est transcrit par les lettres *я* et *ю* dans les terminaisons.

« Ø » signifie l'absence de terminaison.

Table des matières

Remerciements	I
Signes conventionnels utilisés.....	III
Abréviations	VI
Introduction	4
Chapitre 1. Évolution de la langue ukrainienne à travers l’histoire	6
1.1 Introduction	6
1.2 Informations générales sur la langue ukrainienne	7
1.3 Étapes de l’évolution de l’ukrainien.....	8
1.3.1 Du slave commun aux dialectes proto-ukrainiens	8
1.3.2 La période archaïque de l’ancien ukrainien	9
1.3.3 La période médiane de l’ancien ukrainien	10
1.3.4 L’ukrainien au XIX ^e siècle (1798-1900)	13
1.3.5 L’ukrainien de 1900 à 1922	19
1.3.6 L’ukrainien de 1922 à 1939	25
1.3.7 L’ukrainien de 1939 à l’indépendance de l’Ukraine (1991).....	31
1.3.8 L’ukrainien de 1991 à nos jours.....	37
1.4 Conclusion du chapitre 1	44
Chapitre 2. Outils de TAL au service des linguistes	46
2.1 Introduction	46
2.2 L’approche empirique et les outils stochastiques pour l’ukrainien.....	47
2.2.1 Sketch Engine	48
2.2.2 TreeTagger et RNNTagger	52
2.3 L’approche formelle.....	56
2.4 Outils formels pour l’ukrainien	58
2.4.1 VESUM.....	59
2.4.2 GRAC.....	60
2.4.3 LanguageTool API NLP UK	62
2.5 La plateforme NooJ	64
2.5.1 Les dictionnaires NooJ.....	64
2.5.2 Les grammaires NooJ	66
2.5.3 Nos appréciations.....	72
2.6 Conclusion du chapitre 2	73
Chapitre 3. Grammaire morphologique de la langue ukrainienne	75

3.1 Introduction	75
3.2 Les classes d'ALU dans la langue ukrainienne	75
3.2.1 Définition d'une classe grammaticale	75
3.2.2 Division en classes et en groupes de classes	76
3.3 Description des classes grammaticales.....	79
3.3.1 Le nom	79
3.3.2 L'adjectif	90
3.3.3 Le numéral.....	98
3.3.4 Le pronom	101
3.3.5 Le verbe	104
3.3.6 L'adverbe	113
3.3.7 La préposition.....	118
3.3.8 La conjonction	120
3.3.9 La particule	122
3.3.10 L'interjection	125
3.4 Conclusion du chapitre 3	127
Chapitre 4. Le module ukrainien pour NooJ.....	129
4.1 Introduction	129
4.2 Description du fichier « _properties.def » pour le module ukrainien	129
4.2.1 Le choix du jeu d'étiquettes	129
4.2.2 Affichage des éléments	130
4.2.3 Choix des catégories NooJ pour le module ukrainien	131
4.2.4 Déclaration des variables pour les flexions	132
4.3 Dictionnaires ukrainiens pour le module ukrainien de NooJ.....	132
4.3.1 Description du dictionnaire principal	132
4.3.2 Description du dictionnaire secondaire	134
4.3.3 Les ALU dans le dictionnaire du module ukrainien	135
4.3.4 Les grammaires flexionnelles et dérivationnelles	136
4.3.5 Évaluation des dictionnaires	153
4.4 Grammaires morphologiques pour le module ukrainien de NooJ	155
4.4.1 Formalisation des participes avec une grammaire morphologique.....	156
4.4.2 Traitement des verbes avec le postfixe -ся	158
4.4.3 Formalisation des formes verbales orales	160
4.4.4 Formalisation des noms diminutifs ou affectifs	161

4.4.5	Formalisation des adjectifs d'appartenance	163
4.4.6	Formalisation des adjectifs relatifs à partir des noms propres	164
4.4.7	Formalisation des adjectifs et des adverbes, partitifs, excessifs, diminutifs	166
4.4.8	Formalisation du degré de comparaison des adjectifs et des adverbes	168
4.4.9	Formalisation des formes courtes des adjectifs.....	171
4.4.10	Formalisation des formes négatives	172
4.4.11	Formalisation des adverbes et des pronoms indéfinis et négatifs.....	173
4.4.12	Formalisation des alternances au début des mots	174
4.4.13	Formalisation des mots en alphabet latin ou en russe	176
4.4.14	Formalisation des noms propres.....	178
4.4.15	Évaluation de l'ensemble des grammaires morphologiques	178
4.5	Grammaires syntaxiques de désambiguïsation pour le module ukrainien	180
4.5.1	Typologie des homonymes grammaticaux.....	180
4.5.2	Grammaires syntaxiques pour désambiguïser les homonymes flexionnels	181
4.5.3	Désambiguïsation des homoformes et des homonymes transposés	189
4.5.4	Évaluation de l'ensemble des grammaires syntaxiques et perspectives	193
4.6	Utilisations pratiques envisagées.....	194
4.6.1	Utilisation pédagogique	194
4.6.2	Reconnaissance des entités nommées	197
4.6.3	Étude des variantes de la langue ukrainienne	197
4.7	Conclusion du chapitre 4	199
	Conclusion	200
	Liste des figures.....	202
	Liste des tableaux.....	205
	Annexes	207
	Bibliographie	220

Introduction

Ce travail s'intéresse à la formalisation de la langue ukrainienne grâce à la plateforme NooJ.

La langue ukrainienne est très peu décrite dans le monde occidental, alors que c'est la langue officielle d'un pays européen qui compte plus de 45 millions d'habitants et qui est représenté dans plusieurs institutions mondiales. L'ukrainien est également étudié dans plusieurs universités d'Europe.

De plus, dans le contexte actuel de la guerre, le pays se retrouve sur le devant de la scène internationale et au centre de l'intérêt non seulement des médias, mais aussi des chercheurs qui étudient le discours des acteurs principaux de la vie politique en Ukraine.

De ce fait, la formalisation de l'ukrainien à travers un outil informatique pourra trouver plusieurs applications pratiques. Elle permettra notamment de faire l'analyse morphosyntaxique et sémantique approfondie des corpus, et jouera un rôle non seulement dans le développement des applications de TAL (par exemple, extracteurs d'entités nommées, terminologie, traduction automatique, correcteur d'orthographe, etc.), et aussi dans le domaine de l'enseignement assisté par ordinateur (EAO).

En plus de préparer le module ukrainien pour NooJ, nous avons voulu décrire la grammaire de la langue ukrainienne pour pouvoir utiliser nos travaux dans les cours que nous dispensons à l'INALCO ; nous y assurons les cours de thème et de version et, depuis peu, les cours de grammaire pratique.

Dans cette optique, nous avons décidé d'étudier la langue ukrainienne à travers son histoire, son orthographe et sa grammaire, pour appréhender ses particularités et comprendre à quels types de difficultés nous pourrions être confrontée lors de sa formalisation avec un outil informatique.

Ainsi, nous avons fait le point sur la grammaire de la langue ukrainienne pour mieux comprendre les classes grammaticales et déterminer la meilleure façon de formaliser tel ou tel phénomène linguistique. Nous avons étudié l'histoire pour découvrir d'où viennent les nombreuses variantes dans la langue ukrainienne et de quelle façon nous pourrions les inclure dans notre projet.

Ensuite, nous avons décidé d'étudier la plateforme NooJ, que nous avons découverte lors d'un séminaire, d'examiner ses performances et sa fiabilité, de la comparer avec d'autres outils qui permettent de formaliser une langue naturelle. Nous avons découvert que, dans cette optique, NooJ s'avère être le meilleur choix, car il intègre des outils de formalisation (description des catégories, construction des dictionnaires, des grammaires flexionnelles et syntaxiques) et de recherche, permettant ainsi de travailler avec n'importe quel texte ou corpus. NooJ contient également des applications pédagogiques qui peuvent être utilisées en séance de travaux pratiques.

De plus, il s'agit d'un outil formel qui a fait ses preuves, un outil fiable, capable de décrire avec précision tout phénomène linguistique.

Dans cette thèse, nous décrivons d'abord la base théorique sur laquelle sont fondées nos recherches, puis la grammaire morphologique de la langue ukrainienne, avant de décrire son implémentation dans NooJ et de proposer quelques pistes pour de futures recherches.

Notre travail se présente en quatre parties. Dans le premier chapitre, nous parlerons des origines de la langue ukrainienne, de ses normes orthographiques, de sa grammaire, de son

vocabulaire, et des changements que cette langue a subis. Nous nous demanderons s'il est possible de décrire toutes ses particularités et ses phénomènes avec NooJ.

Dans le deuxième chapitre, nous étudierons deux approches qui permettent de concevoir les outils de TAL : l'approche empirique, basée sur les données, et l'approche formelle, basée sur les règles. Nous étudierons attentivement les outils existants pour l'ukrainien, basés sur ces deux approches, à savoir Sketch Engine, TreeTagger et RNNTagger (outils issus de l'approche empirique) et VESUM, GRAC, LangageTool API NLP UK (outils issus de l'approche formelle). Nous y décrirons aussi la méthodologie de NooJ.

Le troisième chapitre sera consacré à la morphologie de la langue ukrainienne. Nous nous limiterons à décrire les phénomènes qu'il faudrait prendre en compte lors de la phase d'implémentation.

Dans le quatrième chapitre, nous décrirons le module ukrainien pour NooJ (version V.1.3) que nous avons construit lors de nos recherches. Ce module est composé d'un dictionnaire principal, de deux dictionnaires secondaires, de quelques grammaires morphologiques qui permettent de reconnaître les formes absentes du dictionnaire et de quelques grammaires syntaxiques qui lèvent les ambiguïtés. Ces ressources seront décrites et évaluées. Notons que nous nous sommes limitée dans ce travail à préparer un modèle qui couvre la langue standard.

Nous comptons continuer ce travail pour améliorer nos ressources et les utiliser en pratique. De ce fait, à la fin de ce quatrième chapitre, nous présenterons quelques cas pratiques que nous pourrions explorer.

Pour terminer, nous ferons un bilan général de notre travail et dégagerons quelques pistes à exploiter dans nos futures recherches.

Chapitre 1. Évolution de la langue ukrainienne à travers l'histoire

1.1 Introduction

La langue ukrainienne moderne est composée de plusieurs couches qui se sont formées au cours de l'histoire : la langue normative (normalisée par l'Académie des sciences de l'Ukraine), le « sourjyk » (un sociolecte, apparu suite à l'interférence de la langue russe), les dialectes régionaux et la langue de la diaspora. Toutes ces couches interagissent et se reflètent dans la littérature, mais aussi sur Internet (blogues, forums, réseaux sociaux), à la télévision et dans la presse.

Dans ce chapitre, nous aimerions nous concentrer sur les événements historiques et géopolitiques qui ont eu un impact sur la langue ukrainienne moderne. Cela nous permettra alors de mieux percevoir les particularités morphologiques, syntaxiques et orthographiques que nous devrions formaliser avec NooJ. Par ailleurs, nous ne visons pas à décrire l'histoire de l'Ukraine de manière exhaustive : nous donnerons simplement des éléments nécessaires à la compréhension de l'évolution de la langue et du contexte dans lequel les linguistes ukrainiens ont travaillé.

Nous présenterons les événements historiques période par période. Pour chaque période, nous donnerons un bref aperçu des normes orthographiques. Puis, pour les périodes les plus anciennes, nous listerons des ouvrages qui reflètent la langue ukrainienne, et pour les périodes les plus récentes, nous présenterons des dictionnaires et des grammaires, en détaillant davantage les ouvrages importants qui servent des références pour les études linguistiques actuelles.

Notons que pour examiner l'évolution de la langue ukrainienne à travers l'histoire, nous avons utilisé plusieurs ouvrages.

Ainsi, pour décrire le contexte historique, nous nous sommes appuyée sur les descriptions données dans l'*Encyclopédie d'histoire de l'Ukraine* (2019). Puis, nous avons croisé ces informations avec les publications d'Ivan Ohienko, Iouri Cheveliov, Anna Tchavarha, Iryna Dmytrychyn, Boulyk-Verkhola S. et Verkhola A., Nina Virtchenko¹⁰.

Pour étudier les origines de la langue et son évolution diachronique, nous avons utilisé la *Grande Encyclopédie académique de l'Ukraine* (2022) et les ouvrages de Cheveliov et Ohienko, cités précédemment, ainsi que les travaux de Stepan Bezenko, Vitaliy Roussaniv's'kyy, Vassyl' Nimtchouk et Nataliia Pouriaïeva, Mariia Brous, Viktoriia Badenkova, Laryssa Holoïoukh et bien d'autres, que nous mentionnerons au fur et à mesure¹¹.

Ensuite, pour donner un aperçu des dictionnaires et des grammaires, nous avons effectué des recherches dans les bibliothèques ukrainiennes en ligne¹² qui donnent accès à de nombreux ouvrages. Ainsi, nous avons pu consulter chaque ouvrage que nous allons présenter plus loin.

¹⁰ (Ohienko 1927 [45], 1949 [46]), (Cheveliov 1962 [64], 1987 [65]), (Tchavarha 2017 [69]), (Dmytrychyn 2018 [93]), (Boulyk-Verkhola et Verkhola 2018 [7]), (Virtchenko 2011 [13]).

¹¹ (Bezenko 1960 [2]), (Roussaniv's'kyy 1985 [55], 2001 [56]), (Nimtchouk et Pouriaïeva 2004 [43]), (Brous 2016 [5]), (Badenkova 2018 [1]), (Holoïoukh 2021 [15]).

¹² 1) *Електронна бібліотека "УКРАЇНІКА"* [Bibliothèque en ligne « Ukrainienne »], une ressource intégrée à la Bibliothèque nationale d'Ukraine Vernad's'kyy <http://irbis-nbuv.gov.ua/cgi-bin/ua/elib.exe?q=&C21COM=F&I21DBN=SIMPLE&P21DBN=UKRLIB> ; 2) *Електронна бібліотека «Чтиво»* [Bibliothèque en ligne « Lecture »] <https://chtyvo.org.ua>.

1.2 Informations générales sur la langue ukrainienne

Aujourd'hui, la langue ukrainienne compte environ 45 millions de locuteurs à travers le monde, ce qui lui permet d'être classée parmi les 20 langues les plus parlées (Encyclopédie 2022 : *Українська мова*).

L'ukrainien est une langue flexionnelle, ce qui signifie que la terminaison d'une unité lexicale peut changer en fonction de son rapport avec les autres éléments de la phrase.

L'ukrainien a son propre alphabet cyrillique de 33 lettres.

Selon la classification généalogique, l'ukrainien fait partie du groupe oriental des langues slaves, qui font partie de la famille des langues indoeuropéennes. Par ailleurs, la langue ukrainienne a des caractéristiques phonétiques et grammaticales communes à plusieurs langues slaves modernes : 29 traits avec le biélorussien et le haut sorabe, 27 avec le bas sorabe¹³, 23 avec le tchèque et le slovaque, 21 avec le polonais, 20 avec le serbe et le macédonien, 18 avec le slovène et 11 avec le russe (Encyclopédie 2022 : *Українська мова*).

Notons que l'ethnonyme « ukrainien », qui désigne le peuple ukrainien et sa langue, est apparu au XII^e siècle, mais ne fut fixé définitivement qu'au XIX^e siècle ; jusque-là prédominaient les ethnonymes « roussien » ou « ruthène », qui indiquaient le lien avec la Rous' kyïvienne ou les ethnonymes « petit-russe » et « russe méridional », qui indiquaient le lien incontestable et la subordination à l'Empire russe.

Victoria Saïdi (2009) [128] souligne que l'ethnonyme « ruthène », en ukrainien « *русинський* » /*roussyns'kyj*/, a été abandonné, car il ne reflétait pas vraiment l'identité du peuple ukrainien, vu qu'en historiographie européenne on se servait de ce nom pour désigner uniquement la population de la Galicie orientale. L'ethnonyme « roussien », en ukrainien « *руський* » /*rous'kyj*/, prêtait à confusion avec l'ethnonyme « russe », en russe « *русский* » /*rousskiy*/, or il s'agit d'identités nationales différentes.

Quant aux ethnonymes « petit-russe », en russe « *малороссийский* » /*malorossiyskiy*/ et « russe méridional », en russe « *южнорусский* » /*iujnorousskiy*/, ils sont apparus lors de l'expansion de l'Empire russe à la fin du XVII^e siècle dans le cadre d'un projet visant à anéantir l'identité ukrainienne. Dès lors, la langue dominante a reçu l'appellation de « langue russe », en russe « *русский язык* » /*rousskiy iazyk*/, ou de « grand-russe », en russe « *великорусский язык* » /*velikorousskiy iazyk*/, tandis que l'ukrainien a reçu l'appellation de « dialecte », traduction du russe « *наречие* » /*narechie*/.

L'ukrainien tire ses origines des dialectes proto-ukrainiens qui ont remplacé le slave commun¹⁴ au VI^e siècle. Selon la périodisation proposée par Michael Moser dans *l'Encyclopédie de l'histoire* (Encyclopédie 2019: *Українська мова*), les **dialectes proto-ukrainiens** se développèrent de 600 jusqu'en 988, année de la christianisation massive de la Rous' kyïvienne, puis évoluèrent vers **l'ancien ukrainien**. L'évolution de l'ancien ukrainien (langue écrite) et des anciens dialectes ukrainiens vers l'ukrainien moderne dura sept siècles. La **période archaïque** dura de 988 jusqu'à 1349, date à laquelle la principauté de Galicie-Volhynie, héritière de la Rous' kyïvienne, fut annexée

¹³ Le haut et le bas sorabe sont des langues slaves minoritaires parlées en Allemagne dans les régions de Lusace et de Brandebourg.

¹⁴ Le slave commun est appelé aussi proto-slave.

au royaume de Pologne¹⁵. **La période médiane** débuta au milieu du XIV^e siècle et s'acheva à la fin du XVIII^e siècle, en 1798, année de la parution du premier texte en **ukrainien moderne**.

Mais les normes orthographiques et grammaticales de la langue ukrainienne moderne ont subi plusieurs changements depuis le début du XIX^e siècle. De plus, pendant la période soviétique, une partie du vocabulaire a été exclue de l'usage, car jugée « nationaliste » et « contre-révolutionnaire ». Cependant, après l'indépendance de l'Ukraine, ce vocabulaire a été en partie réintégré, faisant apparaître des variantes. Afin de mieux comprendre ces phénomènes, il serait intéressant de faire une parenthèse sur chacune des étapes d'évolution de l'ukrainien.

1.3 Étapes de l'évolution de l'ukrainien

1.3.1 Du slave commun aux dialectes proto-ukrainiens

Au début de notre ère, sur les territoires qui s'étendaient des Carpates au Dniro central, habitaient des peuples qui parlaient le slave commun issu de la branche « satem » de l'indo-européen commun. Cette période du slave commun dura environ 2 000 ans et s'acheva approximativement vers l'an 600, laissant place à de nombreux dialectes (Ohienko 1949 [46]). Dans son *Histoire de la langue ukrainienne*, Ohienko démontre que des dialectes bien distincts sont apparus dès le VI^e siècle et ont donné naissance à trois langues slaves : l'ukrainien, le biélorussien et le russe. Ces dialectes proto-ukrainiens sont mentionnés dans les récits des géographes Ibn Khordadbeh (844-848) et Ibn Rustah (903), du chroniqueur Ibn Fadlan (922), de l'historien Maçoudi (mort en 956), de l'écrivain Ibn al-Nadim (mort en 990), ou encore dans *De administrando imperio* de l'empereur byzantin Constantin VII Porphyrogénète (Pivtorak 1993 [50], Nimtchouk 1985 [42]). Ces œuvres évoquent les noms des tribus, des princes, des toponymes, mais aussi les noms des objets et des aliments, ce qui permet d'étudier les particularités des langues de l'époque.

La théorie selon laquelle la langue ukrainienne est issue des dialectes proto-ukrainiens est décrite en détail dans les travaux d'Ahatanhel Kryms'kyy, Stepan Smal'-Stots'kyy, Kost' Nimtchynov, Ivan Ohienko, Hryhoriy Pivtorak, Oleksa Horbatch, Iuriy Cheveliov, Vitaliy Roussaniv's'kyy, Vassyl' Nimtchouk¹⁶.

Notons que pendant la période soviétique prédominait la version selon laquelle la langue ukrainienne ne descendait pas directement du slave commun, mais qu'il y avait eu d'abord le slave oriental commun, descendant de l'ancien slave commun et ancêtre direct du vieux-russe et de l'ancien ukrainien. Selon cette théorie, le slave oriental commun se forma vers le VI^e siècle et continua son existence vers le IX^e siècle, puis fut remplacé entre les IX^e et XIII^e siècles par le vieux-russe, l'ancêtre de l'ukrainien, du biélorussien et du russe. Cette théorie est décrite par le linguiste russe Aleksey Chakhmatov (1916) [63], puis reprise par tous les linguistes soviétiques, comme Boytchouk (Bezpal'ko *et alii* 1962, pp. 17-46 [3]) ou Jovtobrioukh (Jovtobrioukh *et alii* 1980, pp. 5-6 [30]), Boulakhov's'kyy (Boulakhov's'kyy 1956 [6], pp. 7-11). Selon cette hypothèse, chaque tribu slave de l'Est avait un dialecte peu différent des dialectes de ses voisins, puis lorsque l'État de la Rous'

¹⁵ Après la montée en puissance de la principauté de Moscou au début du XI^e siècle et les invasions tatares, Kyïv perd de son importance et le centre politique se déplace vers l'ouest où le royaume de Galicie-Volhynie se forme.

¹⁶ (Kryms'kyy 1922 [34]), (Smal'-Stots'kyy 1923 [58], 1933 [59]), (Ohienko 1949 [46]), (Pivtorak 1993 [50], 2001 [51]), (Cheveliov 1994 [66], 2002 [67]), (Roussaniv's'kyy 2001 [56]), (Nimtchouk 2015 [44]), (Horbatch 1993 [20]).

kyïvienne est né, les dialectes se sont rapprochés. Les chercheurs ukrainiens rejettent actuellement cette théorie, car elle néglige artificiellement toutes les preuves du contraire et se concentre principalement sur les phénomènes liés aux trois langues en survolant rapidement leurs différences.

1.3.2 La période archaïque de l'ancien ukrainien

Entre 988 et 1349, sur les territoires de la Rous' kyïvienne se développèrent deux langues écrites : le slavon d'Église (la langue des textes ecclésiastiques, liturgiques et des Évangiles) et l'ancien ukrainien (la langue administrative, juridique et littéraire). Le slavon d'Église utilisa d'abord l'alphabet glagolitique¹⁷, puis très rapidement l'alphabet cyrillique¹⁸, tandis que l'ancien ukrainien utilisait uniquement l'alphabet cyrillique. Ces deux langues furent parfois mélangées dans le même texte, notamment dans les récits historiques, appelés « chroniques ». Les langues écrites se distinguaient des dialectes par leur structure et leur vocabulaire, mais elles retraçaient parfaitement les particularités de ces derniers, car elles en empruntaient de nombreux éléments. De ce fait, les chercheurs considèrent que même le slavon d'Église ukrainien se différenciait du slavon d'Église des autres territoires.

Les textes les plus connus de cette époque sont :

- *Ізборник Святослава* [*le Recueil de Sviatoslav*] de 1073 ;
- *Ізборник* [*le Recueil*] de 1076 ;
- *Руська правда* [*la Vérité roussienne*] du XI^e siècle ;
- *Повчання Володимира Мономаха* [*l'Enseignement de Volodymyr Monomakh*] fin du XI^e - début du XII^e siècle ;
- *Слово о полку Ігоревім* [*le Dit de la campagne d'Igor*] du XII^e siècle ;
- *Повість врем'яних літ* [*le Récit des temps passés*] du XII^e siècle ;
- *Галицько-Волинський літопис* [*les Chroniques de Galicie et Volhynie*] du XIII^e siècle ;
- *Галицьке Євангеліє* [*l'Évangile de Galicie*] de 1144 ;
- *Луцьке Євангеліє* [*l'Évangile de Loutsk*] du XIV^e siècle.

L'ancien ukrainien se distinguait déjà des autres langues slaves par son système phonétique et grammatical, son lexique et son orthographe. Les textes ont été analysés par plusieurs chercheurs¹⁹ qui ont révélé de nombreuses caractéristiques propres à la langue ukrainienne moderne :

- la lettre *z* pour désigner le phonème fricatif glottal voisé *h* ;
- le dédoublement de *o* et *e* étymologiques après le déclin de *ъ* et *ь* faibles, ce qui entraînera l'alternance des voyelles *i/e* et *i/o* dans les syllabes fermées/ouvertes ;
- la lettre *o* initiale à la place des *a* ou *ε*, notamment dans les noms propres ;
- la mouillure de la lettre *ц* ;
- l'alternance de *в* et *у/оу* au début des mots ;
- la mouillure du suffixe adjectival *-ськ-* ;
- la terminaison *-мо* au présent, à la première personne du pluriel du verbe *бути* [*être*] → *єсмо*, qui s'est ensuite étendue à tous les verbes ;
- la terminaison *-у* pour les noms masculins au génitif ;
- la terminaison *-ові, -еві, -єві* au datif ;

¹⁷ L'alphabet glagolitique fut inventé par Cyrille et Méthode en 862-863 pour traduire la Bible en vieux-slave.

¹⁸ L'alphabet cyrillique fut développé par les élèves de Cyrille et Méthode à partir de l'alphabet grec, dans lequel on ajouta des lettres glagolitiques pour retranscrire les sons absents.

¹⁹ (Bevzenko 1960 [2]), (Nimtchouk 1985 [42]), (Pivtorak 1993 [50], 2001 [51]), (Cheveliov 1994 [66], 2002 [67]), (Roussaniv's'kyi 2001 [56]), (Pavlenko 2010 [49]), (Brous 2016 [5]), (Holoïoukh 2021 [15]).

- le vocatif ;
- la terminaison *-mь* au présent pour les verbes à la troisième personne du singulier ou du pluriel indicatif qui formera plus tard la deuxième conjugaison ;
- la disparition sporadique de *-mь* à la troisième personne du singulier au présent de l'indicatif de certains verbes qui formeront plus tard la première conjugaison ;
- la position de la particule *-ся* avant le verbe ;
- le vocabulaire proprement ukrainien qui n'existait pas dans d'autres langues slaves, y compris le biélorussien ou le russe (mois, objets du quotidien, noms des espaces, étendues naturelles).

1.3.3 La période médiane de l'ancien ukrainien

Les linguistes ukrainiens situent la période médiane entre 1349 et 1798. Pendant cette longue période, les territoires ukrainiens furent répartis entre des puissances étrangères. Notons quelques étapes historiques qui nous permettront de comprendre l'évolution de la langue durant cette période.

Le contexte historique

Au XIV^e siècle, les terres ukrainiennes furent partagées entre le Grand-Duché de Lituanie (les régions de Kyïv et de Tchernihiv, la Volhynie et le nord-est de la Podolie), la principauté de Moldavie (Bucovine), le royaume de Hongrie (Transcarpatie) et le royaume de Pologne (Galicie). La Galicie d'ailleurs resta sous la domination polonaise jusqu'en 1772, date à laquelle elle fut rattachée à l'Autriche lors du premier partage de la Pologne (Holoïoukh 2021 [15]).

En 1569, après la signature de l'union de Lublin entre la Lituanie et la Pologne, les terres ukrainiennes furent à nouveau redistribuées : la Volhynie, la Podolie et la région de Kyïv furent rattachées à la Pologne.

En 1654, après le traité de Pereyaslav, une partie des terres ukrainiennes, et notamment les régions de Tchernihiv, de Poltava, de Soumy, de Kyïv et de Tcherkassy, entrèrent sous la protection de la Moscovie²⁰. Une autre partie des terres ukrainiennes, qui s'étendait autour du Dniro du nord au sud, reçut le statut d'État indépendant sous le nom de Hetmanat, mais cet État perdit son indépendance en 1764. Toutes les régions d'Ukraine sous la domination de l'Empire russe reçurent le nom de « Petite Russie ».

Quant à la Podolie, elle devint indépendante quelques décennies après la révolte de Petro Dorochenko en 1667, avant d'être attachée à la Pologne-Lituanie en 1699, puis à l'Empire russe en 1793, lors du deuxième partage de la Pologne. La Volhynie, elle aussi, fut rattachée à la Russie deux ans plus tard lors du deuxième partage de la Pologne.

La situation linguistique, les ouvrages et l'orthographe

À la fin de cette période, les territoires de l'Ukraine actuelle étaient partagés en deux parties : rive gauche du Dniro sous la domination de l'Empire russe et rive droite sous la domination des royaumes occidentaux.

Lors des partages des terres ukrainiennes, l'ancien ukrainien et le slavon d'Ukraine subirent, sous l'influence des langues dominantes, des modifications disparates.

²⁰ Au XV^e siècle, l'appellation « *Россія* » /*Rossia*/, c'est-à-dire « la Russie », est apparue pour désigner la Moscovie ; ce nom est devenu officiel seulement sous Pierre le Grand.

Ainsi, dans le Grand-Duché de Lituanie et sur les terres sous la domination de la principauté de Moldavie, l'ancien ukrainien, ou, comme on l'appelait à l'époque, le roussien, devint la langue officielle. Il se développa librement et s'enrichit de son propre vocabulaire administratif et domestique (Ohienko 1949 [46]).

Dans le royaume de Pologne, la langue officielle fut d'abord le latin. En parallèle, dans les régions nouvellement rattachées jusqu'au XVI^e siècle, l'ancien ukrainien continua d'être en usage. De plus, l'invention de l'imprimerie au XV^e siècle donna un élan considérable au développement de la culture ukrainienne, car elle permit d'avoir des manuels scolaires en ukrainien et, de ce fait, d'instaurer l'éducation en ukrainien. Les premiers livres en slavon pour les Ukrainiens et les Bélarussiens furent publiés en 1491 dans l'imprimerie cracovienne de Schweipolt Fiol. En 1574, à Lviv, Ivan Fedorov imprima un livre pour l'école, *Буквар [l'Abécédaire]*, composé de l'alphabet et de textes de lecture. Ce manuel est considéré comme le premier livre imprimé en ukrainien (Nimtchouk 1985 [42]).

Au milieu du XVI^e siècle, le polonais devint la langue officielle, la noblesse ukrainienne de ces régions se convertit au catholicisme et adopta le polonais, seul le peuple continua de parler sa langue. Pendant cette période, l'ukrainien s'imprégna de latinismes, puis de polonismes.

En 1596, après l'union de Brest²¹, sur les territoires sous domination polonaise, la langue ukrainienne devint le symbole de la rébellion contre la polonisation et la religion catholique. L'esprit national se ressentit encore plus après le traité d'Androusovo²² et, de ce fait, le gouvernement des Deux Nations décida de limiter l'utilisation de l'ukrainien. Ainsi, en 1696, le Parlement de Pologne-Lituanie²³ promulgua un décret qui ordonna le remplacement de l'ukrainien par le polonais dans les tribunaux et les institutions administratives (Boulyk-Verkhola 2018 [7]).

En Transcarpatie, où la langue administrative était le latin, l'ukrainien se développa en parallèle jusqu'à l'union d'Oujhorod²⁴, signée en 1646. À partir de cette date, l'ukrainien fut progressivement éliminé de toutes les sphères de la vie publique et religieuse (Encyclopédie d'histoire 2019 : *Українська мова*).

Sur les territoires contrôlés par la Moscovie, la langue dominante était le russe. À partir du XVII^e siècle, comme ailleurs, l'ukrainien fut banni de l'enseignement et de toutes les instances publiques et religieuses. Le premier décret restrictif contre la langue ukrainienne fut promulgué en 1622 par le tsar Michel Romanov sur proposition de son père Philarète, patriarche de Moscou, exigeant de brûler tous les exemplaires de *Євангеліє Ставровецького [l'Évangile de Stavrovets'kyy]* sortis en Ukraine. En 1690, le Conseil patriarcal de l'Église orthodoxe de Moscou prononça un anathème contre les livres des hommes d'Église et des écrivains ukrainiens Petro Mohyla, Kyrylo Stavrovets'kyy, Lazar Baranovytsch, Antoniyy Radyvylovs'kyy, Epyphaniyy Slavynets'kyy, Ioanykiy

²¹ L'union de Brest (1596) – unification d'une partie de la métropole du patriarcat œcuménique de Kyïv avec l'Église catholique romaine, reconnaissance de la suprématie du pape et acceptation de la doctrine catholique. Suite à cette unification naquit l'Église gréco-catholique d'Ukraine.

²² Le traité d'Androusovo (1667) – une trêve qui mit la fin la guerre entre la Russie et la République des Deux Nations (Pologne-Lituanie), entraînant le partage des territoires ukrainiens.

²³ En polonais : Sejm walny I Rzeczypospolitej Rzeczypospolitej Obojga Narodów : Korony Królestwa Polskiego i Wielkiego Księstwa Litewskiego.

²⁴ L'union d'Oujhorod (1646) fut signée par soixante-trois prêtres orthodoxes de Transcarpatie qui décidèrent de se placer sous le protectorat de l'Église catholique. Suite à cette signature naquit le diocèse gréco-catholique de Moukatchevo.

Galiatovs'kyy. En 1729, Pierre II ordonna de réécrire en russe tous les textes administratifs ukrainiens. En 1763, Catherine II interdit l'enseignement en ukrainien à l'académie Mohyla de Kyïv. En 1769, le Conseil patriarcal de l'Église orthodoxe de Moscou interdit d'imprimer des manuels scolaires en ukrainien et ordonna de confisquer les manuels existants. En 1775, la Sitch zaporogue²⁵ fut dissoute et les écoles ukrainiennes sur les territoires cosaques furent fermées (Boulyk-Verkhola 2018 [7]).

Les chercheurs²⁶ qui ont travaillé sur cette période soulignent que la langue ukrainienne est parfaitement représentée dans les textes administratifs, ecclésiastiques ou littéraires suivants :

- les textes administratifs du Grand-Duché de Lituanie du XIV^e au début du XVI^e : actes de donation et d'achat, testaments, plaintes de citoyens ;
- *Судебник [le Code judiciaire]* (1468) de Casimir IV Jagellon, grand-duc de Lituanie et roi de Pologne ;
- *Литовський Статут [le Statut de Lituanie]* de 1529 et de 1566 ;
- *Пересопницьке Євангеліє [l'Évangile de Peresopnytsia]*²⁷ (1556-1561, Volhynie) de Mykhaylo Vassylevytch et du moine Hryhoriy ;
- les œuvres polémiques d'Ivan Vychens'kyy, notamment *Книжка [le Livre]* de 1600 et ses nombreuses épîtres écrites au début du XVII^e siècle ;
- l'œuvre polémique *Палінодія, або Книга оборони [Palinodia, ou le livre de la défense]* (1622) de Zakhariia Kopystens'kyy ;
- les ballades folkloriques telles que *Самійло Кішка [Samiylo Kichka]*, *Маруся Богуславка [Maroussia Bohouslavka]*, *Хмельницький і Барабаш [Khmelnys'kyy et Barabach]* (fin du XVI^e-milieu du XVII^e siècle) ;
- la poésie²⁸ de Herassym et Meletiy Smotryts'kyy, Kyrylo Tranquillion Stavrovets'kyy, Ielysey Pletenets'kyy, Kassian Sakovytych, Tarasiy Zemka et Ian Joravnyts'kyy ;
- les œuvres dramaturgiques²⁹ de Pamva Berinda, Kassian Sakovytych, Iakoub Gavatovytych ;
- les œuvres de Hryhoriy Skovoroda.

Dans ces œuvres, en plus de la fixation des phénomènes relevés pendant la période précédente, les chercheurs³⁰ listent des caractéristiques supplémentaires, parmi lesquelles :

- l'alternance des lettres *в* et *у* en début de mot ;
- l'apparition de la lettre *и* à la place de *ѣ* ;
- l'apparition de la lettre *и* à la place de *е* ;
- l'apparition de la lettre *о* après les chuintantes ;
- l'ajout de la lettre *в* en début de mot, notamment avant le *о* ;

²⁵ La Sitch zaporogue (en ukr. *Запорізька січ*), – centre militaire et administratif des Cosaques, armée libre ukrainienne, dans le sud de l'Ukraine actuelle, fut fondée au milieu du XVI^e siècle. Elle fut à l'origine de la création de l'Hetmanat en 1649.

²⁶ (Bevenko 1960 [2]), (Nimtchouk 1985 [42]), (Pivtorak 1993 [50], 2001 [51]), (Cheveliov 1994 [66], 2002 [67]), (Roussaniv's'kyy 2001 [56]), (Pavlenko 2010 [49]), (Brous 2016 [5]), (Holoïoukh 2021 [15]).

²⁷ L'Évangile de Peresopnytsia est l'une des premières traductions d'un texte biblique dans la langue littéraire. Lors de leur investiture, les présidents ukrainiens prêtent serment à la nation non seulement sur la Constitution, mais aussi sur l'Évangile de Peresopnytsia.

²⁸ Les premières œuvres poétiques ukrainiennes sont apparues à la fin du XVII^e.

²⁹ Les premières œuvres dramaturgiques ukrainiennes sont apparues également à la fin du XVII^e.

³⁰ (Hrytsenko 2017 [17,18]), (Bevenko 1960 [2]), (Cheveliov 2002 [67]), (Nimtchouk 1985 [42]), (Nimtchouk et Pouraïeva 2004 [43]), (Roussaniv's'kyy 2001 [56]), (Holoïoukh 2021 [15]) et autres.

- le remplacement de *л* par *љ*, puis par *в* à la fin des verbes au passé masculin et dans la position entre le *о* et une autre consonne ;
- le dédoublement de la consonne molle devant la voyelle finale ;
- le durcissement de certaines consonnes devant les voyelles (écriture de *ы* au lieu de *и*) ;
- le durcissement des labiales à la fin des mots ;
- le phonème /g/ est retranscrit par deux consonnes *кз* dans les mots empruntés ;
- les alternances *з/з, к/ц, х/с* ;
- l'apparition des suffixes *-цьк-, -зьк-* ;
- les constructions avec le verbe *имати [avoir]* + infinitif dans le sens de « devoir faire quelque chose » ;
- l'utilisation du plus-que-parfait (le verbe *бути [être]* au passé, suivi d'un autre verbe au passé)
- l'apparition du futur synthétique pour les verbes imperfectifs par ajout de la terminaison à l'infinitif ;
- l'apparition des suffixes dérivatifs *-щина, -ина* et du suffixe *-ко* pour les prénoms diminutifs masculins ;
- la terminaison *-ею* à l'instrumental féminin après les chuintantes et les consonnes molles ;
- l'utilisation de la particule *нехай [que]* à l'impératif de la troisième personne du singulier et du pluriel ;
- l'utilisation très répandue de la terminaison *-мо* à la première personne du pluriel au présent de l'indicatif ;
- contrairement à la syntaxe de la langue russe de l'époque, l'hypotaxe l'emporte nettement sur la parataxe ;
- le vocabulaire proprement ukrainien ;
- les polonismes, les latinismes, les germanismes.

Ajoutons qu'à la fin de cette période deux orthographes existaient : l'orthographe du slavon, normalisée par Meletiy Smotryts'kyj en 1619 dans sa grammaire *Грамматики славѣнскиѧ правилое Сѣнтагма*³¹ [*le Syntagme correct de la grammaire du slavon*], et l'orthographe de la langue littéraire et normative « *grajdanka* »³², introduite par décret de Pierre le Grand en 1708. Ce décret autorisait l'utilisation de l'ancienne orthographe uniquement pour les textes ecclésiastiques. « *Grajdanka* » avait pour but de rapprocher la langue littéraire du russe, elle omettait certaines lettres du cyrillique, considérées comme inutiles pour le russe, mais importantes pour l'ukrainien et le biélorussien, et de ce fait servait à rapprocher toutes les langues qui utilisaient l'alphabet cyrillique (Ohienko 1949 [46]). Ainsi, la lettre *ѣ* et les accents diacritiques, qui étaient adaptés à l'ukrainien, ont disparu.

1.3.4 L'ukrainien au XIX^e siècle (1798-1900)

La période moderne a commencé en 1798, l'année de la publication de *Енеїда [l'Énéide]* d'Ivan Kotl'arevs'kyj³³, qui a été le premier à écrire une grande œuvre littéraire en langue populaire. Le vocabulaire de « l'Énéide » est composé de plus de 7 000 mots et 400 locutions proprement ukrainiens, répartis en 70 groupes thématiques (Holoïoukh 2021 [15]).

³¹ Cette grammaire a été publiée dans la ville de Jewie, aujourd'hui Vievis en Lituanie.

³² En russe : *гражданка*.

³³ L'œuvre a été publiée pour la première fois à Saint-Petersbourg sans autorisation de l'auteur sous le nom *Енеїда. На малоросійській языкъ перелиціюванна І. Котляревскимъ [L'Énéide. Réécrite en petit-russe par I. Kotl'arevs'kyj]*.

Le contexte historique et linguistique

Au XIX^e siècle, les territoires ukrainiens ont été partagés entre l'Empire russe (rive gauche) et l'Empire austro-hongrois (rive droite).

En Ukraine de la rive gauche, l'enseignement secondaire et supérieur se faisait en russe. En 1740-1748, il y avait 786 écoles secondaires sur la rive gauche, et au début du XIX^e siècle, il n'y en avait plus aucune (Holoïoukh 2021 [15]). La langue ukrainienne était enseignée uniquement dans les écoles primaires des diacres. En 1817, l'Académie Mohyla de Kyïv, qui était le centre de l'éducation ukrainienne, a été fermée, de même que les écoles du dimanche pour adultes en 1862.

En 1863, Piotr Valouïev, le ministre des Affaires étrangères de l'Empire russe, a affirmé que la langue ukrainienne n'existait pas comme une langue à part entière, et, par une circulaire, il a interdit de publier en ukrainien les manuels scolaires, la littérature populaire et les livres à contenu religieux. En 1876, Alexandre II a signé le décret d'Ems qui a réitéré l'interdiction de publication, à l'exclusion des livres du type « belles-lettres » et des documents historiques, et l'a élargie aux représentations théâtrales et aux lectures publiques en langue ukrainienne, ainsi qu'à l'importation de livres en ukrainien publiés à l'étranger et à la conservation des livres ukrainiens dans les bibliothèques scolaires. En outre, selon ce décret les instituteurs devaient être contrôlés et toute personne soupçonnée d'être ukrainophile devait être transférée pour travailler en dehors de l'Ukraine. Ce décret a été modifié en 1881 par Alexandre III, qui a autorisé l'impression des dictionnaires en ukrainien et a levé l'interdiction des représentations théâtrales, à condition que les troupes ne fussent pas exclusivement ukrainiennes (Cheveliov 1987 [65]).

Le statut de la langue ukrainienne a été dégradé, elle a été qualifiée de langue des paysans et des domestiques asservis, impropre à une littérature respectable (Ohiienko 1949 [46]). Tous ceux qui parlaient ukrainien étaient considérés comme non éduqués ou opposants à l'Empire russe. Les écrivains, les historiens, les publicistes et les hommes politiques étaient persécutés, mis en prison, envoyés au bagne.

Holoïoukh cite Mykola Kostomarov, éminent historien, ethnographe et écrivain ukrainien :

« L'Ukraine était silencieuse. Et son peuple était méprisé. Le nom de « khokhol », que les Moscovites donnaient aux Cosaques pour leur mèche, est devenu synonyme d'imbécile. La langue poétique de l'Ukraine est devenue un sujet de mépris et de moquerie. Souvent, les Ukrainiens eux-mêmes rougissaient lorsque leur prononciation révélait leur origine méridionale. L'histoire de l'Ukraine a été soit oubliée, soit déformée, conformément aux conformément aux « bonnes intentions » et aux opinions du gouvernement» (Holoïoukh 2021, p. 53 [15])³⁴.

En Ukraine de la rive droite, sous l'Empire austro-hongrois, l'enseignement était assuré en allemand. À partir de 1815, le polonais est devenu la deuxième langue d'enseignement et le tchèque la troisième. En 1818, 383 écoles ukrainiennes ont été fondées. Les Constitutions de l'État autrichien de 1849 et de 1867 garantissaient l'égalité de toutes les langues parlées dans les écoles, les organes gouvernementaux et la vie publique. En 1867, l'ukrainien devint l'une des langues officielles du

³⁴ Je traduis : « Україна замовкла. І її народність була віддана зневазі. Ім'я хохол, яке москалі давали козакам від їх оселедців, стало синонімом дурня. Поетична мова України стала предметом зневаги і глумління. Нерідко самі українці червоніли, коли вимова їх виявляла південне походження. Українська історія була або забута, або представлена у спотвореному вигляді відповідно до благих цілей і поглядів уряду. »

Parlement d'Autriche-Hongrie et la deuxième langue du Conseil de Galicie (Boulyk-Verkhola et Verkhola 2018 [7]).

Bien qu'en Galicie, la langue ukrainienne n'ait pas été interdite, il n'existait pas d'écoles exclusives en ukrainien, car le polonais était enseigné au même niveau que l'ukrainien (Holoïoukh 2021 [15]).

Au début du XIX^e siècle, sur tous les territoires de l'Ukraine actuelle, se sont développées en parallèle plusieurs variantes littéraires : l'ukrainien du Dnipro central³⁵, le galicien³⁶, le bucovinien³⁷, le roussyn³⁸. Hrytsenko (1993) [16] distingue deux branches du roussyn : le roussyn transcarpatien³⁹ et le roussyn des régions de Batchka et de Syrmie où a émigré la population de la Transcarpatie ukrainienne. La variante du Dnipro central évoluait très rapidement, car les dialectes méridionaux, mais aussi les dialectes du Nord s'y interpénétraient, aussi serait-il plus judicieux d'appeler cette variante l'ukrainien de la rive gauche.

Chacune de ces variantes englobait plusieurs dialectes et était influencée par une ou parfois plusieurs langues dominantes : dans les régions du Centre et l'Est, c'était le russe, en Galicie le polonais et l'allemand, en Bucovine le roumain et l'allemand, en Transcarpatie le hongrois et l'allemand. La standardisation de la langue écrite se faisait grâce aux textes publiés qui étaient perçus comme des modèles linguistiques (Ostaptchouk 2013 [48]).

Ainsi, au début du XIX^e siècle, les œuvres d'Ivan Kotl'arevs'kyy, Hryhoriy Kvitka-Osnovianenko, Ievhen Hrebinka, Petro Houlak-Artemovs'kyy, Amvroziy Metlyns'kyy sont devenues des modèles pour **l'ukrainien de la rive gauche**. La deuxième moitié du XIX^e siècle a été marquée par les œuvres de Panteleymon Koulich, Marko Vovtchok, Panas Myrnyy, Lessia Oukraïnka, Mykhaylo Kotsioubyns'kyy, Ivan Netchouy-Levyts'kyy, Ivan Karpenko-Karyy, Marko Kropyvnytskyy. Taras Chevtchenko, qui n'était pas seulement un écrivain, mais un véritable prophète de la nation ukrainienne, est à l'origine d'une grande avancée pour la langue ukrainienne et pour le renouveau national. Ivan Ohienko (Ohienko 1961, p. 23 [47]) a écrit :

« La langue de Chevtchenko est devenue la pierre angulaire de notre langue littéraire. Cela s'est produit parce que Chevtchenko lui-même était le poète le plus puissant de notre pays et que ses œuvres ont captivé toute l'Ukraine : tout le monde les a lues, les a apprises par cœur et a ainsi appris la langue. »⁴⁰

³⁵ *Середньоподніпровський*, basé principalement sur les dialectes des régions du Dnipro central (de Kyïv, Tcherkassy, Dnipro, Poltava) et de Sloboda (Kharkiv, Soumy, Donetsk, Louhansk).

³⁶ *Галицький*, basé sur les dialectes parlés dans les régions d'Ivano-Frankivsk, Lviv, une partie de la région de Ternopil et une partie de la région de Cracovie, qui se trouve en Pologne actuelle, avec les ajouts du slavon d'église, du russe, du polonais.

³⁷ *Буковинський*, basé principalement sur le dialecte houtsoul, couvrait la région actuelle de Tchernivtsi en Ukraine et de Suceava en Roumanie.

³⁸ *Русинський*, basé sur les dialectes de Transcarpatie ukrainienne et d'une partie de la région de Prešov en Slovaquie actuelle, mais aussi sur les dialectes des régions de Batchka (Bačka) et de Syrmie (Srem) en Serbie et Croatie actuelles.

³⁹ *Закарпатський*, basé sur les dialectes des Carpates ukrainiennes et d'une partie de la région de Prešov en Slovaquie actuelle.

⁴⁰ Je traduis : « Шевченкова мова стала в нас за наріжний камінь нашої літературної мови. Сталося це, тому що сам Шевченко був у нас найсильнішим поетом, і його твори захопили всю Україну : їх усі читали, завчали напам'ять, і вже тим самим навчалися й мови. »

Le galicien a été marqué par Markyian Chachkevych, Iakiv Holovats'kyi, Ivan Vahylevych, les représentants du courant *Руська трійця* [*La Triade roussienne*] et les rédacteurs de l'édition *Русалка Дністрівська* [*La Sirène du Dniestr*], puis par Ivan Franko, Mykhaylo Pavlyk, Vassyl' Stefanyk.

La variante bucovinienne est représentée dans les textes de Iuriy Fedkovych, Ol'ha Kobylianska et Ievheniia Iarochynska et **le roussyn** – dans les œuvres d'Oleksandr Doukhnovych, Anatoliy Kralyts'kyi, Ievhen Fentsyk, Vassyl' Tchopoi, Avgoustyn Volochyn, Oleksandr Pavlovych.

Les auteurs utilisaient le vocabulaire et les formes grammaticales propres à leur région et tentaient d'adopter l'orthographe pour correspondre au mieux au système phonétique de la langue parlée. Mais ces variantes n'étaient pas confinées dans leurs régions : les intellectuels ukrainiens se réunissaient et leurs cercles avaient de forts liens entre eux. Ainsi, en Galicie et en Ukraine de la rive gauche, est apparu le mouvement de renaissance nationale « Boudyteli »⁴¹, qui signifie « ceux qui réveillent ». Et c'est à l'intérieur de ce mouvement que l'idée d'une langue unique pour tous les Ukrainiens est née.

En Galicie, Bucovine et Transcarpatie, deux courants linguistiques supplémentaires sont apparus. Le premier regroupait les moscophiles qui défendaient l'unité nationale, culturelle et linguistique avec la Russie, tandis que le second regroupait les populistes qui défendaient l'unité de toutes les terres ukrainiennes et le développement d'une langue ukrainienne unique basée sur les dialectes régionaux.

Les moscophiles ont initié une langue littéraire artificielle, appelée « iazytchie »⁴². Elle a été créée en combinant les dialectes, la langue ecclésiastique et le russe. Cette langue n'avait pas de règles clairement établies et était incompréhensible pour les paysans.

Les orthographes

Dans ce contexte, la normalisation de la langue et de son orthographe n'était pas évidente, et c'est la raison pour laquelle de nombreux essais de standardisation de l'orthographe ont vu le jour.

En 1818, dans sa grammaire *Граматика малоросійського нарѣчя* [*La Grammaire du dialecte petit-russe*], publiée à Saint-Pétersbourg, Oleksandr Pavlovs'kyi a initié la première orthographe phonétique. Elle a été utilisée d'abord dans la région de Kharkiv, puis dans les autres régions du Nord et du Centre.

En 1827, dans la préface du recueil *Малоросійськіє песни* [*les Chansons petit-russes*], publié à Moscou, Mykhaylo Maksymovych a initié l'orthographe étymologique, connue aussi sous le nom de « maksymovytchivka »⁴³. Elle n'était pas très populaire en Ukraine de la rive gauche, mais a trouvé beaucoup de partisans en Bucovine, Galicie et Transcarpatie.

En 1834, Iossyp Lozyns'kyi a publié à Lviv, dans le journal *Gazeta Lwowska*, l'article en polonais *O wprowadzeniu abecadła polskiego do piśmiennictwa ruskiego* [*Sur l'introduction de l'alphabet polonais dans l'écriture roussienne*], dans lequel il proposait d'adopter l'alphabet polonais pour l'ukrainien. Cela a provoqué de fortes réactions en Galicie et finalement le projet de « latynka »⁴⁴ a

⁴¹ En ukr. : *Будителі*.

⁴² En ukr. : *язичіє*.

⁴³ En ukr. : *максимовичівка*.

⁴⁴ En ukr. : *латинка*.

été abandonné. En 1859, Aguenor Golouchows'kyy a tenté à nouveau d'introduire l'alphabet latin pour l'ukrainien, mais cette tentative a également échoué.

En 1837, à Budapest, dans le recueil *Русалка Днѣстровая [La Sirène du Dniestr]*, Markian Chachkevych a initié l'orthographe appelée « chachkevychivka »⁴⁵. Elle a trouvé peu de partisans en Galicie, mais a été largement utilisée en Ukraine de la rive gauche.

En 1856, Panteleïmon Koulich a initié dans le recueil *Записки о Южной Руси [Notes sur la Rous' méridionale]*, publié à Saint-Pétersbourg, l'orthographe phonétique appelée « koulichivka »⁴⁶. Puis, dans sa *Граматика [Grammaire]* (Saint-Pétersbourg, 1857), il donna une description détaillée des règles orthographiques et grammaticales. Cette orthographe était très répandue en Ukraine de la rive gauche.

En 1857, dans le recueil *Ужинок рідного поля [la Récolte de la terre natale]*, publié à Moscou, Mykola Hatsouk introduisit l'orthographe appelée « hatsoukivka »⁴⁷. Il la reprit dans l'abécédaire *Українська абетка [l'Alphabet ukrainien]* (Moscou, 1861). Cette orthographe fut promue dans les territoires ukrainiens sous domination russe, mais à cause de sa complexité elle ne fut pas populaire.

En 1861, dans son *Букварь южнорусскій [Abécédaire du roussien méridional]*, publié à Saint-Pétersbourg, Chevtchenko normalisa « l'orthographe phonétique de Sloboda »⁴⁸, qui fut utilisée sur les territoires de la rive gauche.

En 1873, sur les territoires ukrainiens de la rive gauche, se répandit une nouvelle orthographe phonétique, introduite dans le premier volume du recueil *Записки Юго-Западного отдела Императорского Русского географического общества [Notes de la Division du Sud-Ouest de la Société géographique impériale]*, publié à Kyïv. Parmi les linguistes qui lancèrent cette orthographe figuraient Pavlo Jytets'kyy, Volodymyr Antonovych, Mykhaylo Drahomanov, Konstantyn Mykhaltchouk, et d'autres.

En 1877-1878, dans le recueil *Громада [La Communauté]*, publié à Genève, on utilisa l'orthographe qu'on connaît sous le nom de « drahomnivka »⁴⁹ ou « hertsegovynka »⁵⁰, introduite par Mykhaylo Drahomanov. Cette orthographe ne fut pas très populaire, son utilisation se limita à quelques publications à Genève et en Galicie.

En 1886, Ievhen Jelexhiv'skyy introduisit dans *Малорусско-німецький словар [le Dictionnaire petit-russe – allemand]*, publié à Lviv, l'orthographe appelée « jelexhivka »⁵¹. Cette orthographe fut utilisée dans l'Ukraine de la rive droite.

En 1893, dans leur grammaire *Руска граматика [la Grammaire roussienne]*, publiée à Lviv, Stepan Smal'-Stots'kyy et Fedir Gartner normalisèrent l'orthographe phonétique galicienne. Ils approfondirent le travail entrepris par Jelexhiv'skyy. Cette orthographe devint officielle dans les écoles en Bucovine et Galicie sous domination austro-hongroise et trouva également ses partisans dans l'Ukraine de la rive gauche. Elle fut rééditée par la suite trois fois, en 1907, 1922 et 1927.

⁴⁵ En ukr. : шашкевичівка.

⁴⁶ En ukr. : кулішівка.

⁴⁷ En ukr. : гацуківка.

⁴⁸ En ukr. : Слобожанський фонетичний правопис.

⁴⁹ En ukr. : драгоманівка.

⁵⁰ En ukr. : герцеговинка ; cette orthographe a reçu cette appellation car elle se rapprochait de l'orthographe serbe.

⁵¹ En ukr. : желехівка.

Les grammaires et les dictionnaires

Outre les grammaires citées plus haut, plusieurs autres grammaires ont été publiées pendant cette période. Nous proposons de faire une courte revue des publications les plus connues.

En 1823, à Peremychl, Ivan Mohyl'nytckyy rédigea la première grammaire de la langue ukrainienne en Galicie, *Граматика языка словена-руського* [*Grammaire de la langue slavo-roussienne*]. Dans sa préface, il justifia l'indépendance et les origines de la langue ukrainienne.

En 1830, à Budapest, fut publiée *Grammatica slavo-ruthena* de Mykhaylo Loutchkay, un linguiste de Transcarpatie. Dans son ouvrage, publié en latin, l'auteur compara l'ukrainien et le vieux-slave.

En 1834, à Peremychl, parut *Граматика руської або малоруської мови в Галичині* [*Grammaire de la langue roussienne ou du petit-russe en Galicie*] de Iossyp Levyts'kyu. Cette grammaire fut publiée en allemand et visait à étudier « iazytchie ». Dans son travail, Levyts'kyu essaya d'unifier la langue ukrainienne de la rive gauche et celle de la rive droite, notamment en y ajoutant des éléments des deux variantes.

En 1849, à Lviv, fut publiée *Грамматика руского языка* [*Grammaire de la langue roussienne*] de Iakiv Holovats'kyu. Cette grammaire est la plus exhaustive de toutes les grammaires écrites précédemment. L'auteur y passe en revue les phénomènes phonétiques et grammaticaux qui englobaient toutes les variantes de la langue ukrainienne littéraire à cette époque, puis il analyse les phénomènes propres aux dialectes galiciens.

En Galicie, plusieurs grammaires furent publiées pour être utilisées dans le cadre scolaire. Parmi ces grammaires les plus répandues :

- *Граматика малоруського языка для шкіл парафіяльних в Галиції* [*Grammaire du petit-russe pour les écoles paroissiales de Galicie*] de Théophile Hlyns'kyu (Lviv, 1845) ;
- *Граматика руського языка* [*Grammaire de la langue roussienne*] de Mykhaylo Ossadtsa (Lviv, 1862) ;
- *Граматыка руска для ужитку в школах людових в Галичині* [*Grammaire de la langue roussienne pour les écoles publiques de Galicie*] d'Omelian Partyts'kyu (Lviv, 1873) ;
- *Граматика руского языка для школь середныхъ* [*Grammaire de la langue roussienne pour les écoles secondaires*] d'Omelian Ohonovs'kyu (Lviv, 1889).

En Ukraine de la rive gauche, dans la première moitié du XIX^e siècle, la plupart des études en linguistique visaient à défendre la langue ukrainienne et à démontrer qu'il s'agissait d'une langue indépendante et non d'un dialecte du russe ou du polonais. Nous pouvons citer *Двѣ русскія народности* [*Deux nations roussiennes*] de Mykola Kostomarov (in *Основа*, Pétersbourg, 1861), plusieurs publications de Mykhaylo Maksymovytch, dont *Ответные письма Погдину* [*Lettres-réponses à M. Pogodine*] (in *Русская беседа, Échanges russes*), Moscou, 1856).

À la même époque, Oleksandr Potebnia, membre correspondant de l'Académie des Sciences de Saint-Pétersbourg, publia plusieurs études. Outre un grand nombre de publications concernant la linguistique en général ou la langue russe, Potebnia publia quelques travaux dédiés à la langue ukrainienne, et notamment *Заметки о малорусском наречии* [*Notes sur le dialecte petit-russe*] (Voronej, 1871) et *Руководство к обучению грамоте, составленное для малорусских вечерних школ* [*Manuel destiné à l'enseignement de la lecture et de l'écriture, élaboré pour les écoles du soir petit-russes*] (édition spéciale de *Київська старовина* [*L'Antiquité de Kyïv*] (Kyïv, 1899).

En 1876, Pavlo Jytets'kyj publia à Kyïv un ouvrage linguistique approfondi sur l'histoire de la phonétique de la langue ukrainienne, *Очерк звуковой истории малорусского наречия [Aperçu de l'histoire des phonétiques du dialecte petit-roussien]*.

Pendant cette période, parurent plusieurs dictionnaires bilingues, tels que :

- *Le Dictionnaire russo-ukrainien* de Mykhaylo Levchenko (1874) qui contient 7 500 mots russes et 9 000 mots ukrainiens ;
- *Словарь російсько-український [Dictionnaire russe-ukrainien]* en quatre volumes de Mykola Umanets' et Mykhaylo Komarov (Lviv, 1893-1898) qui contient 37 000 mots ;
- *Русско-малоросійський словарь [Dictionnaire petit-russe]* en deux volumes de Ievhen Tymtchenko (Kyïv, 1898-1899).

Quelques dictionnaires descriptifs furent également produits, et notamment :

- *[Словарь малорусского наречия [Dictionnaire du dialecte petit-russe]* d'Oleksandr Afanassiev-Tchoujbyns'kyj (Saint-Pétersbourg, 1855) ;
- *Опытъ южнорусскаго словаря [Essai d'un dictionnaire du russe méridional]* de Kalenyk Cheykovs'kyj, Volume 1 (Kyïv, 1861) et Volume 5⁵² (Moscou, 1893) ;
- *Словарь малоросійских идиомов [Dictionnaire des idiomes du petit-russe]* de Mykola Zakrevsky (Moscou, 1861) ;
- *Словниця української (або югової руської) мови [Dictionnaire de la langue ukrainienne (ou du roussien méridional)]* de Fortounat Piskounov (Odessa, 1873).

1.3.5 L'ukrainien de 1900 à 1922

Au début du XX^e siècle, vers 1900, les territoires ukrainiens furent divisés entre trois États : la Russie, l'Autriche et la Hongrie (les deux derniers étaient réunis au sein de l'Autriche-Hongrie). Environ 85 % des Ukrainiens vivaient dans l'Empire russe, 13 % en Galicie orientale et en Bucovine, annexées à l'Autriche, et 2 % en Transcarpatie sous domination hongroise (Cheveliov 1987 [65]).

Un mouvement de renouveau national, apparu à la fin du XIX^e siècle, grandissait, et l'idée de retrouver une Ukraine libre et indépendante se propageait aussi bien dans l'Ukraine de la rive gauche que dans l'Ukraine de la rive droite. Mais les événements historiques de cette période empêchèrent le peuple ukrainien d'aller au bout de ses aspirations. Aussi est-il important de les rappeler.

Le contexte historique

En 1905-1907, la crise économique et la défaite lors de la guerre russo-japonaise de 1904-1905 provoquèrent une révolution dans l'Empire russe. En Ukraine de la rive gauche, des soulèvements paysans eurent lieu dans 64 régions (sur 94). Le mouvement fut amplifié par des insurrections dans les usines et dans les régiments militaires. Le 17 octobre 1905, pour apaiser la situation, le tsar Nicolas II signa un manifeste annonçant de grandes réformes concernant les droits individuels (les libertés d'expression, de réunion et d'association), le suffrage, le rôle de la Douma dans l'élaboration des lois et le contrôle du budget.

À la fin de l'année 1905, à Kyïv, fut fondé le Parti radical démocrate ukrainien⁵³. Ce parti avait pour but d'obtenir une large autonomie pour l'Ukraine. En 1908, le parti fut dissous et une partie de

⁵² Les autres manuscrits ont brûlé dans un incendie, et n'ont donc pas été édités.

⁵³ En ukrainien : *Українська демократично-радикальна партія (УДРП)*.

ses membres ont formé l'Union des progressistes ukrainiens. Cette union a été créée pour défendre les intérêts ukrainiens, notamment l'accès à l'enseignement en ukrainien, l'utilisation de la langue ukrainienne dans les tribunaux, les églises, la promotion de la culture ukrainienne dans l'Empire russe, mais aussi à l'étranger.

La première guerre mondiale (1914-1918) donna un nouveau virage au mouvement national ukrainien. Le 1^{er} août 1914, à Lviv, fut créé le Conseil suprême ukrainien⁵⁴, qui a fait le choix de s'allier à l'Autriche-Hongrie dans la guerre contre l'Empire russe.

Le 4 août 1914, toujours à Lviv, des hommes politiques ukrainiens exilés de l'Empire russe créèrent l'Union de libération de l'Ukraine⁵⁵. Les membres de cette union prévoyaient la création d'un État ukrainien indépendant.

En septembre 1914, l'armée russe envahit une grande partie de la Galicie orientale, puis en mars 1916 la Bucovine et la Volhynie. Quelques mois après, ces territoires furent repris par l'armée polonaise et ses alliés.

Le 5 mai 1915, le Conseil suprême ukrainien fut réorganisé en Conseil général de l'Ukraine, composé de 31 délégués (dont les députés des partis politiques de Galicie, de Bucovine et de l'Ukraine de la rive gauche). Le Conseil général annonça sa volonté de bâtir un État ukrainien indépendant sur les terres ukrainiennes faisant partie de l'Empire russe. Il s'engagea également à obtenir l'autonomie des terres ukrainiennes faisant partie de l'Autriche-Hongrie.

Le 5 novembre 1916, Vienne et Berlin s'accordèrent sur la création du royaume de Pologne et prévoyèrent d'élargir les frontières du royaume aux territoires ukrainiens pris aux Russes.

Le 30 mai 1917, le Conseil général déclara au Parlement autrichien que les terres de l'ancienne principauté de Galicie-Volhynie étaient ukrainiennes et qu'elles ne pouvaient donc pas faire partie de l'État polonais.

À **Kyïv**, en mars, après le début de la révolution bolchevique russe de février 1917, le Conseil central d'Ukraine⁵⁶ fut créé. Il avertit dès le mois de juin le gouvernement provisoire de la Russie de son intention de créer une République autonome d'Ukraine. Après la prise du pouvoir par les bolcheviks en Russie en octobre 1917, le Conseil central proclama d'abord, le 20 novembre, l'autonomie de la République populaire d'Ukraine à l'intérieur de l'État russe. Puis, le 22 janvier 1918, n'ayant pas obtenu les garanties d'autonomie, il proclama, l'indépendance de **la République populaire d'Ukraine**.

Presque au même moment, le 12 décembre 1917, à **Kharkiv**, fut créée **la République populaire ukrainienne des Soviets**, membre de l'URSS. Elle fut proclamée indépendante en mars 1918.

Peu après, le 29 avril 1918, Pavlo Skoropats'kyy, aidé par l'armée allemande, réussit un coup d'État et créa un État monarchique, le **Hetmanat**, avec **Kyïv** pour capitale. Les territoires de l'Hetmanat s'étendaient de la Transnistrie, la région de Kholm et la Podlachie (en Pologne actuelle), situées à l'ouest, jusqu'aux régions de Belgorod et Voronej (en Russie actuelle), situées à l'est, ainsi que des régions de Brest et Gomel (au Bélarus actuel), de Briansk (en Russie actuelle), situées au nord, jusqu'aux régions de Taganrog, Rostov, Kouban, Stavropol (en Russie actuelle), et de Crimée,

⁵⁴ En ukrainien : *Головна Українська Рада*.

⁵⁵ En ukrainien : *Союз Визволення України*.

⁵⁶ En ukrainien : *Українська Центральна Рада*.

situées au sud. Une carte allemande de 1918 retrace bien les frontières délimitées par le 80^e gouvernement de Skoropats'kyy. Olena Boyko (2023) [80] a publié la carte que nous avons reprise dans la Figure 1.

Mais le Hetmanat n'a pas duré. En novembre 1918, Simon Petlioura fut nommé commandant en chef de l'armée de la République populaire d'Ukraine. Le 14 décembre 1918, après plusieurs semaines de combats, l'armée de Petlioura s'empara de Kyïv, mettant fin à l'existence du Hetmanat.

Figure 1. Carte allemande de l'Ukraine de 1918



Première source: Central State Archives of Foreign Archival Ucrainica (TsDAZU)

Peu avant ces événements, le 19 octobre 1918, l'Assemblée constituante des membres ukrainiens du Parlement autrichien, réunie à **Lviv**, proclama la création d'un État indépendant – la **République populaire d'Ukraine occidentale** – sur l'ensemble des territoires ethniques ukrainiens de Galicie, de Bucovine et de Transcarpatie.

À **Kharkiv**, le 6 janvier 1919, le gouvernement provisoire des ouvriers et des paysans d'Ukraine se réunit et approuva le nouveau nom de l'État : la **République socialiste soviétique d'Ukraine**.

Le 22 janvier 1919, la République populaire d'Ukraine et la République populaire d'Ukraine occidentale s'unirent en un seul État Indépendant – **l'Ukraine**, dont les frontières furent inscrites dans les documents du Conseil central. Lors de la validation des territoires de l'Ukraine, il fut tenu compte du principe ethnique (les terres à majorité ukrainienne furent considérées comme ukrainiennes).

Très rapidement, l'Ukraine fut confrontée à l'armée soviétique d'un côté et à l'armée polonaise de l'autre. Affaiblie par les combats sur plusieurs fronts et en l'absence d'alliances solides, l'Ukraine indépendante subit une défaite.

À la fin de 1920, le gouvernement de la République populaire d'Ukraine se déplaça en exil⁵⁷, où il nomma le premier président, Simon Petlioura, et ouvrit des institutions ukrainiennes (écoles, bibliothèques, églises, publications) dans plusieurs villes européennes, notamment Varsovie, Paris, Weimar, Munich, etc.

À la fin de la Première Guerre mondiale, les terres ukrainiennes furent partagées ainsi : la Bucovine fut cédée à la Roumanie par le traité de Saint-Germain (septembre 1919) et par le traité de Sèvres (août 1920), la Transcarpatie fut rattachée à la Tchécoslovaquie par le traité de Trianon (juin 1920), la Galicie, la Volhynie occidentale et la Polésie furent cédées à la Pologne par le traité de Riga (mars 1921), et cette sécession fut approuvée par décision de la conférence des ambassadeurs de l'Entente (mars 1923).

La République socialiste soviétique d'Ukraine perdit alors son indépendance et intégra l'Union soviétique le 30 décembre 1922.

Le contexte linguistique

Ces événements historiques ont eu une incidence directe sur la langue.

Le décret d'Alexandre II était toujours en vigueur en Ukraine de la rive gauche dans les années 1900 : il n'y avait aucune publication en ukrainien et l'enseignement se faisait en russe. Les seuls livres qu'on pouvait trouver en ukrainien avaient été publiés en Galicie et importés illégalement. C'est ainsi que le vocabulaire galicien a pu se propager en Ukraine orientale.

Par ailleurs, le 24 novembre 1905 et le 26 avril 1906, toujours dans l'Empire russe, de nouvelles règles temporaires entrèrent en vigueur : en particulier, la censure des publications en ukrainien fut levée. Les publicistes ukrainiens profitèrent de cette brèche et relancèrent la presse périodique.

Le relâchement de la censure fit espérer l'ouverture d'écoles ukrainiennes, mais malheureusement cela n'arriva pas : de 1905 jusqu'à 1914, aucune école ukrainienne ne fut ouverte (Cheveliov 1987 [65]).

En Autriche-Hongrie, contrairement à l'Ukraine de la rive gauche, la langue ukrainienne n'était pas interdite. Cependant, pour défendre leur droit constitutionnel, les Ukrainiens ont dû affronter l'administration autrichienne ou hongroise, qui limitait l'utilisation de la langue, notamment dans le domaine de l'enseignement.

Quant aux publications périodiques, leur nombre était considérable, notamment en Galicie, où même les intellectuels de la rive gauche étaient publiés.

De ce fait, la Galicie est devenue le centre du développement de la langue ukrainienne littéraire. Au cours des deux premières décennies du XX^e siècle, l'interaction entre le galicien et l'ukrainien de la rive gauche s'intensifia : ce dernier s'enrichit d'un vocabulaire significatif et de formes grammaticales empruntées du galicien, tandis que dans les publications galiciennes apparurent des russismes, et des participes actifs au présent, inhabituels pour le galicien (Holoïoukh 2021 [15]).

Entre 1918 et 1920, lorsque la langue ukrainienne eut obtenu le statut de langue d'État, la terminologie sociopolitique se développa considérablement.

⁵⁷ Le Conseil de la République populaire d'Ukraine en exil cessa de fonctionner le 15 mars 1992 seulement, après avoir adopté une résolution de transfert des pouvoirs aux autorités de l'Ukraine indépendante.

Le 1^{er} août 1918, l'Hetman Pavlo Skoropats'kyi approuva une loi sur l'apprentissage obligatoire de la langue ukrainienne, de l'histoire et de la géographie de l'Ukraine, sur tous les territoires du Hetmanat. De ce fait, plusieurs millions d'exemplaires de manuels scolaires ukrainiens furent publiés. En parallèle, des départements de langue, de littérature, de culture, d'histoire et de droit ukrainiens furent créés dans les universités de Kyïv, de Kharkiv et d'Odessa. Le gouvernement créa également l'Académie ukrainienne des sciences, la Bibliothèque nationale, les Archives nationales, la Galerie nationale d'art, le Théâtre national et le Musée d'histoire.

Les orthographes

Pendant cette période, il y eut plusieurs tentatives de normalisation de l'orthographe.

En 1900, à l'Académie nationale des sciences de Lviv, une commission linguistique fut chargée de préparer une nouvelle orthographe ukrainienne. Participèrent à ce projet plusieurs linguistes et enseignants, parmi lesquels des Ukrainiens de la rive gauche : Oleksandr Kolessa, Mykhaylo Pavlyk, Volodymyr Hnatiouk, Illia Kokoroudz, Kyrylo Stoudyns'kyi, Stepan Smal'-Stots'kyi, Ivan Franko, Mykhaylo Hrouchevs'kyi, Ahatanhel Kryms'kyi, ainsi que des linguistes de la rive droite : Volodymyr Dorochenko et Ievhen Timochenko. La commission rédigea en 1904 « L'orthographe roussienne avec un dictionnaire »⁵⁸ basée sur la « jelekhivka ». Cette orthographe fut validée par l'Académie nationale des sciences.

En 1908-1909, à Kyïv, sous la rédaction de Borys Hrintchenko, fut publié *Словарь української мови [Dictionnaire de la langue ukrainienne]*. Il a été rédigé à partir de la « koulivivka », en tenant compte de certaines règles de l'orthographe galicienne. Hrintchenko a proposé de développer un système orthographique unique pour l'ensemble de l'Ukraine afin d'éviter les normes orthographiques multiples, qui nuisaient au développement de la langue littéraire ukrainienne. La « hrintchenivka » fut adoptée dans toutes les éditions et publications ukrainiennes.

En 1918-1919, Ivan Ohienko reprit et améliora la « hrintchenivka » dans le recueil *Головніші правила українського правопису [Principales règles de l'orthographe ukrainienne]*, préparé à la demande de la commission linguistique de l'Académie des sciences de l'Ukraine. Le recueil fut validé par la commission et publié en 1919 à Kyïv. En 1920, la commission révisa cette orthographe en ajoutant quelques corrections mineures. Le gouvernement approuva également son utilisation sur l'ensemble du territoire ukrainien. En 1921, à Kyïv, une version révisée de l'orthographe fut publiée dans une édition à grand tirage.

En 1922, à l'initiative de la Société scientifique Chevtchenko de Lviv, la commission linguistique élaborait une orthographe *Правописні правила [Règles d'orthographe]*, basée sur les *Principales règles de l'orthographe ukrainienne* publiées à Kyïv en 1921. Cela permit de rapprocher l'orthographe galicienne et l'orthographe kyïvienne. La principale différence entre les deux orthographes resta l'écriture des mots d'origine étrangère.

Les grammaires et les dictionnaires

Sous l'Empire russe, après la révolution de 1905, parurent quelques grammaires pour les écoles :

- *Коротка грамати́ка української мови [Grammaire abrégée de la langue ukrainienne]* de Petro Zaloznyy (Poltava, 1906) ;

⁵⁸ En ukr. : *Руска правопись зі словарцем.*

- *Коротка граматики для школи [Grammaire abrégée pour l'école]* de Havryl Cherstiouk (Poltava, 1907) ;
- *Українська граматики [Grammaire ukrainienne]* de Ievhen Tymtchenko (Kyïv, 1907) ;
- *Граматики української мови [Grammaire de la langue ukrainienne]* d'Ivan Netchouy-Levyts'kyu (Kyïv, 1914).

Toutefois, ces grammaires ne furent utilisées que pour l'enseignement à domicile, car les écoles ukrainiennes étaient interdites.

En Ukraine de la rive droite, la situation n'était pas propice non plus. De ce fait, peu de grammaires furent publiées pendant cette période.

En 1901, à Oujhorod, Avhoustyn Volochyn publia *Методическая граматики угро-русского литературного языка для народных школ [Grammaire méthodique de la langue littéraire ougro-roussienne pour les écoles publiques]*. Cette grammaire contenait des éléments lexicaux de la variante roussyne de la langue ukrainienne et beaucoup de russismes car Volochyn était russophile. La grammaire a été rééditée à Oujhorod en 1919 sous le titre *Методическая граматики карпато-русского языка для народных школ [Grammaire méthodique de la langue littéraire carpatho-roussienne pour les écoles publiques]* et en 1923 sous le titre *Методична граматики карпато-русского языка для низших клас народных школ [Grammaire méthodique de la langue littéraire carpatho-roussienne pour les écoles primaires publiques]*.

En 1913, à Vienne, a été publiée en allemand *Grammatik der ruthenischen (ukrainischen) Sprache*, et en 1914, en ukrainien *Граматики руської мови [Grammaire de la langue roussienne]* de Stepan Smal'-Stots'kyu et Fedir Gartner. Dans leur ouvrage, les auteurs rejetaient la théorie du slave oriental commun qui prédominait dans la linguistique russe et soutenaient que l'ukrainien descendait directement du slave.

Après 1917, des écoles ukrainiennes ont commencé à s'ouvrir et le besoin en grammaires a augmenté. Ainsi, de nombreuses grammaires scolaires ont été éditées partout en Ukraine, dont la plus répandue, *Початкова граматики української мови [la Grammaire élémentaire de la langue ukrainienne]* d'Olena Kourylo (Kyïv, 1917), qui a été rééditée dix fois entre 1917 et 1926 (Encyclopédie 2022 : *Курило Олена Борисівна*).

Plusieurs grammaires pour les autodidactes, les étudiants et le public initié ont été également publiées, dont :

- *Украинская граматики [la Grammaire ukrainienne]* d'Ahatanhel Kryms'kyu (Moscou, 1907) ;
- *Курс практичної граматики української мови [Grammaire pratique de l'ukrainien]* (Rastatt, 1918) et *Граматики української мови для самонавчання та в допомогу шкільній науці [Grammaire de la langue ukrainienne pour l'auto-apprentissage et pour compléter les études scolaires]* (Berlin, 1921) de Vassyl' Simovytsch ;
- *Українська граматики [Grammaire ukrainienne]* (Kyïv, 1918) et *Курс украинского языка [Cours de langue ukrainienne]* (Kyïv, 1918) d'Ivan Ohienko.

Au début du XX^e siècle, plusieurs dictionnaires furent édités, parmi lesquels *Словарь української мови [Dictionnaire de la langue ukrainienne]* de Hrintchenko de 1907-1908, que nous avons déjà cité plus haut, publié à Kyïv en quatre volumes. C'était à cette époque l'un des dictionnaires les plus complets jamais réalisés pour l'ukrainien ; il comprend 68 000 mots. Les mots ukrainiens y sont expliqués par des équivalents russes ou des phrases descriptives. La grande majorité des mots sont illustrés par des phrases. Ce dictionnaire contient des variantes

orthographiques, les dérivations, les mots composés, ainsi que beaucoup de termes et de métaphores.

Un autre dictionnaire remarquable a été préparé par Zeon Kouzelia et Mykola Tchaykovs'kyy, et publié à Tchernivtsi en 1910. Il s'agit du *Словар чужих слів : 2 000 слів чужого походження в українській мові* [Dictionnaire des mots étrangers : 2 000 mots d'origine étrangère dans la langue ukrainienne]. C'est un des premiers dictionnaires qui expliquent des mots d'origine étrangère (étymologie, contexte d'utilisation, orthographe en alphabet latin).

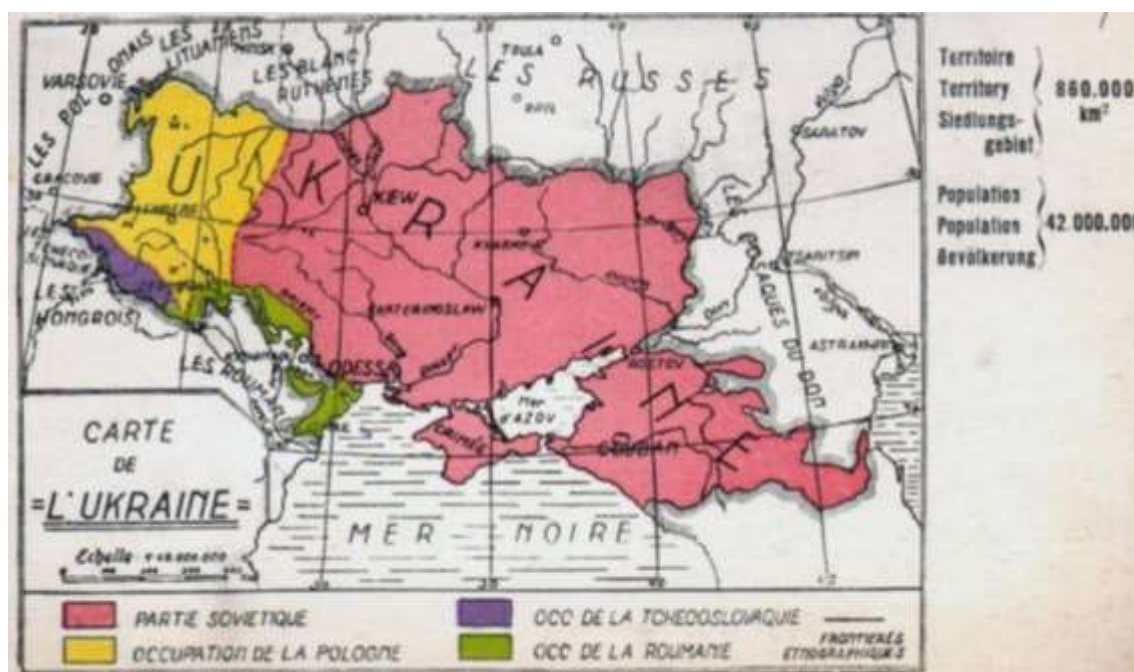
Pendant cette période, en plus des grammaires, plusieurs études furent publiées, notamment dans le domaine de la morphologie (Vassyl' Simovytych, Mykola Hrouns'kyy, Mykola Soulyma, Olena Kourylo, Serhiy Smeretchyns'kyy), de la dérivation (Stepan Smal'-Stots'kyy), de la lexicologie et de l'étymologie (Vassyl' Simovytych, Hryhoriy Ilins'kyy, Petro Bouzouk, Ivan Franko, Ivan Nechuy-Levyts'ky), de la syntaxe (Mykola Soulyma, Vassyl' Simovytych, Mykola Hrouns'kyy, Serhiy Smeretchyns'kyy), de la phonétique (Olaf Brok, Stepan Smal-Stots'kyy, Alexandre Thomson), et de la dialectologie (Ivan Zilyns'kyy, Ivan Franko).

1.3.6 L'ukrainien de 1922 à 1939

Le contexte historique et linguistique

Comme nous l'avons mentionné précédemment, pendant cette période, les terres ukrainiennes étaient partagées entre quatre pays : la République socialiste soviétique d'Ukraine, qui faisait partie de l'URSS, la Pologne, la Roumanie et la Tchécoslovaquie. Une carte postale contenant une carte de l'Ukraine, éditée en Belgique dans les années 1930, illustre l'article de Hrabovs'kyy (2020) [22] et permet de se rendre compte de la géographie ukrainienne de cette période (Figure 2).

Figure 2. Carte postale belge des années 1930 contenant une carte de l'Ukraine



Première source : Central State Archives of Foreign Archival Ucrainica (TsDAZU)

En Union soviétique, les premières décennies furent marquées par l'ukrainisation, visant à renforcer le sentiment national et, par ce biais, à rompre avec l'héritage tsariste. Cette politique permit l'utilisation généralisée de la langue ukrainienne, y compris dans l'administration.

L'enseignement se faisait aussi en ukrainien : 2,5 millions d'enfants ont eu accès aux études en ukrainien en 1927 (Badenkova 2018 [1]). Les publications périodiques fleurissaient, la littérature et la critique littéraire se développaient également.

À Kharkiv, en janvier 1926, fut créée l'Académie libre de littérature prolétarienne. L'organisation se fixa pour objectif la création d'une nouvelle littérature s'ouvrant à la culture de l'Europe occidentale. Le leader de l'organisation, Mykola Khvyliovyy, lança même le slogan « Loin de Moscou ! » (Hrabovs'kyy 2020 [22]). Mais l'Académie fut dissoute le 28 janvier 1928 et ses membres furent réprimés.

L'ukrainisation finit par susciter des inquiétudes à Moscou et, en 1932, elle fut interdite par un décret du Comité central du parti. Dès la fin des années 1920, toute manifestation de patriotisme national était réprimée : les intellectuels et les ecclésiastiques ukrainiens, mais aussi les simples citoyens soupçonnés de patriotisme furent envoyés dans des camps de travail ou fusillés.

Cela allait même au-delà des sanctions individuelles, car toute la nation a été concernée par la terreur communiste : en 1932-1933, le régime soviétique organisa le « Holodomor »⁵⁹, l'extermination par la famine de la campagne ukrainienne, qui était hostile à l'Union soviétique. Les Ukrainiens qualifient cet acte de génocide de la nation ukrainienne.

Le 18 novembre 1933, le plénum du Comité central du Parti communiste ukrainien adopta une résolution visant à mettre fin à la politique d'ukrainisation. Il déclara officiellement que l'ukrainisation menaçait l'existence de l'URSS unifiée.

Cette résolution fut suivie de nouvelles purges de l'élite ukrainienne, qui s'enchaînèrent jusqu'à la Seconde Guerre mondiale.

Ainsi, avant 1938, 259 écrivains ukrainiens furent publiés, et après 1938, seulement 36 (Virtchenko 2018 [13]).

Le personnel de l'Institut de linguistique de l'Académie nationale des sciences fut liquidé complètement d'abord dans les années 1920, puis à nouveau presque complètement dans les années 1930 : en 1938, il ne restait que deux linguistes sur l'ensemble du personnel (Cheveliov 1987 [65]).

Le russe fut réintroduit dans les écoles. À partir de 1933, des universités furent également ouvertes en Ukraine, mais dans la plupart, l'enseignement y était dispensé en russe. La presse fut aussi russifiée.

Dans les ouvrages scientifiques russes, le nom « Ukraine » fut à nouveau remplacé par « Южная Россия » [*Russie méridionale*] et « Малороссия » [*Petite-Russie*] (Ohienko 1949 [46]).

En Pologne, le traité de Riga de 1921 et l'accord de la conférence des ambassadeurs de 1923 garantissaient aux Ukrainiens ayant intégré la Pologne la protection de leur langue et de leur culture, mais la réalité était tout autre. Le statut juridique de la langue ukrainienne fut déterminé en partie par des traités internationaux et en partie par la Constitution polonaise et les lois de l'État. En 1927,

⁵⁹ En ukr. : *Голодомор* (étymologiquement « *famine* » + « *tuer* »). Aujourd'hui, il n'existe pas de consensus sur l'origine de cette famine : les historiens s'accordent à dire qu'il s'agit d'une famine organisée, toutefois un petit groupe soutient qu'il s'agit de circonstances malheureuses et l'attribue à de mauvaises récoltes et à une mauvaise planification de l'économie. Plusieurs estimations des pertes dues à la famine ont été faites. Les plus pertinentes, celles d'Oleh Volovyna (USA) et de l'Académie nationale des sciences d'Ukraine, estiment les pertes à 4,5 millions au total (3,9 millions de morts et 600 000 enfants non nés).

le ministère des Affaires militaires publia un décret qui stipulait que toutes les autorités et institutions militaires sur l'ensemble du territoire de la Pologne devaient travailler exclusivement et sans exception en polonais.

Une publication de 1928 du ministère de l'Éducation stipula que, dans les écoles pour les minorités nationales, la langue maternelle pouvait être utilisée sur un pied d'égalité dans les échanges internes, excluant ainsi l'ukrainien de toute communication externe. Le même document spécifiait que les écoles pouvaient être polonaises ou bilingues. Pour ouvrir une école bilingue, au moins 150 parents devaient déposer une demande ; pour ouvrir une école professionnelle bilingue, au moins 40 % des élèves devaient être ukrainiens et, là encore, un nombre suffisant de demandes de parents devait être recueilli (Cheveliov 1987 [65]).

Ces démarches visaient à limiter l'utilisation de la langue ukrainienne et étaient le premier pas du projet plus vaste de polonisation du peuple ukrainien, car les idées nationales, tout comme en Ukraine soviétique, était jugées dangereuses pour l'unité de l'État.

Ainsi, entre 1923 et 1927, le nombre des écoles primaires où l'enseignement était dispensé en ukrainien diminua considérablement, passant de 2 450 à 804. L'enseignement dans l'université de Lviv se faisait également en polonais, y compris les cours d'ukrainien moderne (Cheveliov 1987 [65]).

Le 12 juillet 1927, lors d'un congrès à Prague, le Groupe de la jeunesse nationale ukrainienne et la Légion des nationalistes ukrainiens, deux groupes qui existaient sur les territoires ukrainiens hors de l'Union soviétique, fusionnèrent pour former l'Union des organisations nationalistes ukrainiennes. L'Union voulait réunir ses moyens pour promouvoir l'identité ukrainienne (Encyclopédie de l'histoire 2019 : ОУН).

Les années 1930 furent marquées par des actions policières illégales de « pacification », qui consistaient à réprimer les « nationalistes ukrainiens » en détruisant leurs lieux de rencontre, entre autres. Ainsi, en juillet-septembre 1930, 191 actes de sabotage furent perpétrés dans les communes ukrainiennes de Galicie. Pendant la deuxième vague, d'après les sources ukrainiennes, environ 750 villages furent soumis à la « pacification », dont 250 subirent de lourdes pertes matérielles et des destructions. En septembre-novembre, plusieurs hommes politiques, ou appartenant à diverses organisations, furent arrêtés (Encyclopédie de l'histoire 2019 : *Пацифікація*).

Une autre attaque contre l'identité nationale sur les territoires de la Pologne eut lieu en 1937-1938, quand de nombreuses églises ukrainiennes, accusées de liens avec l'Ukraine soviétique, furent détruites en Podlachie et dans la région de Kholm.

En Roumanie, après l'annexion de la Bucovine, les termes du traité de Saint-Germain (1919) furent acceptés. Un paragraphe stipulait que la Roumanie accordait aux minorités nationales le droit d'utiliser leur langue, que ce soit oralement ou par écrit, de créer, gérer et contrôler des institutions caritatives, religieuses et civiques, de créer des écoles et d'autres établissements d'enseignement privés autorisés à utiliser leur propre langue, afin que, dans les villes comptant un nombre significatif de citoyens issus des minorités nationales, l'enseignement en langue maternelle des minorités puisse être dispensé dans les écoles primaires publiques (Encyclopédie de l'histoire 2019: *Сен-Жерменський договір* et *Конституція Румунії*).

Mais dans le traité de Sèvres de 1920, qui reconnut pour la première fois l'appartenance de toute la Bucovine à la Roumanie, il n'y avait plus aucune mention concernant les droits des minorités et leurs langues. Les obligations que la Roumanie assumait en signant le traité de Saint-Germain ne furent pas incorporées dans la Constitution du 28 mars 1923, ni dans celle du 27 février 1938.

Dès les années 1920, le gouvernement commença activement une politique de romanisation des Ukrainiens. Ainsi, il insista pour que les paysans roumains puissent s'installer en Bucovine en leur donnant les terres confisquées aux églises ukrainiennes pour diminuer la concentration des Ukrainiens dans les mêmes villages (Cheveliov 1987 [65]).

En 1924, le gouvernement édicta une loi obligeant tous les Roumains ayant perdu leur langue maternelle à scolariser leurs enfants uniquement dans des écoles roumaines (Virtchenko 2018 [13]). Ainsi, en 1927, il n'y avait plus aucune école ukrainienne en Bucovine (Cheveliov 1987 [65]). En 1934, le ministère de l'Éducation promulgua un décret visant à renvoyer tous les enseignants ukrainiens qui demandaient le retour de la langue ukrainienne dans les écoles, « pour hostilité à l'État et au peuple roumain » (Virtchenko 2018 [13]).

Bien que la presse et l'édition en ukrainien ne fussent pas totalement interdites, certaines maisons d'édition furent fermées, et celles qui continuaient de publier furent soumises à une censure sévère. Entre 1918 et 1928, seulement 39 livres furent publiés en ukrainien (Cheveliov 1987 [65]).

L'Église orthodoxe a été entièrement reprise par les Roumains en 1921, sous le nom d'Église orthodoxe grecque (Cheveliov 1987 [65]).

En Tchécoslovaquie, le traité de Trianon (1920)⁶⁰ et la première Constitution de du 29 février 1920 garantirent l'autonomie de la Transcarpatie et permirent au peuple ukrainien de constituer son propre Parlement⁶¹, qui pouvait promulguer des lois en matière linguistique, éducative et religieuse, en matière d'administration interne et dans d'autres domaines déterminés par les lois de la Tchécoslovaquie (Traité de Trianon 1920, Tchavarha 2017 [69]).

Cependant, le Parlement fut élu seulement le 13 février 1939, à Khoust, après l'entrée en vigueur de la Constitution du 22 novembre 1938 qui élargissait l'autonomie de la Transcarpatie.

La première Constitution stipulait que, dans les villes où un nombre important de minorités nationales vivaient, l'enseignement public devait pouvoir se faire dans la langue de la minorité, mais en même temps, la langue tchécoslovaque⁶² devait être une matière d'étude obligatoire (Encyclopédie de l'histoire 2019 : *Генеральний статут про організацію і адміністрацію Підкарпатської Русі*).

Dans le traité de Trianon, la Transcarpatie était définie en tant que « territoires roussyns », et le peuple comme « Roussyns du sud des Carpates »⁶³. Cet ethnonyme fut repris du Statut général de la Transcarpatie, adopté le 18 novembre 1919, dans lequel il était noté que, parallèlement au nom « Transcarpatie », il serait possible d'utiliser le nom « Roussynie », et que la langue ethnique serait la langue officielle (Encyclopédie de l'histoire 2019 : *Генеральний статут про організацію і адміністрацію Підкарпатської Русі*).

Cela permit au gouvernement de la Tchécoslovaquie d'introduire l'idée que l'ethnie des « Roussyns » était une ethnie à part entière, et la langue roussyne une langue à part entière et non

⁶⁰ https://wwi.lib.byu.edu/index.php/Treaty_of_Trianon

⁶¹ En tchèque : *sněm*.

⁶² Il ne s'agit pas d'une langue réelle mais d'un terme politique qui désignait les langues de la majorité ethnique.

⁶³ Dans [Traité de Trianon, 1920] Ruthenian territory et Ruthenians to the south of the Carpathians, traduction en tchèque de « Ruthenians » – « *Rusíni* ».

une variante de l'ukrainien, pour écarter toute envie de rapprochement avec les autres territoires ukrainiens. Ainsi, un malaise autour de la langue nationale s'installa sur ces territoires.

Dans un tel contexte, trois courants linguistiques se développèrent : « les russophiles », « les ukrainophiles » et « les nationalistes roussyns ». Chaque courant déclarait sa langue, c'est-à-dire le russe, l'ukrainien ou le roussyn, langue officielle. N'ayant pas trouvé de consensus au sujet de la langue, les autorités de Transcarpatie écrivirent une lettre demandant de trancher. Le gouvernement tchécoslovaque ne voulut pas se prononcer officiellement en faveur de telle ou telle langue, mais recommanda d'utiliser le galicien, avec l'orthographe étymologique plus familière à la population de Transcarpatie.

Mais la politique du gouvernement changea très rapidement. En 1933 puis en 1936, il déconseilla d'utiliser l'adjectif « ukrainien » au sujet du peuple de Transcarpatie et de sa langue (Cheveliov 1987 [65], Tchavarha 2017 [69]).

À l'époque, la personnalité emblématique de la Transcarpatie était Avgoustyn Volochyn, grand défenseur de la langue ukrainienne. Il était rédacteur ou corédacteur de plusieurs éditions périodiques en ukrainien. Il fonda également l'organisation éducative et politique *Prosvita*⁶⁴ et a été le cofondateur de l'organisation éditoriale *Ounino*⁶⁵. En 1938, il est devenu ministre de la Santé puis Premier ministre de l'Ukraine des Carpates, et, le 15 mars, président de l'Ukraine des Carpates.

Les orthographes

Au milieu des années 1930, la langue ukrainienne avait besoin de normalisation, car son orthographe ne correspondait pas complètement à la langue qui évoluait. Mais à cause du chaos qui régnait dans l'Ukraine soviétique et de l'absence de corpus linguistique en Ukraine occidentale, il n'y avait pas grand monde pour faire ce travail. Cheveliov (1987) [65] décrit ainsi la situation de la linguistique ukrainienne au cours de cette période :

« Le chaos linguistique a eu un impact majeur sur les écoles. Les enseignants étaient extrêmement confus et effrayés, les élèves étaient désorientés. C'était un crime de ne pas suivre le nouveau programme, et il n'y avait aucun moyen de le suivre parce qu'il n'y avait pas de consignes. Il semblait que l'instabilité soit inhérente à la langue ukrainienne [...]. Le prestige de la langue ukrainienne, déjà faible, est tombé encore plus bas. »

Dans ce contexte, il était urgent d'établir des règles d'orthographe claires et adaptées. Ainsi, en juillet 1925, à l'initiative du Conseil des commissaires du peuple et de l'Académie des sciences de la République socialiste soviétique d'Ukraine, une Commission d'État chargée de rédiger une nouvelle orthographe ukrainienne fut créée. Ce projet d'orthographe fut validé par l'Académie le 1^{er} septembre 1926 et envoyé dans toutes les instances administratives de la République socialiste soviétique d'Ukraine et dans les régions ukrainiennes des pays voisins pour avoir un retour.

Afin d'éliminer les divergences entre les systèmes orthographiques ukrainiens existants (ukrainien occidental et ukrainien oriental) et de mettre au point une orthographe unique pour l'ensemble de l'Ukraine, il fut décidé d'organiser une conférence à Kharkiv. Y furent invités des délégués de Galicie, de Bucovine, de Volhynie et de Transcarpatie. Au cours des discussions, le projet d'orthographe ukrainienne de 1926 subit quelques modifications, en particulier dans les paragraphes

⁶⁴ En ukr. : *Просвіта* [lumières].

⁶⁵ En ukr. : *Уніо*.

concernant les règles de transcription des mots étrangers et des noms propres. Cette orthographe fut ée approuvée à Kyiv le 6 septembre 1928.

Après la conférence, cette orthographe commença à être popularisée et diffusée en Galicie et en Transcarpatie, qui ne faisaient pas partie de l'Ukraine soviétique à l'époque. En Galicie, Volhynie et Bucovine, cette orthographe fut tout de suite adoptée, et elle resta en vigueur jusqu'en 1939. Elle a été aussi adoptée par les Ukrainiens en exil et est toujours utilisée par la diaspora ukrainienne.

En Ukraine soviétique, cette orthographe resta officielle jusqu'en 1933. Mais après l'arrêt de l'ukrainisation, la Commission d'orthographe, dirigée par Andriy Khvyliia, qualifia l'orthographe de Kharkiv de « nationaliste » et ordonna d'éliminer tous les éléments bourgeois venus de Galicie. Ainsi, la lettre « r » fut supprimée de l'alphabet et la terminologie scientifique ukrainienne fut révisée et harmonisée avec le nouveau dictionnaire russo-ukrainien de 1927. Les 15 modèles de dérivation morphologie propres à la langue ukrainienne furent supprimés et remplacés par des modèles russes. L'orthographe fut approuvée par Khvyliia le 5 septembre 1933.

Les grammaires et les dictionnaires

Cheveliov (1962) [66] remarque qu'en Ukraine, au début de la période soviétique, deux écoles linguistiques existaient : les ethnographes et les puristes. Les ethnographes plaçaient au cœur de leurs recherches le peuple ukrainien et ses chansons, ses ballades et ses contes populaires. L'école ethnographique regroupait Ahatanhel Kryms'kyi, Ievhen Tymtchenko, Olena Kourylo (premiers travaux), Mykola Hladkyi, Serhiy Smeretchyns'kyi.

Les puristes tentaient de « nettoyer » la langue ukrainienne de tous les emprunts lexicaux, grammaticaux (affixes et flexions) ou syntaxiques. Oleksa Syniavs'kyi, Mykola Sulyma, Mykola Nakonechnyy et Olena Kourylo (œuvres de la deuxième période) faisaient partie de cette école. Cheveliov (1962) [66] considérait que Vsevolod Hantsov et Andriy Nikovs'kyi en étaient également proches.

Quelques linguistes n'appartenaient pas à ces écoles, parmi eux Panteleimon Kovaliv, Mykola Hrouns'kyi, Konstiantyn Nimtchynov.

Pendant cette période, suite à l'instauration de l'ukrainisation, plusieurs manuels scolaires, mais aussi des grammaires, des dictionnaires bilingues ou encore des dictionnaires terminologiques ont été publiés.

Mais dès la fin des années 1920, la méthodologie du Parti commença à dominer la linguistique ukrainienne. Les autorités publièrent dans des revues linguistiques des listes des mots et des formes à ne pas utiliser, et donnèrent des instructions sur les mots et formes par lesquels il fallait les remplacer pour rapprocher l'ukrainien du russe. L'article d'Andriy Khvyliia *Знищити коріння українського націоналізму на мовному фронті* [*Détruire les racines du nationalisme ukrainien sur le front linguistique*], paru d'abord dans la revue *Ukraïns'kyi bolchovyk* [*Bolchevik d'Ukraine*] en 1933, puis réédité sous forme de brochure à part, ou encore l'article d'Oleksandr Finkel *За більшовицьку пильність на фронті творення української радянської культури* [*Pour une vigilance bolchevique sur le front de la création de la culture soviétique ukrainienne*], paru dans la revue *Movoznavstvo* [*Linguistique*] en 1934, sont de bons exemples de l'ingérence du Parti dans la linguistique. Dans ces articles, les auteurs dénoncent les puristes qui inventaient un vocabulaire artificiel pour éloigner la langue ukrainienne du russe.

Le Parti allait jusqu'à interdire les travaux déjà publiés. Ainsi, les trois volumes du dictionnaire *Російсько-український академічний словник* [*Dictionnaire académique russo-ukrainien*] (Kyiv : Tchervonyy chliah, 1924-1933) sous la rédaction d'Ahatanhel Kryms'kyi et Serhiy Iefremov ont été

saisis et le quatrième manuscrit prêt pour la publication fut détruit. Le même sort fut réservé aux dictionnaires terminologiques préparés par l'Académie des sciences d'Ukraine, eux aussi détruits. Le dictionnaire *Українсько-російський словник : за новим правописом* [Dictionnaire russo-ukrainien : selon la nouvelle orthographe] d'Ovsiy Izumov (Kharkiv, Kyiv, 1930) fut interdit.

Dans de telles conditions, de nombreux chercheurs ont cessé d'écrire (Ohienko 1949 [46]).

Le nouveau dictionnaire russo-ukrainien académique conventionnel *Російсько-український словник* fut publié en 1937 à Kyiv, sous la rédaction de Pavlo Moustiatsa et de Stepan Vassylevs'kyy. La langue ukrainienne du dictionnaire fut purgée de tous les mots nationalistes, remplacés par des emprunts au russe.

Le développement de la lexicographie ukrainienne trouva refuge dans l'émigration. Ainsi, pendant cette période, furent publiés :

- *Медичний латинсько-український словник* [Dictionnaire médical latin-ukrainien] de Martiryi Halyn (Prague, 1926) ;
- *Німецько-український кишеньковий словник* [Dictionnaire de poche allemand-ukrainien] de Halyna Nakonechna (Leipzig, 1939) ;
- *Німецький та український військовий і летунський словник* [Dictionnaire militaire et aéronautique allemand-ukrainien] en deux volumes d'Ivan Il'nits'kyy-Zankovytsch (Berlin, 1939) ;
- ainsi que les très nombreuses études linguistiques de Stepan Smal'-Stots'kyy, Ivan Ohienko, Oleksandr Kolessa, Vassyl' Simovytsch.

1.3.7 L'ukrainien de 1939 à l'indépendance de l'Ukraine (1991)

Pendant cette période, les Ukrainiens tentèrent de restaurer leur indépendance, sans y parvenir. Rappelons ces événements dans l'ordre chronologique.

Le contexte historique et linguistique pendant la guerre

Le 14 mars 1939, l'armée hongroise, alliée à l'Allemagne nazie, entra dans les territoires de la Transcarpatie. Le lendemain, le gouvernement de la Transcarpatie en profita pour déclarer l'indépendance de l'Ukraine des Carpates, mais le 18 mars, tous ces territoires furent envahis par les Hongrois, et le gouvernement de la Transcarpatie fut poussé à l'exil.

Le 23 août 1939, l'Allemagne et l'URSS signèrent le pacte Molotov-Ribbentrop de non-agression, qui décidait du partage de la Pologne entre les deux pays.

Peu après la signature du pacte, le 1^{er} septembre, l'armée allemande envahit la Pologne, marquant ainsi le début de la Seconde Guerre mondiale.

Le 17 septembre, l'armée soviétique, à son tour, entra en Pologne de l'Est. Ainsi, en novembre 1939, la Galicie, la Volhynie occidentale et la Polésie furent rattachées à la République socialiste soviétique d'Ukraine.

La Bucovine et la Bessarabie, qui faisaient partie de la Roumanie, subirent le même sort : ces régions furent rattachées à la République socialiste soviétique d'Ukraine en août 1940.

Après cet événement, les autorités soviétiques dessolèrent la Société scientifique Chevtchenko de Lviv. De nombreux membres de l'organisation furent arrêtés, tout comme les personnalités des autres régions rattachées. Certains d'entre eux purent émigrer et participer à la vie scientifique de la diaspora ukrainienne, notamment en France, en Allemagne, aux États-Unis, au Canada, ou encore en Australie.

Après la déclaration de guerre de l'Allemagne à l'URSS le 22 juin 1941, l'*Organisation des nationalistes ukrainiens*⁶⁶ (OUN), sous la direction de Stepan Bandera, se mobilisa sur les territoires de l'Ukraine occidentale. L'OUN prévoyait de libérer les terres ukrainiennes des nazis et des Soviétiques et de restaurer l'indépendance de l'Ukraine. Le 30 juin 1941, les troupes nazies entrèrent à Lviv, et la fraction OUN-B profita du départ des troupes soviétiques pour créer le Conseil national ukrainien de Lviv et proclamer la restauration de l'État ukrainien. Mais très rapidement les Allemands écrasèrent cette initiative en mettant dans des camps de concentration tous les dirigeants de l'organisation. Par la suite, le mouvement continua la lutte clandestine. Ses membres élargirent aussi les rangs de l'*Armée insurrectionnelle ukrainienne (UPA)*⁶⁷.

Le 19 septembre 1941, les troupes nazies entrèrent à Kyïv. Le 5 octobre 1941, la fraction OUN-M organisa le Conseil national ukrainien de Kyïv, à l'instar du Conseil central du peuple de l'Ukraine des Carpates (1938) et du Conseil national ukrainien de Lviv (juin 1941). Sous l'égide du Conseil national ukrainien, l'Université de Kyïv, l'Institut polytechnique, l'Institut de l'industrie alimentaire et l'Institut de médecine, ainsi que les écoles de Kyïv, rouvrirent leurs portes. Plusieurs organisations furent créées, parmi lesquelles l'Union des écrivains ukrainiens et la Croix-Rouge ukrainienne. Les éditions périodiques ukrainiennes furent également relancées. Mais, le 27 novembre 1941, les autorités nazies statuèrent que le Conseil national ukrainien faisait de l'ingérence dans la vie politique et, de ce fait, engagèrent des répressions à l'encontre de ses membres.

En 1941, à Oujhorod, à l'initiative du gouvernement hongrois, fut créée la Société des sciences de Transcarpatie. Dirigée par le linguiste Ivan Garaïda, elle fut très active. Grâce aux efforts de ses membres, plusieurs éditions furent créées. En 1944, la Société fut dissoute par les Soviétiques et Garaïda fut arrêté. Il mourut en prison peu après son arrestation.

À la fin de la guerre, l'URSS récupéra presque tous les territoires ethniques de l'Ukraine : à la conférence de Yalta du 4 à 11 février 1945, les Alliés approuvèrent les frontières entre la Pologne et la République socialiste soviétique d'Ukraine selon la ligne Curzon, établie en 1919. Les Alliés ont aussi approuvé la reprise des territoires que l'URSS s'était vu accorder lors de la signature du pacte Molotov-Ribbentrop.

Le 29 juin 1945, l'URSS et la Tchécoslovaquie signèrent un accord sur la sécession de la Transcarpatie et sa réunification avec l'URSS.

Le 15 septembre 1947, est entré en vigueur le traité de paix de Paris, selon lequel la Roumanie cédait à l'Union soviétique la Bessarabie, la Bucovine du Nord, cinq îles sur le bras de Chilia et l'Île aux serpents.

En 1954, la Crimée fut rattachée à la République socialiste soviétique d'Ukraine. C'est alors que la délimitation des frontières de l'Ukraine soviétique fut achevée. En 1991, ces frontières sont devenues celles de l'Ukraine indépendante.

La carte de l'Ukraine, prise sur Wikipédia, illustre bien la géographie ukrainienne de cette période (Figure 3).

⁶⁶ En ukr. : *Організація українських націоналістів (ОУН)*. L'organisation, créée en 1929, s'est scindée en février 1940 en OUN-B, sous le commandement de Stepan Bandera, et OUN-M, sous le commandement d'Andriy Mel'nyk.

⁶⁷ *Українська повстанська армія (УПА) [Armée ukrainienne insurgée] (1941-1954)*. Cette armée a été organisée pour combattre les nazis, les Soviétiques et l'armée polonaise. Les premières unités de l'UPA ont été formées dans la région de Rivne par Taras Boulba-Borovets' le 28 juin 1941.

Figure 3. Évolution des frontières de l'Ukraine entre 1922 et 1954



Source : <https://commons.wikimedia.org/wiki/File:Ukraine-growth-fr.png?uselang=fr>

Le contexte historique et linguistique après la guerre

En 1953, après la mort de Staline et la nomination de Nikita Khrouchtchev⁶⁸ à la tête du Parti communiste de l'URSS, une nouvelle vague d'ukrainisation commença, car il fallait « dépoloniser » les populations des régions récemment annexées. Cette ukrainisation prépara le terrain au mouvement dissident culturel et littéraire des *Soixantards*⁶⁹, qui a marqué la société ukrainienne. Les *Soixantards* représentaient la génération des jeunes personnes qui luttèrent contre la russification et voulaient convaincre le Parti de préserver et de développer la culture et la langue ukrainiennes. Parmi eux, il y avait des écrivains, des poètes, des linguistes et des critiques littéraires, ainsi que des réalisateurs et des peintres (Ivan Dziouba, Vassyl' Stous, poètes, linguistes et critiques littéraires, Viatcheslav Tchornovil, critique littéraire et homme politique, Vassyl' Symonenko, poète et critique littéraire, Pavlo Tytchyna, Ivan Dratch, etc.).

Le 14 octobre 1964, à Moscou, au Comité central du Parti, le pouvoir fut renversé. Ce renversement entraîna la destitution de Khrouchtchev de son poste de premier secrétaire. Cela marqua le durcissement du régime communiste, qui entraîna à son tour le renforcement de la censure, la restauration du totalitarisme et l'écrasement de toute idée dissidente.

⁶⁸ De 1938 à 1949, il dirigea le comité central du Parti communiste de l'Ukraine (avec une interruption de mars à décembre 1947) et fut l'initiateur de la politique de la terreur sur les terres de l'Ukraine occidentale. Au poste de premier secrétaire du Parti de 1953 à 1964, il mena une politique de déstalinisation et d'adoucissement du régime totalitaire.

⁶⁹ En ukr. : *Шістдесятники* signifie littéralement « ceux des années 60 ».

De nouvelles mesures restrictives contre la langue ukrainienne furent prises sous Leonid Brejnev⁷⁰. Ainsi, en 1970, le ministère de l'Éducation de l'URSS décréta l'obligation de rédiger et de soutenir toutes les thèses uniquement en russe.

En 1985, Mikhaïl Gorbatchev commença une politique de restructuration du pouvoir qui s'accompagnait de la démocratisation et de la transparence, ce qui permit une décentralisation du pouvoir. Ainsi, en 1989, le Conseil central de la République socialiste soviétique d'Ukraine adopta une loi sur les langues, selon laquelle l'ukrainien devenait la seule langue nationale de la République.

Le 16 juillet 1990, le Conseil central de la RSS d'Ukraine adopta la Déclaration de la souveraineté de l'État ukrainien. Suite à cette déclaration, le poste de président de la RSS d'Ukraine fut créé, et des élections se sont tenues le 1^{er} décembre. Un référendum sur l'indépendance fut organisé le même jour, et plus de 90 % des votants se prononcèrent en faveur de l'indépendance de l'Ukraine.

L'indépendance fut actée le 24 août 1991.

Les orthographes

Au début de cette période, en 1942, le Conseil central de la RSS d'Ukraine demanda à l'Académie des sciences, évacuée à Oufa, d'harmoniser l'orthographe ukrainienne. En effet, en raison de l'annulation des normes orthographiques de 1928, il y avait beaucoup de confusions dans la presse, la littérature et les manuels. Le Parti donna des directives pour cette nouvelle orthographe : ne pas s'écarter inutilement des normes existantes, se conformer à la prononciation des populations, limiter les variantes orthographiques, réduire les exceptions, se rapprocher de l'orthographe russe (Ohiienko 1949 [46], Nimtchouk et Pouriaïeva 2004 [43]).

La nouvelle version fut prête en octobre, sous la direction de Leonid Boulakhovs'kyi, mais elle ne fut approuvée par la Commission et le Conseil des commissaires du peuple que le 8 mai 1945, et publiée en 1946. Peu de changements furent apportés à l'orthographe de 1933 : quelques règles furent simplifiées (écriture des mots empruntés, des nombres, terminaisons au génitif), et les règles de ponctuation et d'écriture de certains mots en attaché et en majuscules furent légèrement modifiées, à l'instar des règles orthographiques russes.

En 1959, quelques modifications mineures furent ajoutées à l'orthographe de 1946. Cette nouvelle version, publiée en 1960, devint l'édition officielle jusqu'en 1990.

Notons que pendant cette période, il a été décidé par la directive du Parti de rééditer les principaux ouvrages de l'héritage ukrainien selon les normes orthographiques de 1946 et de 1960.

Quant à l'orthographe de la diaspora, plusieurs manuels d'orthographe furent publiés, mais leurs auteurs n'apportèrent aucun changement aux règles orthographiques de 1928, ils ajoutèrent seulement des explications aux règles. Nous pouvons citer :

- *Український правопис [Orthographe ukrainienne]* d'Ivan Zilins'kyi (Cracovie, 1941) ;
- *Основні правила українського правопису [Principales Règles de l'orthographe ukrainienne]* de Yuriy Cherekh (Cheveliov) (Munich, 1946) ;
- *Український правопис [Orthographe ukrainienne]* de Petro Oksanencko (Augsbourg, 1946), ou de Iaroslav Roudnyts'kyi (Winnipeg, 1949).

⁷⁰ Au poste de premier secrétaire du Parti de 1964 à 1982.

Les grammaires et les dictionnaires

Pendant la guerre, l'activité linguistique ne fut pas complètement arrêtée et des linguistes continuaient leur travail.

Ainsi, à Oujhorod, en 1941, fut publiée *Граматика руського языка [Grammaire de la langue roussienne]* d'Ivan Garaïda. Cette grammaire devint le manuel officiel des écoles ukrainiennes en Hongrie et servit de recueil des normes orthographiques pour les éditions de la Société des sciences de Transcarpatie. Il faut noter que cette grammaire a gardé l'orthographe étymologique à l'instar de la langue ancienne, alors que l'ukrainien moderne adoptait l'orthographe phonétique, de sorte que la nette différence entre les deux écritures donnait une raison aux partisans du mouvement roussyn d'affirmer que le roussyn était une langue à part.

Comme nous l'avons noté plus haut, après la dissolution de la Société scientifique Chevtchenko, le noyau des recherches en linguistique s'est déplacé de la Galicie à l'étranger, où de nombreux ouvrages sur la morphologie, la syntaxe, l'histoire de la langue et la dialectologie furent publiés. Un certain nombre de dictionnaires bilingues, mais aussi de dictionnaires terminologiques, virent aussi le jour. Parmi les ouvrages parus pendant cette période, les plus importants sont :

- *Граматика української мови [Grammaire de la langue ukrainienne]* de Panteleymon Kovaliv (Munich, 1946) ;
- *Граматика української мови : посібник для гімназії [Grammaire de la langue ukrainienne : manuel pour le secondaire]* de Iouriï Cherekh (Cheveliov) et D. Kyslytsia (Munich, 1947) ;
- *Граматика української мови [Grammaire de la langue ukrainienne]* en deux volumes de Dmytro Kyslytsia (Toronto, 1963-1974) ;
- les ouvrages d'Ivan Ohienko tels que *Українська літературна мова [Histoire de la langue littéraire ukrainienne]* (Winnipeg, 1949), *Граматичні основи літературної мови [Fondements grammaticaux de la langue littéraire ukrainienne]* (Saskatoon, 1951) et *Як говорити й писати по-літературному [Comment parler et écrire en ukrainien littéraire]* (Winnipeg, 1958) ;
- les ouvrages de Iouriï Cheveliov, tels que *Синтаксис сучасної української літературної мови. Просте речення [Syntaxe de la langue ukrainienne. La phrase simple]* (Munich, 1951), *Нарис сучасної української літературної мови [Aperçu de la langue littéraire ukrainienne moderne]* (Munich, 1951), *Історична фонологія української мови [Phonologie historique de la langue ukrainienne]* (Heidelberg, 1979).

À Kyïv, les publications en linguistique furent relancées dans les années 1950. Elles furent gérées par l'Institut linguistique de l'Académie des sciences de la RSS d'Ukraine, situé à Kyïv. Tout était centralisé : l'Académie de Kyïv fut chargée de la coordination des travaux menés par le département de linguistique de l'Académie des sciences de Lviv, ainsi que par les départements linguistiques des universités et des instituts pédagogiques de différentes villes d'Ukraine (Houmeniouk 2014 [26]).

Les dictionnaires académiques les plus importants de l'époque soviétique d'après-guerre sont :

- *Академічний російсько-український словник [Dictionnaire académique russe-ukrainien]* sous la rédaction d'Ahatanhel Kryms'kyï et de Serhiï Iefremov (Moscou, 1948) ;
- *Українсько-російський словник [Dictionnaire ukrainien-russe]* en six volumes sous la rédaction d'Illia Kyrytchenko (Kyïv, 1953-1963) ;
- *Польсько-український словник [Dictionnaire polonais-ukrainien]* en deux volumes sous la rédaction de Loukiïa Houmets'ka (Kyïv, 1958) ;
- *Українсько-російський і російсько-український словник власних імен людей [Dictionnaire ukrainien-russe et russe-ukrainien des noms propres de personnes]* sous la rédaction d'Illia Kyrytchenko (Kyïv, 1964) ;

- *Українсько-російський словник [Dictionnaire russe-ukrainien]* sous la rédaction d'Ivan Bilodid (Kyïv, 1968) en trois volumes ;
- *Словник української мови [Dictionnaire de la langue ukrainienne]* sous la rédaction d'Ivan Bilodid (1970-1980) en onze volumes ;
- *Етимологічний словник української мови [Dictionnaire étymologique de la langue ukrainienne]*, prévu en sept volumes, sous la rédaction d'Oleksandr Melnytchuk (les volumes 1 à 3 de 1982 à 1989, les volumes 4 à 6 de 2003 à 2012 ; le volume 7 est toujours en préparation).

Dans ces dictionnaires, le vocabulaire ukrainien trop éloigné du russe a été évité. Ainsi, dans les dictionnaires russe-ukrainien de 1948 et ukrainien-russe de 1953-1963, une grande partie du vocabulaire proprement ukrainien a été soit totalement supprimée, soit reléguée en deuxième position. Dans *le Dictionnaire de la langue ukrainienne* (1970-1980), les mots proprement ukrainiens étaient assortis des mentions « dépassé », « régional », « dialectal », « rare » pour limiter leur utilisation. Par ailleurs, ce dictionnaire contient des emprunts russes ou des unités lexicales formées grâce aux règles dérivationnelles de la langue russe.

Pendant cette période, quelques autres dictionnaires bilingues furent publiés (ukrainien-bélarussien, russe-moldave-ukrainien, anglais-ukrainien-anglais, français-ukrainien-français allemand-ukrainien, ukrainien-italien).

Словник української мови (СУМ-11) [Dictionnaire de la langue ukrainienne (SUM-11)], sous la rédaction d'Ivan Bilodid (1970-1980), en onze volumes, est le premier dictionnaire explicatif de la langue ukrainienne. Presque toutes les recherches en lexicographie ukrainienne se sont basées par la suite sur ce dictionnaire, qui donne des exemples d'emploi du vocabulaire en contexte (extraits d'œuvres littéraires).

Plusieurs grammaires furent également publiées à Kyïv pendant cette période. Les plus importantes sont :

- la grammaire académique en deux volumes, *Курс сучасної української літературної мови [Cours d'ukrainien littéraire moderne]* sous la rédaction de Leonid Bulakhovs'kyï (Kyïv, 1951) ; le premier volume est composé de cinq parties (introduction, lexique, phonétique, morphologie, accentuation) et le deuxième volume de trois parties (phrase simple, phrase composée, ponctuation) ;
- le *Курс сучасної української літературної мови [Cours d'ukrainien littéraire moderne]* de Mykhaylo Jovtobrioukh et Borys Koulyk en deux volumes (Kyïv, 1959) ; le premier volume est composé de cinq parties (introduction, vocabulaire et phraséologie, phonétique, écriture et orthographe, morphologie) et le deuxième volume de trois parties (syntagme, phrase simple, phrase composée) ;
- l'ouvrage académique collectif *Сучасна українська літературна мова [Langue littéraire ukrainienne moderne]* sous la rédaction d'Ivan Bilodid en cinq volumes (Kyïv, 1969-1973) ; il s'agit d'une description complète de la structure phonétique (volume 1) et grammaticale (volume 2), de la syntaxe (volume 3), de la composition lexicale (volume 4) et du système stylistique (volume 5) de la langue ukrainienne ; ce travail a été pendant très longtemps la référence pour toutes les recherches en linguistique ukrainienne ;
- l'ouvrage académique *Структура українського дієслова [Structure du verbe ukrainien]* de Vitaliy Roussaniv's'kyï (Kyïv, 1971) ; ce travail est l'une des premières tentatives visant à examiner le verbe d'un point de vue synchronique et diachronique ; ainsi sont étudiés les morphèmes verbaux et l'aspect des verbes ;
- l'ouvrage académique *Фразеологія української мови [Phraséologie de la langue ukrainienne]* de Laryssa Skrypnyk (1973) ; dans ce travail, l'auteur analyse différents types d'unités

phraséologiques, explique leurs relations systémiques et identifie les principales sources de la phraséologie ;

- *Іменник в українській мові [Le Substantif dans la langue ukrainienne]* d'Ivan Matviias (Kyïv, 1974) ; dans cet ouvrage, l'auteur étudie les catégories grammaticales du nom et sa déclinaison, il compare les normes littéraires et les variantes dialectales ; ce travail est précieux pour quiconque étudie les dialectes ukrainiens ;
- *Будова слова і словотвір [La structure des mots et la dérivation]* de Volodymyr Horpynytsch (1977) ; dans cet ouvrage, l'auteur étudie les morphèmes et les schémas dérivationnels ;
- l'ouvrage académique *Словотвір сучасної української літературної мови [La dérivation dans la langue ukrainienne littéraire moderne]* sous la rédaction de Mykhaylo Jovtobrioukh (1979) ; dans ce travail sont étudiés les types, les moyens et la sémantique des dérivations ;
- *Морфеміка української мови [Les Morphèmes de la langue ukrainienne]* d'Olena Bezpoïasko et Kateryna Horodens'ka (1987) ; dans cette monographie, sont étudiées les caractéristiques fonctionnelles et sémantiques des morphèmes dérivationnels et les tendances d'évolution du système dérivationnel ukrainien ;
- l'ouvrage académique *Частини мови в семантико-граматичному аспекті [Les Classes grammaticales dans leur aspect sémantique et grammatical]* d'Ivan Vykhovanets' (1988) ; dans l'ouvrage, sont étudiés les critères de classification des classes grammaticales, la description des classes et des catégories grammaticales, la valence sémantique et syntaxique.

1.3.8 L'ukrainien de 1991 à nos jours

La situation linguistique

Depuis l'indépendance, l'ukrainien s'utilise dans toutes les sphères de la vie : l'éducation, les sciences, la presse écrite, la télévision, la radio, la politique, l'économie, l'administration, la justice, l'internet, l'art et la culture. Par conséquent, la langue s'enrichit d'un nouveau vocabulaire, lié à divers domaines.

Notons aussi que le nombre de locuteurs natifs a augmenté, surtout depuis l'agression de la Russie contre l'Ukraine, car de nombreuses personnes refusent de parler la même langue que les envahisseurs. Cela suscite de nouveaux besoins en termes d'enseignement et de TAL.

Cependant, la qualité de la langue parlée s'est dégradée, surtout dans certaines régions. Cette dégradation est le résultat de décennies de négligence de l'ukrainien, ce qui a entraîné une méconnaissance de la langue par la population. D'autre part, pendant une longue période, l'ukrainien a été utilisé presque exclusivement comme langue parlée à la maison, dans laquelle s'invitaient de nombreux dialectismes et russismes.

L'utilisation en ukrainien de russismes et de normes grammaticales russes est à l'origine de l'apparition d'un hybride des deux langues, appelé le « sourjyk »⁷¹. Il s'agit d'une langue spontanée et non normalisée, apparue dans la situation de bilinguisme. En général, le « sourjyk » est qualifié de sociolecte, employé par des locuteurs non éduqués qui ne maîtrisent aucune des deux langues.

Après la libéralisation de la langue ukrainienne, cette langue hybride s'est retrouvée dans l'espace public, c'est-à-dire à la télévision, dans la presse et sur internet. Le « sourjyk » est apparu sur la scène (par exemple, Vierka Serdioutchka, diffusée largement à la télévision), dans la littérature

⁷¹ En ukr. : *суржик*, qui signifie « mélange », autres variantes de transcription : « sourjik » (transcription du russe), « surzhik » / « surzhyk » (tradition anglophone) et « suržyk » (translittération internationale scientifique).

(Artem Tchapy, Mykhaylo Brynykh, Serhiy Saïgon, Sofiia Androukhovytch, Yuriy Androukhovytch, Svitlana Pyrkalo, Oksana Zaboujko, etc.), au cinéma (à notre avis, l'exemple le plus parlant est *Колу падають дерева* [Quand les arbres tombent] de Maryssia Nikitiouk (2018), sur internet (blogues, forums, YouTube, presse en ligne).

Le « sourjyk » est beaucoup étudié depuis l'indépendance de l'Ukraine. Des articles paraissent fréquemment dans des éditions périodiques telles que *Мовознавство* [La Linguistique], *Мова і культура* [La Langue et la Culture], *Мова і суспільство* [La Langue et la Société], *Наука і освіта* [les Sciences et l'Éducation]. Deux études plus approfondies ont été également publiées (Serbens'ka 1994 [57], Massenko 2011 [40]). On compte aussi des publications consacrées au « sourjyk » en dehors de l'Ukraine (Bilaniuk 1997 [75], 2004 [76], 2010 [76]), (Flier 1998 [98]), (Sériot 2005 [134]), (Artyushkina 2022 [73]).

L'ouverture de l'Ukraine au monde a également entraîné un flux incontrôlé d'emprunts à des langues étrangères, notamment des anglicismes (Holoïoukh 2021 [15]).

L'expansion du « sourjyk » et l'apparition des anglicismes ont fait revenir les tendances puristes dans la linguistique ukrainienne. Les autres raisons du retour du purisme sont l'écartement du vocabulaire proprement ukrainien des dictionnaires et le remplacement des normes orthographiques et grammaticales traditionnelles par des normes qui ont rapproché l'ukrainien du russe pendant la période soviétique (Tatarenko 2008 [61], Marynenko 2018 [39]).

Ces tendances puristes se font sentir dans la lexicographie ukrainienne et dans la morphologie dérivationnelle, mais aussi dans les éditions des normes orthographiques récentes (Holoïoukh 2021 [15]).

Dans ce processus de réhabilitation de la langue ukrainienne, la diaspora a joué un rôle important. Il faut noter que la langue de la diaspora est restée préservée des russismes imposés par la linguistique soviétique, et qu'elle se fonde sur la variante galicienne, car lors des premières vagues⁷², les immigrés venaient principalement de l'Ukraine occidentale.

Après l'indépendance, les échanges entre l'Ukraine et la diaspora s'intensifiaient : six congrès mondiaux ukrainiens⁷³ furent organisés, et, en Ukraine, la littérature et la presse périodique de la diaspora commencèrent à circuler.

Ainsi, la langue de la diaspora a commencé à s'imposer en Ukraine. Les interactions entre l'ukrainien moderne et l'ukrainien de la diaspora ont bien été décrites dans (Taranenko 2013 [56]), (Kozatchevs'ka et Sydorenko 2013 [32]).

Par ailleurs, en feuilletant de récentes publications ukrainiennes en linguistique, nous avons remarqué que le nouveau vocabulaire, le « sourjyk » et la langue de diaspora suscitent un grand intérêt de la part des linguistes ukrainiens. Nous constatons aussi un regain d'intérêt certain pour les textes littéraires des XVIII^e et XIX^e siècles dans leurs langues d'origine.

⁷² Dans l'Encyclopédie d'histoire ukrainienne (Encyclopédie 2019 : *Українська діаспора*), les quatre vagues de l'immigration ukrainienne sont décrites. La première vague couvre la période allant des deux dernières décennies du XIX^e siècle jusqu'au début de la Première Guerre mondiale (1914) ; la deuxième, la période de l'entre-deux-guerres (1918-1939) ; la troisième, la période suivant la Seconde Guerre mondiale (elle a été la plus intense immédiatement après la guerre) ; la quatrième, celle des départs isolés du territoire de l'Ukraine dans les années 1970 et 1980.

⁷³ Cette information se trouve sur le site officiel du Congrès : <https://www.ukrainianworldcongress.org/ua/istoriya-sku>

Les orthographes

Dans les années 1990, quand les restrictions concernant la langue ukrainienne furent complètement levées, il devint nécessaire de revoir les normes orthographiques. Ainsi, dans les nouvelles éditions de 1990, les modifications suivantes ont été apportées : la lettre *ʹ* a été rétablie, le vocatif, qui était facultatif dans la version précédente, est devenu obligatoire, et quelques autres modifications ont été apportées, notamment au niveau de l'écriture des noms d'origine étrangère.

En 1993, l'orthographe de 1990 fut rééditée avec quelques légères corrections, par exemple l'utilisation des lettres *i* et *u* et dans les noms propres. Les règles subirent encore quelques modifications dans les rééditions de 1997, de 1999 (appelées « le projet »), 2000, 2003, 2007, 2012, 2015. On y a clarifié les règles d'écriture des mots composés, l'utilisation de la majuscule, l'utilisation de *-y* (*-ю*) au génitif des noms de la deuxième déclinaison, l'utilisation du vocatif, et les règles de ponctuation.

En 2019 est sortie une nouvelle norme, qui a repris certaines normes de l'orthographe de 1928. Il s'agit des modifications suivantes :

- l'écriture du mot *проект* [*projet*] et de ses dérivés avec *є* et non avec *e* ;
- l'introduction de la terminaison alternative pour les mots féminins en *-ть* et pour quelques autres mots féminins avec consonne finale ;
- les prénoms féminins en *-ь* ou en labiale sont devenus variables ;
- les règles de l'écriture de *нів-* avec des substantifs ont été modifiées ;
- quelques ajustements des écritures des noms empruntés et des mots

Les grammaires et les dictionnaires

Dans cette section, nous présentons une liste commentée des principaux travaux des dernières décennies (par ordre chronologique) que nous avons consultés pour préparer le module NooJ pour la langue ukrainienne.

1) *Граматика української мови. Морфологія* [*Grammaire de la langue ukrainienne. Morphologie*] d'Olena Bezpoïasko, Kateryna Horodens'ka, Vitaliy Roussaniv'skyi (Kyïv, 1993). Cette grammaire est composée de dix parties : la première décrit l'objet de la morphologie, les suivantes décrivent neuf classes grammaticales (le nom, l'adjectif, le numéral, le verbe, l'adverbe, la préposition, la conjonction, la particule, l'interjection).

2) *Граматика української мови. Синтаксис* [*Grammaire de la langue ukrainienne. Syntaxe*] d'Ivan Vykhovanets' (Kyïv, 1993). Dans cet ouvrage sont étudiés les différents types d'unités syntaxiques et leurs fonctions, les relations syntaxiques, les axes syntagmatiques et l'organisation de la phrase.

3) *Сучасна українська літературна мова* [*Langue littéraire ukrainienne moderne*] sous la rédaction d'Arnol'd Hrychtchenko (Kyïv, 1993). Cette grammaire est composée des chapitres suivants : la phonétique et la phonologie, l'orthoépique, le graphisme et l'orthographe, la lexicologie, la phraséologie, la lexicographie, le système dérivationnel et la morphologie de la langue ukrainienne.

4) *Сучасна українська літературна мова* [*Langue littéraire ukrainienne moderne*] sous la rédaction de Mariia Pliouchtch (Kyïv, 1994). Après l'introduction, dans laquelle sont expliquées les origines de l'ukrainien, son développement et les normes orthographiques et grammaticales en vigueur, les auteurs étudient la phonologie, la phonétique et l'orthoépique, l'orthographe, la lexicologie, la structure morphologique du mot, la dérivation, la grammaire morphologique, la syntaxe et la ponctuation.

5) *Сучасна українська літературна мова. Морфеміка. Словотвір. Морфологія [Langue littéraire ukrainienne moderne. Les Morphèmes. Les Dérivations. La Morphophonologie]* de Volodymyr Horpynytsch (Kyïv, 1999). Ce manuel traite des phénomènes morphémiques, morphologiques, morphophonologiques et dérivationnels de la langue ukrainienne. Il couvre certaines problématiques qui n'ont pas encore été analysées, ou qui ont été peu analysées, dans la linguistique ukrainienne (valence, morphonologie, synthèse dérivationnelle).

6) *Морфологія української мови: підручник [Morphologie de la langue ukrainienne: un manuel]* de Volodymyr Horpynytsch (Kyïv, 2004). Dans ce manuel, après une introduction sur la morphologie, l'auteur donne une description détaillée et argumentée de chaque classe grammaticale : le nom, l'adjectif, le numéral, le pronom, le verbe, l'adverbe, le « staninvyk » (la classe des unités qui expriment l'état), le modal, la préposition, la conjonction, la particule, l'interjection, la conversion.

7) La grammaire académique *Теоретична морфологія української мови [Morphologie théorique de la langue ukrainienne]* d'Ivan Vyhovanets' et Kateryna Horodens'ka (Kyïv, 2004). Dans cet ouvrage, les classes grammaticales, les catégories grammaticales et leurs valeurs sont étudiées à la fois d'un point de vue théorique (la définition des morphèmes, l'élaboration des méthodes de la classification, de la catégorisation, la description des principes de la conversion) et d'un point de vue descriptive (approche synchronique). La grammaire est composée de neuf chapitres : introduction à la morphologie théorique, le nom, l'adjectif, le numéral, les mots pronominaux, le verbe, l'adverbe, les mots-outils, les mots-phrases.

8) *Граматика української мови. Морфеміка. Словотвір. Морфологія [Grammaire de la langue ukrainienne. Les Morphèmes. Les Dérivations. La Morphologie]* de Mariia Pliouchtch (Kyïv, 2005). La grammaire est composée de trois chapitres. Dans le chapitre *Morphèmes* sont étudiés le radical, la base, les affixes. Dans le chapitre *Dérivations* sont étudiés les différents types de dérivation, puis les dérivations propres à chaque classe grammaticale. Dans le chapitre *Morphologie* sont étudiées les classes suivantes : le nom, l'adjectif, le numéral, le pronom, le verbe, l'adverbe, la préposition, la conjonction, la particule, l'interjection.

9) *Фразеологія сучасної української мови [Phraséologie de la langue ukrainienne]* de Viktor Oujtchenko (Louhansk, 2005). L'auteur étudie la sémantique et la structure des unités phraséologiques de la langue ukrainienne, leur formation, leur sémantique et l'étymologie, leur fonctionnement. Une attention particulière est accordée aux composantes proprement ukrainiennes mais aussi aux emprunts et aux innovations phraséologiques.

10) *Український синтаксис: теоретико-прикладний аспект [Syntaxe ukrainienne: aspects théoriques et applications]* d'Anatolii Zahnitko (Donetsk, 2009). L'ouvrage est composé de deux parties : théorique et pratique. Dans la partie théorique, sont étudiés l'ensemble des unités syntaxiques, les types de liens syntaxiques, les problèmes liés à l'analyse d'une phrase et d'un texte. Dans la partie pratique sont analysés des axes syntagmatiques et des schémas de phrases.

11) *Сучасна українська літературна мова: Лексикологія. Фонетика [Langue littéraire ukrainienne moderne: Lexicologie. Phonétique]* de A.K. Moyssiienko, O.V. Bas-Kononenko, V.V. Bondarenko (Kyïv, 2010). Le manuel couvre les questions les plus importantes de la lexicologie, de la phraséologie et de la phonétique de la langue ukrainienne. Dans le chapitre *Lexicologie* sont décrits les aspects suivants : le vocabulaire de la langue ukrainienne, les relations entre les lexèmes, la structure sémantique d'un mot, les types de dérivation, la polysémie, la classification des unités phraséologiques. Dans le chapitre *Phonétique*, sont décrites : la phonétique, la phonologie et la prosodie de la langue ukrainienne.

12) *Теоретична грамати́ка сучасної української мови. Морфоло́гія. Синтаксис [Grammaire théorique de la langue ukrainienne. La Morphologie. La Syntaxe]* d'Anatoliy Zahnitko (Donetsk, 2011). L'ouvrage est composé de deux parties. Dans la partie *Morphologie* sont étudiées les classes grammaticales, leurs catégories, les catégories grammaticales transversales, les relations catégorielles des classes grammaticales. Dans la partie *Syntaxe*, sont étudiés le syntagme, les types de phrases, leur classification, leur structure, les axes syntagmatiques, la prosodie de la phrase, le texte en tant qu'unité syntaxique, la ponctuation.

13) *Грамати́ка сучасної української літературної мови. Морфоло́гія [Grammaire de la langue littéraire ukrainienne moderne. La Morphologie]* d'Ivan Vykhovanets', Kateryna Horodens'ka, Anatoliy Zahnitko (Kyïv, 2017). Dans la première partie, *Questions fondamentales de la morphologie*, sont décrits les principes des classifications des classes grammaticales, les types de classes grammaticales, la conversion (changement de catégorie grammaticale), les notions de catégorie grammaticale, grammème, lexème, mot, forme fléchie, morphème, propriété grammaticale, et paradigme. Dans les chapitres suivants, les classes grammaticales (le nom, l'adjectif, le numéral, le verbe, l'adverbe, le mot-outil, le mot-phrase) sont étudiées en profondeur.

Les dictionnaires

Après l'indépendance de l'Ukraine, plusieurs dictionnaires interdits pendant la période soviétique ont été réédités, parmi lesquels *Російсько-український академічний словник [Le Dictionnaire académique russo-ukrainien]* d'Ahatanhel Kryms'kyu et Serhiy Iefremov et *Словарь української мови [Dictionnaire de la langue ukrainienne]* de Borys Hrintchenko. Ont été également réédités, dans la collection *Із словникової спадщини [Patrimoine lexicographique]*, plusieurs dictionnaires terminologiques de l'Académie nationale qui avaient été retirés ou n'avaient jamais pu être publiés⁷⁴.

Nombre de nouveaux dictionnaires ont également vu le jour, parmi lesquels des dictionnaires explicatifs, étymologiques, phraséologiques, orthographiques, terminologiques, bilingues, de synonymes, d'antonymes, de paronymes, d'épithètes, de toponymes, de néologismes, d'emprunts, de dialectes, d'argot, de la langue de tel ou tel poète.

Tatarenko (2013) [62] souligne l'absence manifeste de normes unifiées pour ces dictionnaires : certains d'entre eux ont été rédigés dans la continuité de la tradition soviétique, d'autres au contraire ont exclu tout le vocabulaire soviétique et même ont été rédigés selon les règles orthographiques de 1928.

Nous ne ferons pas la liste de ces dictionnaires car elle serait extrêmement longue, nous citerons seulement ceux que nous avons utilisés ou que nous pourrions utiliser éventuellement lors de nos futures recherches.

1) *Практичний словник синонімів української мови [Dictionnaire pratique des synonymes de la langue ukrainienne]* de Sviatoslav Karavans'kyu (Kyïv, 1998). La première édition contenait environ 15 000 groupes synonymiques, tandis que la cinquième (Lviv, 2014) en contient plus de 20 000. Le dictionnaire contient également des termes et des expressions de la langue ukrainienne qui étaient négligés, ainsi que des néologismes. Il utilise l'orthographe de 1928. Cet ouvrage est très intéressant pour étudier notamment la langue de la diaspora.

⁷⁴ Ces informations sont données ici : <https://encyclopedia.kyiv.ua/terminolohichni-slovnyky>

2) *Російсько-український словник складної лексики [Dictionnaire russo-ukrainien du lexique composé]* de Sviatoslav Karavans'kyj (Kyïv, 1998). Il contient environ 40 000 traductions de mots, de formes grammaticales et d'expressions idiomatiques russes. Pour la première fois dans l'histoire de la lexicographie ukrainienne, à chaque entrée verbale dans le dictionnaire on trouve une traduction du participe russe. En outre, le dictionnaire propose des traductions d'unités phraséologiques, de proverbes, de citations tirées d'œuvres littéraires.

3) Le dictionnaire académique *Словник афіксальних морфем української мови [Dictionnaire des affixes de la langue ukrainienne]* de Klymenko N.F., Karpilovs'ka E.A., Karpilovs'kyj V.S., Nedozyj T.I. (Kyïv, 1998). Il contient 817 affixes (145 préfixes et 672 suffixes) de la langue littéraire ukrainienne moderne. Les affixes sont classés par classes grammaticales. Ce dictionnaire nous a été utile lors de la préparation des paradigmes ou des grammaires dérivationnelles.

4) Le dictionnaire académique *Словник українських морфем [Dictionnaire des morphèmes ukrainiens]* de Laryssa Poliouha (Lviv, 2001). Il contient environ 40 000 mots de divers domaines, et chaque entrée du dictionnaire est divisée en morphèmes. Y sont soulignés des lexèmes formés de manière irrégulière, des alternances de voyelles et de consonnes. Nous avons utilisé cet ouvrage pour préparer les paradigmes flexionnels et dérivationnels.

5) Le dictionnaire académique *Граматичний словник української мови: сполучники [Dictionnaire grammatical de la langue ukrainienne: conjonctions]* de Kateryna Horodens'ka (Kherson, 2001). Il contient 328 conjonctions. Pour chaque entrée lexicale, l'auteur indique l'origine et la fonction et donne plusieurs exemples tirés de différents textes. Nous avons consulté ce dictionnaire lors de la préparation des grammaires de désambiguïsation.

6) Le dictionnaire académique *Словник фразеологізмів української мови [Dictionnaire des phraséologismes de la langue ukrainienne]* sous la rédaction de Vira Bilonjenko, Iryna Hnatiouk, Valentina Diatchuk, Nadiia Nerovnia, Tetiana Fedorenko (Kyïv, 2003). Il contient les 7 922 unités phraséologiques les plus courantes. Les auteurs fournissent une explication complète, les dérivations et des exemples pour chaque entrée du dictionnaire. Cet ouvrage serait utile pour établir la liste des collocations de la langue ukrainienne.

7) Le dictionnaire académique *Граматичний словник української літературної мови. Словозміна [Dictionnaire grammatical de la langue littéraire ukrainienne. Les paradigmes flexionnels]* de Kryts'ka V.I., Nedozyj T.I., Orlova L.V., Pouzdyrieva T.K., Romaniouk Y.V. (Kyïv, 2011). Il contient environ 140 000 mots. Pour chacun sont donnés les paradigmes et les schémas accentuels, qui sont décrits à la fin de l'ouvrage. Nous avons utilisé ce dictionnaire pour préparer les paradigmes pour NooJ.

8) Un nouveau dictionnaire, *Словник української мови (СУМ-20) [Dictionnaire de la langue ukrainienne (SUM-20⁷⁵)]*, est en cours de préparation par l'Académie des sciences depuis 2010 ; le 13^{ème} volume a été publié en 2022. Ce dictionnaire a pour ambition de couvrir toute la langue ukrainienne ; de ce fait il est très utile pour vérifier le vocabulaire absent des autres ressources.

Les dictionnaires en ligne

Depuis les deux dernières décennies, plusieurs dictionnaires ont été mis en ligne. Il est important de noter que, contrairement à un dictionnaire électronique, dans lequel l'utilisateur a accès à la base entière, qu'il peut télécharger et utiliser selon les modalités de la licence, un

⁷⁵ Il ne s'agit pas de la transcription de l'abréviation du nom en ukrainien, mais du nom officiel en latin.

dictionnaire en ligne permet de faire des recherches dans la base uniquement sur son site (onglet *Recherche*).

1) Ainsi, les 13 premiers volumes du dictionnaire *СУМ-20 [SUM-20]* sont disponibles sur le site sum20ua.com (© 2023). Ce dictionnaire couvre actuellement 127 763 entrées lexicales, soit les mêmes informations que sa version papier. Ce site est géré par la Fondation pour la langue et l'information ukrainiennes de l'Académie nationale des sciences d'Ukraine.

2) La Fondation gère également un autre site, *Український лінгвістичний портал [Portail linguistique ukrainien]* <https://lcorp.ulif.org.ua/dictua/> (© 2023). Le dictionnaire disponible sur ce site est composé de cinq modules :

- *Paradigme*, ce volet est basé sur *Орфографічний словник української мови [Dictionnaire orthographique de la langue ukrainienne]* (Kyïv, 2009) ;
- *Transcription (en anglais)*, ce volet est basé sur *Орфоепічний словник української мови [Dictionnaire orthoépique de la langue ukrainienne]* en deux volumes (Kyïv, 2001-2003) ;
- *Synonymie*, ce volet est basé sur *Словника синонімів української мови [Dictionnaire des synonymes de la langue ukrainienne]* en deux volumes (Kyïv, 1999-2000) ;
- *Antonymie*, ce volet est basé sur *Словника антонімів української мови [Dictionnaire des antonymes de la langue ukrainienne]* (Kyïv, 1999) ;
- *Phraséologie*, ce volet est basé sur deux dictionnaires : *Фразеологічний словник української мови [Dictionnaire des phraséologismes de la langue ukrainienne]* en deux volumes (Kyïv, 1999) et un autre qui porte le même nom, mais qui a été édité en 2003 à Kyïv.

Le portail couvre 261 000 entrées du dictionnaire.

3) Le dictionnaire *СУМ-11 [SUM-11] de 1970-1980* est disponible sur le site <http://www.inmo.org.ua/sum.html> (© 2023). Il est géré par l'Institut de linguistique Oleksandr Potebnia de l'Académie nationale des sciences d'Ukraine. L'utilisateur peut y trouver les mêmes articles que dans la version papier.

4) Un autre dictionnaire en ligne qui utilise *СУМ-11* est hébergé par le site <http://sum.in.ua/> (©2018). Dans la version en ligne, les noms des auteurs et les noms des éditions pour chaque citation ne sont pas abrégés, la lettre *r* a été intégrée, malgré son absence dans la version papier. Le dictionnaire propose également le volet *Dictionnaire libre*, complété par les utilisateurs. Il contient actuellement 507 entrées.

5) Une collection de nombreux dictionnaires (dont *СУМ-11*, *СУМ-20*, le *dictionnaire électronique ВЕСУМ [VESUM]* et autres) est accessible sur le site <https://goroh.pp.ua>, (© 2023). Le site est composé de douze volets :

- *Explicitations* (251 000 entrées, avec des exemples tirés des textes) ;
- *Paradigme* (441 000 entrées, y compris le paradigme « 0 » pour les entrées invariables) ;
- *Transcription* (330 000 entrées, avec leur transcription en alphabet cyrillique) ;
- *Rimes* (3 000 000 entrées) ;
- *Synonymie* (109 000 entrées) ;
- *Traductions* (ua-en, ua-dt, ua-ru, ru-ua, ua-ge) ;
- *Fréquence* (260 000 entrées) ;
- *Étymologie* (260 000 entrées) ;
- *Phraséologie* (260 000 entrées) ;
- *Morphèmes* (817 entrées) ;
- *Anagrammes* (absence d'informations sur le nombre d'entrées) ;
- *Utilisation* (explication des règles, du « sourjyk », des erreurs expliquées).

Points négatifs de ce dictionnaire : pour les volets *Transcription, Rimes, Traductions, Fréquence, Anagrammes*, les sources ne sont pas données et il n'y a aucune description de la méthodologie.

6) *Словарь української мови [Dictionnaire de la langue ukrainienne]* de Borys Hrinchenko (1908-1909) a été mis en ligne par Oleksandr Oksymets' sur le site <http://hrinchenko.com> (© 2010). Il contient 68 000 entrées de la première édition papier.

7) Il existe également un grand nombre de dictionnaires bilingues en ligne, parmi lesquels les plus intéressants sont r2u.org.ua (ua-ru-ua), (© 2020), et e2u.org.ua (ua-en-ua), (© 2019), car ils permettent de faire des recherches dans plusieurs dictionnaires papier référencés dans l'onglet *Recherche*, y compris des dictionnaires de la diaspora ou des dictionnaires édités au début du XX^e siècle. Ainsi, r2u.org.ua englobe 30 dictionnaires et e2u.org.ua, 19.

1.4 Conclusion du chapitre 1

Dans ce chapitre nous avons montré combien la langue ukrainienne a évolué au cours de l'histoire. Cette évolution a duré jusqu'à la période la plus récente. Depuis le XIX^e siècle, les normes orthographiques et grammaticales ont été modifiées à de multiples reprises. De ce fait, lire un livre écrit au XIX^e siècle ou même au début de XX^e siècle n'est pas évident pour un non-initié. De plus, pendant la période soviétique, de nombreux ouvrages ont été réécrits en ukrainien selon les normes orthographiques de 1946 ou de 1960. Cela a lissé les particularités de la langue ukrainienne. Actuellement, les chercheurs reprennent les textes originaux et examinent leurs lexiques ainsi que leurs particularités morphologiques et syntaxiques, pour revenir aux traditions ukrainiennes.

Mais est-il possible d'automatiser ces recherches ? Est-ce que NooJ pourrait analyser les textes anciens ? Les comparer aux textes récents ?

Nous avons constaté qu'il existe en ukrainien des variantes orthographiques et grammaticales. Et nous savons qu'une telle variété, si elle n'était pas prise en compte « à la base » par les logiciels TAL, poserait des problèmes lors de l'analyse automatique des textes, car un grand nombre d'unités lexicales ne seraient pas reconnues ou, éventuellement, seraient mal étiquetées, et de ce fait les statistiques construites sur un tel étiquetage seraient faussées.

C'est pourquoi il serait intéressant de voir dans quelle mesure NooJ pourrait résoudre ce problème. Pourrions-nous lier toutes les variantes au même lemme ? Sera-t-il possible de reconnaître la période ou l'appartenance d'un texte à partir de ses normes orthographiques particulières ?

Nous avons vu aussi que les normes de l'ukrainien ont été parfois transgressées au contact de la langue russe, ce qui a fait apparaître le « sourjyk ». Bien qu'il soit compliqué d'établir des règles propres au « sourjyk », les linguistes arrivent à repérer quelques tendances. De ce fait, serait-il envisageable de retracer ces transgressions avec NooJ ?

Et, si nous continuons sur la même lancée : peut-on formaliser aussi la langue de la diaspora ? Avons-nous les moyens de le faire ou est-ce un défi irréalisable ? Et si oui, peut-on « normaliser » les textes en langue de la diaspora ?

Nous pourrions tester ces hypothèses, puis mesurer l'importance de l'investissement et le volume de travail à accomplir dans le chapitre dédié au module ukrainien pour NooJ. Mais nous savons d'emblée que nous ne pourrions pas traiter tous ces sujets de manière exhaustive dans le cadre d'une thèse, car cela supposerait un travail approfondi sur chaque question et prendrait beaucoup trop de temps. Par ailleurs, à l'issue de nos tests, nous pourrions envisager d'y revenir plus tard et de mener à bien les recherches entamées.

Notons aussi que les normes orthographiques, les grammaires et les dictionnaires que nous avons passés en revue nous serviront de base pour préparer le module ukrainien pour NooJ ou pour mener nos futures recherches.

Chapitre 2. Outils de TAL au service des linguistes

2.1 Introduction

Bien que l'intérêt pour la langue ukrainienne ait fortement augmenté ces dernières années, elle reste peu décrite et formalisée dans le monde occidental. Les outils informatiques libres d'accès restent aussi peu nombreux et ne répondent pas forcément à tous les besoins que peut avoir un étudiant ou un chercheur en linguistique. La majorité des outils sont développés avec l'approche empirique et n'ont pas vraiment de bases linguistiques solides. De plus, la fiabilité des outils disponibles doit être remise en question, car ils contiennent trop d'erreurs.

De ce fait, la formalisation de l'ukrainien à travers un outil informatique fiable et motivé linguistiquement nous paraît un défi important.

Depuis les années 1950, de nombreux outils de TAL capables d'analyser des données textuelles ont vu le jour. Ces outils contiennent souvent des ressources pour plusieurs langues et peuvent combiner de multiples fonctionnalités :

- afficher la fréquence des unités (occurrences, lemmes, parties de discours) ;
- trouver des concordances (exemples d'utilisation en contexte) des unités uniques ou en collocation ;
- trouver des collocations dans des corpus parallèles ;
- faire des recherches par champ lexical (analyse sémantique) ;
- étiqueter le texte avec des informations (analyse morphologique, syntaxique, sémantique) ;
- intégrer ou créer des dictionnaires électroniques ;
- créer ses propres corpus ;
- générer un texte ou résumer des documents ;
- traduire automatiquement les mots ou les phrases ;
- reconnaître et générer la parole.

Un outil qui donne accès à toutes ces fonctionnalités et qui, de plus, est simple d'utilisation serait un outil idéal pour un linguiste. Il est également important de souligner que, pour qu'un outil de traitement de corpus fonctionne bien, il faut qu'il puisse cerner les unités textuelles, leurs formes grammaticales, leur rôle dans la phrase, leur sémantique, c'est-à-dire que la compréhension de l'unité textuelle doit s'approcher le plus possible de la lecture telle que peut la faire un humain. Or, pour cela, il faut que cet outil sache lever des ambiguïtés (car une occurrence textuelle peut être interprétée différemment dans des syntagmes différents), reconnaître les noms propres, les mots composés, maîtriser la flexion et la dérivation, établir les liens syntaxiques entre les mots à l'intérieur de la phrase ou même à l'intérieur d'un texte, repérer le champ lexical correctement, et regrouper les variantes orthographiques des mêmes mots. Toutes ces tâches sont au cœur de la recherche actuelle en TAL.

Il existe deux approches :

- empirique, qui s'appuie sur les données extraites de grands corpus. Cette approche est utilisée pour développer des outils stochastiques⁷⁶ ;

⁷⁶ Sous le nom d'outils stochastiques, nous entendons les outils statistiques, les outils probabilistes et les réseaux de neurones.

- formaliste, qui s'appuie sur les dictionnaires et des règles de grammaire construites par des linguistes. Cette approche est utilisée pour développer des outils formels.

Dans ce chapitre, nous allons donner un aperçu rapide de ces deux approches et des quelques outils qui en résultent.

Puis nous décrivons les possibilités qu'offre la plateforme NooJ⁷⁷ que nous avons sélectionnée pour formaliser la langue ukrainienne.

2.2 L'approche empirique et les outils stochastiques pour l'ukrainien

Comme nous l'avons mentionné dans l'introduction, les outils stochastiques sont développés selon une approche empirique. Dans les outils stochastiques, les connaissances sur la langue sont obtenues automatiquement à partir des corpus d'apprentissage qui ont été fournis au système. Ces corpus d'apprentissage ont été préalablement étiquetés, principalement à la main. Ils sont composés de multiples textes supposés représentatifs des relations qu'une unité textuelle peut avoir dans une phrase avec les autres unités, permettant ainsi de classer cette unité en lui attribuant une étiquette morphologique, syntaxique ou sémantique.

L'étiquetage automatique se fait ensuite à l'aide d'un modèle mathématique, entraîné sur le corpus d'apprentissage, puis validé par une vérification avec un second corpus. Le modèle établit des règles en fonction du contexte dans lequel l'unité a été recensée. Une unité textuelle est traitée par de tels outils comme une séquence d'octets. Le contexte est très souvent limité à quelques unités textuelles proches, car si le contexte est trop grand, les calculs sont irréalisables à grande échelle en raison de la complexité combinatoire. À cela s'ajoute le fait que les contextes d'usage ne peuvent pas être tous décrits à l'avance, du fait de la créativité des locuteurs (Poibeau 2014 [125]).

Pour attribuer une étiquette à chaque unité textuelle, les outils stochastiques procèdent d'abord au découpage du texte en séquences de formes graphiques (tokénisation⁷⁸, détection de limite de la phrase), puis ils les comparent aux données de leur base de connaissances en utilisant les processus de lemmatisation⁷⁹ et de racinisation⁸⁰.

Depuis la naissance de ces outils, leur architecture ainsi que les algorithmes qui sont utilisés pour traiter les données ont beaucoup évolué : nous sommes passés de simples modules statistiques et probabilistes, dont les résultats sont souvent peu satisfaisants, à des systèmes complexes et beaucoup plus performants et plus rapides, comme ceux qui utilisent des réseaux neuronaux.

Les outils issus de l'approche empirique sont réputés peu coûteux et rapides à mettre en place et, de ce fait, ils sont majoritaires sur le marché mondial des outils de TAL, ce qui explique un certain engouement pour ces outils : les grandes entreprises telles que Google, Amazon, Microsoft ou IBM investissent beaucoup dans les outils stochastiques. Pourtant, ces outils ne sont pas sans failles :

⁷⁷ NooJ est un environnement de développement linguistique gratuit et open source, développé par Max Silberstein en 2003. Les cadres théorique et méthodologique de NooJ sont décrits dans (Silberstein 2015 [135]). Les ressources linguistiques et le manuel d'utilisation sont disponibles pour le téléchargement sur le site officiel de l'association : <https://nooj.univ-fcomte.fr/>

⁷⁸ Processus qui segmente la phrase (sac de mots) en une liste d'items (*tokens* en anglais). On parle d'itémisation.

⁷⁹ Processus qui ramène une occurrence textuelle à sa forme du dictionnaire.

⁸⁰ Processus qui consiste à déterminer la racine, le préfixe et la fin du mot.

- ils ont besoin de grands corpus de données, or, certaines langues, y compris l'ukrainien, manquent de corpus étiquetés en libre accès, donc les entreprises souhaitant développer un tel outil doivent embaucher du personnel pour étiqueter d'abord leurs corpus d'apprentissage ;
- les corpus d'apprentissage en libre accès contiennent des erreurs au niveau de l'étiquetage, mais aussi des fautes d'orthographe, car souvent, ils étaient scannés, reconnus automatiquement, puis étiquetés par des personnes pas toujours qualifiées pour ce travail (Volokh et Neumann 2011 [150], Dickinson 2015 [92], Silberztein 2019 [137], 2024 [138]) ;
- ils présentent de fausses règles de grammaire basées sur le principe du contexte similaire que ces outils utilisent pour annoter les textes (Silberztein 2024 [138]) ;
- les jeux d'étiquettes sont limités (Silberztein 2024 [121]) ;
- les homonymes et les locutions ne sont pas traités correctement, car ce type de logiciel n'est pas capable de transformer correctement un syntagme (Silberztein 2024 [138]) ;
- les évaluateurs qui présentent les résultats de performance de ces outils ne sont souvent pas compétents en linguistique, aussi ne peuvent-ils pas réellement comprendre l'intérêt qu'un tel outil peut avoir pour un linguiste (Boitet 2024 [78]). Ainsi, un outil affichant 98 % de précision peut s'avérer inadéquat pour un utilisateur qui cherche à faire l'analyse d'un discours.

Certains outils de TAL qui s'appuient sur l'approche empirique permettent de traiter les textes en ukrainien. Nous examinerons dans cette partie les trois outils les plus utilisés : Sketch Engine, TreeTagger et RNNTagger. Bien que ces outils permettent une prise en main assez rapide et offrent de nombreuses possibilités, ils présentent des défauts qui, à notre avis, ne permettent pas de produire des résultats fiables. Le but de cette section est de repérer ces défauts pour appuyer notre volonté de concevoir le module ukrainien pour NooJ.

2.2.1 Sketch Engine

Description de l'outil

Sketch Engine est une plateforme pour la linguistique développée par Adam Kilgarriff, Pavel Rychly, Pavel Smrz et David Tugwell en 2004 [110]. Elle est disponible en ligne sur <https://app.sketchengine.eu>. La plateforme héberge plus de 700 corpus de textes dans plus de 100 langues. Son histoire ainsi que toute la documentation relative à son utilisation sont bien décrites sur le site. Elle permet de faire l'étiquetage morphologique et sémantique, l'analyse statistique, la recherche des mots selon différents critères (mot, lemme, phrase, suite de caractères), la recherche de collocations, la construction de graphiques, l'affichage de nuages de mots, la création de dictionnaires à partir des textes étiquetés, l'extraction de termes bilingues, l'affichage de mots-clés et des n-grammes⁸¹ les plus fréquents. L'outil est gratuit pendant 30 jours, puis il faut payer un abonnement annuel.

Le corpus ukrainien Ukrainian Web 2020 and 2014 (ukTenTen20) a été créé en 2014, puis complété en 2020. Il est constitué d'une grande collection de textes web, de genres variés. Il est possible de télécharger le corpus existant sur son ordinateur et de créer ses propres corpus. La description du corpus est donnée sur la page de l'outil.

⁸¹ Un n-gramme est un ensemble des éléments successifs dans un texte.

Nous trouvons également le lemme *калатайти* au lieu de *калатати*. Dans notre texte, le verbe a été utilisé à l’impératif singulier. Pour lemmatiser ce verbe, il fallait d’abord enlever la terminaison de l’impératif –*й*, puis ajouter la terminaison de l’infinif –*ти*.

Autre exemple, les noms propres affectifs *Павлусь* et *Настуся* ont été catégorisés comme venant des verbes (des lemmes verbaux) *павлитися* et *наститися*, et le nom masculin commun au pluriel *боєприпаси* a été lemmatisé comme verbe *боєприпасти*.

→ Erreurs d’étiquetage morphologique

Sketch Engine offre la possibilité de chercher des mots en contexte, puis d’afficher les étiquettes de ces mots en contexte. Après avoir fait quelques tests, nous avons remarqué un très grand nombre d’erreurs.

Sur la Figure 5, nous pouvons voir le texte que nous avons utilisé pour tester Sketch Engine. Ici, il y a 2 phrases composées de 45 mots.

Figure 5. Texte utilisé pour tester les performances des outils

Павлусь Малинка був гарний з лиця, мов чорнявий та кучерявий бог Аполлон. Настуся Самусівна була пишна на вроду, мов Афродіта, котра тільки що вихопилась з морської піни на хвилях, але ще не помолилася богу, не вмилась і не обтерла гаразд рушником морської піни на виду⁸⁴.

[Pavlous Malynka était aussi beau que le dieu Apollon aux cheveux noirs et bouclés. Nastoussia Samoussivna était d’une beauté éblouissante comme Aphrodite, qui venait d’émerger de l’écume des vagues, mais n’avait pas encore prié Dieu, ne s’était pas lavé le visage et ne l’avait pas bien essuyé avec une serviette.]

Dans la Figure 6, nous avons mis *en bleu* les lemmes erronés et *en rouge* les étiquettes erronées. Nous pouvons voir que 16 mots ont reçu des étiquettes morphologiques erronées et 8, des étiquettes avec des lemmes erronés.

Figure 6. Erreurs d’étiquetage morphologique dans Sketch Engine

Павлусь/*павлитися/Ncmsnn* Малинка/*малинк/Mlcmnsn* був/*бути/Vapis-sm* гарний/*гарний/Afpmsnf*
 з/з/*Spsg* лиця/*лиць/Ncnsn,/,/X* мов/*мова/Css* чорнявий/*чорнявий/Ao-msnf* та/*та/Ccs*
 кучерявий/*кучерявий/Ao-msnf* бог/*бог/Ncmsny* Аполлон/*аполлон/Ao-msns./,/X* Настуся/*наститися/Ao-*
fsns Самусівна/*самусівний/Ncfsnn* була/*бути/Vapis-sf* пишна/*пишний/Аp-fsns-ep* на/*на/Spsa*
 вроду/*врода/Ncfsan,/,/X* мов/*мова/Css* Афродіта/*афродіт/Ncfsnn,/,/X* котра/*котрий/Pr--f-sna*
 тільки/*тільки/Q* що/*що/Pr--nnsnn* вихопилась/*вихопитися/Vmpip3s* з/з/*Spsg* морської/*морський/Ao-fsgf*
 піни/*піна/Ncfsn* на/*на/Spsl* хвилях/*хвиля/Ncfpln,/,/X* але/*але/Ccs* ще/*ще/R* не/*не/Q*
 помолилася/*помолитися/Vmpip3s* богу/*бог/Ncfsan,/,/X* не/*не/Q* вмилась/*вмитися/Vmeis-sf* і/*і/Ccs*
 не/*не/Q* обтерла/*обтерти/Vmeis-sf* гаразд/*гаразд/Spsi* рушником/*рушник/Ncnsin*
 морської/*морський/Ao-fsgf* піни/*піна/Ncfsn* на/*на/Spsl* виду/*вид/Ncmsln*

→ Erreurs liées aux fautes d’orthographe

Généralement, les outils stochastiques sont programmés pour reconnaître tous les mots du texte traité. Alors par défaut, ils traitent les mots avec une orthographe erronée comme des mots corrects, c’est-à-dire qu’ils cherchent à leur attribuer une étiquette et/ou un lemme.

Pour vérifier cela, nous avons intégré dans notre corpus une phrase composée de 6 mots, dont 3 fautes d’orthographe :

⁸⁴ *Без пумтя [Les Inutiles] d’Ivan Nechouy-Levyts’kyu (1900).*

Сестра *читає* книгу, а бр~~я~~*т* *слихає*.

au lieu de :

Сестра *читає* книгу, а бр~~а~~*т* *слу*хає. [La sœur lit un livre et le frère écoute].

Les formes erronées en question n'existent pas, autrement dit ces mots ne veulent rien dire. Pourtant, Sketch Engine les a étiquetés de la façon suivante :

читає/V-рiрЗs-/ч~~ет~~ати/
бр~~я~~т/Ncmsnn/бр~~я~~/
слихає/V-рiрЗs-/сл~~их~~ати/

Donc, si nous avons voulu construire un dictionnaire à partir de ces données, nous y aurions intégré des entrées lexicales erronées, et si nous avons voulu faire des statistiques, elles auraient été faussées.

Notons qu'un outil formel calcule le lemme de chaque mot en consultant un dictionnaire construit avec soin et des règles morphologiques décrites précisément, et s'il ne le trouve pas, alors il l'identifie comme forme inconnue, comme c'est le cas pour NooJ.

→ Gestion des mots avec l'orthographe parallèle

Nous avons voulu également vérifier comment sont traités les mots avec des variantes orthographiques. Nous avons vu qu'ils ont été lemmatisés avec des lemmes différents. Ainsi, pour les variantes *грунт* – *грунт* [la terre]⁸⁵, au lieu d'avoir un seul lemme dans Sketch Engine, il y a deux lemmes indépendants. Cela fausse l'analyse de la fréquence, car la fréquence du lemme *грунт* est ainsi sous-évaluée.

→ Gestion des locutions

Toutes les langues ont des locutions, c'est-à-dire un groupe de mots qui constituent une seule unité lexicale avec son propre sens, et qui se comporte grammaticalement et syntaxiquement comme un mot unique. En ukrainien, par exemple, nombreux sont les adverbes ou les prépositions qui sont composés justement de plusieurs mots. Nous avons voulu tester si Sketch Engine est capable de distinguer ces locutions. Pour cela, nous avons pris deux locutions adverbiales, *кінець кінцем* [finalement] et *нога в ногу* [en même temps], et nous avons constaté que ces deux locutions ne sont pas reconnues (voir Figure 7).

Figure 7. Étiquetage des locutions *нога в ногу* et *кінець кінцем* dans Sketch Engine



Ici, la première locution est traitée comme la combinaison « nom au nominatif + préposition + nom à l'accusatif » et la seconde comme « nom au nominatif + nom à l'instrumental ».

⁸⁵ Dans les textes, on peut trouver « *грунт* » selon la norme orthographique en cours, et « *грунт* » selon les normes orthographiques de 1933 à 1990.

Nos appréciations

Nous arrivons à la conclusion que nous ne pouvons pas utiliser Sketch Engine pour l'ukrainien à cause des défauts suivants :

- erreurs dans l'étiquetage morphologique ;
- attribution de lemmes erronés ;
- étiquetage de mots erronés comme si c'étaient des mots existants ;
- absence de lien entre certaines variantes orthographiques et morphologiques ;
- non-traitement des locutions.

2.2.2 TreeTagger et RNNTagger

Étant donné que le modèle de RFTagger utilisé par Sketch Engine pour le module ukrainien n'est pas libre d'accès, nous avons voulu vérifier comment fonctionnent d'autres outils stochastiques développés par le même auteur, à savoir TreeTagger et RNNTagger. Ces deux outils sont très prisés dans le monde du TAL et ont des modèles ukrainiens libres d'accès.

Description des outils

TreeTagger est l'un des premiers outils développés par Helmut Schmid (1994 [130], 1995 [131]). Cet outil permet d'annoter chaque unité d'un texte avec des informations sur la classe grammaticale et le lemme. La version actuelle a des ressources pour 28 langues, y compris pour l'ukrainien. L'outil est disponible sur <https://www.cis.lmu.de/~schmid/tools/TreeTagger/>.

Il existe une interface en ligne pour 17 langues⁸⁶, mais l'ukrainien ne fait pas partie de cette liste. On peut travailler avec l'ukrainien en installant le module sous Linux. Les instructions sont données sur le site de l'outil.

TreeTagger utilise le jeu d'étiquettes Universal Dependencies PoS tagset. Après l'étiquetage du texte, il affiche le résultat en trois colonnes : forme fléchie, classe grammaticale, lemme. Il fait également une reconnaissance partielle des unités inconnues : il attribue l'étiquette « classe grammaticale » à toutes les unités, mais quand il ne trouve pas le lemme, il affiche l'étiquette <unknown>.

RNNTagger utilise des techniques plus récentes fondées sur l'utilisation de réseaux neuronaux. Il a été développé par Helmut Schmid en 2019 (Schmid 2019 [132]). RNNTagger permet d'annoter des textes avec des informations sur la classe grammaticale, les catégories morphosyntaxiques et les lemmes. Il supporte 30 langues modernes, y compris l'ukrainien, et 11 langues anciennes. L'outil est disponible sur <https://www.cis.lmu.de/~schmid/tools/RNNTagger/>.

Sur le site officiel, il est noté que, par comparaison avec TreeTagger, le nouvel outil calcule l'étiquetage avec une plus grande précision et lemmatise tous les mots. En revanche, le fichier de paramètres est plus important, il est plus lent, nécessite PyTorch, et les lemmes de mots inconnus sont devinés. De ce fait, le résultat n'est pas toujours fiable⁸⁷.

⁸⁶ L'application a été développée par le Centre de Traitement automatique du Langage (CENTAL), Université catholique de Louvain. Disponible sur <https://cental.uclouvain.be/treetagger>

⁸⁷ Voir <https://www.cis.uni-muenchen.de/~schmid/tools/RNNTagger> (Schmid 2019 [132]).

Dans RNNTagger, le jeu d'étiquettes Universal Dependencies PoS tagset est combiné avec les jeux MULTTEXT-East Ukrainian qui produisent des informations morphosyntaxiques très détaillées, comme pour Sketch Engine.

Les erreurs dans les outils

→ Erreurs de lemmatisation et d'étiquetage dans TreeTagger

Pour nos tests, nous avons utilisé le même texte, composé de 10 434 formes fléchies, que pour les tests sur Sketch Engine. Après avoir étiqueté le texte avec TreeTagger, nous avons relevé des anomalies.

Tout d'abord, nous avons détecté 315 erreurs d'étiquetage, puis nous avons remarqué que 557 formes fléchies n'ont pas été ramenées à leur lemme, alors qu'il s'agit de formes appartenant au vocabulaire courant. De plus, nous avons remarqué que parfois les formes fléchies d'un même lemme (qui par ailleurs n'a pas été déterminé) ont des étiquettes de « classe grammaticale » différentes. Nous pouvons voir cela dans le Tableau 1.

Tableau 1. Étiquettes POS pour *Настуся* dans TreeTagger

Forme fléchie	Classe grammaticale	Lemme
.	SENT (fin de phrase)	.
Настусю	PROPN (nom propre)	<unknown>
Настуся	VERB (verbe)	<unknown>
у	ADP (préposition)	у
Настусі	NOUN (nom commun)	<unknown>

L'auteur a probablement utilisé une liste de fins de mot dans laquelle *-ся* a été noté comme un suffixe verbal. Or, ici, nous avons la racine *Настус-* et des terminaisons de nom féminin *-ю* pour le vocatif, *-я* pour le nominatif et *-і* pour le génitif (ou locatif, ou datif). Ce type d'erreurs n'aurait pas été produit avec un outil formel qui utiliserait des règles décrites avec une grande précision (comme celui que nous présenterons plus loin).

Il est également surprenant que les formes fléchies d'un mot donné ne soient pas toujours ramenées au même lemme, comme on peut le voir dans les Tableaux 2 et 3.

Tableau 2. Lemmatisation de *Київщині* et *Київщини* dans Sketch Engine

Forme fléchie	Classe grammaticale	Lemme
і	CCONJ (conjonction)	І
на	ADP (préposition)	На
Київщині	PROPN (nom propre)	Київщина
Педагогів	NOUN (nom commun)	Педагог
Київщини	PROPN (nom propre)	<unknown>

Dans le Tableau 2, le lemme de *Київщині* (locatif précédé par la préposition *на*) a été correctement calculé, alors que le lemme de *Київщини* (génitif sg., ici précédé par un nom au génitif pl.) est incorrect. Nous pouvons supposer que le corpus d'entraînement n'était pas suffisamment grand et ne contenait pas cette combinaison, donc le mot a été seulement détecté comme nom propre (car il commence par une lettre majuscule).

Tableau 3. Lemmatisation de *Настусина* et *Настусині* dans Sketch Engine

Forme fléchie	Classe grammaticale	Lemme
Настусина	NOUN (nom commun)	Настусин
Настусині	ADJ (adjectif)	Настусиній

Dans le Tableau 3, il s'agit de deux formes du même mot : *Настусин*, qui par ailleurs est un adjectif d'appartenance, et non pas un nom comme il est noté dans le premier exemple. Dans le deuxième exemple, l'étiquette « classe grammaticale » a été produite correctement, mais pas le lemme, car *Настусиній* est aussi une forme fléchie du mot *Настусин*. En utilisant un outil qui intégrera (les informations tirées) des dictionnaires construits avec précision par des linguistes compétents, nous éviterons ce type d'erreurs. De plus, les outils formels offrent la possibilité de décrire les dérivations, ce qui est exclu avec un outil stochastique, car la dérivation amène des changements bien plus complexes, par exemple l'ajout de suffixes, de préfixes, l'alternance ou la disparition de lettres et, parfois, le changement de classe grammaticale. Dans notre exemple, les formes fléchies *Настусина*, *Настусині* doivent être associées au lemme *Настусин* ou, mieux encore, au lemme *Настуся*, car l'adjectif *Настусин* est dérivé du nom *Настуся*. Il s'agit d'un adjectif d'appartenance qui signifie *appartient à Nastoussia*.

Nous remarquons aussi d'autres erreurs d'étiquetage. Par exemple, la forme *нема* a été étiquetée comme verbe et lemmatisée *немати*. Or, le verbe *немати* n'existe pas. Il existe éventuellement le verbe *мати*, qui peut être précédé par la particule *не*. En réalité, le mot *нема* doit être étiqueté comme prédicat (cette étiquette n'existe pas dans le jeu d'étiquettes qui a été utilisé par l'outil) et être associé au lemme *нема*. Ce type d'erreurs, tout comme les erreurs précédentes, ne serait pas produit par un outil formel qui utiliserait des informations produites par des linguistes compétents.

→ Erreurs de lemmatisation et d'étiquetage dans RNNTagger

Puisque RNNTagger utilise des techniques beaucoup plus récentes, nous avons voulu vérifier s'il peut permettre d'obtenir de meilleurs résultats pour l'ukrainien. Après l'avoir testé avec notre texte, nous avons constaté que toutes les formes ont été lemmatisées. Parmi ces formes, 284 ont été étiquetées incorrectement, ce qui correspond à une précision de 97,8 %. Nous pouvons voir les erreurs les plus courantes dans le Tableau 4.

Tableau 4. Erreurs de lemmatisation de RNNTagger

	Forme fléchie	Lemme RNNTagger	Étiquette RNNTagger	Lemme et étiquette corrigés
1	Настуся	настися	verbe	Настуся [Nastoussia] , nom propre féminin
2	Ніцці	Ніцець	nom commun masculin	Ніцца [Nice] , nom propre féminin
3	кудою	куда	nom féminin	кудою [où] , adverbe ou conjonction
4	змалку	змалка	nom féminin	змалку [dès l'enfance] , adverbe
5	трубами-димарями	трубами-димар	nom masculin	труба-димар [cheminée] , nom masculin
6	шатрами	шатр	nom masculin	шатер [chapiteau] , nom masculin
7	льоду	льод	nom masculin	лід [glace] , nom masculin
8	матері	матеря	nom féminin	мати [mère] , nom féminin
9	царівен	царівен	nom masculin	царівна [princesse] , nom féminin
10	неня	нінь	nom masculin	неня [maman] , nom féminin
11	тільки	тільки	particule	тільки що [à l'instant] , adverbe (locution)
12	що	що	particule	тільки що [à l'instant] , adverbe (locution)

Ici, les deux premières formes fléchies sont des noms propres (un diminutif de prénom féminin ukrainien et un nom de ville française) que l'outil n'a pas pu reconnaître.

Les deux formes suivantes ont été analysées comme étant des occurrences de nom féminin, probablement parce que l’outil a détecté les fins de mots semblables aux terminaisons des noms féminins, bien qu’il n’existe pas de racine nominale ou verbale *куд-* et *з мал(к)-* pouvant justifier cet étiquetage.

La forme fléchie *трубами-димарями*, composée de deux parties indépendantes et variables, a été mal lemmatisée, car seule la deuxième partie a été ramenée à la racine.

La sixième et la septième forme ont été ramenées à la racine sans prendre en compte les phénomènes tels que l’alternance de voyelles *-e-/-i-* et la mobilité de la lettre *-e-*.

La forme *мамери* a été ramenée à la racine comme un nom féminin ordinaire, alors qu’il s’agit d’une exception.

La neuvième et la dixième forme ont été étiquetées comme des noms masculins alors qu’il s’agit de noms féminins : le premier nom a été utilisé dans le texte au génitif pluriel et, de ce fait, il n’a pas de terminaison, comme c’est aussi le cas des noms masculins au nominatif, tandis que le second nom a été utilisé dans le texte au nominatif et a la même terminaison que les noms masculins au génitif.

Les deux dernières formes ont été étiquetées comme des particules, alors qu’il s’agit d’un adverbe, *тільки що*, composé de deux mots. En effet, les mots *тільки* et *що*, traités séparément, peuvent être particules, adverbes ou conjonctions. Et même si on ne traitait pas les locutions, ici ces unités auraient dû être décrites comme des adverbes et non comme des particules.

Nous avons aussi étiqueté avec RNNTagger la même phrase que lorsque nous avons testé les performances de Sketch Engine, étiqueté par RFTagger (voir Figure 5). Nous avons constaté que les erreurs restent très nombreuses : sur 45 mots, 11 ont reçu une mauvaise étiquette. Nous pouvons voir le résultat de ce test dans le Tableau 5. Les erreurs sont marquées en rouge.

Tableau 5. Erreurs d’étiquetage dans RNNTagger

	Forme fléchie	Étiquette	Lemme
1	Малинка	Npmsgy	Малинко
2	був	Vap ^{is} -sm	Бути
3	чорнявий	Ao-msnf	Чорнявий
4	кучерявий	Ao-msnf	Кучерявий
5	Настуся	Ncfsnn	Настуся
6	Самусівна	Npmsgy	Самусівн
7	була	Vap ^{is} -sf	бути
8	тільки	Q	тільки
9	що	Q	що
10	морської	Ao-fsgf	морський
11	богу	Ncmsgn	біг

Nos appréciations

Nous pouvons conclure de ces tests que TreeTagger et RNNTager ne conviennent pas pour la langue ukrainienne, car tout comme la plateforme Sketch Engine, ils produisent trop d’erreurs :

- erreurs dans l’étiquetage morphologique ;
- attribution d’un lemme erroné ;
- erreurs dans l’étiquetage des noms propres ;
- erreurs dans l’étiquetage des noms composés, des noms avec alternance de voyelles ou voyelles mobiles, et des exceptions ;
- erreurs lors de la levée des ambiguïtés ;
- non-traitement des locutions.

2.3 L'approche formelle

Dans les systèmes qui utilisent l'approche formelle, les connaissances linguistiques ont été préalablement décrites sous la forme de règles (grammaires formelles⁸⁸), puis fournies à la machine qui les utilise pour analyser le texte et produire des structures de phrases conformes à la grammaire ou bien vérifier si une phrase appartient à un langage donné. Les équipes qui développent ces outils sont constituées d'informaticiens qui s'occupent de la partie technique et de linguistes qui analysent les phénomènes de la langue, construisent des dictionnaires et des règles pouvant formaliser tous les phénomènes relatifs à la morphologie et à la syntaxe d'une langue.

Une grammaire formelle se construit typiquement en relation avec un dictionnaire de lemmes. Les grammaires sont conçues pour décrire parfaitement une langue. Généralement, on veut qu'elles soient très lisibles par les linguistes, ce qui permet à ceux-ci de les modifier sans avoir des connaissances approfondies en informatique.

Les outils formels permettent de décrire non seulement une unité linguistique, sa classe grammaticale, sa sémantique, ses formes fléchies, mais aussi la structure d'un syntagme ou d'une phrase. Un analyseur morphosyntaxique formel est capable de faire un étiquetage détaillé et précis qui produit le lemme d'une unité textuelle ainsi que toutes ses propriétés grammaticales, syntaxiques ou sémantiques. Il est également capable d'analyser la dérivation, de faire la substitution orthographique et la désambiguïsation motivée.

L'approche formelle a été initiée par Noam Chomsky, qui a décrit dans ses travaux le principe des grammaires génératives (1956 [82], 1957 [83], 1958 [84], 1959a [85], 1959b [86], 1963 [87,88], 1968 [89], 1984 [90]).

« Une grammaire générative est un ensemble de règles de réécriture constituées de deux membres : chaque membre gauche se réécrit en un membre droit. Chaque grammaire générative contient un et un seul symbole de départ » (Silberztein 2015 [136]).

Une grammaire générative est composée de 4 éléments. Elle est généralement présentée ainsi :

$\langle \Sigma, N, R, S \rangle$

« Σ » dénote l'ensemble des symboles terminaux « T », c'est-à-dire des unités qui ne peuvent pas être subdivisées en éléments plus petits, par exemple un élément du vocabulaire.

« N » dénote l'ensemble des symboles non terminaux ou auxiliaires « AUX », qui correspondent aux étiquettes des classes grammaticales ou syntaxiques (par exemple NOM pour le nom ou GN pour le groupe nominal).

« R » dénote l'ensemble des règles de réécriture (R pour « rules »).

« S » dénote l'axiome, ou symbole de départ (S pour « sentence »).

Chomsky a décrit et hiérarchisé quatre types de grammaires génératives (Chomsky 1957 [83]). Chacun de ces types de grammaires permet de décrire un type de langage (c'est-à-dire certains phénomènes linguistiques) et peut être traité par un algorithme capable de vérifier si une séquence donnée appartient ou non à ce langage.

⁸⁸ « Une grammaire formelle est un ensemble de règles qui permet de reconnaître les phrases d'un langage et/ou de produire à volonté toutes les phrases du langage » (Silberztein 2015 [135]).

Les grammaires du **type 0** sont **non restreintes**. Dans une règle $\alpha \rightarrow \beta$, la chaîne α doit contenir au moins un symbole non terminal, et la chaîne β n'est pas restreinte. Les langages générés par ces grammaires peuvent être reconnus par des **machines de Turing**.

Les grammaires du **type 1** sont « non raccourcissantes », c'est-à-dire que, dans une règle $\alpha \rightarrow \beta$, la partie gauche n'est pas plus longue que la partie droite : $|\alpha| \leq |\beta|$. Avec cette définition, on ne peut pas générer la chaîne vide (notée ϵ), ce qui est très gênant. Pour pouvoir le faire, on permet d'avoir dans une grammaire de type 1 la règle $S \rightarrow \epsilon$, à condition que le symbole S n'apparaisse dans aucune partie droite β . Ces grammaires sont dites « de type 1 généralisé », et peuvent être traitées par des **automates linéaires bornés**. Malheureusement, il n'y a pas de structure arborescente canoniquement associée à la génération (ou à l'analyse) d'une chaîne associée à une grammaire G_1 de type 1. Mais on peut toujours construire à partir de G_1 une grammaire G_2 équivalente, dite « contextuelle », dont les règles sont de forme $\gamma A \delta \rightarrow \gamma \beta \delta$ (γ et δ sont les contextes gauche et droit et la règle noyau est $A \rightarrow \beta$), avec de même la possibilité d'inclure la règle $S \rightarrow \epsilon$, à condition cette fois-ci que le symbole S n'apparaisse dans aucun contexte et dans aucune partie droite β .

On peut noter une telle règle contextuelle $A \rightarrow \beta / \gamma _ \delta$, pour faire apparaître son noyau, correspondant à un sous-arbre $A (b_1, b_2, \dots, b_n)$, si $\beta = b_1 b_2, \dots, b_n$. Par exemple, on pourra définir en NooJ une règle GN Singulier \rightarrow Det N Singulier, notée aussi GN \rightarrow Det N / _Singulier ($\gamma = \epsilon$, $\delta =$ Singulier).

Les grammaires du **type 2** sont dites « indépendantes du contexte » ou « hors contexte ». Toutes les règles sont de forme $A \rightarrow \beta$, où A est un symbole non terminal. Il n'y a pas de « restrictions ϵ », car on peut toujours « normaliser » une grammaire G_2 de ce type en une grammaire équivalente G_2' « contextuelle sans contexte », c'est-à-dire pouvant contenir la règle $S \rightarrow \epsilon$, à condition que le symbole S n'apparaisse dans aucune partie droite β et que $\beta \neq \epsilon$ pour tout β . Ces grammaires peuvent être traitées par des **automates à pile**.

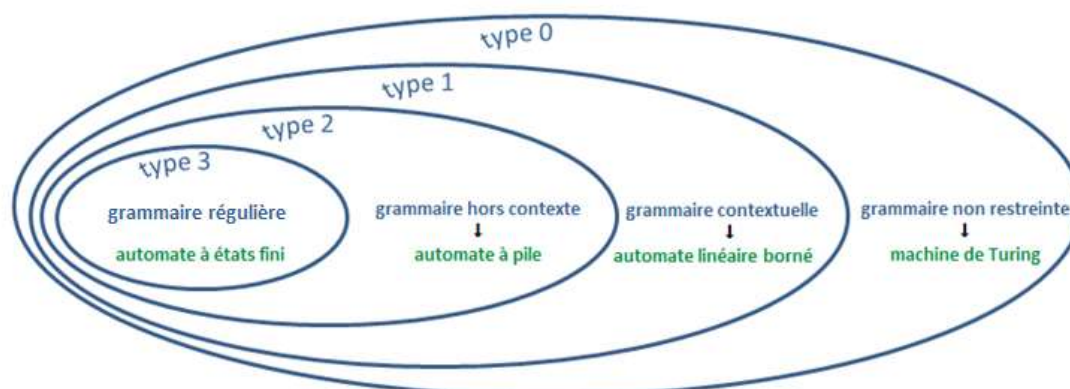
Les grammaires du **type 3** sont appelées grammaires **régulières**⁸⁹. Dans l'équation $A \rightarrow \beta$, le membre gauche n'est constitué que d'un symbole auxiliaire et le membre à droite peut contenir soit un seul symbole non terminal, soit un symbole terminal suivi d'un symbole non terminal, soit le mot vide. Ces grammaires peuvent être traitées par des **automates à états finis**.

Le type 0 englobe le type 1, qui englobe le type 2, qui, à son tour, englobe le type 3. De ce fait, toutes les machines capables de traiter les grammaires de type 0 seront capables de traiter aussi les grammaires de type 1, 2, 3. Les machines capables de traiter les grammaires de type 1 seront capables de traiter les grammaires de type 2 et 3. Les machines capables de traiter les grammaires de type 2 seront capables de traiter également les grammaires de type 3.

La hiérarchie des grammaires génératives est appelée aussi hiérarchie de Chomsky-Schützenberger. Marcel-Paul Schützenberger, connu pour ses travaux en informatique théorique et combinatoire, a aussi joué un rôle important dans le développement de la théorie des langages formels. Plusieurs théorèmes sur les langages formels portent les noms de ces deux scientifiques (Chomsky et Schützenberger 1963 [88], 1968 [89]). Cette hiérarchie peut être représentée par le schéma de la Figure 8. Nous donnerons des exemples de ces types de grammaires dans la section 2.5.2 (Les grammaires NooJ).

⁸⁹ Le terme original est « grammaire linéaire à droite », si les règles sont de forme $A \rightarrow b C$ ou $A \rightarrow b$ ou $A \rightarrow \epsilon$ (comme dit dans ce paragraphe), et « grammaire linéaire à gauche », si les règles sont de forme $A \rightarrow B c$ ou $A \rightarrow c$ ou $A \rightarrow \epsilon$.

Figure 8. Hiérarchie de Chomsky-Schützenberger



Notons aussi que les grammaires régulières sont plus simples à décrire, mais elles ne peuvent pas couvrir tous les phénomènes linguistiques, et que les grammaires non restreintes sont les plus puissantes. Ainsi, les phénomènes linguistiques qui ne pourront pas être décrits par les grammaires du type 3 devront être décrits par des grammaires du type 2, ou du type 1 ; si ce n'est pas possible, il faudra alors utiliser des grammaires du type 0.

Les outils linguistiques les plus connus qui s'appuient sur la hiérarchie de Chomsky-Schützenberger sont :

- Xerox Finite State (XFST) qui permet de construire des grammaires régulières ;
- Backus Naur Form (BNF), Yet Another Compiler Compiler (YACC), Generalized Phrase Structure Grammar (GPSG) qui permettent de construire des grammaires hors contexte ;
- Lexical Functional Grammar (LFG) qui permet de construire des grammaires contextuelles ;
- Head-Driven Phrase Structure Grammar (HPSG) qui permet de construire des grammaires non restreintes.

Chacun des formalismes cités plus haut permet la description d'un certain nombre de phénomènes linguistiques, mais si nous voulions décrire une langue naturelle en entier en utilisant un seul de ces dispositifs, nous serions vite limités. Pour bien faire, nous aurions besoin de les combiner, or techniquement ce n'est pas possible, car ils ne sont pas compatibles. Il faudrait donc utiliser une plateforme qui permette de décrire ces 4 types de grammaires et d'utiliser les algorithmes qui les traitent. De cette idée est née la plateforme NooJ (Silberstein 2015 [136]). Nous détaillerons la plateforme NooJ dans la section 2.5 après avoir donné un aperçu des outils formels existant pour l'ukrainien.

2.4 Outils formels pour l'ukrainien

En linguistique ukrainienne, quelques outils linguistiques issus de recherches formalistes ont gagné en notoriété. Nous commencerons par décrire VESUM⁹⁰, un grand dictionnaire électronique de la langue ukrainienne, conçu pour être utilisé en TAL, notamment pour être intégré à un correcteur orthographique ou encore à un analyseur textuel. Dans un deuxième temps, nous

⁹⁰ VESUM – Великий електронний словник української мови (ВЕСУМ) [Grand dictionnaire de la langue ukrainienne].

décrivons GRAC⁹¹, un grand corpus annoté en amont avec un outil formel. Enfin, nous parlerons de LanguageTool API NLP UK, un analyseur textuel.

2.4.1 VESUM

Description de l'outil

VESUM est un dictionnaire électronique qui inclut les formes fléchies. Il est le fruit du travail de Vassyl' Rysin et Andriy Starko. Dans sa version actuelle⁹², il compte plus de 418 000 lemmes, ce qui correspond à 6,5 millions de formes fléchies. Le dictionnaire inclut plus de 54 000 noms propres, des abréviations, des mots non standard (autour de 11 000) et il est mis à jour régulièrement. Pour créer ce dictionnaire, les auteurs ont utilisé plusieurs sources de référence en linguistique ukrainienne, dont deux dictionnaires papier, un dictionnaire en ligne, et deux grammaires. Le vocabulaire récent a été intégré aux dictionnaires grâce aux éditions spéciales de l'Institut de la langue ukrainienne NANU et aussi suite à l'ajout de mots repérés dans des corpus GRAC et BruK (Starko et Rysin 2020 [60]).

Les entrées du dictionnaire ont des étiquettes syntaxiques (classes et catégories grammaticales) et parfois sémantiques et stylistiques, par exemple l'étiquette « vulg » pour le langage vulgaire. Le dictionnaire contient également les codes des paradigmes morphologiques et des règles de génération des formes fléchies.

On peut avoir accès à ce dictionnaire via le portail <https://vesum.nlp.net.ua>, qui permet de faire une requête par lemme et d'accéder à toutes ses formes fléchies et étiquetées.

Les entrées de dictionnaire sont représentées de la façon suivante⁹³ :

близнюк /n20.a.p.ke.<

Il s'agit du lemme *близнюк* [*jumeau*] suivi du code du paradigme après le symbole « /: n20. » – nom masculin de la deuxième déclinaison, « a. » – terminaison *-a* au génitif singulier, « p. » – il existe des formes fléchies au pluriel, « ke. » - terminaison au vocatif *-e*, précédé d'une alternance de *-к-*.

Le paradigme est décrit dans le fichier « n2.aff ».

Après traitement, le mot *близнюк* produit les formes fléchies suivantes :

близнюк noun:anim:m:v_naz	
близнюка noun:anim:m:v_rod	близнюкові noun:anim:m:v_dav
близнюку noun:anim:m:v_dav	близнюка noun:anim:m:v_zna
близнюком noun:anim:m:v_oru	близнюкові noun:anim:m:v_mis
близнюку noun:anim:m:v_mis	близнюче noun:anim:m:v_kly
близнюки noun:anim:p:v_naz	близнюків noun:anim:p:v_rod
близнюкам noun:anim:p:v_dav	близнюків noun:anim:p:v_zna
близнюки noun:anim:p:v_zna:rare # з в/у	близнюками noun:anim:p:v_oru
близнюках noun:anim:p:v_mis	близнюки noun:anim:p:v_kly

Les formes sont étiquetées comme « nom animé ». Pour chaque forme, le cas est indiqué. Par ailleurs, seules les formes au singulier ont l'étiquette du genre et seules les formes au pluriel ont

⁹¹ GRAC – General regionally annotated corpus of Ukrainian.

⁹² À la date du 20/04/2023.

⁹³ Cet exemple est donné dans (Starko et Rysin 2020 [60]).

l'étiquette du nombre. L'étiquette « # з в/у » signifie que cette forme peut être utilisée après les prépositions з, в, у.

Certaines entrées ont aussi des étiquettes sémantiques ou morphologiques. Par exemple, l'entrée *проект* [*projet*] a une étiquette « ua_1992 » qui signifie qu'il s'agit d'une orthographe de 1992. Ou encore *Козак* [*Kozak*] a une étiquette « prop:geo # місто + річка » qui signifie qu'il s'agit d'un nom propre qui peut désigner un toponyme (soit une ville, soit une rivière) et une étiquette « lname » qui signifie qu'il s'agit d'un nom de famille.

Ce dictionnaire est utilisé par plusieurs outils de TAL, et notamment LanguageTool (correcteur et traducteur, analyseur morphologique), ispell-uk (module de correction d'orthographe sous Linux), BruK (corpus étiquetés), GRAC (corpus avec un module d'étiquetage automatique), pymorphy2 (analyseur morphologique), moteur de recherche Apache Lucene (pour Wikipedia depuis 2017). Nous nous intéresserons à certains d'entre eux plus bas.

En analysant le dictionnaire, nous constatons que le lien entre certaines formes des unités linguistiques n'est pas fait. Ainsi, les verbes perfectif (PF) et imperfectif (IPF) du couple aspectuel verbal, comme *мити* (IPF)/*помити* (PF) [*laver*], ne sont pas liés entre eux. Les formes réflexives dérivées avec le postfixe *-ся* ne sont pas reliées non plus aux verbes de base.

Les variantes orthographiques d'un mot ne sont pas non plus reliées. De même, les formes dérivées qui ont des entrées différentes, comme les noms de métier féminins dérivés des noms de métier au masculin, par exemple *лікар* (masculin) – *лікарка* (féminin) [*médecin*], ou encore les formes du comparatif ou du superlatif des adjectifs, dérivées des adjectifs qualitatifs, par exemple *гарний* – *гарніший* – *найгарніший* [*beau* – *plus beau* – *le plus beau*].

Nos appréciations

Nous regrettons que le lien entre les formes différentes ne soit pas fait dans le dictionnaire, ce qui pourrait être intéressant pour développer des outils d'apprentissage de langue assisté par ordinateur, car l'apprentissage de l'aspect est une des grandes difficultés des étudiants qui ne maîtrisent aucune langue slave. De plus, comme nous l'avons noté plus haut, le lien entre les formes dérivées et surtout entre les variantes orthographiques d'un même mot est nécessaire pour produire une bonne analyse morphosyntaxique.

Malgré ces quelques remarques, nous considérons que VESUM est un dictionnaire fiable et facile à prendre en main. Les auteurs ont publié plusieurs articles décrivant leur travail (Starko et Rysin 2020 [60], 2022 [141], Starko 2020 [139], Sychak et Starko 2022 [147]). Leur travail est décrit également en détail sur le site de référence.

2.4.2 GRAC

Description de l'outil

GRAC est une collection de textes ukrainiens composée de 130 000 textes de genres différents qui sont datés de 1816 à la période la plus récente. La version actuelle, V.17, contient 10 860 605 mots (6 661 169 lemmes)⁹⁴. L'étiquetage est basé sur les dictionnaires VESUM et des règles formelles (Shvedova 2020). Actuellement les ambiguïtés ne sont pas levées, mais cela fait partie du projet.

⁹⁴ Ces informations sont affichées à https://parasol.vmguest.uni-jena.de/grac_crystal/#dashboard?corpname=grac17.

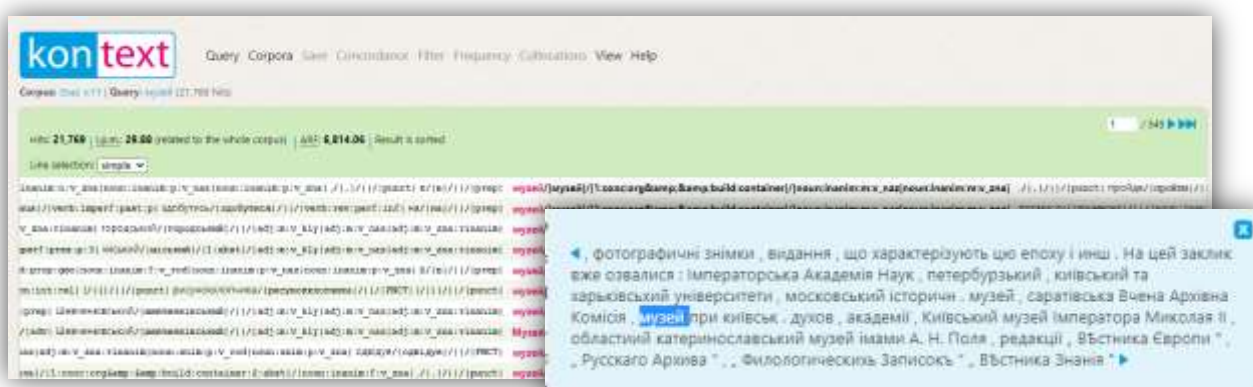
Les utilisateurs non-initiés au TAL peuvent explorer le corpus grâce à deux interfaces web :

- KonTex⁹⁵ donne accès aux versions 10 et 11 ;
- NoSketch Engine⁹⁶ donne accès aux versions de 10 à 17.

L'interface KonTex en mode « visiteur » permet de faire des recherches simples ou combinées par lemme, par forme fléchiée, par forme non complète, ou par étiquette (morphologique ou sémantique). La forme recherchée est affichée en contexte. En revanche, si l'on demande l'affichage de toutes les étiquettes, la lecture du résultat devient compliquée. Voir l'exemple sur la Figure 9.

Il est également possible de faire des recherches dans tout le corpus et dans des parties spécifiques du corpus. Pour avoir accès à des fonctions supplémentaires, il faut s'authentifier. Il n'est pas possible de s'enregistrer en ligne pour le moment.

Figure 9. Recherche du mot *музеї* [musée] dans GRAC-11, interface KonText



L'interface NoSketch Engine offre plus de possibilités car, en sus des mêmes recherches que celles que l'outil précédent propose, elle permet de faire des recherches par phrase, d'afficher la liste de fréquences, de faire de l'extraction de terminologie et de créer ses propres corpus à partir des corpus existants. De plus, la lecture des résultats est plus facile. Voir l'exemple dans les Figures 10 et 11.

Une autre interface donne accès à l'outil qui permet de construire des graphiques de fréquence basé sur GRAC⁹⁷. Elle donne la possibilité de construire des graphes de fréquences d'utilisation des mots par année, par auteur ou par œuvre. Les exemples sont décrits sur le site de GRAC⁹⁸.

⁹⁵ L'interface est accessible à <https://parasol.vmguest.uni-jena.de/grac/corpora/corplist>.

⁹⁶ L'interface est accessible à https://parasol.vmguest.uni-jena.de/grac_crystal/#dashboard?corpname=grac17.

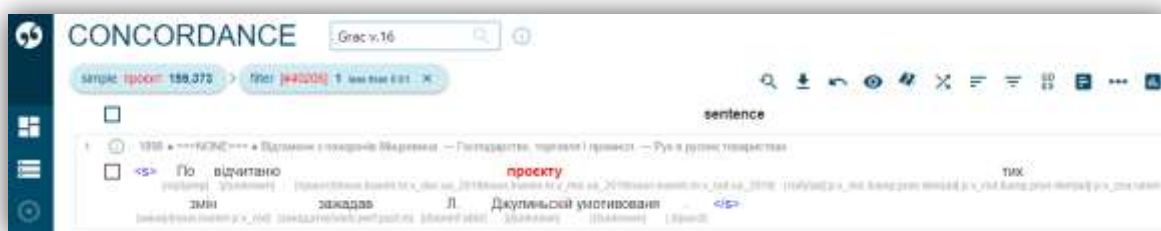
⁹⁷ Interface CQL (le langage de requête Cassandra Query Language) pour GRAC, auteur Tymofij Nikolajenko, accessible à https://parasol.vmguest.uni-jena.de/grac_batch/NGyears.html.

⁹⁸ Les exemples sont accessibles à <https://uacorpus.org/Kyiv/ua/poshuk-u-grak/grafiki-cql>.

Figure 10. Recherche du mot *проект* [projet] (orthographe de 1992) dans GRAC-16, interface NoSketch Engine



Figure 11. Recherche du mot *проект* [projet] (orthographe de 2019) dans GRAC-16, interface NoSketch Engine



Les auteurs de GRAC ont écrit plusieurs articles, décrivant en détail ses performances (Shvedova 2020 [143], Starko 2020 [139], Starko, Rysin, Shvedova 2021a [140], Shvedova, Rysin, Starko 2021b [144], Shvedova M., von Waldenfels R. 2021 [145], Starko et Rysin 2022 [141], Shvedova et alii 2022 [146]).

Nos appréciations

Nous considérons que malgré l'absence de levée des ambiguïtés et une lecture quelque peu compliquée, ces deux portails peuvent être appréciés à juste titre par les chercheurs. En revanche, ils ne seront peut-être pas suffisants pour les étudiants qui apprennent l'ukrainien et qui n'ont pas encore un bon niveau en langue. Nous trouvons également dommage que les variantes orthographiques des mêmes mots, ainsi que les formes dérivées, soient associées à des lemmes indépendants dans cet outil.

2.4.3 LanguageTool API NLP UK

Description du projet

Le projet LanguageTool a été initié en 2003 par Daniel Naber (Naber 2003 [119]). Il est composé de deux outils.

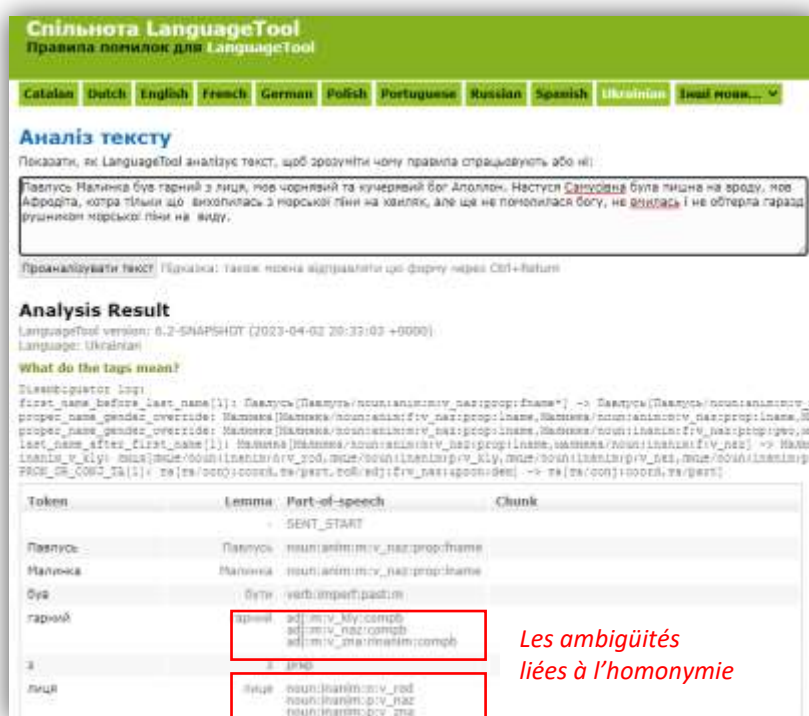
Le premier outil permet de créer un correcteur d'orthographe, de grammaire et de style. Il est disponible gratuitement en ligne à <https://languagetool.org/fr> et supporte actuellement 31 langues et quelques variantes régionales, notamment pour l'anglais ou l'espagnol. Pour la plupart des

languages, y compris l’ukrainien, la détection d’erreurs se fait à partir du dictionnaire et des règles⁹⁹. Toutefois, certaines langues utilisent également des bibliothèques de n-grammes¹⁰⁰.

Le deuxième outil permet de faire l’étiquetage morphologique et sémantique des textes et supporte 25 langues, dont l’ukrainien. Le module ukrainien a été préparé par Dmytro Tchaplyns’kyy, Andriy Rysin et Vasyi Strako. L’outil utilise le dictionnaire VESUM et des grammaires régulières qui sont traitées par des automates à états finis¹⁰¹.

Les utilisateurs initiés peuvent utiliser la version écrite en Groovy, les instructions d’installation et d’utilisation sont données sur la page de l’outil¹⁰². La version Groovy utilise un module automatique de levée des ambiguïtés liées à l’homonymie et de détection des mots inconnus, basé sur les statistiques du corpus BRUK¹⁰³ (les auteurs affirment que la précision de la levée des ambiguïtés est actuellement de ~94 %, et que la précision de détection des lemmes est de 99 %¹⁰⁴).

Figure 12. Analyseur de texte avec LanguageTool



Il existe également une interface web, présentée sur la Figure 12, pour des utilisateurs non-initiés, qui permet l’étiquetage d’un texte court (1 500 signes)¹⁰⁵. L’utilisateur peut également y consulter et tester les 1 133 règles décrites par les auteurs.

⁹⁹ Les exemples sont accessibles à <https://dev.languagetool.org/development-overview>

¹⁰⁰ Ces bibliothèques sont disponibles à <https://languagetool.org/download/ngram-data>

¹⁰¹ Pour plus de précision, voir <https://r2u.org.ua/articles/vesum>

¹⁰² La page de l’outil est https://github.com/brown-uk/nlp_uk/tree/master

¹⁰³ Le corpus Brown de langue ukrainienne disponible à <https://github.com/brown-uk/corpus>

¹⁰⁴ En l’absence de possibilité de tester l’outil personnellement, nous avons obtenu ces informations par nos échanges avec Andriy Rysin le 21 mars 2023.

¹⁰⁵ L’interface est accessible à <https://community.languagetool.org/analysis?lang=uk>

Cette version a quelques règles de désambiguïsation, notamment pour les noms propres, mais les ambiguïtés liées à l'homonymie des cas n'y sont pas levées.

Nos appréciations

Le point positif de l'analyseur morphosyntaxique LanguageTool en ligne est la rapidité de sa prise en main et la possibilité d'analyser ses propres textes. Toutefois, il ne peut être utile que ponctuellement, car il ne lève pas les ambiguïtés et ne permet pas d'analyser de grands textes, donc il serait très laborieux de l'utiliser pour faire des études de textes longs.

Quant à la version Groovy, elle est compliquée à mettre en œuvre, et donc seules les personnes formées seront capables de l'utiliser.

2.5 La plateforme NooJ

NooJ est une plateforme linguistique basée sur l'approche formelle. Actuellement il existe des ressources libres d'accès pour 32 langues¹⁰⁶.

La plateforme a été développée par Max Silberztein spécialement pour la linguistique descriptive. L'auteur a fait le choix de ne pas utiliser un seul formalisme mais d'en développer plusieurs, qui sont capables de décrire de nombreux phénomènes linguistiques. Ainsi, contrairement aux autres outils formels que nous avons présentés dans la section 2.3, avec NooJ on peut construire des grammaires de tous les types de la hiérarchie de Chomsky-Schützenberger, c'est-à-dire des grammaires régulières, hors contexte, contextuelles et non restreintes, qui sont interconnectées (chacune peut inclure les éléments d'une autre). De plus, NooJ ajoute à ces grammaires un certain nombre de mécanismes spécifiques pour faciliter la description linguistique. De ce fait, la plateforme permet de décrire aisément des phénomènes orthographiques, morphologiques (flexionnels, dérivationnels et agglutinatifs), syntaxiques (locaux, en constituants) et sémantiques (calculs prédicatifs et transformationnels).

NooJ propose une architecture « en cascade »¹⁰⁷, c'est-à-dire que le système est composé de plusieurs modules qui, tout en communiquant entre eux, traitent le texte niveau par niveau en commençant par le niveau alphabétique et en allant jusqu'à l'analyse sémantique. L'auteur a décrit la méthodologie en détail dans (Silberztein 2015 [136]).

NooJ contient aussi des outils de développement pour créer et partager des ressources linguistiques de grande taille, tels que des dictionnaires électroniques et des corpus annotés. La plateforme contient également des outils pour produire des concordances, des textes annotés, ainsi que diverses statistiques.

Le moteur de NooJ peut être intégré à des applications de TAL, telles que des outils de conjugaison et de déclinaison, des extracteurs d'entités nommées, des traducteurs automatiques, des générateurs de textes, des paraphraseurs, etc.

2.5.1 Les dictionnaires NooJ

La plateforme NooJ permet de décrire l'ensemble du vocabulaire d'une langue naturelle. Selon la méthodologie NooJ, le vocabulaire est composé d'un nombre fini d'unités linguistiques atomiques

¹⁰⁶ Les ressources sont disponibles sur le site <https://nooj.univ-fcomte.fr/resources.html>

¹⁰⁷ Terme utilisé par Max Silberztein dans (Silberztein 2015 [135]).

(ALU) que l'on peut classer en quatre catégories : les affixes, les mots simples, les mots composés et les expressions (Silberztein 2015 [136]). Ce principe permet de classer chaque élément lexical d'une langue naturelle, et donc de décrire son vocabulaire formellement. Certaines ALU seront rangées dans le dictionnaire, tandis que d'autres seront représentées par des grammaires morphologiques ou syntaxiques, ce que nous verrons dans la section suivante.

Dans un dictionnaire NooJ, les lemmes sont associés à un certain nombre de propriétés morphosyntaxiques (par exemple « NOUN », « feminine », « inanimate ») et sémantiques (par exemple « Toponyme »). L'utilisateur peut créer des propriétés par lui-même, en les déclarant dans le fichier « properties.def » qui définit les codes (les catégories, leurs propriétés et les valeurs de ces propriétés, les descriptions sémantiques). Ces propriétés pourront ensuite être utilisées aussi bien dans le dictionnaire que dans les grammaires NooJ. Ainsi, l'utilisateur est libre de décrire les catégories grammaticales par des abréviations (« V » pour verbe) ou des noms entiers (« VERB »), il peut également créer des propriétés particulières, par exemple « Couple= » pour décrire le couple aspectuel du verbe.

Certaines entrées d'un dictionnaire NooJ peuvent être associées à un certain nombre de paradigmes morphologiques, de deux types, « FLX » et « DRV ».

Les paradigmes du type « FLX » permettent de relier une entrée lexicale à un ensemble de formes fléchies. Par exemple, pour le verbe *enseigner*, le paradigme « FLX » permet la description de toutes ses formes conjuguées : *enseigne, enseignes, enseignons, enseignez, enseignent, etc.*

Les paradigmes du type « DRV » permettent de relier une entrée à une forme dérivée, ce qui implique des modifications plus profondes, comme le changement de la classe flexionnelle ou le changement de la catégorie grammaticale. Par exemple, pour le verbe *enseigner*, un paradigme « DRV » pourrait permettre de décrire sa forme nominale *enseignant* ; et un autre paradigme « DRV » décrirait sa forme nominale *enseignement*.

Prenons pour exemple une entrée du dictionnaire du module ukrainien pour NooJ :

```
читати,VERB+Couple="читати/прочитати"+FLX=ЧИТАТИ+DRV=ПРО:ЧИТАТИ_PF
```

Ici l'entrée *читати* [*lire*] est un verbe qui fait partie du couple aspectuel *читати/прочитати* qui se décline selon le modèle flexionnel « ЧИТАТИ », a une forme dérivée décrite par le modèle « ПРО », et cette forme dérivée elle-même se décline selon le modèle flexionnel « ЧИТАТИ _PF ».

NooJ permet aussi de décrire les variantes orthographiques. Dans toutes les langues, certains mots peuvent avoir des orthographes parallèles, comme *clé* et *clef*, ou *île* et *isle* en français. De même manière, il existe des mots avec orthographes parallèles en ukrainien, comme *проект* et *проект* [*projet*] que nous avons déjà cité précédemment. Dans NooJ, l'utilisateur a la possibilité de relier directement dans les dictionnaires comme ceci :

```
проект,проект,NOUN+Masculine+Common+Inanimate+FLX=ЛИСТ
```

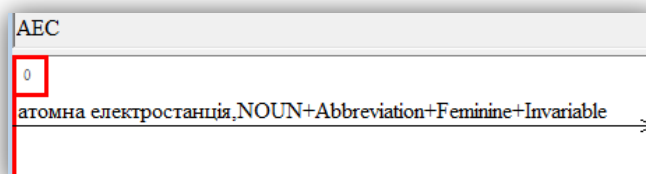
Cette entrée représente le fait que les formes fléchies de l'entrée *проект* seront étiquetées en tant que formes fléchies du lemme *проект*, et lors de la recherche des occurrences de « <проект> » toutes les formes fléchies, aussi bien celles avec « е » que celles avec « е », seront trouvées.

Nous pouvons également de cette façon décrire les abréviations :

```
АЕС,атомна електростанція,NOUN+Abbreviation+Feminine+Invariable
```

Ainsi, les deux formes sont reliées lors d'une recherche. De plus, les personnes qui travailleront avec les textes pourront facilement comprendre les sens de l'abréviation (voir Figure 13).

Figure 13. Annotation de l'abréviation AEC par NooJ



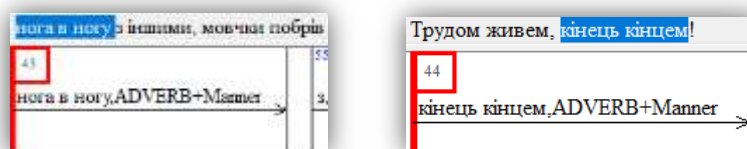
NooJ permet également de décrire l'orthographe des mots composés pouvant s'écrire en un seul mot, avec un trait d'union ou en deux mots. Par exemple, en ukrainien, avant 2019, le mot *мініспідниця* [*mini-jupe*] s'écrivait avec un trait d'union : *міні-спідниця*, alors que maintenant, on doit l'écrire en un seul mot. Dans les textes actuels, on peut trouver ces deux formes, d'autant plus que le correcteur d'orthographe de Word n'est pas mis à jour, les gens continuent donc d'utiliser l'orthographe antérieure. Dans NooJ, nous pouvons lier ces formes en ajoutant le caractère spécial « _ » :

міні_спідниця,NOUN+Feminine+Inanimate+FLX=ГУЛЯ+UNAMB

La propriété UNAMB indique que les deux constituants du mot composé ne doivent être reconnus indépendamment.

La propriété UNAMB est également utile pour décrire les locutions comme les adverbes *кінець кінцем* [*finalement*] ou encore *нога в ногу* [*en même temps*] mentionnés dans la section 2.2. Ainsi, en associant ces entrées à cette propriété, NooJ les traitera comme une seule entité, et non comme une séquence de mots. Nous pouvons voir le résultat dans la Figure 14.

Figure 14. Annotation des locutions *кінець кінцем* et *нога в ногу* par NooJ



Notons aussi que les dictionnaires NooJ peuvent être exportés dans le format .txt (avec les formes fléchies ou sans) et, de ce fait, pourront être réutilisés par diverses applications de TAL.

2.5.2 Les grammaires NooJ

NooJ contient un éditeur textuel et un éditeur graphique permettant de construire des grammaires de trois niveaux linguistiques :

- des grammaires flexionnelles et dérivationnelles, qui ont pour but de décrire les formes d'un lemme ;
- des grammaires morphologiques, qui reconnaissent des séquences des lettres et produisent des lemmes absents du dictionnaire et leurs formes ;

- des grammaires syntaxiques qui reconnaissent des séquences de mots, y compris des expressions figées et des entités nommées. Ces grammaires sont aussi utilisées pour lever les ambiguïtés et produire des transformations de phrases¹⁰⁸.

Pour chacun de ces trois niveaux linguistiques, les utilisateurs peuvent construire les quatre types de grammaires de la hiérarchie de Chomsky-Schützenberger (voir Figure 8).

Grammaires régulières

Pour formaliser de nombreux phénomènes linguistiques, on peut utiliser des grammaires régulières.

Sur la Figure 15, nous donnons des expressions régulières linéaires, écrites selon les normes PERL, qui décrivent un mot et ses variantes (1), les formes fléchies d'un mot (2), un terme et ses variantes (3).

Figure 15. Expressions régulières (PERL)

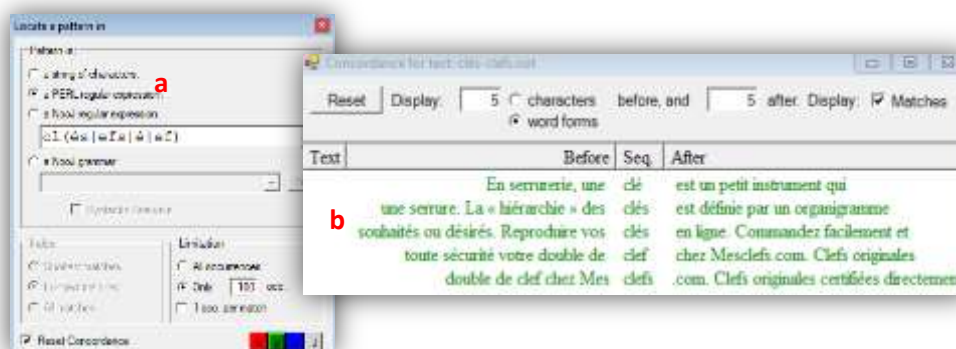
```
(1) cl(é|ef)
(2) enseign(er|e|es|ons|ez|ent)
(3) carte(bancaire|bancaire|bleue|visa|master)
```

Les grammaires régulières peuvent être enrichies par des productions qu'on ajoute après le symbole « / » comme ceci :

cl(é|ef)/NOM

Dans NooJ, les expressions régulières présentées en Figure 15 peuvent être utilisées pour faire les recherches dans le texte (voir Figure 16). Dans notre exemple, nous cherchons toutes les formes du lemme *clé*.

Figure 16. Recherche par expression régulière (a) dans NooJ et résultat de recherche (b)



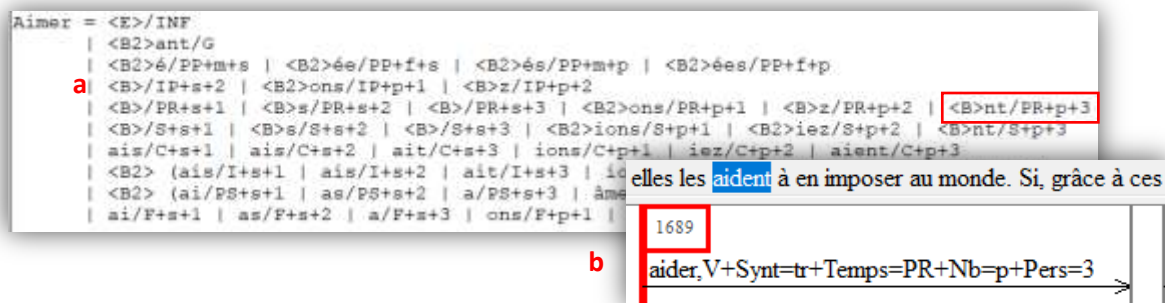
NooJ offre aussi la possibilité de construire des grammaires qui contiennent des expressions régulières à production, c'est-à-dire d'associer aux formes reconnues des analyses qui seront produites lors de l'application de ces grammaires. Ainsi, par exemple, ce type de grammaires est utilisé pour construire les grammaires flexionnelles « FLX » et dérivationnelles « DRV ».

¹⁰⁸ La possibilité d'écrire des transformations a une application pratique pour la traduction, pour la construction de correcteurs d'orthographe, ou de paraphraseurs, de chatbots et d'outils similaires.

Sur la Figure 17 (a), nous pouvons voir une grammaire « FLX » qui décrit la conjugaison des verbes du premier groupe pour le français. L'opérateur indique le nombre de caractères à remplacer par les caractères qui suivent. Après « / » sont notées les propriétés pour la catégorie « V » (verbe), étiquettes encodées pour le français dans le fichier « properties.def ».

En associant cette grammaire à une entrée du dictionnaire de la façon suivante : aider,V+FLX=Aimer, toutes les formes fléchies décrites dans la grammaire seront reconnues dans le texte (Figure 17 (b)).

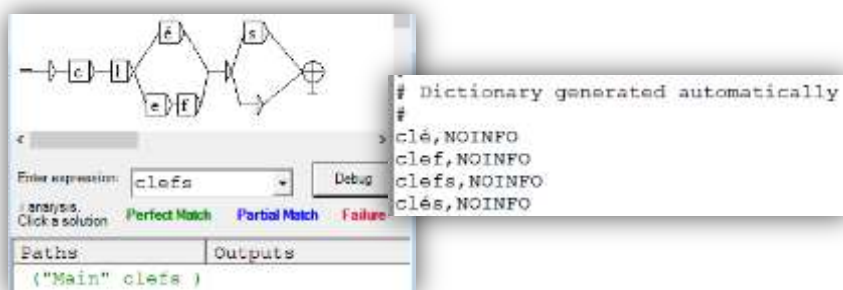
Figure 17. Grammaire flexionnelle (a) dans NooJ et annotation (b) dans le texte qui en résulte



Graphes

Grâce à son éditeur graphique, NooJ permet aussi de décrire des grammaires sous forme graphique¹⁰⁹, comme sur la Figure 18¹¹⁰. À l'aide de ce graphe, les formes *clé*, *clef*, *clés*, *clefs* seront reconnues dans le texte.

Figure 18. Graphe à états finis dans NooJ



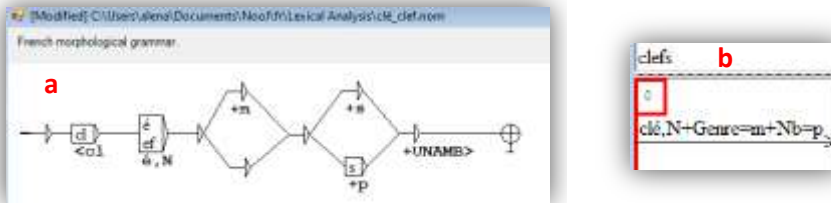
NooJ permet de construire des graphes à production¹¹¹. Tout comme pour les grammaires régulières à production, ces graphes permettent d'associer les formes reconnues à des annotations qui représentent leurs analyses. Les machines qui permettent de traiter des grammaires régulières à production et des graphes à production sont des transducteurs à états finis.

¹⁰⁹ Un graphe à l'état fini décrit la transformation d'un ou plusieurs états initiaux (désignés par la flèche entrante) à l'un ou plusieurs états finaux.

¹¹⁰ Dans NooJ, les graphes à état fini produisent des étiquettes vides.

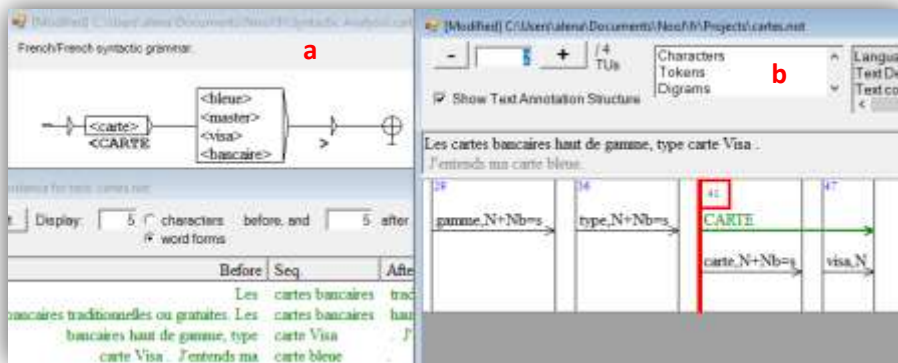
¹¹¹ Un transducteur d'états finis décrit une transformation à partir d'un ou plusieurs états initiaux en un ou plusieurs états finaux et ajoute des informations à la sortie.

Figure 19. Graphe morphologique à production (a) dans NooJ et annotation (b) dans le texte qui en résulte



Sur la Figure 19, nous pouvons voir un transducteur à états finis qui décrit les variantes orthographiques et les formes fléchies du lemme *clé*. En sortie, cette grammaire attribuera les annotations suivantes : le lemme *clé*, la catégorie « N » (pour le « nom »), les propriétés « m » (pour le « masculin ») et « s » (pour le « singulier ») ou p (pour le « pluriel »).

Figure 20. Graphe syntaxique à production (a) dans NooJ et annotation (b) dans le texte qui en résulte



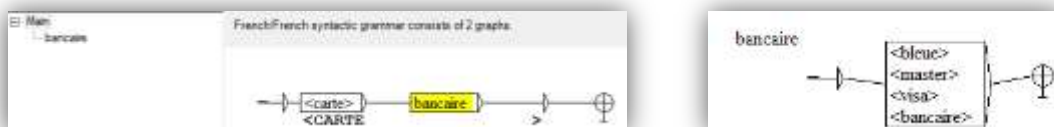
Un autre exemple de grammaire à production est présenté dans la Figure 20. Ici il s'agit d'un graphe syntaxique équivalent à la grammaire syntaxique régulière (3) de la Figure 15.

Les crochets « < > » signifient qu'il faut prendre en compte toutes les formes associées à une entrée lexicale. En appliquant cette grammaire syntaxique à un corpus, nous pouvons retrouver toutes les formes décrites dans la grammaire et les annoter « CARTE ».

Graphes imbriqués

NooJ permet également de construire des grammaires hors-contexte. Ces grammaires peuvent être représentées par des graphes qui contiennent des graphes imbriqués, comme dans la Figure 21. Les références à des graphes imbriqués sont colorées en jaune par défaut.

Figure 21. Grammaire avec des graphes imbriqués



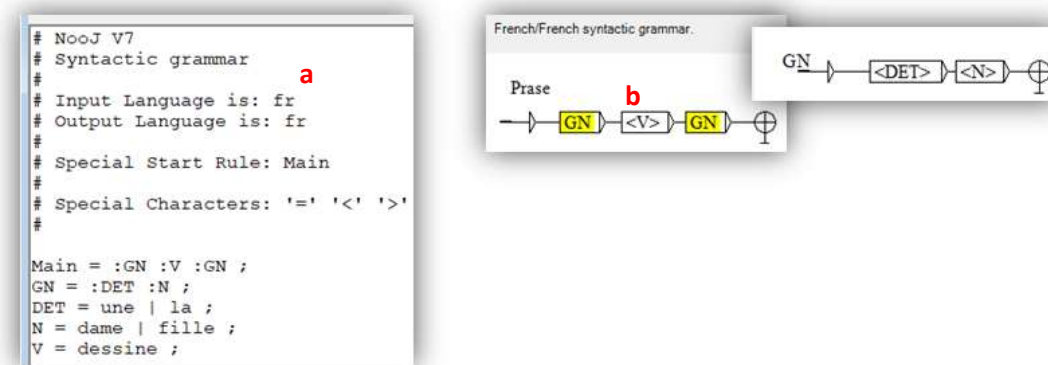
Dans cet exemple, la grammaire contient deux graphes. Le graphe principal est composé de deux nœuds, `<carte>` et `bancaire`, tandis que le dernier graphe décrit les variantes du deuxième composant.

Les grammaires hors-contexte (et leur représentation graphique) sont particulièrement utiles lorsqu'il faut décrire des expressions complexes, par exemple les dates ou l'âge.

Grammaires hors-contexte

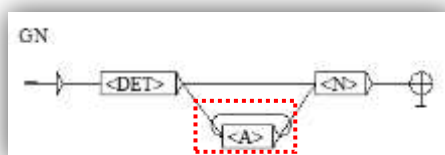
Comme pour les grammaires régulières, NooJ permet de construire des grammaires hors-contexte soit sous forme de graphes, soit sous forme textuelle (voir Figure 22).

Figure 22. Grammaires hors contexte dans NooJ : éditeur de règles (a) et éditeur graphique (b)



Ces grammaires peuvent contenir des symboles lexicaux ou syntaxiques¹¹². Ainsi, pour décrire la phrase « *la dame dessine une belle jeune fille* », on peut ajouter pour les adjectifs le symbole syntaxique `<A>`, pour les déterminants le symbole `<DET>` et pour les noms le symbole `<N>` (voir Figure 23).

Figure 23. Alternative à la règle principale



Ce graphe représente le fait que l'adjectif (`<A>`) peut être employé dans le groupe nominal ; de plus cet adjectif peut être répété un nombre indéfini de fois grâce à la boucle au-dessus du nœud.

Cette grammaire peut aussi être représentée de cette façon dans NooJ :

Main = `<DET> <A>* <N>`.

Voici un autre exemple de grammaire hors-contexte :

Phrase = `:GN <V> :GN` ;

GN = `<DET> <N> | que :Phrase` ;

¹¹² Par le terme « extension », nous sous-entendons une alternative à la règle principale.

Cette grammaire est représentée graphiquement sur la Figure 24. Dans cet exemple, la règle Phrase est définie à partir d'elle-même ; on dit qu'une telle règle est récursive.

Figure 24. Grammaire hors-contexte représentée graphiquement



Les grammaires hors-contexte peuvent être utilisées pour décrire différents types de syntagmes ou de phrases. En appliquant une telle grammaire à un corpus, nous pouvons trouver toutes les phrases qui correspondent aux schémas décrits dans la grammaire.

Grammaires contextuelles

Les grammaires contextuelles du type 1 de la hiérarchie de Chomsky-Schützenberger sont les grammaires dans lesquelles les règles de réécriture sont fixées en fonction du contexte.

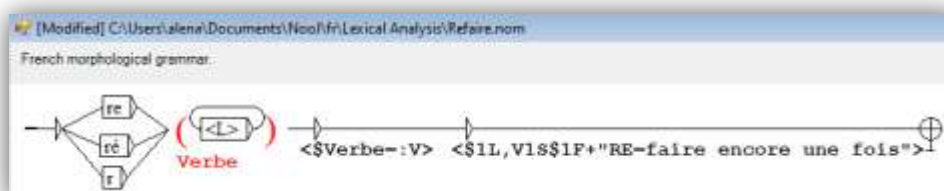
NooJ permet de construire des grammaires contextuelles (type 1 dans la hiérarchie de Chomsky-Schützenberger). Par exemple :

Main = (re | ré | r) (Suffixe <L>*) / <\$Suffixe=:V> ;

Cette grammaire reconnaît dans un premier temps toutes les formes qui commencent par « re », « ré » ou « r », puis vérifie que la séquence de lettres (<L>) suivant le préfixe correspond à une forme verbale. La grammaire reconnaît ainsi les formes *refaire*, *récrire* et *rappeler*, mais pas la forme *riz*, puisque « iz » n'est pas une forme verbale.

La Figure 25 représente la même grammaire, cette fois décrite graphiquement.

Figure 25. Grammaire contextuelle dans NooJ



Cette grammaire morphologique contextuelle à production attribue aux formes reconnues les propriétés du verbe de base, ainsi que la propriété « +RE=faire encore une fois ». Les grammaires contextuelles seront utiles pour analyser des formes inconnues, lever des ambiguïtés, notamment en vérifiant les accords dans le groupe nominal, traiter les reduplications, attribuer des valeurs pour faire des extractions du contenu, etc.

Grammaires non restreintes

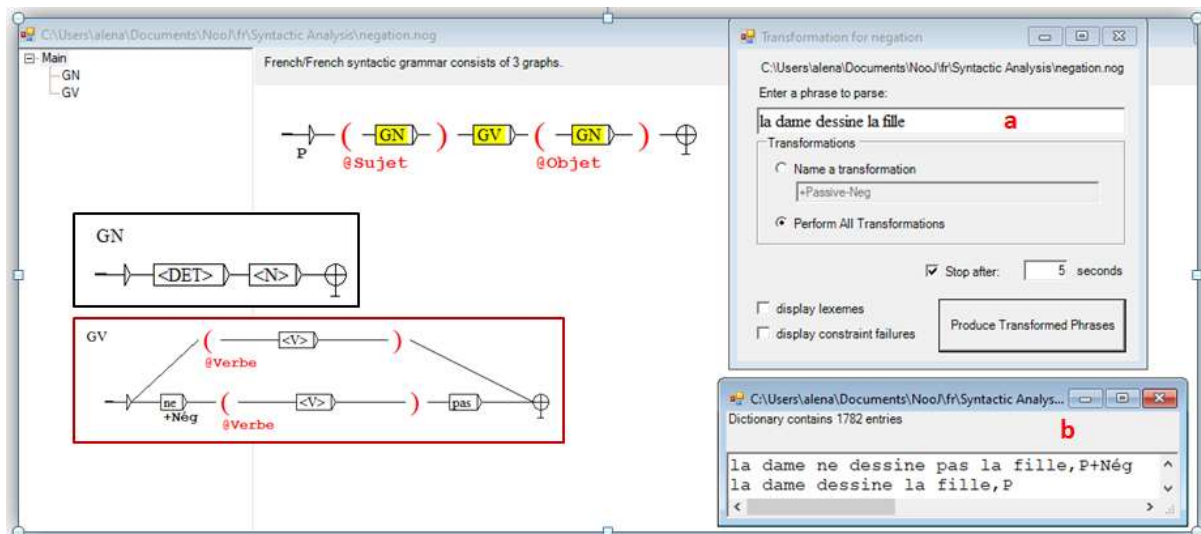
Les grammaires non restreintes (de type 0 de la hiérarchie de Chomsky-Schützenberger) permettent de décrire des langages récursivement énumérables, et donc n'importe quelle langue naturelle. Elles peuvent être traitées par des machines de Turing.

Un exemple de grammaire non restreinte NooJ est présenté sur la Figure 26. Les grammaires non restreintes NooJ utilisent des variables (ici : @Sujet, @Verbe et @Objet) qui permettent de

référencer divers éléments de la phrase. Si l'on soumet à NooJ la phrase « *la dame dessine la fille* » (voir Figure 26 a), nous obtenons en sortie deux phrases : l'affirmative et la négative (annotée « +Nég », voir Figure 26 b).

Les grammaires non restreintes donnent à NooJ la possibilité de décrire toutes sortes de transformations, par exemple le passage de la phrase au singulier à la phrase au pluriel, de la forme active à la forme passive, du présent au passé, ou encore, pour les langues slaves, de l'imperfectif au perfectif, etc. De ce fait, les grammaires de ce type pourront être utilisées pour produire des paraphrases automatiquement, développer des correcteurs d'orthographe, des chatbots, des logiciels de traduction automatique, ou pour effectuer des analyses sémantiques, etc.

Figure 26. Exemple de grammaire non restreinte NooJ



2.5.3 Nos appréciations

La plateforme NooJ répond aux exigences d'un linguiste qui cherche à décrire une langue formellement grâce à un outil de développement informatique, car elle offre toutes les fonctionnalités pour cela.

- La plateforme NooJ permet d'importer des ressources existantes ou de créer ses propres ressources, telles que des dictionnaires ou des corpus annotés.
- Elle permet également de formaliser tout le vocabulaire de façon optimale, parce qu'elle donne la possibilité d'associer les mots avec leurs variantes, de décrire les locutions, les abréviations. Elle permet également de relier les formes dérivées à une entrée lexicale.
- Elle permet de construire et de combiner des grammaires des quatre types de la hiérarchie de Chomsky-Schützenberger, ce qui offre une grande souplesse et permet de décrire n'importe quelle langue naturelle et n'importe quel phénomène linguistique.
- Avec NooJ, l'utilisateur est libre de créer des propriétés à volonté, de tester des hypothèses linguistiques, d'annoter tous les textes, y compris des textes très volumineux.
- Les éditeurs de grammaires (textuelles et graphiques) de NooJ sont faciles à utiliser et ne demandent pas de connaissances particulières en informatique.

Finalement, NooJ peut être utilisé pour développer des logiciels de TAL, tels que des outils d'extraction, de conjugaison ou de déclinaison, des correcteurs d'orthographe, des traducteurs automatiques, des générateurs de textes, des paraphraseurs, des chatbots, etc.

La plateforme est gratuite et peut être téléchargée sur le site de référence <https://nooj.univ-comte.fr>. Sur le même site, les ressources pour plusieurs langues sont disponibles, y compris pour

l'ukrainien. Le forum de l'association NooJ met en relation tous les utilisateurs, ce qui permet d'avoir un soutien non négligeable.

2.6 Conclusion du chapitre 2

Dans cette partie, nous avons étudié quelques outils stochastiques et formels développés pour la langue ukrainienne, nous avons pu voir comment ils fonctionnent et quelles possibilités ils nous offrent.

Nous avons relevé que les outils stochastiques sont actuellement majoritaires. Cela est dû au fait qu'ils sont plus faciles et plus rapides à mettre en place. Ils effectuent l'étiquetage et la lemmatisation de textes grâce à des prédictions issues de leurs bases de connaissances. Ils affichent de bons résultats, mais comme nous l'avons vu, ces résultats ne sont pas toujours corrects et, lors de nos tests, nous avons constaté beaucoup trop d'erreurs, notamment des confusions entre les classes grammaticales, de mauvaises prédictions des mots inconnus, ou des lemmatisations erronées. L'ensemble de ces erreurs compromet la fiabilité des analyses produites par ces outils.

Les outils formels impliquent un important travail des linguistes et, de ce fait, peuvent s'avérer coûteux et longs à mettre en place, mais ils offrent aussi un grand nombre d'avantages¹¹³. Nous pouvons en citer quelques-uns :

- les règles sont testables et modifiables par les utilisateurs ;
- les erreurs sont faciles à détecter et à corriger ;
- les ambiguïtés peuvent être levées avec une très grande précision ;
- les formes dérivées, y compris celles qui prennent en compte la palatalisation et les variantes phonétiques, peuvent être reliées ;
- on peut effectuer et combiner des analyses à plusieurs niveaux : morphologique, syntaxique, sémantique et statistique.

Une plateforme formelle bien conçue, enrichie par des ressources linguistiques (dictionnaires, paradigmes, dérivations, mécanismes de désambiguïsation syntaxique), peut afficher d'une part de meilleurs résultats (supérieurs à 99 %), et d'autre part, éviter les erreurs produites par les outils stochastiques. Et même si des erreurs, dues à une règle mal construite, peuvent s'y incruster, chaque utilisateur sera capable de les corriger.

L'approche formelle attire l'attention d'un certain nombre de chercheurs à travers le monde et les outils qui en résultent ont été développés également pour la langue ukrainienne, ainsi de VESUM, GRAC, LanguageTool API NLP UK. Mais nous ne pouvons les utiliser que ponctuellement, car les possibilités qu'ils offrent sont limitées, ce qui nous a motivée à utiliser la plateforme NooJ pour décrire l'ukrainien.

Nous utiliserons donc la plateforme NooJ pour :

- préparer des dictionnaires ;
- décrire la morphologie flexionnelle et lever les ambiguïtés ;
- décrire les dérivations et, de ce fait, inclure les formes dérivées dans l'analyse ;
- déterminer notre propre jeu d'étiquettes ;
- extraire des données spécifiques ;
- annoter et partager les corpus ;

¹¹³ Ces avantages sont détaillés dans l'ouvrage collectif *Linguistic Resources for Natural Language Processing* sous la direction de Max Silberztein (2024 [121]).

- décrire les transformations des phrases ukrainiennes.

L'ensemble des ressources linguistiques que nous développerons avec NooJ pourra s'avérer très utile, aussi bien pour les linguistes qui étudient la structure de la langue que pour les chercheurs en sciences humaines (historiens, littéraires, psychologues, sociologues, etc.) qui ont besoin d'analyser leur corpus de textes.

Chapitre 3. Grammaire morphologique de la langue ukrainienne

3.1 Introduction

Description des unités linguistiques atomiques (ALU)¹¹⁴ a toujours été au centre des études en linguistique ukrainienne, car elle permet de mieux comprendre des phénomènes propres à la langue, d'étudier comment les pensées sont exprimées et structurées et, par conséquent, de faire des analyses approfondies de textes, de préparer des dictionnaires, de comparer avec une autre langue et ses corpus, et de préparer des logiciels de TAL. Dans ce chapitre, nous allons d'abord étudier la classification des ALU en ukrainien, puis nous proposerons un aperçu de chaque classe d'ALU.

Notons cependant que nous ne ferons pas une étude approfondie de la grammaire ukrainienne, nous décrirons seulement les phénomènes qui seront formalisés dans le module ukrainien NooJ. Ainsi, nous parlerons de la classification morphologique et de la classification sémantico-fonctionnelle des ALU, ce qui nous permettra d'établir des étiquettes précises et détaillées. De plus, la classification morphologique nous aidera à préparer des schémas pour construire non seulement des grammaires flexionnelles (FLX) et dérivationnelles (DRV), mais aussi des grammaires morphologiques productives qui seront capables de produire des formes fléchies des ALU absentes du dictionnaire. Quant à la classification syntaxique, elle pourra nous donner des idées pour construire des grammaires transformationnelles et des grammaires syntaxiques de désambiguïsation.

3.2 Les classes d'ALU dans la langue ukrainienne

3.2.1 Définition d'une classe grammaticale

Les ALU sont généralement étudiées en tant qu'unités appartenant à une classe. Les noms et les nombres de classes varient d'une langue à l'autre. Dans la tradition française, les classes d'ALU sont appelées « classes grammaticales » ; dans la tradition anglophone ce sont des « *parts of speech* » [*parties du discours*], ou POS, et dans la tradition ukrainienne « *частини мови* » [*parties du discours*].

Traditionnellement, une classe grammaticale est définie comme un ensemble d'ALU dont la nature est déterminée par des caractéristiques substantielles et distinctives par rapport aux autres unités. Les ALU de la même classe ont majoritairement les mêmes propriétés syntaxiques, morphologiques, sémantiques et dérivationnelles. Cela veut dire que les ALU de la même classe apparaissent dans des situations syntaxiques similaires, entrent de la même façon en relation avec les autres éléments du syntagme (distinction syntaxique), ont les mêmes catégories grammaticales (distinction morphologique), restituent le contenu sémantique de la même façon (distinction sémantique), et sont souvent formées selon les critères propres à leur classe (distinction dérivationnelle).

Notons que l'ukrainien est une langue flexionnelle ; cela signifie que la terminaison d'une ALU peut changer sous l'influence de facteurs grammaticaux en créant des formes fléchies. Une forme fléchie se distingue de son lemme (la forme du dictionnaire) principalement par la terminaison, mais des phénomènes plus complexes peuvent s'y ajouter, par exemple, la modification de la racine du

¹¹⁴ Pour décrire un élément du vocabulaire, nous reprenons le terme « ALU » proposé par Max Silberstein (Silberstein 2015 [135]).

lemme ou le rajout d'un affixe. L'ensemble des formes fléchies disponibles dans une langue pour un lemme constitue un paradigme, et des paradigmes similaires constituent une classe flexionnelle. Une classe grammaticale composée d'ALU réparties dans des classes flexionnelles est appelée classe grammaticale variable, et une classe grammaticale composée des ALU qui n'ont pas de formes fléchies est appelée classe grammaticale invariable.

Ainsi, la distinction morphologique se fait surtout pour des ALU variables. Cependant, certaines classes grammaticales invariables peuvent aussi être distinguées selon les critères morpho-dérivationnels, comme c'est le cas des adverbes qualificatifs formés par dérivation des adjectifs grâce à un morphème invariable *-o*. Par exemple, l'adverbe *погано* [mal] est formé à partir de l'adjectif *поганий* [mauvais] suite au remplacement de la terminaison adjectivale *-ий* par la terminaison adverbiale *-o*.

Généralement, à l'intérieur d'une classe, on distingue les unités primaires, qui forment le noyau de la classe et qui ont toutes les propriétés de la classe en question, et les unités secondaires, dépourvues d'une ou plusieurs propriétés de la classe.

Les ALU d'une classe peuvent changer leur position dans la phrase et entrer dans des relations syntaxiques différentes avec les autres éléments, et de ce fait passer d'une classe grammaticale à une autre. Ce phénomène est appelé la conversion. La conversion peut être partielle (uniquement syntaxique) ou complète (selon tous les critères), ce que nous pouvons voir dans le Tableau 6.

Tableau 6. Conversion d'une classe à une autre

	Classe 1	Classe 2
Conversion partielle	<i>Хворий</i> чоловік відпочиває. [Un homme malade se repose.]	<i>Хворий</i> відпочиває. [Le malade se repose.]
Conversion complète	<i>Діти стали в коло.</i> [Les enfants se sont mis en cercle .]	<i>Автомобіль стоїть коло¹¹⁵ хати.</i> [L'automobile est garée à côté de la maison.]

L'exemple typique de la conversion partielle est la substantivation des adjectifs. Dans le premier exemple du Tableau 6, la classe 1 de l'ALU *хворий*, c'est l'adjectif « malade » qui s'utilise en tant que qualifiant du nom *чоловік* [homme]. Dans la deuxième colonne, la même ALU *хворий* est utilisée en tant que sujet et joue donc le rôle du nom « le malade », mais ne perd pas son sens, ni ses caractéristiques morphologiques.

Dans le deuxième exemple du Tableau 6, la classe 1 de l'ALU *коло* est le nom qui signifie « le cercle », dans la deuxième colonne, la même ALU est utilisée en tant que préposition signifiant « à côté de ». En passant dans la classe des prépositions, le nom *коло* perd sa fonction syntaxique (la position dans la phrase), sémantique (son sens d'origine) et morphologique (il devient invariable). Ici, la conversion est donc complète.

Le passage d'une ALU d'une classe à une autre crée des homonymes qu'il faut traiter lors de la création d'un outil de TAL.

3.2.2 Division en classes et en groupes de classes

Traditionnellement, en ukrainien, on distingue dix classes d'ALU :

¹¹⁵ L'étymologie de la préposition *коло* est décrite dans le *Dictionnaire étymologique de la langue ukrainienne* (1992, t. 2, p. 515). Il y est indiqué que la préposition *коло* est formée sur la base de l'accusatif singulier du nom *коло* [cercle].

1. le nom ;
2. l'adjectif ;
3. le numéral ;
4. le pronom ;
5. le verbe ;
6. l'adverbe ;
7. la préposition ;
8. la conjonction ;
9. la particule ;
10. l'interjection.

Le nom, l'adjectif, le numéral, le pronom et le verbe sont des classes variables, tandis que l'adverbe, la préposition, la conjonction, la particule et l'interjection sont des classes invariables.

Généralement, les grammairiens ukrainiens proposent des classifications hiérarchisées où les ALU sont divisées en groupes de classes :

- principales et secondaires ;
- cardinales (ou indépendantes) et fonctionnelles (ou dépendantes) ;
- classes des mots et classes des mots-morphèmes ou des mots-outils.

La séparation en classes principales et secondaires se fait selon l'importance de l'ALU dans la phrase. Ainsi, pour la langue ukrainienne, on distingue deux classes principales : le nom et le verbe qui jouent le rôle principal dans la phrase et ont leurs propres catégories grammaticales par rapport aux autres classes qui se regroupent autour du nom et du verbe (Vykhovanets' et Horodens'ka 2004, p. 12 [10]).

La division en classes des ALU cardinales et fonctionnelles n'est pas unanime. Par exemple, Horpynytych (2004, p. 21 [21]), Kostoussiak (2012, p. 21 [33]), Vykhovanets' (1988, p. 14 [9]) proposent 4 classes cardinales : le nom, le verbe, l'adjectif et l'adverbe, tandis que Pliouchtch (2010, p. 86 [52]), Horodens'ka (Vykhovanets' et Horodens'ka 2004, p. 44 [10]), Roussaniv's'kyi (Bezpoïasko et alii 1993, p. 12 [4]), Vykhovanets' dans ses travaux plus récents (Vykhovanets' 2012, p. 5 [11]) proposent 5 classes, en y ajoutant le numéral.

Les mots-morphèmes ou les mots-outils regroupent les prépositions, les conjonctions, les particules et les interjections.

Dans les grammaires modernes apparaît une tendance à multiplier les classes, c'est-à-dire à définir comme classe indépendante un ensemble d'ALU qui était autrefois considéré comme une sous-classe. Ainsi, dans le groupe des mots-morphèmes, certains linguistes distinguent des classes supplémentaires :

11. les mots modaux¹¹⁶, qui déterminent le point de vue du locuteur sur le discours ;
12. les mots prédicatifs¹¹⁷, qui jouent le rôle du prédicat dans la phrase ;
13. la catégorie d'état ou l'ensemble des mots qui expriment l'état¹¹⁸ ;

¹¹⁶ Je traduis *модальник* dans (Horpynytych 2004, p. 241 [19]), (Jovtobrioukh 1965, p. 412 [29]).

¹¹⁷ Je traduis *присудкове слово* sur goroh.pp.ua, *предикативне слово* dans (Koval' 2019 [32]), *предикат* dans SUM-20, (Kavera 2001 [31]).

¹¹⁸ Je traduis *слова категорії стану (станівник)* dans (Horpynytych 2004, p. 235 [19]), (Jovtobrioukh 1965, p. 335 [29]), (Vykhovanets' 1988, p. 119 [9]).

14. les mots-liens¹¹⁹, qui instaurent un lien entre le sujet et le prédicat ;
15. les mots-phrases¹²⁰, c'est-à-dire les mots ou les locutions qui expriment en eux-mêmes une pensée et remplissent le rôle d'une phrase d'un point de vue syntaxique ;
16. les inserts¹²¹, qui expriment l'attitude du locuteur à l'égard de l'énoncé et qui font partie d'une phrase sans en être les membres.

Par ailleurs, la même ALU appartenant à la classe des mots-outils peut être classée en tant que mot modal dans une grammaire, mais mot prédicatif dans une autre, et correspondre à l'adverbe prédicatif dans une troisième. Cette diversité des interprétations vient du rôle que l'ALU en question peut jouer dans la phrase.

Les avis divergent aussi sur l'interprétation des pronoms. Traditionnellement, cette classe est considérée comme une classe à part entière. Cependant, certains linguistes ukrainiens (Bezpoïasko *et alii* 1993, p. 12 [4], Vykhoanets' 1988, pp. 49, 162, 195 [9]) divisent les pronoms en trois sous-classes :

- le nom pronominal comme sous-classe des noms ;
- l'adjectif pronominal comme sous-classe des adjectifs ;
- l'adverbe pronominal comme sous-classe des adverbes.

Notons que, dans le dernier dictionnaire académique SUM-20, sont répertoriées dix classes traditionnelles. Dans le même dictionnaire, la classe des pronoms est composée des pronoms nominaux et des pronoms-adjectifs, les ALU qui étaient considérées autrefois comme des pronoms adverbiaux sont classées comme des adverbes, et tous les mots-morphèmes supplémentaires (de 11 à 16) sont répartis aussi entre les classes traditionnelles.

Par ailleurs, dans SUM-20, on propose des entrées séparées pour les homonymes sémantiques et des paragraphes séparés à l'intérieur de la même entrée si l'ALU joue des rôles particuliers dans une phrase avec la mention « dans le rôle de ». Par exemple, pour l'ALU *коса*, il y a une entrée qui correspond au nom « tresse », une deuxième pour le nom qui signifie « bande de sable », une troisième pour le nom qui signifie « fauche », une quatrième pour le nom qui signifie « ethnie africaine » et une cinquième pour le nom dialectique qui signifie « rate ». Et pour l'ALU *бути*, il y a une seule entrée, mais avec quinze paragraphes : neuf expliquent l'utilisation de l'ALU en tant que verbe, deux en tant que verbe auxiliaire, un en tant que mot-lien, un en tant que particule (*було*), un en tant que mot prédicatif (*буде* qui correspond à la troisième personne au futur) et un en tant qu'insert¹²² (*було*).

Notons que certains dictionnaires actuels proposent des entrées séparées pour les ALU dont une seule forme peut jouer un rôle différent. Ainsi, le dictionnaire en ligne *Горох*, disponible sur <https://goroh.pp.ua>, propose une entrée différente pour *було* en tant que mot-insert¹²³ et une entrée supplémentaire pour *буде* en tant que mot prédicatif¹²⁴.

¹¹⁹ Je traduis *зв'язка* dans (Vykhovanets' et Horodens'ka 2004, p. 363 [10]), (Jovtobrioukh 1965, p. 411 [29]), (Volkova et Maslo 2012, p. 196 [14]).

¹²⁰ Je traduis *слово-речення* dans (Vykhovanets' *et alii* 2017, p. 752 [12]), (Vykhovanets' et Horodens'ka 2004, p. 13 [10]).

¹²¹ Je traduis *вставне слово* sur goroh.pp.ua

¹²² Voir <https://sum20ua.com/Entry/index?wordid=6191&page=226>

¹²³ Voir <https://goroh.pp.ua/Словозміна/було>

¹²⁴ Voir <https://goroh.pp.ua/Словозміна/буде>

3.3 Description des classes grammaticales

Dans cette partie, nous décrivons dans l'ordre proposé traditionnellement dans les grammaires ukrainiennes¹²⁵ les classes grammaticales et leurs propriétés que nous allons par la suite décrire avec NooJ. Nous ne nous arrêterons pas sur les inserts et les mots prédicatifs, car ils méritent une étude plus approfondie et cela dépasserait les limites de nos recherches. Sachant que si, plus tard, nous voulions décrire une ALU d'un point de vue différent, cela ne nous bloquerait pas car, comme nous l'avons vu dans la section précédente, NooJ dispose des outils tels que l'éditeur de grammaires syntaxiques et l'éditeur de grammaires morphologiques productives qui permettent d'ajouter des étiquettes supplémentaires ou de remplacer les anciennes étiquettes par des nouvelles.

3.3.1 Le nom

Description des noms

En ukrainien, les ALU catégorisées dans la classe des noms sont décrites comme des unités qui désignent le substantif (ce qui existe), c'est-à-dire¹²⁶ :

- des objets concrets, comme
телефон [téléphone], автомобіль [voiture], стіл [table] ;
- des êtres vivants, comme
жінка [femme], кіт [chat], Оксана [Oksana (prénom)] ;
- des phénomènes naturels, comme
вітер [vent], дощ [pluie], повінь [inondation] ;
- des espaces, comme
річка [rivière], всесвіт [univers], Європа [Europe] ;
- des propriétés, comme
добота [gentillesse], ввічливість [politesse], страх [peur] ;
- des matières, comme
бронза [bronze], бетон [béton] ;
- des actions, sans tenir compte des propriétés inhérentes à l'action, comme
читання [lecture], біг [course].

Les noms expriment des catégories grammaticales telles que :

- le genre (masculin, féminin, neutre) ;
- le nombre (singulier et pluriel) ;
- le cas (le nominatif, l'accusatif, le génitif, le locatif, le datif, l'instrumental et le vocatif) ;
- l'animé/inanimé (l'animé désigne un être vivant, et l'inanimé désigne un objet ou un concept).

Les noms peuvent avoir des formes dérivées :

- péjorative, qui ajoute des nuances de grandeur et/ou de dépréciation ;

¹²⁵ Cette classification est aussi en adéquation avec SUM-20, le dernier dictionnaire de l'Académie des sciences de l'Ukraine.

¹²⁶ Cette liste est donnée à titre d'exemple et non en tant que liste exhaustive.

- diminutive, qui ajoute des nuances de petitesse et/ou d'affection, d'appréciation ou de compassion.

Ces nuances sont exprimées par les suffixes. Certains de ces suffixes peuvent permettre la création de noms péjoratifs et diminutifs à partir de noms existants.

Pour les péjoratifs, il s'agit des suffixes *-ур-* (*-юр-*), *-ар-* (*-яр-*), *-ак-* (*-як-*), *-иц-*, *-иськ-*, *-уг-*, qui, d'ailleurs, peuvent être combinés¹²⁷ :

- *ніс* [nez] → *носюра* [un énorme nez] ;
- *бик* [taureau] → *бичура* [un énorme taureau] ;
- *слон* [éléphant] → *слоняра*, *слоняка* [un énorme éléphant maladroit] ;
- *вуж* [couleuvre] → *вужака* [une énorme couleuvre] ;
- *камінь* [pierre] → *каменяка* [une énorme pierre] ;
- *баба* [grand-mère ou femme (familier)] → *бабуряка* [méchante grand-mère], *бабусько*, *бабуще* [une énorme et/ou mauvaise femme ou grand-mère] ;
- *ціп* [chaîne] → *ціпуга*, *ціпище*, *ціписько* [une énorme chaîne].

Pour les diminutifs, il s'agit des suffixes *-к-*, *-очк-*, *-ечк-* (*-єчк-*), *-ячк-*, *-ичк-*, *-ун'-* (*-юн'-*), *-оньк-*, *-еньк-* (*-єньк-*), *-ус'-* (*-юс'-*), *-ас'-*, *-иц'-* (*-ц'-*)¹²⁸, qui peuvent aussi être combinés :

- *мама* [maman] → *мамка*, *мамочка*, *мамуся* [petite maman (chérie)] ;
- *баба* [grand-mère ou femme (familier)] → *бабуся* [mamie], *бабусечка*, *бабуня* [mamie chérie], *бабонька* [petite dame] ;
- *верба* [saule] → *вербичка*, *вербиця* [petit saule] ;
- *кума* [marraine] → *кумася*, *кумця*, *кумасенька* [marraine chérie] ;
- *тато* [papa] → *татусь*, *татонько*, *татунь*, *татуньо*, *татусенько*, *татусь*, *татусьо* [petit papa (chéri)] ;
- *весілля* [mariage] → *весіллячко* [un bon petit mariage] ;
- *горе* [chagrin] → *горенько*, *горечко* [un tel chagrin].

En ukrainien, les noms peuvent être divisés également en groupes de noms communs et de noms propres qui désignent quelque chose d'unique, comme un toponyme, un nom de famille, un nom d'institution, etc. Les noms propres s'écrivent avec une majuscule.

Le genre

En ukrainien, le genre est déterminé selon les critères sémantiques, morphologiques ou syntaxiques. Ainsi, le genre des personnes est déterminé principalement selon le critère sémantique, à quelques exceptions près, dans lesquelles le critère morphologique et/ou syntaxique doit aussi être pris en compte. De ce fait, les noms communs :

- qui désignent un homme seront du genre masculin, comme *брат* [frère], *батько* [père], *Максим* [Maxime] ;
- qui désignent une femme seront du genre féminin, comme *мати* [mère], *Ірина* [Iryna], *бабуся* [grand-mère] ;
- qui désignent un petit-être à condition qu'ils se terminent par *-я* ou *-ча*, *-ша*, *-ца*, comme *дитя* [enfant], *дівча* [fillette], *хлоп'я* [garçonnet].

¹²⁷ Par exemple, *-уряк-* = *-ур-* + *-як-*.

¹²⁸ Les suffixes *-ун'-* (*-юн'-*), *-ус'-* (*-юс'-*), *-ул'-*, *-ус'-* (*-юс'-*), *-ас'*, *-иц'-* (*-ц'-*) sont suivis des lettres *-я*, *-ь*, *-е* au nominatif singulier et du signe mou *-ь*, lorsqu'il y a une double suffixation.

Par ailleurs, parfois le même nom peut désigner un homme ou une femme, comme c'est le cas des noms qui ont la terminaison *-a*, comme *сирота* [*orphelin(e)*], *бідолаха* [*malheureux(se)*], les noms de famille ou certains prénoms. Dans ces cas, la distinction du genre se fait selon le critère syntaxique en contexte.

Les critères morphologiques et syntaxiques sont aussi pris en compte pour :

- les noms ayant la terminaison *-o* qui désignent un homme ou une femme, mais qui ont une signification péjorative ; ils peuvent être employés au masculin (pour désigner un homme) ou au neutre (pour désigner un homme ou une femme), comme *цей базикало* [*ce bavard*] (*m.*), *це базикало* [*ce bavard*] (*n.*), *ця базикало* [*cette bavarde*] (*f.*)¹²⁹;
- les noms ayant la terminaison *-e*, précédée par le suffixe péjoratif *-иц-* qui désignent une femme ; ils peuvent être employés au féminin ou au neutre, comme *та бабуице* [*cette mère*] (*f.*), *велике бабуице* [*grande mère*] (*n.*)¹³⁰.

Quant aux autres noms, leur catégorie grammaticale de genre est exprimée principalement morphologiquement, et plus précisément par la terminaison, ou par la combinaison du suffixe et de la terminaison. Toutefois, le critère sémantique peut être pris en compte ; par exemple, le genre du nom dont la forme en question est dérivée. Nous avons relevé également quelques cas où le critère syntaxique doit être pris en compte.

Ainsi, les noms masculins sont ceux qui au nominatif singulier :

- se terminent par une consonne ou par le signe mou *-ь* (on dit qu'ils ont la terminaison « zéro »), comme *стіл* [*table*], *стілець* [*chaise*] ;
- qui sont dérivés de noms masculins à l'aide des suffixes *-ц-*, *-ин-*, *-уг-* (*-юг-*), *-аг-* (*-яг-*), *-ур-* (*-юр-*), *-ар-* (*-яр-*), *-ук-* (*-юк-*), *-ак-* (*-як-*), et ont la terminaison *-а*, *-я*, comme *вітряга* [*vent*], *котяра* [*chat*], *пюра* [*chien*] ;
- qui sont dérivés de noms masculins à l'aide des suffixes *-еньк-*, *-оньк-*, *-ицьк-*, et ont la terminaison *-о*, comme *місяченько* [*lune*], *соколонько* [*faucou*], *сніжисько* [*neige*] ;
- qui sont dérivés de noms masculins à l'aide du suffixe *-иц-*, et ont la terminaison *-е*, comme *голосице* [*voix*], *домице* [*maison*], *сомице* [*poisson-chat*] ;
- qui sont calqués sur des noms étrangers, se terminent par une voyelle, et restent invariables ; ils désignent
 - o des animaux, comme *кенгуру* [*kangourou*],
 - o des vents, comme *сирокко* [*syroko*],

¹²⁹ Cet exemple a été pris sur le site officiel de la langue ukrainienne :

http://ukrainskamova.com/publ/chinnij_pravopis/morfologija_i_pravopis/rid_imennikiv/6-1-0-88?ssid=023554777375715632414&id=4&a=1&ajax=891

¹³⁰ Cet exemple a été tiré de SUM-20 : <https://sum20ua.com/Entry/index?wordid=1963&page=88>

- des noms propres d'origine étrangère qui correspondent aux noms communs masculins, comme

Огайо (штат) [Ohio (état)], Па-де-Кале (департамент) [Pas-de-Calais (département)] ;

- qui désignent des abréviations dont le mot principal est au masculin, comme *СРСР [URSS]*.

Les noms féminins sont ceux qui au nominatif singulier :

- ont les terminaisons *-а, -я*, comme *хата [maison], кухня [cuisine]* ;
- se terminent par une chuintante (*-ж, -ч, -ш, -щ*) et ont la terminaison « zéro » (petit nombre de mots), comme *упряж [harnais], ніч [nuit], суміш [mélange], нехворощ [(nom d'une plante)]* ;
- se terminent par *-в, -ф, -б, -р* et ont une terminaison « zéro » (petit nombre de mots), comme *крав [sang], верф [chantier naval], праматір [aïeule], Об [Ob]* ;
- se terminent par *-ь*, et
 - par (et ont le) suffixe *-ість, -рть, -ість*, comme *радість [joie], смерть [mort], далечинь [distance]*,
 - désignent certains animaux, comme *рись [lynx], лань [daim], кефаль [mulet]* ;
- se terminent par une voyelle autre que *-а* ou une consonne, calqués sur des noms étrangers, qui restent invariables et qui désignent
 - certains fruits et légumes, comme *броколи [brocoli]*,
 - certains animaux et noms communs, comme *цече [tsé-tsé], салями [salami]*,
 - des langues, comme *гінді [hindi]*,
 - des noms propres d'origine étrangère qui correspondent au nom commun féminin, comme *Огайо (річка) [Ohio (rivière)]* ;
- qui désignent des abréviations dont le mot principal est au féminin, comme *УРСР [RSS d'Ukraine]*.

Les noms neutres sont ceux qui au nominatif singulier :

- ont la terminaison *-о* et n'ont pas de suffixe, comme *вікно [fenêtre], село [village], добро [bonté]* ;
- ont la terminaison *-о* et les suffixes *-ств-, -ств-, -ств-* (il s'agit de noms abstraits), comme *новаторство [innovation], господарство [patrimoine]* ;
- ont la terminaison *-о* et les suffixes *-еньк-, -оньк-*, et qui sont dérivés de noms neutres *личенько [visage], крилонько [aile]* ;

- ont la terminaison *-o* et le suffixe *-иськ-*, et qui sont dérivés de noms neutres ou féminins, comme
чудисько [monstre], лаписько [patte] ;
- ont la terminaison *-e* et n'ont pas de suffixe, comme
море [mer], поле [champ], селище [bourgade] ;
- ont la terminaison *-e* et le suffixe diminutif *-ц-*, comme
віконце [fenêtre], човенце [bateau] ;
- ont la terminaison *-e*, le suffixe *-ущ-* et qui désignent des mots communs tels que des lieux (y compris des noms propres) ou des objets de la vie quotidienne, comme
сховище [stockage], пасовище [pâturage], днище [fond] ;
- ont la terminaison *-e*, et le suffixe *-ущ-* et qui sont dérivés de noms neutres ou féminins inanimés et d'un groupe de mots masculins inanimés, comme
картоплище [pommes de terre], дубище [chêne], гніздище [nid] ;
- se terminent en *-стя* et relèvent de la catégorie des inanimés, comme
листя [feuillage], щастя [bonheur], зап'ястя [poignet] ;
- ont la terminaison *-я*, précédée par une consonne double, comme
навчання [formation], бадилля [grandes herbes sèches], безлюддя [désolation] ;
- ont la terminaison *-я*, précédée soit par la consonne *р*, soit par une consonne labiale et l'apostrophe, comme
багатослів'я [verbosité], безвір'я [mécréance], ім'я [prénom] ;
- se terminent par une voyelle, sont calqués sur des noms étrangers, restent invariables et désignent
 - des villes, comme
Нансі [Nancy], Баку [Bakou],
 - des objets ou notions abstraites, comme
кімоно [kimono], кіно [cinéma] ;
- des abréviations invariables dont le mot principal est au neutre, ou des abréviations non traduites, comme
МВС [Ministère des Affaires Intérieures], ЮНЕСКО [UNESCO].

Notons également que certains noms inanimés qui ont la terminaison « zéro » et qui se terminent en chuintante ou en consonne + *ь* peuvent être employés au masculin ou au féminin, comme

дрож [frisson], купіль [bol], фальш [tromperie], харч [nourriture].

Pour ces ALU, la distinction du genre se fait selon le critère syntaxique.

Le nombre

Le nombre est une catégorie grammaticale fondée sur la quantité, suivant que le mot est employé pour désigner un objet ou un concept unique (le singulier) ou une pluralité (le pluriel). En ukrainien, cette catégorie est exprimée morphologiquement par les terminaisons *-и, -і, -ї, -я, -а*. Mais le critère sémantique est aussi pris en compte, notamment pour les noms qui n'expriment qu'un seul genre. Ainsi, nous pouvons identifier les noms :

- dénombrables, qui peuvent exprimer le singulier et le pluriel, comme

стіл – столи [table – tables] ;

- non dénombrables, qui peuvent exprimer uniquement le singulier, car ils ne sont pas soumis à une comptabilité quantitative, puisqu'ils désignent :
 - une substance, comme
молоко [lait],
 - un concept abstrait, comme
радість [joie],
 - un concept collectif, comme
одяг [vêtements],
- des « pluralia tantum », qui peuvent être employés uniquement au pluriel, car ils signifient :
 - la dualité, comme
окуляри [lunettes],
 - un objet composé de plusieurs parties, comme
двери [porte],
 - un concept modulaire, comme
канікули [vacances],
 - des noms propres au pluriel, comme
Карпати [Carpatates],
 - des abréviations indéclinables dont le mot principal est au pluriel, comme
ВМС (Військово-Морські Сили) [Marine (Forces navales)].

Le cas

En ukrainien, les noms ont également la catégorie du cas, qui en comporte 7 : le nominatif, l'accusatif, le génitif, le locatif, le datif, l'instrumental et le vocatif. Chacun de ces cas se caractérise par son rôle syntaxique dans la phrase et possède sa propre terminaison morphologique. La plupart des cas peuvent être précédés par une préposition. Nous proposons de lister ici quelques emplois principaux de chaque cas¹³¹.

Le nominatif s'emploie :

- en tant que sujet dans une phrase avec ou sans verbe, comme dans
Ніч. Холодно. Йде дощ. [Il fait nuit. Il fait froid. La pluie tombe.] ;
- en tant qu'attribut du sujet, comme dans
Київ – моє рідне місто. [Kyiv est ma ville natale.] ;
- après « є » dans l'expression de possession, comme dans
У студентки є підручник. [L'étudiante a un manuel.] ;
- dans la phrase démonstrative, comme dans
Це студентка. [C'est une étudiante.] ;
- dans la construction comparative avec la préposition *ніж*, comme dans
студент молодший ніж студентка [l'étudiant est plus jeune que l'étudiante] ;

¹³¹ Nous prévoyons dans la suite de nos recherches la description complète des cas.

- dans la subordonnée de comparaison, après la conjonction *мов*, comme dans **дівчина гарна, мов Афродіта** [*la jeune fille est belle comme Aphrodite*].

L'accusatif s'emploie pour :

- le complément d'objet direct, comme dans *Я читаю книгу.* [*Je lis un livre.*];
- le complément de lieu avec une préposition qui désigne la direction, comme dans *Я кладу книгу на стіл.* [*Je pose le livre sur la table.*];
- le complément de temps avec certaines prépositions ou sans préposition, comme dans *працювати у суботу* [*travailler le samedi*], *працювати цілий день* [*travailler toute la journée*];
- le complément après certaines autres prépositions (l'objet de la parole, la cause, etc.), comme dans *розповідати про канікули* [*parler des vacances*];
- la construction comparative avec la préposition *за*, comme dans *студент молодший за студентку* [*l'étudiant est plus jeune que l'étudiante*]

Le génitif représente :

- le complément du nom, comme dans *книга студентки* [*livre de l'étudiante*];
- le complément du numéral ; ainsi, les chiffres de 2 à 4 pour les noms féminins sont suivis du génitif singulier, alors que les chiffres de 5 à 20 et les numéraux indéterminés sont suivis du génitif pluriel, comme dans *дві студентки* [*deux étudiantes*], *п'ять студенток*, [*cinq étudiantes*], *багато студенток* [*beaucoup d'étudiantes*];
- le complément d'objet partitif, comme dans *хочу води* [*je veux de l'eau*];
- le complément d'objet de termes abstraits (déverbaux), comme dans *чекати відповіді* [*attendre une réponse*];
- la construction comparative sans préposition ou avec la préposition *від*, comme dans *День став темніше ночі.* [*Le jour est devenu plus court que la nuit.*]¹³², *студент молодший від студентки* [*l'étudiant est plus jeune que l'étudiante*];
- la construction possessive, comme dans *У студентки є підручник.* [*L'étudiante a un manuel.*];
- le complément du mot prädicatif *нема(є)* ou d'un verbe dans une phrase négative, comme dans *У студентки немає підручника.* [*L'étudiante n'a pas de manuel.*], *Студентка не отримала стипендії.* [*L'étudiante n'a pas reçu de bourse.*];
- le complément après un certain nombre de prépositions, comme dans

¹³² L'exemple est tiré du polycopié de *la Grammaire théorique de l'ukrainien*, deuxième année de licence, l'INALCO, 2022-2023, préparée par Oleg Chinkarouk.

книга для **студентки** [livre pour une **étudiante**].

Le locatif s'emploie pour :

- le complément avec certaines prépositions (complément de lieu qui désigne la localisation, complément de temps, complément indiquant des instruments de musique, etc.) comme dans

у школі [à l'école],

на цьому тижні [cette semaine],

зрати на гітарі [jouer de la guitare].

Le datif s'emploie pour :

- le complément d'attribution, comme dans
*Я телефоную **студентці**. [J'appelle une **étudiante**.] ;*
- le complément représentant l'expérient¹³³ (siège du procès), être animé qui éprouve un sentiment, une sensation (physique ou psychologique), dans une phrase asubjectale avec un verbe sans -ся ou avec -ся, ou avec un adverbe (modal ou désignant un état), comme dans
Нам пощастило. [Nous avons eu de la chance.],
*Студентці треба працювати. [Il faut que l'**étudiante** travaille.],*
*Студентці не працюється. [C'est difficile pour l'**étudiante** de se mettre au travail.],*
*Студентці погано. [L'**étudiante** ne se sent pas bien.];*
- l'expression de l'âge, comme dans
*Студентці 20 років. [L'**étudiante** a 20 ans.];*
- le complément du verbe après certaines prépositions secondaires, comme dans
*Вона зробила це **наперекір батькові**. [Elle l'a fait en dépit de son **père**.].*

L'instrumental s'emploie pour :

- le complément de moyen sans préposition з, comme dans
*писати **ручкою** [écrire avec un **stylo**] ;*
- le complément d'accompagnement avec la préposition, comme dans
*хліб з **маслом** [du pain avec du **beurre**] ;*
- le complément du verbe ou du nom, employé après certaines prépositions, comme dans
*Кім сидить під **столом**. [Le chat est assis sous la **table**.];*
- le complément d'agent dans la phrase passive, comme dans
*стаття написана **журналістом** [l'article écrit par un **journaliste**].*

Le vocatif, qui assure la fonction appellative, est employé lorsque l'on s'adresse à quelqu'un, comme dans

*– Оксано, іди сюди. [– **Oksana**, viens ici.].*

¹³³ L'un des principaux rôles sémantiques (ceux des participants du procès) à côté de l'agent, du patient, de l'instrument, etc.

Catégorie grammaticale de l'animé/inanimé

La catégorie grammaticale de l'animé/inanimé répartit les ALU dans les sous-classes des animés (êtres vivants) et des inanimés (objets, concepts abstraits).

Les noms et toutes les classes grammaticales qui s'y accordent n'ont pas de terminaisons bien distinctes pour la catégorie de l'animé/inanimé. Ainsi, à l'accusatif, le groupe nominal (GN) masculin et le GN pluriel qui représentent des animés ont les mêmes terminaisons que lorsqu'ils sont employés au génitif, et le GN masculin et le GN pluriel qui représentent des inanimés ont les mêmes terminaisons que lorsqu'ils sont employés au nominatif.

Prenons en exemple deux phrases, présentées dans le Tableau 7, avec un GN employé à l'accusatif. Ce groupe est composé d'un nom masculin et d'un adjectif, d'abord au singulier, puis au pluriel. Dans la première colonne, le nom masculin de notre GN représente un inanimé ; alors, le nom et l'adjectif ont les mêmes terminaisons qu'au nominatif, tandis que dans la deuxième colonne, celle des animés, le nom et l'adjectif ont les mêmes terminaisons qu'au génitif.

Tableau 7. Le GN à l'accusatif

Singulier	Inanimé		Animé	
	Nominatif	Accusatif	Génitif	Accusatif
	<i>гарний будинок</i>	<i>Я бачу гарний будинок. [Je vois une belle maison.]</i>	<i>гарного чоловіка</i>	<i>Я бачу гарного чоловіка. [Je vois un bel homme.]</i>
Pluriel	<i>гарні будинки</i>	<i>Я бачу гарні будинки. [Je vois des belles maisons.]</i>	<i>гарних чоловіків</i>	<i>Я бачу гарних чоловіків. [Je vois des beaux hommes.]</i>

Ainsi, lorsque nous construirons les ressources pour le module ukrainien, nous devons trouver une solution pour lever les ambiguïtés de ce type.

Les classes flexionnelles du nom

En ukrainien, le paradigme nominal implique les cas et le nombre (singulier et pluriel). Notons que certaines propriétés telles que l'animé/inanimé et le genre (masculin, féminin, neutre), à la différence du paradigme adjectival, n'ont pas de variation à l'intérieur du paradigme nominal¹³⁴.

Traditionnellement, tous les noms sont séparés en quatre classes flexionnelles :

- la première classe regroupe les noms féminins et masculins, avec les terminaisons *-a*, *-я* au nominatif singulier, et se divise en trois groupes :
 - o dur, les noms qui font partie de ce groupe se terminent par *-a*, comme *мама [mama]*, *Микола [Mykola]*,
 - o mou, les noms qui font partie de ce groupe se terminent par *-я*, comme *сім'я [famille]*, *суддя [judge]*,
 - o mixte, les noms qui font partie de ce groupe ont une chuintante (*ж, ч, ш, щ*) comme dernière consonne du radical et la terminaison *-a*, ou se terminent en *-ище* (les noms péjoratifs féminins), comme *калюжа [flaque]*, *афіша, [affiche]*, *бабище [mère]* ;

¹³⁴ Le passage du nom masculin au nom féminin peut se faire uniquement grâce à la dérivation qui suppose des modifications morphologiques plus profondes.

- la deuxième classe regroupe les noms masculins qui au nominatif singulier ont la terminaison « zéro » ou la terminaison *-o*, ainsi que les noms neutres qui au nominatif singulier ont les terminaisons *-o*, *-e*, ou se terminent par *-стя* ou encore par une consonne dédoublée et la terminaison *-я* ; cette classe se divise aussi en trois groupes :

- o dur : les noms qui font partie de ce groupe se terminent par une consonne (hors chuintante et *й*), et la terminaison « zéro » (*-∅*) ou la terminaison *-o*, un certain nombre des noms qui se terminent par *-ар*, *-ур*, *-яр*¹³⁵, comme

стіл [table], небо [ciel], касір – касіри [caissier(s)], маляр – маляри [peintre(s)],

- o mou : les noms qui font partie de ce groupe se terminent par *-ь*, *-й*, et certains noms qui se terminent par *-ар*, *-ур*¹³⁶, pour le masculin, et par *-е*, *-я* pour neutre, comme

стілець [chaise], бібліотекар – бібліотекарі [bibliothécaire(s)], лікар – лікарі [médecin(s)], поле [champ],

- o mixte : ce groupe est constitué des noms masculins ou neutres dont la dernière lettre du radical est une chuintante, mais aussi des noms qui se terminent par *-яр* avec l'accent final, comme

прізвище [nom de famille], ніж [couteau], каменя́р – каменя́рі [taçon(s)] ;

- la troisième classe regroupe le nom *мати [mère]* ainsi que les noms féminins qui au nominatif singulier ont la terminaison « zéro », y compris ceux qui se terminent par *-ь*, comme

кров [sang], ніч [nuit], радість [joie] ;

- la quatrième classe regroupe les noms neutres qui au nominatif singulier ont les terminaisons *-а*, *-я*, et dans lesquels, lors de la déclinaison, apparaissent les suffixes *-ен-*, *-ат-* (*-ят-*), comme

ім'я – імена [prénom(s)], курча – курчата [poussin(s)], хлоп'я – хлоп'ята [garçonnet(s)].

Nous pouvons voir les terminaisons de ces classes dans le Tableau 8.

Dans ce tableau, nous pouvons nous apercevoir que, pour certaines grammèmes, il y a des terminaisons parallèles, car certains noms acceptent plusieurs terminaisons absolument synonymiques, comme c'est le cas du *ном син [fils]* qui peut être fléchi au locatif singulier avec les terminaisons *-і*, *-у* ou *-ові*, tandis que le nom *банк [banque]* uniquement avec la terminaison *-у* et le nom *дім [maison]* uniquement avec la terminaison *-і*. D'ailleurs, souvent, les terminaisons parallèles ne sont pas motivées grammaticalement, mais historiquement.

Notons que, dans la classification traditionnelle, les diverses modifications qui sont apportées au radical, telles que la voyelle mobile, l'alternance de la voyelle, l'alternance consonantique, le dédoublement de la consonne, l'ajout des suffixes, ne sont pas prises en compte. Cette classification n'est pas donc suffisante pour préparer un outil qui serait capable de produire les formes fléchies.

¹³⁵ Il s'agit des noms avec l'accent fixe sur *-ар*, *-ур*, et accent fixe sur le radical pour les noms en *-яр*, hormis quelques exceptions.

¹³⁶ Il s'agit des noms qui se terminent par *-ар*, *-ур*, avec l'accent fixe sur la racine hormis la dernière syllabe ou l'accent final au pluriel.

Par ailleurs, la classification qui prend en compte la composition graphique du lemme et ses formes fléchies et, en conséquence, sépare les flexions synonymiques dans des paradigmes distincts, est décrite dans (Kryts'ka et alii 2011 [35]). Mais cette classification a aussi quelques inconvénients :

- Les paradigmes ne prennent pas en compte la catégorie de l'animé/inanimé ; ainsi, les ALU telles que *колос [épi]* et *цид [voisin]* sont classées dans le même paradigme, bien qu'elles n'aient pas le même accusatif ;
- les paradigmes prennent en compte la catégorie du genre ; ainsi, pour les noms masculins tels que *воєвода [chef militaire]* et les noms féminins tels que *жаба [grenouille]*, on propose des paradigmes différents, alors que ces ALU ont le même schéma flexionnel ;
- la classification prend en compte l'accent mobile des mots, ce qui multiplie considérablement le nombre des paradigmes ; par exemple, rien que pour le nom masculin, l'ouvrage propose 108 paradigmes.

De ce fait, avant de formaliser la déclinaison nominale, il nous faudra préparer une classification qui prenne en compte la composition graphique du lemme et ses formes fléchies, mais éviter les doublons.

Tableau 8. Les classes flexionnelles des noms

Nombre	Cas	I			II			III	IV
		Groupe dur	Groupe mou	Groupe mixte	Groupe dur	Groupe mou	Groupe mixte		
Singulier	Nom.	<i>a</i>	<i>Я</i>	<i>a, e</i>	\emptyset ¹³⁷ , <i>o</i>	<i>ь, ѱ</i> ¹³⁸ , <i>e, я</i>	\emptyset, e	$\emptyset, ь, u$	<i>a, я</i>
	Acc. inan.	<i>y</i>	<i>Ю</i>	<i>y</i>	\emptyset, o	<i>ь, ѱ, e, я</i>	\emptyset, e	$\emptyset, ь, u$	<i>a, я</i>
	Acc. an.				<i>a, y</i>	<i>я, ю</i>	<i>a, y</i>		
	Gén.	<i>u</i>	<i>i, ї</i>	<i>i</i>	<i>a, y</i>	<i>я, ю</i>	<i>a, y</i>	<i>i, u</i>	<i>u, i</i>
	Loc. I	<i>i</i> ¹³⁹	<i>i, ї</i>	<i>i</i>	<i>i, y</i>	<i>i, ю, ї</i>	<i>i, y</i>	<i>l</i>	<i>i</i>
	Loc. II ¹⁴⁰	-	-		<i>ові</i>	<i>єві, єві</i>	<i>єві</i>	-	-
	Dat. I	<i>i</i>	<i>l</i>	<i>y</i>	<i>y</i>	<i>ю</i>	<i>y</i>	<i>l</i>	<i>i</i>
	Dat. II	-	-	-	<i>ові</i>	<i>єві, єві</i>	<i>єві</i>	-	-
	Inst.		<i>e</i> ¹⁴¹ <i>ю,</i>						<i>ам, ям</i>
Voc.	<i>o</i>	<i>e, є, ю</i>	<i>e</i>	<i>e, y</i>	<i>ю, e, я</i>	<i>e, y</i>	<i>e, u</i>	<i>a, я</i>	
Pluriel	Nom.	<i>И</i>	<i>l</i>	<i>i</i>	<i>u</i>	<i>i, ї, я</i>	<i>i, a</i>	<i>l</i>	<i>a</i>
	Acc. inan.	<i>И</i>	<i>l</i>	<i>i</i>	<i>u</i>	<i>i, ї, я</i>	<i>i, a</i>	<i>i, ів</i>	<i>a</i>
	Acc. An	\emptyset	<i>ь, ѱ</i> ²⁷	\emptyset	<i>ів</i>	<i>ів, їв, еї</i>	<i>ів</i>		\emptyset
	Gén.	\emptyset	<i>ь, ѱ</i>	\emptyset	<i>ів</i>	<i>ів, їв, еї, ь</i>	<i>ів</i>	<i>eї</i>	\emptyset
	Loc.	<i>ах</i>	<i>ях</i>	<i>ах</i>	<i>ах</i>	<i>ях</i>	<i>ах</i>	<i>ах, ях</i>	<i>ах</i>
	Dat.	<i>ам</i>	<i>ям</i>	<i>ам</i>	<i>ам</i>	<i>ям</i>	<i>ам</i>	<i>ам, ям</i>	<i>ам</i>
	Inst.	<i>ами</i>	<i>ями</i>	<i>ами</i>	<i>ами</i>	<i>ями</i>	<i>ами</i>	<i>ами, ям</i>	<i>ами</i>
	Voc.	<i>И</i>	<i>l</i>	<i>i</i>	<i>u</i>	<i>i, ї, я</i>	<i>i, a</i>	<i>i</i>	<i>a</i>

¹³⁷ Ce symbole signifie que le radical se termine par une consonne.

¹³⁸ Bien qu'ils ne possèdent pas de terminaison, nous ajoutons dans le tableau les mots dont le radical se termine par *-ь, -ї*, car ceux-ci disparaissent lors de la production des formes fléchies.

¹³⁹ En ukrainien, la lettre « i » apparaît dans la terminaison du groupe dur, bien qu'il s'agisse de la voyelle molle.

¹⁴⁰ Les noms masculins animés ont les terminaisons *-ові, -єві, -єві* pour le locatif et le datif, notés ici comme locatif II et datif II.

¹⁴¹ En ukrainien, la lettre « e » apparaît dans la terminaison du groupe mou, bien qu'il s'agisse d'une voyelle dure.

3.3.2 L'adjectif

Description des adjectifs

En ukrainien, les adjectifs sont considérés comme des unités qui expriment la propriété des noms auxquels ils se rapportent. Dans la phrase, l'adjectif se rapporte soit au nom présent dans la phrase, soit au nom supposé, c'est-à-dire mentionné précédemment dans le discours, avec lequel l'adjectif s'accorde en genre, en nombre et en cas. Ces trois catégories grammaticales de l'adjectif se réalisent syntaxiquement (accord au substantif – noyau du syntagme) et morphologiquement (flexions spécifiques pour chaque grammème de ces trois catégories). Notons que la catégorie de genre de l'adjectif a trois grammèmes (le masculin, le féminin et le neutre), la catégorie de nombre a deux grammèmes (le singulier et le pluriel) et la catégorie de cas a six grammèmes : le nominatif, l'accusatif, le génitif, le datif, l'instrumental et le locatif. Les adjectifs n'ont pas de vocatif, et quand le noyau du syntagme est exprimé par le nom au vocatif, alors l'adjectif s'accorde avec ce nom au nominatif. Les adjectifs masculins et les adjectifs pluriels à l'accusatif expriment aussi la catégorie de l'animé/inanimé qui a deux grammèmes¹⁴².

Les 13 grammèmes mentionnées plus haut forment le système de déclinaisons adjectivales. Notons seulement qu'en ukrainien, certains adjectifs sont invariables et donc n'ont pas de formes fléchies, car ils ont un lemme atypique¹⁴³ pour un adjectif ; de ce fait, ils ne s'accordent pas en genre, en nombre ou en cas avec le substantif dont ils dépendent.

Notons également que, dans la phrase, l'adjectif peut remplir la fonction d'épithète ou d'attribut. Nous pouvons voir deux exemples dans le Tableau 9.

Tableau 9. Exemple d'une épithète et d'un attribut

Épithète	Attribut
Це велика хата. [C'est une grande maison.]	Він став великим . [Il est devenu grand .]

Trois types d'adjectifs

Les adjectifs s'utilisent dans le discours pour exprimer les différentes propriétés attribuées aux êtres vivants, aux objets, aux événements, aux notions, etc. On distingue trois types d'adjectifs :

- les adjectifs qualificatifs qui expriment la qualité, par exemple, *великий* [grand] ;
- les adjectifs relatifs (ou les adjectifs de relation) qui expriment une relation envisagée à travers un objet, un procès ou un lieu, par exemple, *університетський* [universitaire], donc en lien avec l'université, *читацький* [de lecture], donc en lien avec le procès de lecture, *київський* [de lecture], donc en lien avec la ville de Kyïv ;
- les adjectifs d'appartenance qui indiquent le possesseur, par exemple, *мамин* [celui qui appartient à maman].

Ces trois types d'adjectifs peuvent remplir la même fonction dans la phrase, mais ont également leurs différences. Par exemple, seuls les adjectifs qualificatifs peuvent exprimer le niveau d'intensité

¹⁴² Cette catégorie n'est pas exprimée par les adjectifs féminins et neutres au singulier.

¹⁴³ Ce sont des adjectifs du type *бордо* [bordeaux], *люкс* [luxe], *брутто* [brut], etc.

(le comparatif, le superlatif et le superlatif absolu), entrent en lien avec des adverbes de comparaison, ont des antonymes, dérivent vers des adverbes et des noms abstraits.

Les adjectifs qualificatifs, les adjectifs relatifs et les adjectifs d'appartenance formés à partir des noms communs figurent dans les dictionnaires, tandis que les adjectifs d'appartenance formés à partir des noms propres (les noms de famille et les prénoms) et les adjectifs relatifs formés à partir des noms géographiques, qui peuvent être créés facilement, ne sont pas spécifiés dans les dictionnaires¹⁴⁴.

Les adjectifs d'appartenance sont créés par ajout de suffixes spécifiques :

- *-ин- (-ін-)* qu'on ajoute aux noms féminins ou neutres, comme pour *сестра* [sœur] → *сестрин* [qui appartient à la sœur], *Марія* [Mariya] → *Маріїн* [qui appartient à Mariia] ;
- *-ів- (-їв-)* qu'on ajoute aux noms masculins, comme pour *брат* [frère] → *братів* [qui appartient au frère], *Сергій* [Serhiy] → *Сергіїв* [qui appartient à Serhiy].

Lors de la création des adjectifs d'appartenance à partir des noms féminins ou neutres, on observe les alternances consonantiques suivantes :

- *з* → *ж*, comme pour *подруга* [amie] → *подружин* [qui appartient à une amie] ;
- *х* → *ш*, comme pour *свекруха* [belle-mère] → *свекрушин* [qui appartient à la belle-mère] ;
- *к* → *ч*, comme pour *сусідка* [voisine] → *сусідчин* [qui appartient à une voisine].

Quant aux adjectifs relatifs créés à partir des toponymes (noms géographiques), ils sont formés par ajout des suffixes *-ськ-, -івськ-, -анськ-, -енськ-, -инськ-* au radical, comme c'est le cas pour¹⁴⁵

Полтава [Poltava] → *полтавський* [de Poltava],
Семипілки [Semyopilky] → *семипілівський* [de Semyopilky],
Ружа [Rouja] → *ружанський* [de Rouja],
Рівне [Rivne] → *рівненський* [de Rivne],
Ялта [Yalta] → *ялтинський* [de Yalta].

L'alternance de la dernière consonne du radical peut être observée aussi dans les cas suivants :

- *з, з, ж + -ськ-* → *-зьк-*, comme pour *Прага* [Prague] → *празький* [de Prague], *Кавказ* [Caucase] → *кавказький* [de Caucase], *Париж* [Paris] → *паризький* [de Paris] ;
- *к, ц, ч + -ськ-* → *-цьк-*, comme pour *Дрогобич* [Drohobytch] → *дрогобицький* [de Drohobytch], *Вінниця* [Vinnytsia] → *вінницький* [de Vinnytsia], *Вишгородок* [Vychhorodok] → *вишгородоцький* [de Vychhorodok] ;

¹⁴⁴ Seuls les adjectifs les plus fréquents, comme ceux qui sont créés à partir des noms des grandes villes, sont inscrits dans les dictionnaires.

¹⁴⁵ Nous reprenons les exemples proposés dans la grammaire de Iouchtchouk (Iouchtchouk 2004, p. 171 [70]).

- *х, с, ш + ськ* → *-ськ-*, comme pour
Глеваха → *глева**ський***,
Одеса [Odessa] → *оде**ський*** [d'Odessa],
Калуш → *калу**ський*** ;

On constate aussi :

- la disparition du suffixe *-к-*, comme pour
Жмеринка [Jmerynka] → *жмерин**ський*** [de *Jmerynka*] ;
- la disparition du signe mou (*-ь*) final, comme pour
Ірпінь → *ірпін**ський*** ;
- l'alternance de *-о-/-і-*
Лозове → *лозі**вський*** ;
- l'apparition de la voyelle mobile *-е-, ои -о-*, quand le radical se termine par deux consonnes, comme pour
Вашківці [Vachkivtsi] → *вашкі**вецький*** [de *Vachkivtsi*],
Москва [Moscou] → *москов**ський*** [de *Moscou*].

Quant aux noms composés, comme *Біла Церква [Bila Tserkva]*, la première forme reçoit la terminaison *-о-* et les deux formes s'écrivent en un seul mot *білоцерк**івський*** [Bila Tserkva].

Il est important qu'un outil de TAL puisse reconnaître les adjectifs de ce type. Pour cela, il faut soit enrichir le dictionnaire à chaque fois qu'on rencontre un tel adjectif, soit préparer une grammaire qui puisse les reconnaître directement dans le texte.

La forme longue, la forme courte et la forme non contractée des adjectifs

Les adjectifs qualificatifs masculins peuvent avoir une forme longue (déclinable, avec une terminaison composée de deux lettres) et une forme courte (indéclinable, sans terminaison). La forme courte vient de la contraction de la terminaison du nominatif masculin, accompagnée d'un changement de radical : si le radical se termine par deux consonnes, alors la voyelle *-е-* sera ajoutée entre ces deux consonnes, comme pour

*повни**й*** [plein] → *повен* [plein].

Un adjectif court¹⁴⁶ peut être utilisé avec un nom masculin à n'importe quel cas :

- *за зелен **садом*** [derrière le **jardin** vert]. Ici, le nom **jardin** est à l'instrumental ;
- *немає зелен **саду*** [il n'y a pas de **jardin** vert]. Ici, le nom **jardin** est au génitif.

Par ailleurs, la forme courte est utilisée exclusivement dans le langage poétique et folklorique, et certains adjectifs qualificatifs n'en ont pas.

Certains linguistes ukrainiens considèrent que l'opposition entre les adjectifs longs et les adjectifs courts permet de diviser la catégorie en déclinable/indéclinable (Horpynytsch 2004, p. 94 [21]). Il s'agit d'une catégorie dérivative, non classifiante, puisqu'il n'y a aucune différence de sens entre les deux formes. Nous pouvons comparer les adjectifs déclinables/indéclinables dans le Tableau 10.

¹⁴⁶ « Adjectif court » abrège « adjectif à la forme courte ».

Tableau 10. Exemple d'un adjectif court et d'un adjectif long en contexte

Forme courte (sans terminaison)	Forme longue (avec terminaison)
Над горою місяць повен ¹⁴⁷ . [Au-dessus de la montagne, il y a la pleine lune.]	Над горою місяць повнуїї . [Au-dessus de la montagne, il y a la pleine lune.]

Dans la première colonne est notée la phrase du texte folklorique, tandis que dans la seconde, nous remplaçons l'adjectif court par l'adjectif long. Nous constatons que dans les deux cas, les adjectifs jouent le même rôle d'épithète dans la phrase et ont exactement le même sens.

Dans la langue ukrainienne normative, les adjectifs féminins, neutres et pluriels ont la forme contractée : leurs terminaisons au nominatif et à l'accusatif ont perdu une lettre. Toutefois, les formes non contractées sont souvent utilisées dans les textes poétiques et folkloriques. Nous pouvons comparer les deux formes dans le Tableau 11.

Tableau 11. Exemple d'adjectifs contractés et non contractés en contexte

Genre et nombre	Adjectif normatif (contracté)	Adjectif poétique (non contracté)	Traduction
Féminin sg.	свята	святая	<i>sainte</i>
	синя	синяя	<i>bleue</i>
Neutre sg.	святе	святеє	<i>saint</i>
	синє	синєє	<i>bleu</i>
Pluriel	святі	святїї	<i>saints/saintes</i>

Et voici un exemple en contexte :

Тойді, як, Господи, **святая**, На землю правда прилетить.

[Quand, mon Dieu, sur la terre viendra la **sainte** vérité.] (T. Chevtchenko)¹⁴⁸

Les formes contractées existent également pour les adjectifs masculins et neutres employés au locatif (voir Tableau 12, où elles sont notées en bleu).

Traditionnellement, les adjectifs non contractés présentés dans le Tableau 11 et les adjectifs courts ne sont pas inclus dans les classes flexionnelles décrites par les grammaires, vu que leur utilisation n'est pas fréquente. Ils ne sont pas inclus non plus dans les paradigmes qu'on peut voir sur les portails linguistiques <https://goroh.pp.ua>, <https://slovnyk.ua> et <http://www.mova.info>. Par ailleurs, si nous voulons préparer un outil qui puisse couvrir toute la langue, il faut bien évidemment proposer une solution pour reconnaître ces formes.

Les classes flexionnelles des adjectifs

En ukrainien, les adjectifs qualificatifs et relatifs sont répartis entre trois classes flexionnelles : groupe dur (modèle « ВЕЛИКИЙ »), groupe mou (modèle « СИНІЙ ») et groupe en « yod » (modèle « БЕЗКРАЙІ »).

Nous pouvons voir les terminaisons de ces trois classes dans le Tableau 12. La partie variable est séparée du radical par un trait d'union, les changements réguliers sont colorés en vert, les formes non contractées, en rouge et les formes contractées, en bleu.

¹⁴⁷ L'exemple est tiré du site <https://www.pisni.org.ua/songs/7204969.html>

¹⁴⁸ L'exemple est tiré du site <http://litopys.org.ua/shevchenko/shev2120.htm>

Quant aux adjectifs d'appartenance, ils ont la forme courte pour le nominatif et l'accusatif inanimé masculin, alors que pour les autres grammèmes, ils ont les mêmes terminaisons que les adjectifs qualificatifs et relatifs du groupe dur (modèle « ВЕЛИКИЙ »).

Tableau 12. Les trois classes flexionnelles des adjectifs

Nombre	Genre	Cas	Modèle « ВЕЛИКИЙ »	Modèle « СИНІЙ »	Modèle « БЕЗКРАЙ »
Singulier	Masculin	Nominatif	велик- ий	син- ій	безкра- ій
		Accusatif inanimé	велик- ий	син- ій	безкра- ій
		Accusatif animé	велик- ого	син- ього	безкра- його
		Génitif	велик- ого	син- ього	безкра- його
		Locatif	велик- ому	син- ьому	безкра- йому
		Locatif contracté	велик- ім	син- ім	безкра- ім
		Datif	велик- ому	син- ьому	безкра- йому
		Instrumental	велик- им	син- ім	безкра- ім
	Neutre	Nominatif	велик- е	син- є	безкра- є
		Nominatif non contracté	велик- ее	син- єє	–
		Accusatif	велик- е	син- є	безкра- є
		Accusatif non contracté	велик- ее	син- єє	–
		Génitif	велик- ого	син- ього	безкра- його
		Locatif	велик- ому	син- ьому	безкра- йому
		Locatif contracté	велик- ім	син- ім	безкра- ім
		Datif	велик- ому	син- ьому	безкра- йому
	Instrumental	велик- им	син- ім	безкра- ім	
	Féminin	Nominatif	велик- а	син- я	безкра- я
		Nominatif non contracté	велик- ая	син- яя	–
		Accusatif	велик- у	син- ю	Безкраю
		Accusatif non contracté	велик- ую	син- юю	–
Génitif		велик- ої	син- ьої	безкра- йої	
Locatif		велик- ій	син- ій	безкра- ій	
Datif		велик- ій	син- ій	безкра- ій	
Instrumental		велик- ою	син- ьою	безкра- йою	
Pluriel	Nominatif	велик- і	син- і	безкра- ї	
	Nominatif non contracté	велик- ії	син- ії	–	
	Accusatif animé	велик- их	син- іх	безкра- їх	
	Accusatif inanimé	велик- і	син- і	безкра- ї	
	Accusatif non contracté	велик- ії	син- ії	–	
	Génitif	велик- их	син- іх	безкра- їх	
	Locatif	велик- их	син- іх	безкра- їх	
	Datif	велик- им	син- ім	безкра- їм	
Instrumental	велик- ими	син- іми	безкра- їми		

Par ailleurs, les adjectifs d'appartenance qui ont été créés à partir de noms masculins subissent un changement de radical dès qu'une terminaison leur a été attribuée. Ainsi, les adjectifs créés à partir des noms du groupe dur subiront un changement du type 1, ceux créés à partir des noms du groupe mou, un changement du type 2, et à partir des noms en « yod », un changement du type 3. Nous pouvons voir ces trois types de changement dans le Tableau 13.

Tableau 13. Changements dans le radical des adjectifs d'appartenance

Grammèmes	Radical du type 1 « БАТЬКІВ »	Radical du type 2 « ДІДУСІВ »	Radical du type 3 « СЕРГІЇВ »
N. et Acc. inanimé masc.	батьк ів	дідус ів	Сергі їв
Autres grammèmes	батьк ов-	дідус ев-	Сергі єв-

Quant aux adjectifs d'appartenance créés à partir de noms féminins, ils n'ont aucun changement de radical, et se déclinent selon le modèle « ВЕЛИКИЙ ». Nous pouvons voir deux modèles de flexion (« МАМИН » et « БАТЬКІВ ») pour les adjectifs d'appartenance dans le Tableau 14.

Tableau 14. Les modèles de flexion « МАМИН » et « БАТЬКІВ »

Nombre	Genre	Cas	Modèle « МАМИН »	Modèle « БАТЬКІВ »
Singulier	Masculin	Nominatif	мамин	батьків
		Génitif	мамин-ого	батьков-ого
		Accusatif inanimé	мамин	Батьків
		Accusatif animé	мамин-ого	батьков-ого
		Datif	мамин-ому	батьков-ому
		Instrumental	мамин-им	батьков-им
		Locatif	мамин-ому	батьков-ому
		Locatif contracté	мамин-ім	батьков-ім
	Neutre	Nominatif	мамин-е	батьков-е
		Génitif	мамин-ого	батьков-ого
		Accusatif	мамин-е	батьков-е
		Datif	мамин-ому	батьков-ому
		Instrumental	мамин-им	батьков-им
		Locatif	мамин-ому	батьков-ому
		Locatif contracté	мамин-ім	батьков-ім
		Féminin	Nominatif	мамин-а
	Génitif		мамин-ої	батьков-ої
	Accusatif		мамин-у	батьков-у
	Datif		мамин-ій	батьков-ій
	Instrumental		мамин-ою	батьков-ою
	Locatif		мамин-ій	батьков-ій
Pluriel	Nominatif	мамин-і	батьков-і	
	Génitif	мамин-их	батьков-их	
	Accusatif animé	мамин-их	батьков-их	
	Accusatif inanimé	мамин-і	батьков-і	
	Datif	мамин-им	батьков-им	
	Instrumental	мамин-ими	батьков-ими	
	Locatif	мамин-их	батьков-их	

Dans les Tableaux 12 et 14, nous pouvons remarquer que les paradigmes des adjectifs ont beaucoup de flexions homonymiques, car les formes du masculin et du neutre ont les mêmes terminaisons pour tous les cas, hormis le nominatif. De plus, dans certains cas, à l'intérieur du même genre, il y a aussi des flexions homonymiques. Ainsi, lors de la préparation du module ukrainien, il faudrait absolument prendre en compte ce phénomène linguistique et trouver le moyen de lever ces ambiguïtés.

Le comparatif, le superlatif et le superlatif absolu des adjectifs

Comme nous l'avons dit plus tôt, en ukrainien, les adjectifs qualitatifs peuvent exprimer le niveau d'intensité (ou degré de comparaison) et, de ce fait, avoir la forme du comparatif, du superlatif et du superlatif absolu.

Le comparatif indique le degré d'intensité (ou le premier degré de comparaison), il peut avoir la forme synthétique, exprimée par le suffixe *-ш-* ou *-іш-* qu'on ajoute au radical de l'adjectif qualitatif, et la forme analytique, exprimée par l'adverbe *більш* [plus] ou *менш* [moins] suivi d'un adjectif qualitatif.

- partiellement, comme dans
зеленувате яблуко [la pomme **un peu verte**],
напівзелене яблуко [la pomme **à moitié verte**];
- excessivement, comme dans
зеленюще яблуко [la pomme **trop verte**];
- ou encore avec l'affection du locuteur, comme dans
зелененьке яблуко [la **p'tite** pomme **verte**].

Les adjectifs qui expriment une propriété partielle, que nous allons appeler ici « adjectifs partitifs », se forment grâce aux :

- suffixes **-уват-** (-юват-), **-ав-** (-яв-), comme pour
гіркий [amère] → *гіркуватий* [un peu amère],
коричневий [marron] → *коричнюватий* [un peu marron, qui a un reflet marron],
солодкий [sucré] → *солодкавий* [légèrement sucré],
чорний [noir] → *чорнявий* [légèrement noir];
- préfixe **напів-** qui signifie « à moitié », comme pour
сухий [sec] → *напівсухий* [à moitié sec].

Les adjectifs qui expriment une propriété excessive, que nous allons appeler ici « adjectifs excessifs », se forment grâce aux :

- suffixes **-езн-**, **-елезн-**, **-енн-**, **-ющ-**, **-елющ-**, comme pour
товстий [gros] → *товстенний*, *товсте́зний*, *товстеле́зний*, *товстю́щий*,
товстелю́щий [très gros];
- préfixes **над-**, **пра-**, **пре-**, sachant que le préfixe **пре-** peut être combiné avec le suffixe ou non, comme pour
довгий [longue] → *предовжелезний* [excessivement longue],
веселий [joyeux] → *надвеселий* [trop joyeux],
старий [vieux] → *прастарий* [archi-vieux].

En ukrainien, les adjectifs qualitatifs ont aussi des formes diminutives qui expriment soit l'affection ou l'ironie du locuteur, soit l'affection et la petitesse à la fois. Les adjectifs diminutifs sont formés par l'ajout des suffixes **-еньк-**, **-есеньк-**, **-ісіньк-**, **-усіньк-** (**-юсіньк-**) au lexème d'un adjectif, comme pour :

- *малий* [petit] → *маленький*, *малесенький*, *малюсінький* ;
- *давній* [ancien] → *давненький*, *давнесенький*, *давнісінький*.

Les adjectifs qualitatifs ont des antonymes qui sont exprimés :

- soit par un adjectif qui a un lexème complètement différent, comme pour
товстий [gros] → *худий* [maigre] ;
- soit par un morphème **не-** qui ajoute un sens négatif, comme pour
гарний [beau] → *негарний* [pas beau, laid].

Ces derniers sont appelés ici « adjectifs négatifs ».

Dans le Tableau 16, nous pouvons voir quelques exemples¹⁴⁹ des adjectifs décrits ici, en contexte.

Tableau 16. Exemples d'adjectifs partitifs, excessifs et diminutifs en contexte

Forme	Exemple	Traduction
Partielle	<i>Вода в джерелах гіркувата, пахне аптекою.</i>	<i>L'eau dans la source est un peu amère, elle a l'odeur de la pharmacie. (I. Bakhmout)</i>
Excessive	<i>В руках у його був превеликий бутель доброї горілки.</i>	<i>Dans ses mains, il tenait une énorme bouteille d'eau-de-vie. (I. Netchouy-Levyts'kyy)</i>
Diminutive	<i>Його хата була малюсінька світличка з одним вікном.</i>	<i>Sa maison, c'était une toute petite pièce avec une seule fenêtre. (B. Hrintchenko)</i>
Négative	<i>Одне, що в батька було некрасиве, – одяг. Ну такий носив одяг негарний, такий безбарвний, убогий!</i>	<i>La seule chose qui ne fût pas belle chez mon père, c'étaient ses vêtements. Il portait des vêtements si laids, si ternes, misérables ! (O. Dovjenko)</i>

Il est important de noter que les adjectifs partitifs, excessifs et diminutifs, et surtout négatifs, sont souvent exclus des dictionnaires, donc encore une fois, il faut trouver le moyen pour les inclure dans les ressources ukrainiennes pour NooJ.

3.3.3 Le numéral

Description des numéraux

En ukrainien, les ALU catégorisées dans cette classe indiquent un nombre ou un rang.

Les numéraux se divisent en :

- cardinaux qui indiquent les noms de nombre, comme *один [un], два [deux]* ;
- collectifs qui indiquent le nombre d'objets dans leur ensemble, leur unité indivisible, comme *двоє, двійко [les deux], обидва [tous les deux]* ;
- ordinaux qui indiquent le rang et ont les terminaisons adjectivales, comme *перший [premier], другий [deuxième]* ;
- fractionnels qui indiquent les noms des fractions, comme *три цілих і вісім десятих [trois unités et huit dixièmes]* ;
- indéterminés, créés par conversion des adverbes, et qui, de ce fait, deviennent déclinables, comme *багато [beaucoup] → мати багатьох друзів [avoir beaucoup d'amis]*.

Les numéraux s'utilisent avec un nom ou un pronom, sauf dans les équations mathématiques.

Les numéraux peuvent être :

- simples, composés d'une seule ALU, comme *один [un], двадцять [vingt]* ;
- composés de plusieurs ALU, comme *двадцять один [vingt-et-un]*.

¹⁴⁹ Tous ces exemples sont tirés du site <https://goroh.pp.ua>

La déclinaison du numéral

Les numéraux cardinaux, collectifs et ordinaux possèdent la catégorie du cas. Pour quelques numéraux invariables, la catégorie du cas est exprimée syntaxiquement par le contexte, car ils s'utilisent avec les compléments qui expriment le cas, comme dans

*Жменю рубаного коріння залийте **півтора літрами** води і прокип'ятить. [Versez une poignée de racines hachées dans un **litre et demi d'eau** et faites bouillir.] (www.unian.ua).*

Les cardinaux *один [un]*, *два [deux]* et tous les ordinaux possèdent la catégorie de genre (masculin, féminin, et neutre) et de nombre (singulier et pluriel), sauf *два*.

Le cardinal *один [un]* et les ordinaux possèdent également la catégorie de l'animé/inanimé.

En ukrainien, il existe plusieurs schémas de production des formes fléchies de numéraux. Notons au passage que dans les numéraux ordinaux et cardinaux composés, seule la dernière composante est déclinée.

Certains numéraux cardinaux se déclinent comme les noms :

- *мільйон [million]* se décline comme un nom de la classe flexionnelle II, groupe dur ;
- *нуль [zéro]* se décline comme un nom de la classe flexionnelle II, groupe mou ;
- *тисяча [millier]* se décline comme un nom de la classe flexionnelle I, groupe mixte.

Historiquement, les numéraux cardinaux de 5 à 19, les dizaines 20-30 et 50-80, ainsi que les centaines de 200 à 900 à l'instrumental ont deux formes parallèles synonymiques.

Dans la linguistique ukrainienne, on propose quelques schémas de déclinaison pour les numéraux cardinaux. Par exemple, on indique que les nombres 3 et 4 se déclinent comme 2, sauf que le nombre 2 a le nominatif et l'accusatif féminin, alors que 3 et 4 – non, de plus, le signe mou apparaît avant la terminaison dans la déclinaison de 3 et 4 (voir Tableau 18).

On indique également que les nombres de 5 à 20 et le nombre 30 se déclinent selon le même schéma, hormis les numéraux cardinaux 5, 6, 7 et 8, qui subissent des modifications différentes (voir Tableau 19).

Nous avons décidé de refaire les modèles pour prendre en compte les lettres mobiles (voir Tableaux de 17 à 21). Pour établir nos modèles, nous avons consulté les dernières normes (Normes orthographiques 2019, pp. 107-110).

Tableau 17. Le paradigme de 1

Cas	ОДИН (1)			
	Masculin	Neutre	Féminin	Pluriel
Nominatif	<i>один</i>	<i>один-е</i>	<i>один-а</i>	<i>один-і</i>
Accusatif inanimé	<i>один</i>	<i>один-е</i>	<i>один-у</i>	<i>один-і</i>
Accusatif animé	<i>один-ого</i>			<i>один-их</i>
Génitif	<i>один-ого</i>		<i>один-ої, один-ієї</i>	<i>один-их</i>
Locatif	<i>один-ому, один-іму</i>		<i>один-ій</i>	<i>один-ими</i>
Datif	<i>один-ому</i>		<i>один-ій</i>	<i>один-им</i>
Instrumental	<i>один-им</i>		<i>один-ою, один-ією</i>	<i>один-их</i>

Tableau 18. Les paradigmes de 2, 3, 4

Cas	ДВА (2)	ТРИ (3)	ЧОТИРИ (4)
Nominatif m. et n.	<i>дв-а</i>	<i>тр-и</i>	<i>чотир-и</i>
Nominatif f.	<i>дв-і</i>		
Accusatif m. et n.	<i>дв-а</i>	<i>тр-и</i>	<i>чотир-и</i>
Accusatif f.	<i>дв-і</i>		
Accusatif second	<i>дв-ох</i>	<i>тр-ьох</i>	<i>чотир-ьох</i>
Génitif	<i>дв-ох</i>	<i>тр-ьох</i>	<i>чотир-ьох</i>
Locatif	<i>дв-ох</i>	<i>тр-ьох</i>	<i>чотир-ьох</i>
Datif	<i>дв-ом</i>	<i>тр-ьом</i>	<i>чотир-ьом</i>
Instrumental	<i>дв-ома</i>	<i>тр-ьома</i>	<i>чотир-ма</i>

Tableau 19. Les paradigmes de 5, 6, 7, 8, 50

Cas	П'ЯТЬ (5)	ШІСТЬ (6)	СІМ (7)	ВІСІМ (8)	П'ЯТДЕСЯТ (50)
Nominatif	<i>п'ят-ь</i>	<i>шіст-ь</i>	<i>сім</i>	<i>вісім</i>	<i>п'ятдесят</i>
Accusatif	<i>п'ят-ьох</i>	<i>шіст-ьох</i>	<i>сім-ох</i>	<i>вісьм-ох</i>	<i>п'ятдесят-ьох</i>
Génitif	<i>п'ят-и</i>	<i>шіст-и</i>	<i>сім-и</i>	<i>вісьм-и</i>	<i>п'ятдесят-и</i>
Locatif	<i>п'ят-ьох</i>	<i>шіст-ьох</i>	<i>сім-ох</i>	<i>вісьм-ох</i>	<i>п'ятдесят-ьох</i>
Datif	<i>п'ят-и</i>	<i>шіст-и</i>	<i>сім-и</i>	<i>вісьм-и</i>	<i>п'ятдесят-и</i>
Instrumental	<i>п'ят-ьома</i>	<i>шіст-ьома</i>	<i>сім-ома</i>	<i>вісьм-ома</i>	<i>п'ятдесят-ьома</i>

Tableau 20. Les paradigmes de 40 et 90

Cas	СОРОК (40)	ДЕВ'ЯНОСТО (90)
Nominatif	<i>сорок</i>	<i>дев'яност-о</i>
Accusatif	<i>сорок</i>	<i>дев'яност-о</i>
Génitif	<i>сорок-а</i>	<i>дев'яност-а</i>
Locatif	<i>сорок-а</i>	<i>дев'яност-а</i>
Datif	<i>сорок-а</i>	<i>дев'яност-а</i>
Instrumental	<i>сорок-а</i>	<i>дев'яност-а</i>

Tableau 21. Les paradigmes de 200 et 300

Cas	ДВІСТІ (200)	П'ЯТСОТ (500)
Nominatif	<i>дві-сті</i>	<i>п'ят-сот</i>
Accusatif	<i>дві-сті</i>	<i>п'ят-сот</i>
Génitif	<i>двох-сот</i>	<i>п'яти-сот</i>
Locatif	<i>двох-стах</i>	<i>трьох-стах</i>
Datif	<i>двом-стам</i>	<i>п'яти-стам</i>
Instrumental	<i>двома-стами</i>	<i>п'ятьма-стами, п'ятьома-стами</i>

Notons que les numéraux de 9 à 20 et 30 se déclinent selon le modèle « П'ЯТЬ » (5), le numéral 100 selon le modèle « ДЕВ'ЯНОСТО » (90), les numéraux 300 et 400 selon le modèle « ДВІСТІ » (200), les numéraux de 600 à 900 selon le modèle « П'ЯТСОТ » (500).

Les numéraux indéterminés comme *кілька* [quelque] se déclinent selon le modèle « ДВА » (au masculin uniquement) et le numéral *багато* [beaucoup] selon le modèle « ТРИ ».

Les numéraux collectifs de 2 à 20 et le collectif 30 (il n'existe pas de collectifs au-delà) se déclinent comme les numéraux cardinaux correspondants, autrement dit, ils se distinguent uniquement au nominatif et à l'accusatif.

Les numéraux ordinaux se déclinent comme les adjectifs, selon le modèle « СИНІЙ » pour *третій* [troisième], et « ВЕЛИКИЙ » pour les autres (voir Tableau 12).

3.3.4 Le pronom

Description des pronoms

En ukrainien, les ALU qui font partie de la classe grammaticale des pronoms remplissent dans la phrase les mêmes fonctions que le nom, l'adjectif ou le numéral, mais sans nommer directement le substantif et sans indiquer les caractéristiques précises.

En général, les pronoms sont divisés en :

- pronoms personnels (ou pronoms nominaux), qui remplacent dans la phrase un nom et qui désignent quelqu'un à la première et à la deuxième personne ou représentent quelqu'un ou quelque chose à la troisième personne, comme
я [je], *ми* [nous], *він* [il] ;
- pronoms possessifs (ou pronoms-adjectifs possessifs), qui indiquent le possesseur (la première, deuxième ou la troisième personne), comme
мій [mon (mien)], *твій* [ton (tien)], *наш* [notre (nôtre)] ;
- pronoms démonstratifs (ou pronoms-adjectifs démonstratifs) qui indiquent le nom impliqué dans le discours, ou le nom qui a déjà été mentionné avant, ou le nom présent à l'esprit du destinataire, comme
цей [celui-ci], *пар той* [celui-là] ;
- pronoms déterminés (ou pronoms-adjectifs déterminés), qui précisent le sens d'un nom ou d'un groupe nominal en indiquant une caractéristique généralisée, comme
кожен [chaque], *інший* [autre] ;
- pronoms interrogatifs et relatifs, qui sont utilisés pour formuler une interrogation sur l'identité ou la propriété d'un nom, ou bien pour introduire la phrase subordonnée, dans laquelle il faut faire la relation avec quelque chose qui a été mentionné dans la phrase principale¹⁵⁰. On distingue deux types de pronoms interrogatifs et relatifs :
 - o pronoms nominaux, comme
хто [qui], *що* [que],
 - o pronoms-adjectifs, comme
який [quel], *чий* [à qui] ;
- pronoms indéterminés, qui indiquent que la propriété est inconnue du locuteur ou minimisent l'importance de la propriété ; les pronoms de ce type ont été créés par ajout d'une particule au pronom interrogatif, c'est pourquoi ils peuvent être divisés en :
 - o pronoms nominaux, comme
щось [quelque chose], *що-небудь* [n'importe quoi],

¹⁵⁰ Rappelons seulement qu'une question sur l'action et sa qualité est formulée grâce à un adverbe interrogatif.

- pronoms-adjectifs, comme
якийсь [un certain], який-небудь [n'importe lequel] ;
- pronoms négatifs, formés par ajout de la particule *ні-* au pronom interrogatif et qui peuvent être divisés en pronoms nominaux et pronoms-adjectifs, comme
ніхто [personne], нічого [rien], нічий [à personne].

Les déclinaisons des pronoms

Tous les pronoms-adjectifs ont les mêmes catégories grammaticales que les adjectifs (le cas, le nombre et le genre) et les mêmes flexions ; de ce fait, ils sont, pour la plupart, déclinés selon les mêmes modèles que les adjectifs (par exemple, le pronom interrogatif *який [quel]* se décline selon le modèle « ВЕЛИКИЙ » et le pronom possessif *їхній [leur]* – selon le modèle « СИНІЙ »)¹⁵¹, à une différence près : l'absence des formes longues et des formes courtes. Toutefois, certains pronoms-adjectifs ont quelques particularités.

Par exemple, *наш [notre], ваш [votre]* et *кожен [chaque]* ont au nominatif et à l'accusatif inanimé masculin singulier la terminaison « zéro », mais le reste du paradigme est équivalent au paradigme adjectival¹⁵². De plus, *кожен [chaque]* a le *-e-* mobile. Il y a également le possessif *мій [mon, mien]*¹⁵³, le possessif interrogatif *чий [à qui]*, le déterminatif (*у)весь [entier]*, les démonstratifs *цей [celui-ci], той [celui-là]* qui ont la terminaison « zéro » au nominatif masculin singulier, mais ils subissent des modifications différentes (voir Tableau 22).

Tableau 22. Les paradigmes des pronoms-adjectifs

	Genre	Cas	Modèle « НАШ »	Modèle « КОЖЕН »	Modèle « МІЙ »	Modèle « ЧИЙ »	Modèle « ВЕСЬ »	Modèle « ТОЙ »	Modèle « ЦЕЙ »
Singulier	Masculin	Nom.	<i>наш</i>	<i>кожен</i>	<i>мі-ї</i>	<i>чи-ї</i>	<i>весь-ь</i>	<i>т-ої</i>	<i>ц-ей</i>
		Acc. inan.	<i>наш</i>	<i>кожен</i>	<i>мі-ї</i>	<i>чи-ї</i>	<i>весь-ь</i>	<i>т-ої</i>	<i>ц-ей</i>
		Acc an.	<i>наш-ого</i>	<i>кожн-ого</i>	<i>мо-го</i>	<i>чи-його</i>	<i>вс-ього</i>	<i>т-ого</i>	<i>ц-ього</i>
		Gén.	<i>наш-ого</i>	<i>кожн-ого</i>	<i>мо-го</i>	<i>чи-його</i>	<i>вс-ього</i>	<i>т-ого</i>	<i>ц-ього</i>
		Loc.	<i>наш-ому</i>	<i>кожн-ому</i>	<i>мо-єму</i>	<i>чи-йому</i>	<i>вс-ьому</i>	<i>т-ому</i>	<i>ц-ьому</i>
		Loc. contracté	<i>наш-ім</i>	<i>кожн-ім</i>	<i>мо-ім</i>	<i>чи-ім</i>	<i>вс-ім</i>	<i>т-ім</i>	<i>ц-ім</i>
		Dat.	<i>наш-ому</i>	<i>кожн-ому</i>	<i>мо-єму</i>	<i>чи-єму</i>	<i>вс-ьому</i>	<i>т-ому</i>	<i>ц-ьому</i>
		Inst.	<i>наш-им</i>	<i>кожн-им</i>	<i>мо-ім</i>	<i>чи-ім</i>	<i>вс-ім</i>	<i>т-им</i>	<i>ц-им</i>
	Neutre	Nominatif	<i>наш-е</i>	<i>кожн-е</i>	<i>мо-є</i>	<i>чи-є</i>	<i>вс-е</i>	<i>т-ої</i>	<i>ц-е</i>
		Accusatif	<i>наш-е</i>	<i>кожн-е</i>	<i>мі-ї</i>	<i>чи-є</i>	<i>вс-е</i>	<i>т-ої</i>	<i>ц-е</i>
		Gén.	<i>наш-ого</i>	<i>кожн-ого</i>	<i>мо-го</i>	<i>чи-його</i>	<i>вс-ього</i>	<i>т-ого</i>	<i>ц-ього</i>
		Loc.	<i>наш-ому</i>	<i>кожн-ому</i>	<i>мо-єму</i>	<i>чи-йому</i>	<i>вс-ьому</i>	<i>т-ому</i>	<i>ц-ьому</i>
		Loc. contracté	<i>наш-ім</i>	<i>кожн-ім</i>	<i>мо-ім</i>	<i>чи-ім</i>	<i>вс-ім</i>	<i>т-ім</i>	<i>ц-ім</i>
		Dat.	<i>наш-ому</i>	<i>кожн-ому</i>	<i>мо-єму</i>	<i>чи-єму</i>	<i>вс-ьому</i>	<i>т-ому</i>	<i>ц-ьому</i>
Inst.	<i>наш-им</i>	<i>кожн-им</i>	<i>мо-ім</i>	<i>чи-ім</i>	<i>вс-ім</i>	<i>т-им</i>	<i>ц-им</i>		

¹⁵¹ Voir le Tableau 12.

¹⁵² Hormis le nominatif et l'accusatif masculin singulier, *наш [notre], ваш [votre], кожен [chaque]* se déclinent comme les adjectifs du modèle « ВЕЛИКИЙ ».

¹⁵³ Le possessif *мій [ton, tien]* et le possessif réfléchi *свій [à soi, sien]* se déclinent selon le même modèle que le possessif *мій [mon, mien]*.

Genre	Cas	Modèle « НАШ »	Modèle « КОЖН »	Modèle « МІЙ »	Modèle « ЧИЙ »	Modèle « ВЕСЬ »	Modèle « ТОЙ »	Modèle « ЦЕЙ »
Féminin	Nom.	наш-а	кожн-а	мо-я	чи-я	вс-я	т-а	ц-я
	Acc.	наш-у	кожн-у	мо-ю	чи-ю	вс-ю	т-у	ц-ю
	Gén.	наш-ої	кожн-ої	мо-єї	чи-єї	вс-ієї	т-ої	ц-єї
	Loc.	наш-іў	кожн-іў	мо-їў	чи-їў	вс-іў	т-іў	ц-іў
	Dat.	наш-іў	кожн-іў	мо-їў	чи-їў	вс-іў	т-іў	ц-іў
	Inst.	наш-ою	кожн-ою	мо-єю	чи-єю	вс-ією	т-ою	ц-ією
Pluriel	Nom.	наш-і	кожн-і	мо-ї	чи-ї	вс-і	т-і	ц-і
	Acc. inan.	наш-і	кожн-і	мо-ї	чи-ї	вс-і	т-і	ц-і
	Acc an.	наш-ух	кожн-ух	мо-їх	чи-їх	вс-їх	т-ух	ц-ух
	Gén.	наш-ух	кожн-ух	мо-їх	чи-їх	вс-їх	т-ух	ц-ух
	Loc.	наш-ух	кожн-ух	мо-їх	чи-їх	вс-їх	т-ух	ц-ух
	Dat.	наш-ум	кожн-ум	мо-їм	чи-їм	вс-їм	т-ум	ц-ум
	Inst.	наш-уму	кожн-уму	мо-їму	чи-їму	вс-їма	т-уму	ц-уму

Par ailleurs, tous les pronoms personnels et les pronoms interrogatifs *xmo* [qui], *що* [que] ont leurs propres modèles flexionnels. Les pronoms personnels possèdent les catégories de cas, de nombre (sauf le pronom personnel réfléchi *себе*) et de genre (mais uniquement pour la troisième personne du singulier).

Les pronoms nominaux déterminés ou indéterminés possèdent uniquement la catégorie de cas. Nous pouvons voir les déclinaisons des pronoms nominaux personnels dans le Tableau 23 et interrogatifs dans le Tableau 24.

Notons que le pronom personnel réfléchi *себе* se décline sur le modèle « ТИ », mais il n'a pas de nominatif ; dans les pronoms indéterminés seule la première partie se décline selon les modèles des pronoms interrogatifs correspondants, puis les pronoms négatifs se déclinent eux aussi sur les modèles des interrogatifs correspondants.

Tableau 23. Les paradigmes des pronoms personnels

Cas	Modèle « Я »	Modèle « ТИ »	Modèle « ВІН »	Modèle « ВОНО »	Modèle « ВОНА »	Modèle « МИ »	Modèle « ВИ »	Modèle « ВОНИ »
Nom.	я	ти	він	воно	вона	ми	ви	вони
Acc.	мене	тебе	нього, його	нього, його	неї, її	нас	вас	них, їх
Gén.	мене	тебе	нього, його	нього, його	неї, її	нас	вас	них, їх
Loc.	мені	тобі	ньому	ньому	її	нас	вас	них
Dat.	мені	тобі	йому	йому	її	нам	вам	їм
Inst.	мною	тобою	ним	ним	нею	нами	вами	ними

Tableau 24. Les paradigmes des pronoms *xmo* et *що*

Cas	Modèle « ХТО »	Modèle « ЦО »
Nom.	хто	що
Acc.	кого	що
Gén.	кого	чого
Loc.	кому, кім	чому, чім
Dat.	кому	чому
Inst.	ким	чим

3.3.5 Le verbe

Description des verbes

Les ALU qui font partie de la classe grammaticale des verbes expriment une action, un processus ou un état, comme

- читати [lire], робити [faire] (action) ;
- зеленіти [devenir vert], грюкнути (двери) [claquer (une porte)], віяти (вітер) [souffler (le vent)] (processus) ;
- спати [dormir], лежати (книжка) [être posé à plat (livre)] (état).

En ukrainien, les verbes présentent les catégories grammaticales suivantes :

- l'aspect, exprimé par un préfixe ou un suffixe, ou encore par l'accent de mot ; à cette catégorie est liée la notion de couple aspectuel, composé de deux verbes : imperfectif (IPF) – perfectif (PF), qui sont unis par la sémantique, mais le PF focalise l'attention sur le résultat de l'action, tandis que l'IPF focalise l'attention sur son déroulement, comme читати (IPF) – **про-читати** (PF) [lire être en train de lire – lire terminer de lire] ;
- les sous-aspects déterminé et indéterminé, où s'opposent deux verbes imperfectifs de mouvement ; le verbe déterminé désigne un mouvement unidirectionnel dans la direction indiquée et/ou avec un but précis ; le verbe indéterminé désigne un mouvement multidirectionnel ou répétitif, habituel, aller-retour au passé, ou la faculté d'accomplir le mouvement, comme бігати (déterminé) – бігати (indéterminé) [courir] ;
- la voix :
 - o active, exprimée par les verbes transitifs et intransitifs, comme жінка **читає** [une femme lit], чоловік **біжить** [un homme court] ;
 - o passive, exprimée soit par ajout du postfixe -ся à un verbe imperfectif transitif, soit par une construction comportant le verbe auxiliaire бути et le participe passé passif perfectif du verbe principal, soit encore par une construction asubjectale avec l'auxiliaire бути et une forme figée issue du participe passé passif perfectif, comme книжка **читається** [le livre est lu], книжка **буде прочитана** [le livre sera lu], книжку **було прочитано** [le livre a été lu] ;
- les verbes réfléchis, où celui qui fait l'action est le sujet syntaxique, comme дитина **миється** [un enfant se lave] ;
- la personne (1^{re}, 2^e, 3^e), exprimée par des terminaisons, comme я чита-ю [je lis] (la 1^{re}), ти чита-єш [tu lis] (la 2^e), він чита-є [il lit] (la 3^e) ;
- le nombre (singulier et pluriel), exprimé par des terminaisons, comme я чита-ю [je lis] (le singulier), ми чита-ємо [nous lisons] (le pluriel) ;
- le genre (masculin, féminin et neutre au singulier), exprimé par un suffixe et des terminaisons, comme він чита-в [il a lu], вона чита-ла [elle a lu], дитя чита-ло [l'enfant a lu] ;
- le mode :
 - o l'infinitif, exprimé par la terminaison -ти ou -ть (forme dialectique, issue de la variante de l'ukrainien de la rive gauche), comme малювати ou bien малювать [peindre], dans

Вона ще добре вмiла **малювати** квітки на фарфорi.
[Elle était également douée pour **peindre** des fleurs sur de la porcelaine. (I. Netchouy-Levyts'kyu)¹⁵⁴];

- l'indicatif (combiné à la notion du temps), exprimé par des terminaisons, comme *ми чита-емо* nous lisons (ici, on dit que le verbe est au présent de l'indicatif),
- l'impératif, exprimé par une terminaison ou par la particule *хай, най, нехай* [que] suivie du verbe au présent ou au présent-futur, comme *чита-й* [lis], *хай чита-є* [qu'il lise], *нехай прочита-є* [qu'il lise],
- l'hypothétique (combiné à la notion du temps), exprimé :
soit par un suffixe et une terminaison (les mêmes qu'au passé indicatif) et la particule *бу(б)*, comme *ми чита-л-и б* [nous lirions],
soit en ajoutant à la forme précédente l'auxiliaire *бути* au passé, on parle alors de l'hypothétique passé, comme *ми були б чита-л-и* [nous aurions lu] ;
- le temps (mis à part l'hypothétique passé, ne concerne que le mode indicatif) :
 - le présent, exprimé par les terminaisons, comme *ми чита-емо* [nous lisons],
 - le passé, exprimé par un suffixe et des terminaisons, comme *ми чита-л-и* [nous lisions],
 - le plus-que-parfait, exprimé par un verbe principal au passé et le verbe auxiliaire *бути* [être] au passé, comme *спробува-л-а бу-л-а* [tenta, avait tenté], dans *Маринка спробувала була кинутись і до Доротеї, але та відповіла різко.* [Marynka **tenta** de se précipiter vers Dorothee, mais celle-ci lui répondit sèchement.] (M. Staryts'kyu)¹⁵⁵ ;
 - le futur, exprimé soit par des terminaisons, soit par le verbe auxiliaire conjugué *бути* [être] et le verbe principal à l'infinitif, soit par un préfixe et les terminaisons du présent, comme *ми чита-ти-memo* [nous lirons] (futur synthétique imperfectif), *ми будемо читати* [nous lirons] (futur analytique imperfectif), *ми про-чита-емо* [nous lirons] (présent-futur),
 - le participe, exprimé par des suffixes et des terminaisons, comme *спочива-юч-ий* [qui se repose], *посиві-л-ий* [grisonnant], *чита-н-ий* [lu], *взя-т-ий* [pris] ;
 - le gérondif exprimé par des suffixes, comme *чита-ючи* [en lisant], *прочита-вши* [ayant lu] ;
 - la forme verbale asubjectale¹⁵⁶, exprimée par des suffixes, comme

¹⁵⁴ Tiré du site <https://lib.com.ua/uk/book/bez-puttia>

¹⁵⁵ Tiré du site <https://www.ukrlib.com.ua>

¹⁵⁶ Dans les grammaires traditionnelles, ces formes sont appelées « impersonnelles », mais nous préférons suivre Oleg Chinkarouk, qui utilise le terme « forme asubjectale », c'est-à-dire privée de sujet syntaxique, pour

прочита-но [lu], поверну-мо [rendu].

Notons que certaines catégories grammaticales mentionnées plus haut ne sont pas propres à toutes les formes. Par exemple, l'infinifit exprime uniquement l'aspect ; le genre est exprimé seulement par les verbes au passé, au plus-que-parfait et à l'hypothétique ; la forme verbale asubjectale, le participe et le gérondif n'expriment pas non plus toutes les catégories (voir plus bas, dans les sections où ces formes sont décrites).

Du point de vue de sa structure, le verbe ukrainien est composé d'un radical, d'une terminaison et éventuellement d'un suffixe, d'un préfixe et du postfixe *-ся*. Notons que le préfixe, le radical et le suffixe forment la base du verbe. Chaque verbe ukrainien a deux bases : la base de l'infinifit et la base du présent, qui peuvent être identiques pour certains verbes et différentes pour d'autres (voir le Tableau 25).

Syntaxiquement, le verbe peut jouer le rôle du :

- prédicat, comme dans
Жінка читає. [Une femme lit.] ;
- sujet, comme dans
*Критично мислити – не означає довіряти комусь або не довіряти, це вміти вбирати інформацію і оцінювати її.
[L'esprit critique ne consiste pas à faire confiance ou à se méfier de quelqu'un, mais à être capable d'absorber des informations et de les évaluer.] (<https://www.aup.com.ua>) ;*
- complément, comme dans
Жінка закінчила читати. [La femme a fini de lire].

Les verbes peuvent également être classés selon les caractéristiques sémantiques suivantes :

- le déplacement, comme
бігати [courir], ходити [marcher], ставити [poser] ;
- la position, comme
стояти [être debout] ;
- l'état, comme
зеленіти [devenir vert], світати [poindre (pour l'aube)] ;
- le déroulement spécifique (début, fin d'action, refaire une action, etc.), comme
заспівати [se mettre à chanter], доспівати [finir de chanter], перечитати [relire].

L'aspect du verbe

Comme nous l'avons déjà indiqué dans notre article sur la formalisation des verbes ukrainiens avec NooJ (Saint-Joanis, Silberztein 2021 [129]), le terme de l'aspect a été utilisé pour la première fois concernant la langue tchèque en 1603 par Vavrinc Benedikt Nedožerský dans sa *Grammaticae Bohemicae* pour distinguer les verbes simples des verbes dérivés (Mazon 1913 [116]). Ses successeurs ont développé cette notion qui est apparue rapidement dans d'autres langues slaves¹⁵⁷.

éviter toute confusion avec les phrases où l'action n'est pas produite par une personne (comme dans la phrase « *Дощить [Il pleut]* »).

¹⁵⁷ Meletiy Smotryts'kyi, recteur du collège de Kyïv, reprend le terme dans sa *Grammaire slave du syntagme correct* publiée en 1619 (Roussaniv's'kyi 2001 [56]).

C'est une catégorie verbale qui se base sur les diverses façons de concevoir le processus exprimé par le verbe dans le temps (Holt 1943). La valeur actuelle des verbes slaves est exprimée par les verbes perfectifs (PF) et imperfectifs (IPF). Les verbes imperfectifs focalisent l'attention sur le processus même de l'action tandis que les verbes perfectifs focalisent l'attention sur le franchissement de la borne finale du procès (autrement dit, sur le résultat, pédagogiquement parlant).

La plupart des verbes imperfectifs et perfectifs forment des couples aspectuels (IPF/PF) qui permettent d'exprimer la même action ou le même état d'un point de vue différent (durée, répétition, etc.).

En linguistique slave, il existe deux écoles de grammairiens qui ont des avis controversés sur le sujet. La première¹⁵⁸ considère qu'une partie des couples aspectuels sont composés de deux formes du même verbe, c'est-à-dire sémantiquement identiques, mais ayant des formes morphologiques différentes. Ajoutons qu'il y a deux types de formes aspectuelles : les « préfixaux » et les « suffixaux ». Les « préfixaux » ont le même radical, mais se différencient par les préfixes perfectivants (qui donnent au verbe l'aspect perfectif) tandis que les « suffixaux » ont le même radical et le même préfixe et se différencient par les suffixes imperfectivants (qui donnent au verbe l'aspect imperfectif).

Prenons l'exemple du couple *робити* (IPF) – *зробити* (PF) [*faire*]. La forme imperfective n'a ni suffixe ni préfixe, tandis que la forme perfective est composée du radical *робу-* et du préfixe perfectivant *з-*.

Ce couple peut être employé dans les phrases suivantes :

- *Іван робив вправу дві години.* [*Ivan a fait l'exercice pendant deux heures*].
- *Іван зробив вправу за дві години.* [*Ivan a fait l'exercice en deux heures*].

Les linguistes du deuxième groupe¹⁵⁹ voient dans le couple aspectuel une opposition entre deux verbes différents, mais qui ont une sémantique proche, et où le deuxième verbe est dérivé du premier grâce à un préfixe ou un suffixe (Włodarczyk 1984 [153]).

Dans notre exemple *робити* (IPF) – *зробити* (PF), le verbe *зробити* serait ainsi dérivé du verbe *робити* par ajout du préfixe *з-*, qui a ajouté aussi une information sémantique supplémentaire, celle d'un résultat abouti dans une limite temporelle.

Tout à fait exceptionnellement, un couple aspectuel peut être formé à partir de deux verbes morphologiquement distincts, par exemple *говорити* (IPF) – *сказати* (PF) [*dire*] ou *брати* (IPF) – *взяти* (PF) [*prendre*] ; inversement, un verbe peut se comporter dans une phrase comme perfectif et dans une autre comme imperfectif (les verbes bi-aspectuels).

En ukrainien, il existe 9 préfixes perfectivants¹⁶⁰, appelés « vides » depuis Agrell (1908) ou encore « désémantisés » par Guiraud-Weber (1987 [101])¹⁶¹, qui contribuent à la formation de l'aspect perfectif et ne modifient pas le sens du lexème verbal. Toutefois, il existe en tout 37

¹⁵⁸ Par exemple, (Vykhovanets' 1988 [9]), (Roussaniv's'kyi 2001 [56]), (Vykhovanets' et Horodens'ka 2004 [10]), (Pliouchtch 2010 [52]).

¹⁵⁹ Par exemple, (Karcevski 1927 [107]), (Włodarczyk 1984 [153]), (Gwizdecka 2009 [102]).

¹⁶⁰ Horpynytch (2004) [19] donne 7 préfixes plus 2 variantes du préfixe *з-* : *зі-* et *с-*.

¹⁶¹ Nous empruntons cet argument à André et Hélène Włodarczyk (2001[153]).

variantes morphologiques préfixales verbales simples, 3 doubles, ainsi que des variantes triples¹⁶². Un verbe imperfectif simple peut être associé à plusieurs préfixes qui portent des variations sémantiques. Les variantes peuvent être traduites par un même verbe dans une langue étrangère non slave, comme dans l'exemple cité plus haut, ou alors par des verbes différents, par exemple *переробити* [refaire], *підробити* [avoir un job ; falsifier].

Ces variantes peuvent s'associer à leur tour à d'autres verbes pour former des couples aspectuels :

- *переробити* – *переробл-юва-ти* [refaire] ;
- *підробити* – *підробл-юва-ти* [avoir un job, falsifier].

Notons que ces variantes se traduisent en français par un seul et même verbe.

Il va de soi que les verbes *підроблювати* (IPF) [avoir un job ; falsifier] et *зробити* (PF) [faire] ne constitueront jamais un couple aspectuel, à cause de leur éloignement sémantique ; même les paires de verbes *робити* (IPF) [faire] et *підробити* (PF) [avoir un job ; falsifier] ne trouveront jamais un contexte leur permettant de former un couple aspectuel.

De ce fait, pour désigner un couple du type *робити* (IPF) – *зробити* (PF) [faire], nous utiliserons le terme de couple aspectuel dominant.

En général, les couples aspectuels ne sont pas liés dans les outils de TAL¹⁶³, ce qui complique l'analyse et peut corrompre les statistiques.

De plus, la notion d'aspectualité n'est pas évidente à appréhender pour un étudiant qui apprend une langue slave. Ainsi, lorsqu'un étudiant doit faire un exercice de transformation et passer d'une phrase imperfective qui marque le présent à la phrase perfective qui marque le futur ou le passé et vice versa, il doit connaître le couple aspectuel. De la même façon, un logiciel TAL qui fait la traduction ou la paraphrase doit connaître le couple aspectuel. De ce fait, il nous paraît important de faire le lien entre les verbes du couple aspectuel.

La catégorie de déterminé/indéterminé

Dans les langues slaves, les verbes de mouvement, c'est-à-dire les verbes qui expriment la manière de se déplacer, font l'objet d'une attention particulière. Chaque verbe qui exprime cette catégorie est lié à un autre verbe dans un couple : le premier désigne le mouvement unique (le verbe déterminé), comme c'est le cas du verbe *їти* [aller à pied], et le second désigne le mouvement répété, séquencé, ou le mouvement qui désigne un aller-retour (le verbe indéterminé), comme c'est le cas du verbe *ходити* [aller à pied]. Cette dernière opposition est particulièrement importante lorsqu'on doit transformer la phrase qui se déroule au présent en phrase qui s'est déroulée au passé. Par exemple :

- 1) Я **їду** до музею. → 2) Я вже **ходив** до музею.
[Je **vais** au musée → Je **suis** déjà **allé** au musée.]

¹⁶² Nous prenons en compte toutes les variantes : répandues : *в-* (*ві-*), *у-*, *уві-*; *від-* (*віді-*); *ву-*; *до-* (*ді-*); *з-* (*зі-*, *із-*, *с-*); *за-*; *на-*; *над-* (*наді-*); *о-*, *об-* (*обі-*); *під-* (*піді-*); *пере-*; *по-*; *при-*; *про-*; *роз-* (*розі-*); rares : *пред-*, *пре-*, *обез-*, *зне-*; empruntés : *де-* (*дез-*), *ду-* (*дус-*), *ре-*; double préfixe : *не-до-*; *під-на-*; *по-* + n'importe quel autre préfixe. Dans les verbes à triple préfixe, les deux premiers sont le plus souvent *по-* + *на-*, comme, par exemple : *по-на-ви-гадувати* [inventer plusieurs choses à plusieurs reprises].

¹⁶³ Nous avons vérifié cela dans tous les outils stochastiques et formels présentés dans le chapitre 2.

Dans le premier exemple, l'action a lieu au présent ; ici, le verbe désigne un mouvement unidirectionnel dans la direction indiquée dans la phrase. Dans le deuxième exemple, la même action a eu lieu au passé ; ici, le verbe désigne un mouvement dans la direction indiquée et suppose aussi le retour, c'est-à-dire : « j'y suis allé, mais je suis déjà revenu ».

Les verbes de mouvement peuvent être :

- simples, c'est-à-dire sans préfixe, comme *бізму* (déterminé) – *бізаму* (indéterminé) [*courir*],
- préfixés, comme *перепебізму* (PF) – *перепебізаму* (IPF) [*traverser en courant*].

La catégorie de déterminé/indéterminé est exprimée uniquement par les verbes simples. L'ajout du préfixe remplace l'opposition déterminé/indéterminé par l'opposition PF/IPF.

Tous comme les verbes du couple aspectuel, les verbes du couple déterminé/indéterminé ne sont pas liés dans les outils disponibles pour l'ukrainien¹⁶⁴. Cependant, leur lien est indiqué dans SUM-20. Par exemple, pour le verbe *бізаму*, nous trouvons la description suivante :

« **БІГАТИ**, аю, аеш, недок. 1. Те саме, що **бізти** 1, але з позначенням повторюваності, тривалості дії, що відбувається в різний час або в різних напрямках.

[**БІГАТИ**, -аю, -аеш¹⁶⁵, IPF. 1. identique à **бізти** 1, mais désignant la répétition, la durée d'une action qui se déroule à des moments différents ou dans des directions différentes.]¹⁶⁶ »

Nous pensons qu'il serait important d'indiquer également ce lien dans un outil informatique.

Les classes morphologiques des verbes

Traditionnellement, les verbes ukrainiens sont répartis en deux groupes de conjugaison, selon la terminaison de la troisième personne du présent au pluriel (I^{re} conjugaison = -уть (-ють) ; II^e conjugaison = -ать (-ять)) et entre 10 et 14 classes morphologiques¹⁶⁷. Les classes se répartissent entre productives et improductives. Les classes productives s'enrichissent par des nouveaux verbes à la différence des classes improductives¹⁶⁸.

La répartition en classes morphologiques se fait d'après les suffixes et éventuellement d'après le groupe de conjugaison. Le contexte historique de formation des bases de conjugaison est également pris en compte¹⁶⁹. Notons que pour chaque verbe ukrainien, il existe deux bases :

- la base de l'infinitif, qui est utilisée pour produire les formes fléchies du futur synthétique, du passé (indicatif et hypothétique), du plus-que-parfait, du participe passé, du gérondif antérieur ;

¹⁶⁴ Nous avons vérifié cela dans tous les outils stochastiques et formels, présentés dans le chapitre 2.

¹⁶⁵ Indication de conjugaison.

¹⁶⁶ Voir SUM-20 : <https://sum20ua.com/Entry/index?wordid=4235&page=159>

¹⁶⁷ Il n'existe pas de classification en classe morphologique univoque. Dans (Horpyntch 2004 [19]), il y a 10 classes, dans (Roussaniv's'kyy 1985 [55], Pliouchtch 2010 [52]), il y a 12 classes, dans (Vykhovanets' et Horodens'ka 2004 [10]), il y a 14 classes.

¹⁶⁸ Nous n'en dirons pas plus sur cette classification qui se trouve en dehors du périmètre de nos recherches.

¹⁶⁹ Cela signifie que dans la même classe sont rangés des verbes qui ont subi des modifications (alternance consonantique, disparition d'une voyelle) et d'autres dont le radical n'a pas été modifié.

- la base du présent, qui est utilisée pour produire les formes fléchies du présent, du futur-présent et de l'impératif, du participe présent et du gérondif concomitant.

Nous pouvons voir ces classes dans le Tableau 25, établi sur la base des descriptions données dans (Vykhovanets' et Horodens'ka 2004, p. 281 [10]). Notons que les classifications proposées par d'autres linguistes n'ont que quelques divergences.

Bien que la classification (Tableau 25) soit assez complète du point de vue morphologique, elle n'est pas suffisante lorsque nous voulons l'utiliser pour enseigner la conjugaison des verbes ukrainiens à un étranger, car elle ne prend pas en compte certains phénomènes linguistiques, tels que l'alternance consonantique, la voyelle mobile ou le suffixe mobile au passé, ou encore l'apparition d'une consonne dans la base du présent. De plus, elle ne tient pas compte de la production des participes et des formes asubjectales en *-но*, *-мо*, ou même de la production de l'impératif¹⁷⁰.

Par ailleurs, ces phénomènes linguistiques sont pris en compte dans (Kryts'ka et alii 2011 [35]), mais comme pour les noms, cette classification tient compte de l'accent mobile et, en conséquence, multiple les paradigmes. Elle ne tient pas compte non plus de la production des participes et des formes asubjectales en *-но*, *-мо*.

Pour enseigner l'ukrainien, Oleg Chinkarouk a préparé 72 modèles de conjugaison¹⁷¹, dont nous pouvons voir un exemple dans la Figure 27 et que nous allons utiliser pour préparer nos paradigmes verbaux.

Tableau 25. Les classes morphologiques des verbes

Classes	Base de l'infinitif	Base du présent	Exemples	Groupe
1	-ува- -юва-	-уј-, ¹⁷² -юј-	рис- ува -ти – ноч- у -ють мал- юва -ти – мал- ю -ють	I
2	-а- -я-	-ај- -яј-	чит- а -ти – чит- а -ють мін- я -ти – мін- я -ють	I
3	-і-	-іј-	зелен- і -ти – зелен- і -ють	I
4	-ну-	-н-	зи- ну -ти – зи- н -уть	I
5	∅ ¹⁷³	∅	повз-ти – повз- уть	I
6	-ва-	-вај-	ки- ва -ти – ки- ва -ють	
7	-а-(-я-)	∅	рв- а -ти – рв- уть ві- я -ти – ві- ють	I
8	∅	-ј-	чу-ти – чу- ють бу-ти – б'- ють	I
9	-і-	∅	хот- і -ти – хоч- уть	I
10	-ва-	-ј-	да- ва -ти – да- ють	I
11	-и- (-ї-)	∅	бач- и -ти – бач- ать по- ї -ти – по- ять	II
12	-а-	∅	леж- а -ти – леж- ать	II
13	-і-	∅	сид- і -ти – сид- ять	II
14	∅	∅	біг-ти – біж- ать	II

¹⁷⁰ Les verbes de la même classe avec l'accent à l'infinitif sur la racine ou sur le suffixe n'ont pas la même flexion à l'impératif, comme *люб-і-ти* → *люб-і*, *бáрв-и-ти* → *бáрв-∅*.

¹⁷¹ Il s'agit d'un ouvrage en préparation, auquel j'apporterai aussi une contribution.

¹⁷² Ici, « j » marque le son « yod », qui est transcrit par les lettres я et ю dans les terminaisons.

¹⁷³ Ici, « ∅ » signifie l'absence de suffixe.

Figure 27. Exemple de paradigme verbal

IMPERFECTIF		PERFECTIF	
Base du Présent	Base de l'Infinitif	Base du Présent	Base de l'Infinitif
Présent : читá-ю	Prétérit : читá-в	Présent-Futur : прочита-ю	Prétérit : прочита-в
Impératif : читá-й	Futur synthétique : читáти-му	Impératif : прочита-й	
	Futur analytique : буд-у читáти		
Gérondif concomitant : читá-ючи	Gérondif antérieur : читá-вши		Gérondif antérieur : прочита-вши
	Participe passé passif : читá-ний		Participe passé passif : прочита-ний

Le gérondif

Le gérondif est une forme verbale invariable qui sert à décrire certaines circonstances de l'action. En tant que forme verbale, le gérondif exprime uniquement l'aspect et le temps. On distingue :

- le gérondif concomitant, créé à partir de la base du présent du verbe IPF par ajout des suffixes *-ючи (ючи)* pour les verbes de la première conjugaison et *-ачи (ячи)* pour les verbes de la deuxième conjugaison, comme
чита-му [lire] → чита-ючи [en lisant] ;
- le gérondif antérieur¹⁷⁴, créé à partir de la base de l'infinitif du verbe IPF ou PF par ajout des suffixes *-вши, -ши*, comme
чита-му [lire] → чита-вши [ayant lu].

Dans la phrase, le gérondif joue le rôle d'un second prédicat et s'accorde avec ses compléments de la même façon que le verbe dont il est issu :

- | | |
|--|--------------|
| 1. Le verbe : читати книгу (accusatif) дитині (datif) | 1. lire |
| 2. Le gérondif : читаючи книгу (accusatif) дитині (datif) | 2. en lisant |

Dans la plupart des outils en ligne, les gérondifs ne font pas partie du paradigme verbal et sont traités comme des formes différentes. Nous avons vérifié cela sur l'exemple du gérondif *читаючи [en lisant]* dans VESUM (<https://vesum.nlp.net.ua/?w=читаючи>), Goroh (<https://goroh.pp.ua/Словозміна/читаючи>) et Sketch Engine (<https://app.sketchengine.eu/>) et nous pouvons voir les résultats dans le Tableau 26. Dans aucun des outils le lemme du verbe de départ n'est mentionné. De plus, dans Goroh et Sketch Engine, l'aspect n'est pas indiqué.

Tableau 26. Le gérondif *читаючи* dans des outils existants

	Description complète	Catégorie	Lemme
VESUM	<i>читаючи</i> : advp:imperf	gérondif	читаючи
Goroh	<i>читаючи</i> : verbe, gérondif, présent, non réflexif	verbe	читаючи
Sketch Engin	читаючи <i>читаючи/V-ppp-s-/читаючи-в</i>	verbe	читаючи

Par ailleurs, dans le dictionnaire académique SUM-20, les gérondifs ne sont pas donnés en entrée, cela signifie qu'ils sont traités en tant que formes fléchies d'un verbe. Il nous semble aussi

¹⁷⁴ Plutôt que d'utiliser la terminologie traditionnelle, c'est-à-dire « gérondif présent » et « gérondif passé », nous avons emprunté les termes « gérondif concomitant » et « gérondif antérieur » à Oleg Chinkarouk. Nous considérons que ces termes représentent mieux la nature du gérondif pour l'ukrainien.

important de lier le gérondif au paradigme verbal, d'un point de vue linguistique aussi bien que pratique (recherches des concordances, transformation des phrases).

Le participe

Les participes remplissent à la fois les fonctions du verbe (action, processus, état) et de l'adjectif (caractéristique d'un objet). Ils sont formés par ajout d'un suffixe et de la flexion adjectivale à la base de l'infinitif ou à la base du présent du verbe avec lequel les participes gardent un lien sémantique. En ukrainien, on distingue :

- le participe passé passif, dérivé du verbe IPF ou PF par ajout des suffixes *-н-*, *-ен-*, *-м-* à la base l'infinitif, comme
чита-н-уї [lu], *привез-ен-уї* [amené], *пробу-м-уї* [perforé] ;
- participe passé actif, dérivé du verbe PF par ajout du suffixe *-л-* à la base de l'infinitif¹⁷⁵ du verbe PF, comme
згас-л-уї [éteint] dans
[Nous étions assis près du feu éteint.] (L. Oukraïinka) ;
- participe présent actif, dérivé du verbe IPF par ajout des suffixes à la base du présent *-ач-* (*яч*), *-уч-* (*юч*)¹⁷⁶, comme
ви-юч-уї [hurlant] dans
І вони потягли його, виючого, лугами.
[Et il hurle pendant qu'ils le traînaient à travers les prairies]. (H. Tioutiounnyk)

Notons qu'en ukrainien, les participes actifs ne sont pas productifs, c'est-à-dire qu'ils ne sont pas créés à partir de tous les verbes ; de ce fait, ils sont exclus du paradigme verbal. En ukrainien d'aujourd'hui, ils sont d'usage très limité.

Les participes expriment les catégories verbales telles que l'aspect et le temps et les catégories adjectivales telles que le genre, le nombre et le cas. Ils produisent les formes fléchies selon les mêmes schémas que les adjectifs (voir Tableau 12).

Certains linguistes ukrainiens¹⁷⁷ considèrent que le participe exprime la propriété d'un objet qui résulte d'un processus et, de ce fait, le rapprochent d'un adjectif, tandis que d'autres mettent en avant les caractéristiques de l'objet et de ce fait le considèrent comme l'une des formes du verbe¹⁷⁸. Par ailleurs, certains participes se sont détachés du verbe d'origine et sont devenus des adjectifs (Horpynytsch 2004, p. 216 [21]).

Tout comme les gérondifs, les participes ne sont pas inclus dans le paradigme verbal dans les outils disponibles pour le grand public. Nous avons vérifié cela dans VESUM (<https://vesum.nlp.net.ua/?w=читаючи>), Goroh (<https://goroh.pp.ua/Словозміна/читаний>) et Sketch Engine (<https://app.sketchengine.eu>) (voir Tableau 27). Ici, le verbe de départ n'est pas non plus mentionné et, dans Goroh et Sketch Engine, l'aspect n'est pas indiqué.

¹⁷⁵ Par ailleurs, il s'agit d'une forme ancienne, donc d'autres changements sont observés dans la base du verbe, par exemple, la disparition du suffixe verbal.

¹⁷⁶ T. Chevtchenko dans sa poésie a également utilisé les participes, formés grâce au suffixe *-ящ-* (Horpynytsch 2004, p.217 [19]).

¹⁷⁷ Vyhovanets', Cheveliov, Hrychtchenko dans (Horpynytsch 2004 [19]).

¹⁷⁸ Jovtoubrioukh, Roussaniv's'kyj dans (Horpynytsch 2004 [19]).

Tableau 27. Le participe *читаний* dans des outils existants

	Description complète	Catégorie	Lemme
VESUM	<i>читаний</i> : <i>adj:m:v_naz:etadjp:pasv:imperf:compb</i>	adjectif et participe	читаний
Goroh	<i>читаний</i> : verbe, passif, participe, passé, non réflexif	verbe	читаний
Sketch Engine	читаний читаний/A-pmsnf-pp-/читаний-j	adjectif	читаний

Par ailleurs, nous pouvons trouver des participes comme entrées indépendantes dans le dictionnaire académique SUM-20, mais le verbe de départ y est indiqué, comme dans

БАЧЕНИЙ, а, е. Дієпр. пас. до бачити
« БАЧЕНИЙ » [vu], а [f.], е [n.], *participe passif du verbe « бачити » [voir]*

Donc, pour formaliser les participes, nous devons prendre en compte non seulement leurs paradigmes, mais aussi leur lien sémantique.

Les formes asubjectales en -но, -то

Les formes asubjectales en *-но, -то* sont dérivées des verbes de la même façon que les participes, mais elles restent invariables. À l'origine, ces ALU sont d'anciens participes passés passifs courts au nominatif neutre qui se sont détachés du paradigme et se sont transformés en formes verbales indéclinables qui n'expriment que l'aspect, par exemple :

занес-ти [répertorier] → *занес-ен-ий* → *занесе-но*.

Ces formes sont utilisées en tant que prédicat dans une phrase sans sujet syntaxique, mais avec un COD représentant un objet/patient (comme dans une phrase active), comme dans :

670 птахів занесено до Червоної книги України.
[670 oiseaux **sont répertoriés** dans le Livre rouge de l'Ukraine.] (<https://redbook-ua.org>).

Assez rarement, un agent représenté par un complément d'agent à l'instrumental (comme dans une phrase passive) peut être présent.

Par ailleurs, les formes asubjectales ont été intégrées aux paradigmes verbaux dans tous les outils mentionnés plus haut.

Ainsi, lorsque nous intégrerons ces formes aux ressources ukrainiennes pour NooJ, nous devons veiller à le faire aussi.

3.3.6 L'adverbe

Description des adverbes

En ukrainien, l'adverbe est une classe qui regroupe les ALU invariables, utilisées dans le discours pour modifier ou spécifier le sens d'un verbe, d'un adjectif, d'un autre adverbe ou d'un nom. L'adverbe peut être utilisé en tant que :

- spécification d'un procès (en se rapportant à un verbe), comme *добре працювати* [**bien** travailler] ;
- spécification d'une propriété (en se rapportant à un adjectif), comme *завжди гарний* [**toujours** beau] ;
- modification d'une spécification (en se rapportant à un autre adverbe), comme *дуже швидко* [**très** vite] ;

- spécification d'un objet ou d'une notion (en se rapportant à un nom), comme *котлета по-київськи* [la boulette à la manière kiévienne].

Syntaxiquement, les adverbes peuvent remplir les fonctions de :

- complément de temps, comme *зробити спочатку* [faire d'abord] ;
- complément de lieu, comme *жити далеко* [vivre loin] ;
- complément de manière, comme *зробити швидко* [faire vite] ;
- complément de nom, comme *дорога додому* [le chemin de la maison] ;
- prédicat¹⁷⁹, comme *на вулиці холодно* [dehors, il fait froid].

Un syntagme formé d'un verbe et d'un adverbe peut être complété par un nom ou sans préposition :

- *праворуч від хати* [à droite de la maison] ;
- *працювати краще колеги* [travailler mieux qu'un collègue].

Morphologiquement, les adverbes se distinguent par leurs terminaisons invariables, exprimées par les suffixes suivants :

- *-о*, il s'agit des adverbes formés à partir des adjectifs qualificatifs, comme dans *рано* [tôt] ;
- *-и*, comme dans *інколи* [parfois] ;
- *-е*, comme dans *дуже* [très] ;
- *-а*, comme dans *учора* [hier] ;
- *-∅* (suffixe « zéro », c'est-à-dire sans suffixe), comme dans *тепер(∅)* [maintenant] ;
- *-і*, comme dans *сьогодні* [aujourd'hui] ;
- *-ом, -ем, -ою, -єю*, il s'agit des noms à l'instrumental qui sont passés dans la classe des adverbes, comme *часом* [parfois], *бігцем* [en courant], *стрілою* [très vite (comme une flèche)], *пулею* [très vite (comme une balle)] ;

¹⁷⁹ Comme nous l'avons indiqué précédemment dans la section 3.2.2, certains linguistes mettent les adverbes de ce type dans une classe à part.

- *-ськи, -зьки, -цьки*, il s'agit des adverbes formés à partir des adjectifs relatifs, comme *по-українськи* [à l'ukrainienne], *по-французьки* [à la française], *по-турецьки* [à la turque].

Certains adverbes ont été formés par contraction d'une préposition et d'un substantif. Dans ce cas, l'adverbe a gardé la terminaison de la forme du départ, qui est devenue invariable. On peut distinguer les adverbes qui sont composés :

- d'un morphe antéposé qui peut s'écrire séparément de la racine *в-* (*у-*), *на-* + le suffixe *-у* ou \emptyset (suffixe « zéro ») ; il s'agit de la terminaison des noms à l'accusatif qui est devenue le suffixe, comme dans *вгору* [en haut], *у міру* [suffisamment], *на жаль*(\emptyset) [malheureusement];
- de morphes antéposés *до-*, *від-*, *з-*, *с-* + les suffixes *-у*, *-а*, *-і*, il s'agit des terminaisons des noms au génitif, comme dans *доволі* [assez], *спочатку* [d'abord], *звечора* [à partir du soir], *віднизу* [à partir d'en bas] ;
- d'un morphe antéposé *по-* et d'un suffixe *-ому*, *-ьому*, *-єму*, il s'agit des terminaisons des adjectifs au datif qui sont passés dans la classe des adverbes, comme dans *по-доброму* [sans dispute], *по-домашньому* [comme à la maison], *по-твоєму* [à ta façon].

En ukrainien, certains adverbes sont formés à l'aide d'une particule négative qui devient un morphe, comme les adverbes formés à l'aide des morphes antéposés *не-* ou *ні-* :

- *невдало* [sans succès] ;
- *ніколи* [jamais].

Ou encore, les adverbes appelés traditionnellement indéterminés ou indéfinis, formés par contraction d'une particule ou d'une conjonction à un adverbe interrogatif. Ces adverbes s'écrivent soit avec un trait d'union, soit en un seul mot. À la formation des adverbes indéterminés prennent part :

- les morphes postposés *-небудь*, *-сь*, comme dans *як-небудь* [n'importe comment], *якось* [un jour] ;
- les morphes antéposés *будь-*, *казна-*, *де-* et *аби-*, comme dans *абияк* [n'importe comment], *будь-який* [n'importe lequel], *казна-як* [n'importe comment], *деколи* [parfois] ;
- un morphe interposé *-не-*, comme dans *коли-не-коли* [parfois].

Les adverbess formés ainsi ont le sens indéterminé avec l'ajout possible du sens supplémentaire, par exemple, « malaisant » ou « répétition ».

En ukrainien, les adverbess qui sont formés à l'aide des particules négatives ou indéfinies sont très souvent absents des dictionnaires. Ainsi, en préparant un outil de TAL, nous devons trouver un moyen pour les y inclure.

La classification des adverbess

Traditionnellement, dans la linguistique ukrainienne, on considère que les adverbess ne sont pas indépendants sémantiquement, car leur sens est exprimé seulement dans le syntagme.

Prenons l'exemple de l'adverbe *гарно* :

- *гарно працювати* [travailler bien],
- *гарно п'яний* [très soûl].

Dans le premier syntagme, l'adverbe se rapporte au verbe et précise la qualité de l'action. Il est traduit par l'adverbe *bien*. Tandis que, dans le deuxième syntagme, notre adverbe se rapporte à l'adjectif et démontre l'intensité de la propriété de l'adjectif. Et il est traduit différemment par l'adverbe *très*.

Donc, en ukrainien, nous parlons non pas de sous-groupes sémantiques des adverbess, mais plutôt de sous-groupes sémantico-syntaxiques. En résumant les diverses classifications, nous pouvons isoler les sous-groupes suivants :

- adverbess de lieu qui répondent à la question « où ? », comme *жити близько* [vivre à côté] ;
- adverbess de direction qui répondent aux questions « dans quelle direction ? », « où ? » « d'où ? », comme *іти додому* [aller à la maison], *повертатися звідти* [revenir de là-bas] ;
- adverbess de temps qui répondent à la question « à quel moment ? », comme *працювати вдень* [travailler la journée] ;
- adverbess qualificatifs qui peuvent exprimer
 - o la qualité et qui répondent à la question « comment ? », comme *добре працювати* [travailler bien],
 - o la manière et qui répondent à la question « de quelle façon ? », comme *іти швидко* [marcher vite],
- adverbess d'état qui expriment
 - o l'état de l'environnement et qui répondent à la question « il fait comment ? », comme *у кімнаті жарко* [il fait chaud dans la pièce],
 - o l'état de l'être vivant et qui répondent à la question « se sentir comment ? », comme *мені боляче* [j'ai mal] ;
- adverbess d'intensité
 - o simples, ils s'utilisent pour intensifier le sens d'un autre adverbe ou d'un adjectif, comme *дуже багатий* [très riche], *занадто повільно* [trop lentement],
 - o préfixés, ils sont dérivés d'un adverbe suite à la l'ajout d'un préfixe qui exprime l'intensité, comme

працювати забагато [travailler trop] ;

- adverbes de quantité qui répondent à la question « combien ? », comme *мало [peu] ;*
- adverbes de cause qui répondent aux questions « pourquoi ? », « à cause de quoi ? », comme *зробити ненароком [faire quelque chose par accident] ;*
- adverbes de but qui répondent à la question « dans quel but ? », comme *зробити напоказ [faire quelque chose pour se montrer aux autres] ;*
- adverbes de comparaison qui peuvent être remplacés par une tournure *як [comme] + nom*, comme *співати соловейком [chanter comme un rossignol]*, il s'agit des noms à l'instrumental qui ont été adverbialisés ;
- adverbes de liaison qui fonctionnent comme des conjonctions de coordination, comme *також [aussi] ;*
- adverbes modaux¹⁸⁰ qui déterminent le rapport avec la réalité et qui s'utilisent souvent en tant que prédicat. Ces adverbes peuvent exprimer
 - le rapport à la vérité, comme *справді зрозуміти [comprendre véritablement]*,
 - certitude / incertitude, comme *безперечно [sans doute]*,
 - enchaînement logique, comme *зокрема [en particulier]*,
 - évaluation émotionnelle, comme *дивно [bizarrement] ;*
- adverbes interrogatifs qui sont utilisés pour poser des questions, comme *коли? [quand] ;*
- adverbes indéterminés qui sont utilisés pour indiquer la propriété inconnue par le locuteur, ou minimiser l'importance de la propriété, comme *де-небудь [quelque part, peu importe où] ;*
- adverbes de négation, formés par l'ajout de la particule *ні-*, comme *ніколи [jamais]*.

Le comparatif, le superlatif et le superlatif absolu des adverbes

Comme les adjectifs, les adverbes qualificatifs peuvent avoir des degrés de comparaison : le comparatif, le superlatif et le superlatif absolu (voir 3.3.2). Notons que les adverbes ont seulement la forme synthétique qui se termine par le suffixe *-e*. Dans le Tableau 28, nous pouvons voir ces trois formes pour l'adverbe *весело [joyeusement]*.

¹⁸⁰ Comme nous l'avons indiqué précédemment dans la section 3.2.2, certains linguistes mettent les adverbes de ce type dans une classe à part.

Tableau 28. Le comparatif, le superlatif et le superlatif absolu des adverbes

Adverbe qualificatif	Comparatif	Superlatif	Superlatif absolu
<i>весело</i> [joyeusement]	<i>веселіше</i> [plus joyeusement]	<i>найвеселіше</i> [le plus joyeusement]	<i>якнайвеселіше / щонайвеселіше</i> [le plus joyeusement possible]

Les adverbes partitifs, excessifs et diminutifs

En ukrainien, les adverbes comme les adjectifs qualificatifs peuvent avoir des formes dérivées qui ajoutent un sens supplémentaire à l’adverbe de départ :

- le sens de l’incomplet, du partiel, exprimé par
 - o les suffixes *-уват-* (*-юват-*), *-ав-* (*-яв-*), comme dans
мало реу → *малува́то* [vraiment peu],
 - o le préfixe *напів-*, comme dans
голосно fort → *напівголосно* [pas très fort] ;
- le sens de l’excès, exprimé, à la différence des adjectifs, uniquement par le préfixe *пре-*, comme dans
пречудово [très-très bien] ;
- le sens de l’affectif, exprimé par ajout des suffixes *-еньк-*, *-есеньк-*, *-ісіньк-*, *-юсіньк-*, comme dans
гарно [joliment] → *гарне́нько*, *гарне́сенько*, *гарні́сінько*, *гарню́сінько*.

Dans le Tableau 29, nous pouvons voir quelques exemples¹⁸¹ des adverbes décrits dans cette section, en contexte.

Tableau 29. Exemples d’adverbes partitifs, excessifs et diminutifs en contexte

Forme	Exemple	Traduction
Partitive	<i>Мені видається, що сиру малува́то.</i>	<i>Il me semble qu’il n’y a pas tout à fait assez de fromage. (H. Vdovytchenko)</i>
Excessive	<i>Вони пре́чудово знаються на всіх тонкощах відлову.</i>	<i>Ils connaissent très bien toutes les subtilités de la pêche. (I. Franko)</i>
Diminutive	<i>Посідаймо близе́нько та вип’ємо по чарочці повне́нько!</i>	<i>Asseyons-nous l’un à côté de l’autre et buvons un verre que je servirai à ras bord. (P. Myrnyy)</i>

Tout comme les adjectifs partitifs, excessifs et diminutifs, les adverbes correspondants, mais aussi les formes du comparatif, superlatif, superlatif absolu, sont souvent exclus des dictionnaires ; il faudrait donc trouver le moyen pour les inclure.

3.3.7 La préposition

Description des prépositions

Une préposition est une ALU invariable et non autonome, qui sert à exprimer les relations sémantiques et syntaxiques d’un groupe nominal ou pronominal avec un autre élément dans une phrase. Les prépositions précisent le sens des ALU variables dont elles déterminent le cas. Rappelons

¹⁸¹ Tous ces exemples sont tirés du site <https://goroh.pp.ua>

que dans la langue ukrainienne, les cas sont exprimés morphologiquement par les terminaisons, mais ils peuvent être spécifiés par les prépositions. De ce fait, les prépositions jouent un rôle important lors de la désambiguïsation.

En ukrainien, certaines prépositions peuvent spécifier :

- un seul cas, comme
без [sans], біля [à côté de], для [pour], qui sont suivies uniquement d'un syntagme au génitif ;
- plusieurs cas, comme
 - o *на [sur, à]*, qui peut être suivie du locatif ou de l'accusatif,
 - o *під [sous, au-dessous]*, qui peut être suivie de l'instrumental ou de l'accusatif,
 - o *з [de, avec]*, qui peut être suivie de l'instrumental, de l'accusatif ou du génitif.

Notons que, lors de la traduction dans une autre langue, ces prépositions sont traduites différemment.

Classification des prépositions

En fonction de leur origine, les prépositions sont divisées en :

- primaires ou simples, par exemple
над [au-dessus de], в [dans, à], до [jusqu'à, contre], comme dans
Зоня припала обличчям до волосся Юзі.
[Zonia a serré son visage contre les cheveux de louzia]. (L. Oukraïinka) ;
- secondaires ou créées par conversion d'une autre ALU invariable ou d'une forme fléchie d'une ALU variable, par exemple
 - o *коло [à côté de]*, qui coïncide avec le nom au nominatif *коло [cercle]*, comme dans :
Коло Максима стояв п'ятилітній синок Борис.
[Boris, le fils de Maksym âgé de cinq ans, se tenait à côté de lui.] (I. Bahrianyy),
 - o *шляхом [par]*, qui coïncide avec le nom à l'instrumental *шлях [chemin]*, comme dans :
Збагачення мови іде, безперечно, не тільки шляхом прямого утворення нових слів, неологізмів, а й шляхом переосмислення слів давніх.
[L'enrichissement de la langue passe, bien sûr, non seulement par la création directe de nouveaux mots, de néologismes, mais aussi par la réinterprétation de mots anciens.] (M. Ryl's'kyy),
 - o *уздовж [le long de]*, qui coïncide avec l'adverbe *уздовж [sur toute sa longueur]*, comme dans :
Уздовж стін лави, покриті полавочниками з червоного сукна.
[Le long des murs, des bancs recouverts de couvertures en tissu rouge.] (B. Hrintchenko).

Les prépositions apparues par conversion sont très nombreuses, notamment celles apparues suite à la conversion de l'adverbe. En conséquence, il faudra inclure ces homonymes dans la solution que nous allons préparer à cet effet.

Selon leur structure, les prépositions peuvent être :

- simples, c'est-à-dire composées d'un seul mot, comme
над [au-dessus de], коло [à côté de] ;
- dérivées, c'est-à-dire formées par contraction de deux prépositions qui peuvent s'écrire soit en un seul mot, soit avec un tiret, comme

посеред [au milieu], із-за [de derrière] ;

- composées de plusieurs unités, comme

на відміну від [contrairement à], під час [pendant].

Bien que ces dernières soient nombreuses (Vykhovanets' et Horodens'ka 2004, p. 299 [10]), la plupart d'entre elles sont exclues des dictionnaires traditionnels.

Les prépositions peuvent être également classées selon leur expression sémantique dans les sous-groupes de prépositions spéciales, temporelles ou logiques. Dans cette classification, on prend en compte aussi le rôle grammatical et la signification du syntagme avec lequel cette préposition est employée. Faire une telle classification serait intéressant lorsque nous étudierons les prépositions de façon plus approfondie, en particulier le phénomène de la conversion ou de la création de prépositions composées.

3.3.8 La conjonction

Description des conjonctions

Selon le point de vue traditionnel, une conjonction est une ALU invariable, qui sert à lier :

- à l'intérieur d'une phrase, des ALU qui appartiennent à la même classe grammaticale et ont la même fonction syntaxique ;
- la proposition subordonnée à la proposition principale ;
- les propositions coordonnées.

Classification des conjonctions

Selon leur fonction, on distingue :

- les conjonctions de subordination (hypotaxe), qui établissent une hiérarchie entre les propositions d'une phrase composée et lient la proposition subordonnée à la proposition principale (procédé hiérarchisant) ;
- les conjonctions de coordination (parataxe), qui lient des propositions de même niveau (procédé séquentiel).

Selon leur structure, les conjonctions peuvent être :

- simples, c'est-à-dire composées d'un seul mot, comme *і [et] ;*
- dérivées, c'est-à-dire formées par contraction de deux mots (la conjonction, la préposition et la particule, le pronom démonstratif), comme *або (а+бо) [ou bien] ;*
- composées de plusieurs unités, comme *тому що [parce que] ;*
- répétitives, qui s'utilisent devant le premier élément, puis se répètent devant chaque élément ajouté, comme *і ... , і ... , і ... , [et ... , et ... , et ...] ;*
- composées de deux parties : la première partie s'utilise devant le premier élément, puis la deuxième devant l'élément à ajouter, comme *якщо ... , то [si ... alors].*

Bien qu'une conjonction n'exprime pas de sens lexical par elle-même, elle a un sens en contexte. Par exemple, on utilise les conjonctions *i, ŭ, ma*, qui signifient « et », pour lier deux verbes quand on a besoin de montrer que deux actions se déroulent simultanément.

Les linguistes ukrainiens¹⁸² classent les conjonctions dans les groupes sémantico-fonctionnels suivants :

- copulatives, qui marquent la simultanéité ou l'addition, comme *i(ŭ) [et]* ;
- accumulatives, qui marquent l'ajout d'une propriété, comme *ma ще ŭ [et en plus]* ;
- disjonctives, qui marquent un choix, comme *чи [soit]* ;
- adversatives, qui marquent l'opposition, comme *однак [toutefois]* ;
- graduelles, comme *не тільки... але ŭ [non seulement ... mais en plus]* ;
- causales, comme *оскільки [car]* ;
- temporelles, comme *з того часу як [depuis que]* ;
- explicatives, comme *тобто [c'est-à-dire]* ;
- conditionnelles, comme *якби [si]* ;
- conjonction de but, comme *щоб [pour (que)]* ;
- conjonction de concession, comme *хоч [bien que]* ;
- conjonction de comparaison, comme *мов [comme]*.

Certaines conjonctions forment des couples homonymiques avec les ALU d'autres classes. Nous pouvons voir quelques exemples¹⁸³ dans le Tableau 30.

Ainsi, le but d'un linguiste qui formalise la conjonction à travers un outil informatique est de décrire la conjonction et la désambiguïser en contexte.

¹⁸² (Vykhovanets' 1988 [9]), (Zahnitko 1996 [30]), (Vykhovanets' et Horodens'ka 2004 [10]), (Pliouchtch 2010 [52]), (Vykhovanets' et alii 2017 [12]).

¹⁸³ Les exemples sont pris sur le site <https://goroh.pp.ua>

Tableau 30. Les couples homonymiques qui contiennent une conjonction

UL	Conjonction	Autres
ОДНАК	<i>Усмішка гостра, а однак сумна. [Un sourire franc, et pourtant triste.] (L. Oukraïнка)</i>	Adverbe – Куди ти, синку? – спитала мати. – От так, пройдуся, поки снідання, однак без діла, – мовив Корній. [Sa mère lui demande : – Où vas-tu, mon fils ? – Je vais juste me promener pendant que les autres petits-déjeunent, de toute façon , je ne fais rien, répond Korniy.] (L. Oukraïнка)
ТАК	<i>Як тільки літо, так я й в дорогу! [Dès que l'été arrive, je suis sur la route !] (L. Oukraïнка)</i>	Particule – Це ж її хата? – Так , її... [– N'est-ce pas sa maison ? – Oui , c'est sa maison...] (P. Myrnyy)
МОВ	<i>А сивий гетьман, мов сова, Ченцеві зазирає в вічі. [Et l'hetman aux cheveux gris, tel un hibou, regarde le moine dans les yeux.] (T. Chevtchenko)</i>	Génitif pluriel du nom мова [langue] Він знав багато іноземних мов . [Il connaissait de nombreuses langues étrangères.]
А	<i>Є багато на світі учених людей, а поетів мало. [Il y a beaucoup d'érudits dans le monde, mais peu de poètes.] (M. Kotsioubyns'kyu).</i>	Interjection Здорові, Любов Олександрівно! А , і ви тут, Оресте Михайловичу! [Bonjour, Lioubov Olexandrivna ! Ah , vous êtes là aussi, Orest Mykhaylovytch !] (L. Oukraïнка)]

3.3.9 La particule

Description des particules

En ukrainien, les particules sont des ALU non variables, qui servent à renforcer, souligner ou donner des nuances sémantiques à un élément de la phrase, ou à la phrase entière, ou bien à produire des constructions syntaxiques affirmatives, négatives ou interrogatives, ou encore, qui sont utilisées comme auxiliaires pour former l'hypothétique.

Classification des particules

Les particules peuvent être classées dans les groupes sémantico-fonctionnels suivants :

- affirmatives, par exemple
так [oui], еге [oui], авжеж [bien sûr], атож [bien sûr], servent à construire les phrases affirmatives¹⁸⁴, comme dans
– Чому ти, Чіпка, не одружишся? – «Хіба воно краще?» – **Авжеж**, краще.
[Pourquoi, Tchirka, tu ne te maries pas ? – « Mais est-ce que c'est mieux ? » – **Bien sûr**, c'est mieux.] (P. Myrnyy) ;
- négatives, c'est-à-dire celles qui servent à construire les phrases négatives, par exemple
не, ні, ані [ne pas], qui s'utilisent pour des constructions négatives, comme dans
*Сіяло сонце; в небесах **ані** хмариночки.*
[Le soleil brillait, et dans le ciel, il n'y avait **pas** un petit nuage.] (T. Chevtchenko) ;

¹⁸⁴ Les affirmations peuvent être ironiques (*авжеж! [c'est ça !]*) ou interrogatives (*авжеж? [n'est-ce pas ?]*).

- interrogatives, c'est-à-dire celles qui servent à construire les phrases interrogatives, par exemple
чи [est-ce que], хіба [est-ce que (avec une nuance d'étonnement)], що за [quel genre / c'est quoi ce], comme dans
Що за плеск?.. Із ночі, із туману випливають темні байдаки...
[C'est quoi ce clapotis? De la nuit, du brouillard, surgissent des barques sombres.]
(V. Sossioura) ;
- modales :
 - a. *хай [que], нехай [que]* sont utilisées pour exprimer un ordre, un souhait ou un accord ; la particule est suivie d'un sujet (éventuellement) et d'un verbe à la troisième personne au singulier ou au pluriel, comme dans
Коли махну рукою – хай оркестр грає Марсельєзу.
[Quand je ferai un signe avec ma main, que l'orchestre joue la Marseillaise.]
(O. Korniytchouk),
 - b. *бодай [que]* s'utilise dans des phrases de type « malédiction », comme dans
Бодай же вас, цокотухи, та злидні побили.
[Que la misère s'abatte sur vous, les commères.] (Т. Chevtchenko),
 - c. *бу (б) [si]* s'utilise pour former l'hypothétique ; la particule est suivie d'un sujet + soit un verbe au passé, soit un verbe au plus-que-parfait, soit un infinitif. On l'utilise pour exprimer les faits irréels, comme dans
Розказав бу про те лихо, та чи то ж повірять!
[Je parlerais de ce malheur, mais est-ce qu'on me croira ?] (Т. Chevtchenko),
 - d. *щоб [que]* s'utilise aussi pour exprimer la crainte, la probabilité, le fait envisagé ou supposé, la demande, l'ordre ou la malédiction. La particule est suivie d'un sujet + un verbe au passé, comme dans
Українська артистка Верка Сердючка разом зі своєю командою почала вигукувати зі сцени: « Путін, щоб ти здох! ».
[La chanteuse-humoriste ukrainienne Vierka Serdioutchka et son groupe se sont mis à crier sur la scène : « Que tu crèves, Poutine ! »] (unian.ua) ;
- démonstratives, par exemple
це (оце, отце) [ceci], то (ото, отто) [là-bas, c'est], ось [voici], qui pointent un composant de la phrase, comme dans
– Йосипе! – одказав старий ... – Дивися: ото Бог...
[– Yossyp ! – a dit le vieillard... – Regarde, là-bas, c'est Dieu.] (P. Myrnyy) ;
- particules qui apportent diverses nuances au mot ou à la phrase entière, par exemple
 - e. *навіть [même], ж (же) [donc],* qui renforcent le sens d'un des mots de la phrase comme dans
Дядьку, он зима біліє! От тепер же згину я!
[Mon oncle, l'hiver devient blanc ! C'est là que je vais mourir !] (L. Hlibov),
 - f. *як* montre l'étonnement et une forte émotion, comme dans
як гарно ! [que c'est beau !],
 - g. *-но* et *-бо* sont utilisées dans les phrases impératives pour exprimer la menace ou, au contraire, rendre l'ordre moins catégorique, comme dans
підійди-но до мене [approche-toi donc],

сідай-**бо** [*assieds-toi alors*],

○ лише (лиш) [*seulement*], лишень [*seulement*], хоч (хоча) [*au moins*], тільки [*seulement*], qui mettent des limites, comme dans
Батьки та бабуся розібрали усіх дітей, **лише** одна дівчинка стояла на лаві.

[*Les parents et les grands-mères ont récupéré tous les enfants, seule une fille est restée debout sur le banc.*] (M. Vinhranovs'kyu),

○ то, ну, qui renforcent la demande, comme dans
То коли, кажете, прийти?
[*Quand avez-vous dit de venir ?*] (L. Oukraïinka).

Il est important de noter que les particules peuvent être classées dans plusieurs groupes sémantico-fonctionnels ; ainsi, le dictionnaire SUM-20 propose 4 paragraphes pour décrire la particule *би*, 6 paragraphes pour décrire la particule *ото* et 7 paragraphes pour décrire la particule *ну*.

De plus, tout comme les autres mots-outils, les particules peuvent jouer le rôle de conjonctions, d'interjections ou même de prépositions (voir Tableau 31), ce qui complique leur description, en nous forçant à chercher une solution pour les désambiguïser.

Tableau 31. Les couples homonymiques qui contiennent une particule¹⁸⁵

UL	Particule	Autres
бо	<i>Галю! послухай-бо, – промовив менший брат, – слухай-бо, мама принесла чоботи. [Écoute donc, Halia, dit le jeune frère, – Écoute donc, maman a apporté des bottes.] (M. Vovtchok)</i>	Conjonction <i>Швидше будемо писати, бо хочеться спати і вам, і мені. [Écrivons rapidement, car vous et moi avons envie de dormir.] (T. Chevtchenko)</i>
це	<i>– Кого ж ти, сину, маєш на приміті? – спитала мати, – Нимідору .. – Яку це Нимідору? [– A qui penses-tu, mon fils ? demanda la mère, Nymydora... Quelle Nymydora donc ?] (I. Netchouy-Levyts'kyu)</i>	Pronom démonstratif <i>Та це життя було зовсім не безжурне. [Mais cette vie n'était pas du tout sans soucis.] (J. Roni)</i>
за	<i>Що то за люди метушилися оце ввечері біля вашого медпункту? [Quel genre de personnes se sont agitées autour de votre centre médical ce soir ?] (M. Tcherniavs'kyu)</i>	Préposition <i>За день до мого од'їзду я востаннє похапцем передивлявся свої книжки. [Un jour avant mon départ, je me suis empressé de consulter mes livres pour la dernière fois.] (S. Vassyl'tchenko)</i>
ну	<i>Ну хто, хто мені повірить, що я сам, на власні очі бачив той театр Бергоньє. [Mais qui me croira si je dis que j'ai vu de mes propres yeux ce théâtre de Bergonier ?] (V. Nestayko)</i>	Interjection <i>– Ну, спасибі, брате! – кажу я. [– Alors merci, mon frère, dis-je.] (M. Vovtchok)</i>

Selon leur structure, les particules peuvent être :

- simples, c'est-à-dire composées d'un seul mot, comme
чи, хіба, хай, би ;

¹⁸⁵ Les exemples sont tirés du dictionnaire SUM-20 : <https://sum20ua.com>, du dictionnaire SUM : <https://sum.in.ua/> (« *це* » en tant que particule), du site UkrLib : <https://www.ukrlib.com.ua> (« *це* » en tant que pronom).

- composées ou doubles, qui s'écrivent avec un tiret, comme *ОН-ОН, АНІ-НІ, ВСЬОГО-НА-ВСЬОГО* ;
- particules-morphes indépendants qui s'ajoutent à une autre ALU, comme *-НО, -БО*.

Ainsi, en préparant un module ukrainien pour un outil de TAL, il faut veiller à ce que ces ALU soient reconnues correctement, c'est-à-dire les particules *-НО* et *-БО* comme des formes indépendantes et les particules composées ou doubles comme des locutions.

3.3.10 L'interjection

Description des interjections

Dans la linguistique ukrainienne, l'intérêt pour les interjections s'est accru ces dernières décennies. En plus des grammaires traditionnelles, qui proposent un chapitre pour cette classe des ALU, des ouvrages et des articles spécifiques ont été publiés (Rak 2000 [54]), (Kourylo 2000 [37, 38], 2009 [31]), (Houmeniouk 2003 [25]), (Hatsenko 2009a [17], 2009b [18]).

Traditionnellement, les interjections sont traitées en tant qu'ALU périphériques non variables. Toutefois, certains linguistes, par exemple Vykhoivanets' (1988, p. 34 [9]) et Horodens'ka (dans Bezpoïasko *et alii* 1993, p. 100 [4]), les considèrent comme des unités purement syntaxiques. Néanmoins, tous les linguistes s'accordent à dire qu'une interjection n'a pas de liens grammaticaux avec d'autres unités de la phrase et qu'elle n'exprime pas le sens en elle-même, puisque le sens est exprimé simultanément par le contexte, l'intonation, parfois par la mimique et les gestes.

Classification des interjections

Syntaxiquement, les interjections peuvent être :

- des phrases indépendantes qui se terminent principalement par un point d'exclamation, comme dans¹⁸⁶
– *Еге-ге! Я думаю, що ви тут доглядаєте? А ви спати!*
[– **Alors !** Je me demandais ce que vous gardiez par ici ? Et vous ? Vous dormez !] (P. Myrnyy) ;
- des insertions à l'intérieur d'une phrase, entourées par des virgules ; il n'est pas rare que la phrase complète se termine par un point d'exclamation, comme dans
– *Сперечатися за однісіньку репліку, їй-бо, смішно!*
[– Se disputer pour une seule réplique, **je vous jure** que c'est ridicule !] (A. Kryjanivs'kyj) ;
- des prédicats et, dans ce cas, on ne met pas de virgule, comme dans
– *Геть Ларька, геть з покоїв!*
[– **Va-t'en** Lar'ka, **va-t'en** des chambres !] (S. Vassyl'tchenko).

Les interjections peuvent être classées dans les sous-groupes sémantico-fonctionnels suivants :

- les interjections émotives qui expriment un large spectre d'émotions, par exemple *ой* qui exprime l'étonnement, la jubilation, la déception, le mécontentement, la peur ou encore la souffrance, comme dans
Бахнув наосліп з рушниці. – Ой! – скрикнув Остап й вхопивсь за груди.

¹⁸⁶ Ces exemples ont été trouvés sur le site Goroh <https://goroh.pp.ua>.

[Il a tiré avec un fusil sans regarder. – **Ah!** – cria Ostap et il s’attrapa la poitrine]
(M. Kotsioubyns’kyy),

- les interjections « jurons » que l’on utilise pour exprimer la colère, le mécontentement, le désaccord ou encore la déception et le désarroi, comme *чорт* dans
Він почав нишпорити. Чорт! Десь заподілися книжки.
[Il a commencé à fouiller. **Que diable!** Les livres se sont perdus quelque part.]
(M. Kotsioubyns’kyy) ;
- les interjections injonctives qui expriment l’incitation à l’action, l’ordre de faire quelque chose, l’appel, comme *цить* que l’on utilise pour demander à quelqu’un de se taire, comme dans
– **Цить!** Постривай! – шепоче він мені.
[– **Chut!** Attends! – me chuchote-t-il.] (M. Vovtchok) ;
- les onomatopées qui imitent des bruits, comme *бух*, qui imite un bruit sourd dans la phrase
У нас оце якраз яблука падають... Вийдеш на світанку, а вони – бух, бух.
[Chez nous, en ce moment, les pommes tombent. On sort au lever du soleil, et on les entend – **boum, boum.**] (H. Tioutiounnyk) ;
- les locutions de politesse et d’étiquette qui expriment les salutations, les adieux, les demandes et les remerciements, par exemple *будь ласка* [s’il te plaît], *спасибі* [merci], comme dans
– А Явдоху хто з світу звів? – питає гостро Чіпка. – То це я, сину?.. – Уже ж не я... – **Спасибі...**
[– Et qui a envoyé Iavdokha dans l’au-delà? – demande Tchipka franchement. – C’est moi, mon fils? – Certainement pas moi... – **Merci...**] (P. Myrnyy).

Selon leur structure, les interjections peuvent être :

- simples, comme
ой, гей, фу ;
- doubles, comme
ах-ах-ах, га-га-га, кап-кап ;
- allongées, comme
бу-у-ух, гу-у-у, о-о-ох ;
- composées de plusieurs ALU, comme
дякувати Богу [grâce à Dieu], *будьте ласкаві* [s’il vous plaît], *добрий вечір* [bonsoir].

Selon leur origine, les interjections peuvent être :

- liées au verbe, c’est-à-dire qu’elles font référence au bruit produit par l’action exprimée par le verbe et qu’elles coïncident avec la racine du verbe, par exemple *ляпати* [éclabousser, faire du bruit dans l’eau] → *ляп* [plouf], comme dans
Сядемо на березі, ноги у воду зануримо: ляп-ляп!
[Allons-nous asseoir sur la berge et mettre nos pieds dans l’eau : **plouf-plouf!**] (K. Hryb)
- créées par conversion d’un nom, par exemple *лишенько*, qui est le diminutif du nom *лихо* [malheur] et que l’on utilise pour exprimer le désarroi ou la peur, comme dans

І що то він тримає в'юнке та блискуче в руках? Лишенько, та це ж бо залізний ланцюг.

[Qu'est-ce que cette chose qui se tortille et brille dans ces mains ? Oh mon Dieu, c'est une chaîne en métal.] (Ie. Houtsalo)

- créées par conversion d'un verbe, par exemple *подумаєш*, qui est la deuxième personne du singulier présent-futur perfectif du verbe *подумати* [réfléchir], et que l'on utilise pour exprimer l'ironie, l'irrespect, l'indifférence, comme dans
– *А я не боюсь!* – *Помахала рукавичкою і прискорила ходу.* – **Подумаєш**, – *цвіркнув крізь зуби Дмитро і вернувся назад.*
[– Mais je n'ai pas peur ! – Elle a fait un au revoir avec son gant et s'est mise à marcher plus vite. – Vof, a dit à travers les dents Dmytro avant de revenir.] (M. Zaroudnyy)
- créées par conversion d'un pronom, par exemple *це* du pronom démonstratif *цей* [ce], employé au neutre, que l'on utilise pour exprimer l'hésitation, comme dans
– *Ви тут побалакайте... а я... це... піду... – прямує до виходу.*
[– Je vous laisse discuter... et moi... euh... je m'en vais... – il se dirige vers la sortie.] (O. Bahirova)
- créées par conversion d'un adverbe, par exemple *отак*, de l'adverbe *отак* [de cette façon], que l'on utilise pour exprimer la surprise, la désolation, la déception, comme dans
– *Не змогла я зробити це, – мотаю головою.* – **Отак!** – *відповів Борис.*
[– Je n'ai pas pu le faire, – je fais un signe de tête. – Zut ! – a répondu Borys.] (O. Bahirova)
- créées par conversion d'une conjonction, par exemple *а то*, de la conjonction *а то* [et même], pour exprimer la fierté, comme dans
– *Ой, Максик українською мовою заговорив!* – **А то!**
[– Ça alors, Maksyk s'est mis à parler ukrainien ! – Bah évidemment !] (Sur des réseaux sociaux)

La formalisation des interjections par un outil de TAL peut s'avérer une tâche très compliquée, car de nombreuses interjections ont été créées par conversion et, de ce fait, produisent des homonymes grammaticaux et donc potentiellement des ambiguïtés lors de l'étiquetage de texte. Parfois, il est difficile même pour un humain de savoir à quelle classe appartient telle ou telle unité. Par ailleurs, il est du devoir d'un linguiste qui prépare des ressources pour un logiciel TAL de bien décrire ces unités et d'essayer de trouver des solutions pour les désambiguïser.

3.4 Conclusion du chapitre 3

Après avoir étudié attentivement la classification des ALU de la langue ukrainienne, nous avons adopté les dix classes grammaticales traditionnelles qui figurent dans le dernier dictionnaire académique SUM-20 : le nom, l'adjectif, le numéral, le pronom, le verbe, l'adverbe, la préposition, la conjonction, la particule et l'interjection. Ces classes devront s'afficher lors de l'étiquetage des textes.

Toutefois, nous avons vu que certaines ALU ou leur formes fléchies pouvaient jouer des rôles syntaxiques spécifiques et sortir du cadre de leur catégorie initiale, comme c'est le cas de certaines formes utilisées en tant qu'insert ou encore un mot prédicatif. Nous devons prendre cela en compte lorsque nous préparerons les ressources ukrainiennes pour NooJ.

Pour chaque classe variable, nous avons décrit les catégories grammaticales. Certaines de ces catégories pourront être ajoutées dans le dictionnaire, par exemple, le genre ou la catégorie de l'animé/inanimé pour les noms, tandis que les autres feront partie des paradigmes flexionnels, par exemple, le genre ou la catégorie de l'animé/inanimé pour les adjectifs.

Nous avons également vu qu'en ukrainien il y avait un grand nombre d'homonymes grammaticaux, liés aux flexions homonymiques ou créés suite au passage des ALU d'une classe à une autre. Il conviendra donc de les désambiguïser en contexte, en utilisant des grammaires syntaxiques.

En outre, pour chaque classe, nous avons décelé un certain nombre d'informations sémantiques et fonctionnelles qui peuvent être utiles pour un utilisateur lors de l'analyse de texte. Ces informations seront donc à ajouter lors de la création du dictionnaire.

Nous avons vu aussi que certaines classes avaient des ALU composées de plusieurs mots ; il faudra alors veiller à ce que ces ALU ne soient pas traitées comme des combinaisons de mots, mais comme des locutions.

Il ne faudra pas non plus oublier de relier toutes les formes des mêmes unités, comme c'est le cas des gérondifs et des participes, ou encore des diminutifs.

Chapitre 4. Le module ukrainien pour NooJ

4.1 Introduction

Nous avons préparé la première version du module ukrainien pour NooJ qui nous permet d'annoter n'importe quel texte ukrainien. Le module est publié en open-source sur le site Web de NooJ (page "Linguistic Resources"). La version que nous décrivons dans ce travail est la 1.3. Elle est composée d'un dictionnaire principal et d'un dictionnaire secondaire qui sont liés à deux fichiers contenant les paradigmes flexionnels (FLX) et dérivationnels (DRV), 18 grammaires morphologiques qui produisent des lemmes et leurs formes fléchies, et 10 grammaires syntaxiques de désambiguïsation.

Dans ce chapitre, nous expliquerons les étapes de notre travail d'implémentation, en nous appuyant toujours sur la grammaire de la langue ukrainienne.

Nous commencerons par décrire le fichier « `_properties.def` » dans lequel nous avons défini les étiquettes pour le module ukrainien.

Nous décrirons ensuite le dictionnaire et les paradigmes FLX et DRV que nous avons construits pour associer les formes fléchies et dérivées à chaque entrée du dictionnaire.

Après cela, nous donnerons des détails sur les grammaires morphologiques qui nous permettent non seulement de mieux couvrir le vocabulaire ukrainien, mais aussi de relier les formes dérivées aux formes de départ, ce qui est particulièrement intéressant lorsque nous faisons de l'analyse de texte, d'autant plus qu'aucun outil en libre accès ne permet de le faire actuellement.

La dernière section de ce chapitre sera consacrée aux grammaires qui permettent la levée d'une grande partie des ambiguïtés.

4.2 Description du fichier « `_properties.def` » pour le module ukrainien

4.2.1 Le choix du jeu d'étiquettes

Comme nous l'avons déjà mentionné dans le chapitre 2, dans NooJ, le jeu d'étiquettes n'est pas imposé, ainsi l'utilisateur est libre de créer ses étiquettes en mettant en avant les principes qui lui paraissent importants. Créer les étiquettes pour le module NooJ consiste à définir les catégories, leurs propriétés et les valeurs de ces propriétés. Notons qu'ici nous utiliserons le terme de « catégorie » pour désigner la classe syntaxique des ALU (par exemple, « nom »), et « propriété » et « valeur » toutes les autres propriétés et leurs valeurs (par exemple, « cas = génitif », « classe sémantique = toponyme »).

Nous avons utilisé des termes anglais pour toutes les étiquettes de catégories, de propriétés et de leurs valeurs, car cela permettra de lire facilement les annotations sans se référer à chaque fois à un tableau d'équivalences¹⁸⁷. Nous avons écrit les catégories en lettres capitales, puis les propriétés et les valeurs en majuscule. Par exemple, nous avons créé l'étiquette « NOUN » pour la classe

¹⁸⁷ Nous avons utilisé les abréviations seulement pour les étiquettes sémantiques composées de deux mots, comme « L_Name » pour « last name ».

grammaticale du nom, puis la propriété « Case » pour désigner le cas, auquel nous avons associé les valeurs (nominatif, accusatif, génitif, datif, instrumental, locatif, vocatif) comme ceci¹⁸⁸ :

NOUN_Case = Accusative | Dative | Genitive | Instrumental | Nominative | Locative | Vocative;

Rappelons que dès que les catégories, les propriétés et les valeurs sont définies dans le fichier « properties.def », elles peuvent être utilisées dans n'importe quel dictionnaire ou grammaire du module. Par exemple, nous avons associé la catégorie « NOUN » aux entrées du dictionnaire, nous avons utilisé les valeurs de la propriété « Case » dans les grammaires FLX qui décrivent les formes fléchies (Figure 33), mais aussi dans les grammaires morphologiques pour décrire les formes dérivées et leurs flexions (Figure 42), ou encore dans des grammaires syntaxiques de désambiguïisations (Figure 72).

Les noms des catégories, des propriétés et de leurs valeurs s'afficheront dans les annotations du texte utilisées pour faire des recherches dans les textes (voir Figures 28 et 29). Dans l'exemple illustré par la Figure 28, nous entrons dans l'outil de recherche l'expression <NOUN+Genitive> et nous trouverons tous les noms au génitif dans notre texte. Dans l'exemple illustré par la Figure 30, nous voyons une annotation dans laquelle apparaissent les étiquettes « NOUN » et « Genitive ».

Figure 28. Recherches par étiquettes de Nooj

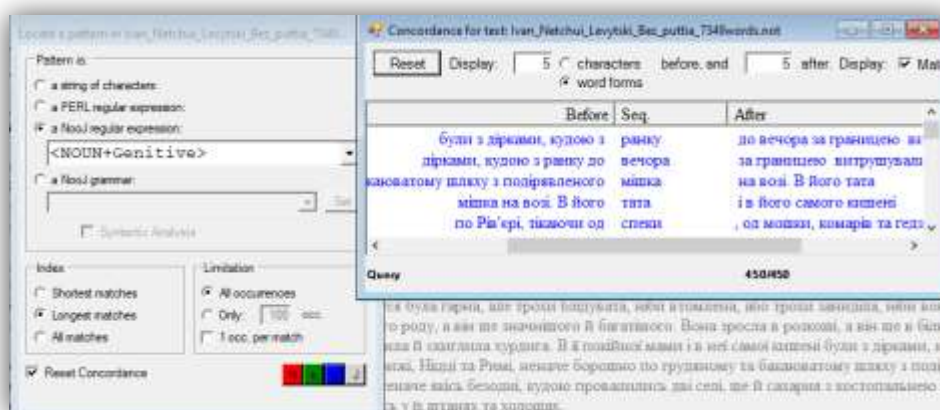
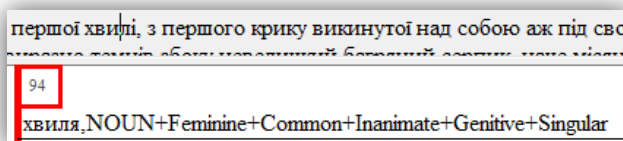


Figure 29. Affichage des étiquettes dans les annotations du texte



4.2.2 Affichage des éléments

Pour afficher les éléments dans le fichier « _properties.def », nous avons adopté le même principe que pour le module du bélarussien¹⁸⁹ car cet affichage nous paraît pratique et transparent pour trouver les informations recherchées très rapidement, puisque :

¹⁸⁸ La méthode de création de tous les fichiers Nooj est décrite dans le manuel disponible lors du téléchargement de l'outil.

- les éléments y sont affichés en ordre alphabétique ;
- les catégories NooJ sont séparées par des commentaires ;
- les catégories, les propriétés et les valeurs sont recensées exhaustivement.

4.2.3 Choix des catégories NooJ pour le module ukrainien

Nous avons inscrit dans le fichier «_properties.def» les 10 classes grammaticales traditionnelles décrites dans le chapitre 3. Nous avons décidé d’y ajouter également deux classes syntaxiques : « PREDICATIVE » pour décrire les mots prädicatifs et « INSERT » pour les mots-inserts (Figure 30).

Figure 30. Description des catégories NooJ dans le module ukrainien

```
#####
##_properties.def
#####
# Count of categories = 10
#
# ADJECTIVE
# ALVERB
# CONJUNCTION
# INSERT
# INTERJECTION
# NOUN
# NUMERAL
# PARTICLE
# PREDICATIVE
# PREPOSITION
# PRONOUN
# VERB
```

Ensuite, pour chaque catégorie, nous avons décrit les propriétés correspondantes. Par exemple, pour l’adjectif, nous avons listé les propriétés et leurs valeurs pour les catégories grammaticales de l’animé/inanimé, du cas, du degré de comparaison, du genre, du nombre. Nous avons créé aussi la propriété « Expression » pour décrire les adjectifs partitifs, excessifs, diminutifs et négatifs. Puis, pour décrire les formes courtes des adjectifs masculins et les formes non contractées des adjectifs féminins, neutres et pluriels, nous avons déclaré la propriété « Form ». Et enfin, pour formaliser les adjectifs qualificatifs, relatifs et les adjectifs d’appartenance, nous avons créé la catégorie « Type ». Nous pouvons voir ces propriétés en Figure 31.

Figure 31. Description des annotations de l’adjectif

```
#####
##### ADJECTIVE #####
#####
# Count of categories = 8
ADJECTIVE_Animation = Animate | Inanimate; # 2
ADJECTIVE_Case = Accusative | Dative | Genitive | Instrumental | Nominative | Locative; # 6
ADJECTIVE_Degree = Comparative | Superlative | Highest_Superlative; # 3
ADJECTIVE_Expression = Diminutive | Excessive | Partitive | Negative; # 4
ADJECTIVE_Form = Short | Uncontracted; # 2
ADJECTIVE_Gender = Feminine | Masculine | Neuter; # 3
ADJECTIVE_Number = Singular | Plural; # 2
ADJECTIVE_Type = Belonging | Qualitative | Relative; # 3
```

Voici la liste de toutes les propriétés que nous avons définies pour le module ukrainien :

- « Animation », décrit la catégorie grammaticale de l’animé/inanimé ;

¹⁸⁹ Le module est préparé par Yury Hetsevich, Sviatlana Hetsevich, Yauheniya Yakubovich, certains fichiers du module sont disponibles sur le site <https://nooj.univ-fcomte.fr/resources.html>

- « Aspect »¹⁹⁰, décrit l'opposition entre le verbe IPF et le verbe PF ;
- « Case », exprime la catégorie grammaticale du cas ;
- « Degree », exprime la catégorie grammaticale du degré de comparaison ;
- « Form », décrit les ALU qui ont une forme particulière ;
- « Expression », décrit les ALU qui ont une propriété sémantique en commun et ont été formées selon le même schéma dérivatif (par exemple, les noms affectifs) ;
- « Gender », exprime la catégorie grammaticale du genre ;
- « Mood », exprime la catégorie grammaticale du mode verbal ;
- « Motion_Type », décrit l'opposition entre le verbe déterminé et le verbe indéterminé ;
- « Number », exprime la catégorie grammaticale du nombre ;
- « Origin », décrit les ALU qui ont été formées par conversion ;
- « Person », exprime la catégorie grammaticale de la personne ;
- « Semantic », décrit les ALU qui partagent les aspects sémantiques ;
- « Structure », décrit les ALU qui ont une structure particulière ;
- « Tense », exprime la catégorie grammaticale du temps ;
- « Transitivity », décrit les verbes transitifs ou intransitifs ;
- « Type », décrit les ALU qui ont la même propriété, les mêmes fonctions syntaxiques, et peuvent avoir les mêmes caractéristiques morphologiques et sémantiques.

Notons que quelques propriétés, comme « Semantic », ne seront pas utilisées dans cette première version¹⁹¹, mais nous les utiliserons plus tard. Par exemple, nous ajouterons des marques sémantiques aux entrées des dictionnaires, comme c'est le cas dans le module russe¹⁹².

4.2.4 Déclaration des variables pour les flexions

Pour que NooJ puisse gérer les contraintes d'accord flexionnel dans les grammaires syntaxiques, il faut déclarer aussi toutes les valeurs produites lors des flexions. Cette déclaration se fait une seule fois pour toutes les catégories. Pour l'ukrainien, nous avons déclaré les valeurs suivantes :

INFLECTION = Accusative | Dative | Genitive | Instrumental | Nominative | Locative |
Vocative | Feminine | Masculine | Neuter | Singular | Plural | Animate | Inanimate |
Uncontracted | 1 | 2 | 3 | Future | Past | Present | Imperfective | Perfective | Anterior |
Concomitant | Asubjectal | Gerund | Indicative | Infinitive | Passive;

4.3 Dictionnaires ukrainiens pour le module ukrainien de NooJ

4.3.1 Description du dictionnaire principal

Pour démarrer le projet, nous avons utilisé le dictionnaire électronique open source version 2.9.1 (2016)¹⁹³, construit par deux informaticiens Maksym Polyakov et Andiy Rysin. Cette version contient autour de 220 000 entrées auxquelles sont associées des classes grammaticales et quelques propriétés grammaticales sous forme de codes (de chiffres et de lettres). Au départ, nous avons

¹⁹⁰ La propriété NooJ « Aspect » a trois valeurs : « Imperfective » qui désigne des verbes imperfectifs, « Perfective » qui désigne des verbes perfectifs, et « Pair » qui décrit le couple aspectuel.

¹⁹¹ À ce jour, les étiquettes sémantiques sont affectées uniquement aux noms qui désignent de petits êtres et des noms propres (les noms de famille, les prénoms, les toponymes).

¹⁹² Le module a été préparé par Vincent Bénet, certains fichiers sont disponibles sur le site <https://nooj.univ-fcomte.fr/resources.html>

¹⁹³ https://github.com/brown-uk/dict_uk

voulu utiliser ce dictionnaire tel quel, mais nous avons remarqué certaines incohérences dans les étiquettes. Par exemple, *боженько* [*bon Dieu*], le nom masculin commun et variable, puis *Мехіко* [*Mexico*], le nom neutre propre et invariable, ainsi que tous les noms de famille en *-ко* sans distinction du genre, sont associés aux mêmes codes. De plus, ce dictionnaire comporte des doublons et aussi des formes fléchies inscrites en tant que lemmes. Par ailleurs, aucun article n'a été publié au sujet de ce dictionnaire qui pouvait appuyer le choix des étiquettes¹⁹⁴. De ce fait, nous avons dû utiliser uniquement la liste des lemmes sans pouvoir prendre en compte les étiquettes.

En premier lieu, nous avons extrait de ce dictionnaire, en utilisant les fonctions d'Excel, tous les doublons et toutes les formes fléchies ; nous avons alors obtenu 175 985 entrées. Puis, nous avons commencé l'étiquetage de chaque entrée, toujours en utilisant les fonctions d'Excel. Lors de cette tâche, nous avons vu que le dictionnaire comportait diverses formes dérivées décrites comme des entrées indépendantes. Nous avons alors supprimé ces entrées dans la version 1.3, car NooJ nous permet de les relier à leur lemme, grâce aux paradigmes DRV, ou encore grâce à des grammaires morphologiques.

De ce fait, nous avons supprimé du dictionnaire :

- les adjectifs et les adverbes comparatifs, superlatifs et superlatifs absolus ; nous avons gardé uniquement les formes avec les racines différentes comme *гарний* [*bon, beau*] → *кращий* [*meilleur, plus beau*] ;
- les formes courtes des adjectifs ;
- les formes non contractées des adjectifs au nominatif et à l'accusatif féminin, neutre et pluriel ;
- les adjectifs et les adverbes partitifs, excessifs, diminutifs et négatifs ;
- les noms diminutifs ; nous avons gardé uniquement les formes peu productives avec l'alternance de voyelles comme *борода* [*barbe*] → *борідка* [*barbichette*] ;
- les formes avec l'orthographe parallèle qu'on peut systématiser (comme l'alternance des formes graphiques au début des mots, par exemple, *одчиняти* et *відчиняти* [*ouvrir*]) ;

Nous avons également remarqué que de nombreux mots-outils étaient absents du dictionnaire, notamment des locutions grammaticales. Alors, nous avons complété ce dictionnaire avec les adverbes, les prépositions et les interjections supplémentaires à partir du dictionnaire en ligne *Горох*¹⁹⁵.

Ainsi, la version actuelle de notre dictionnaire (Ukr_dictionary_V.1.3.dic) est composée de :

- 81 095 noms, dont 535 abréviations ;
- 50 314 adjectifs ;
- 151 numéraux ;
- 103 pronoms ;
- 12 816 entrées verbales ;
- 11 967 adverbes ;
- 175 prépositions ;
- 205 conjonctions ;
- 132 particules ;

¹⁹⁴ Plus tard, le projet a évolué ; des linguistes ont rejoint l'équipe, et le dictionnaire a reçu le nom de VESUM.

¹⁹⁵ Le dictionnaire est disponible sur <https://goroh.pp.ua>, il est présenté comme un portail qui regroupent les entrées en provenance des sources telles que SUM-20 (<https://sum20ua.com>, le grand dictionnaire académique (2015-2023)), SUM (<https://sum.in.ua>, le grand dictionnaire académique (1970-1980)), l'ULIF (<https://lcorp.ulif.org.ua>, portail linguistique de l'Institut des sciences d'Ukraine (2001-2023)), etc.

- 144 interjections ;
- 42 mots prédicatifs ;
- 31 inserts.

Nous pouvons voir les exemples d'entrées pour chaque catégorie dans le Tableau 32.

Tableau 32. Exemples d'entrées du dictionnaire NooJ par catégories

Catégorie NooJ	Description
Adjectif	аварійний,ADJECTIVE+FLX=ВЕЛИКИЙ
Adverbe	активно,ADVERB+Manner
Conjonction	і,CONJUNCTION
Insert	бува,INSERT
Interjection	бульк,INTERJECTION
Nom	Оксана,NOUN+Feminine+Animate+Proper+F_Name+FLX=MAMA
Numéral	вісімдесятий,NUMERAL+Ordinal+FLX=ВЕЛИКИЙ
Particule	ж,PARTICLE
Mot prédicatif	треба,PREDICATIVE
Préposition	у,PREPOSITION
Pronom	ваш,PRONOUN+Possessive+FLX=НАШ
Entrée verbale	читати,VERB+Pair="читати/прочитати"+FLX=ЧИТАТИ+DRV=ПРО:ЧИТАТИ_PF ходить,VERB+Motion+Indeterminate+FLX=ХОДИТИ

Dans ce tableau, chaque entrée est associée à sa catégorie (il s'agit d'une contrainte incontournable), les catégories variables sont associées à un paradigme FLX, les entrées verbales sont associées en plus à un ou plusieurs paradigmes DRV que nous allons décrire dans la section 4.3.3.

Certains lemmes ont été associés à diverses valeurs. Ainsi :

- les adjectifs, les adverbes, les numéraux et les pronoms ont été associés à une propriété « Type » ;
- les noms ont été associés à une des propriétés « Animation », « Gender » (ou « Number » pour les noms plurales tantum), « Type » (ici, il s'agit des noms communs ou propres), puis certains noms ont été associés également à des propriétés sémantiques (la propriété « Semantic »), comme « F_Name » pour les prénoms, « L_Name » pour les noms de famille, « Toponym » pour les toponymes ;
- certains adjectifs, adverbes et noms ont été associés à une propriété « Expression » ;
- les entrées verbales ont été associées à la propriété « Aspect » (6 601 entrées), « Type » (uniquement pour les verbes de mouvement) et à la propriété « Motion_Type » (pour les verbes de mouvement non préfixés).

4.3.2 Description du dictionnaire secondaire

La liste des lemmes que nous avons utilisée comportait également les participes. Comme nous l'avons dit dans le chapitre 3, les grammairiens ukrainiens ne s'accordent pas sur la nature des participes¹⁹⁶. Dans les outils de TAL, ces ALU sont décrites soit comme des adjectifs, soit comme des participes, sans aucun lien avec le verbe de départ, ou encore comme verbes, mais avec le participe masculin en tant que lemme (voir Tableau 23).

¹⁹⁶ Certains participes sont considérés comme adjectifs, d'autres comme d'anciennes formes ne s'utilisant plus dans la langue moderne.

Nous avons décidé de construire une grammaire morphologique capable de produire les participes à partir des verbes. Par ailleurs, nous avons jugé utile d'étiqueter les participes de notre source et de les mettre dans le dictionnaire secondaire. En donnant une faible priorité à ce dictionnaire par rapport à la grammaire morphologique, NooJ pourra analyser ces formes de participe quand notre grammaire ne les reconnaît pas¹⁹⁷.

Comme nous l'avons vu un peu plus haut, chaque entrée du dictionnaire NooJ, doit être associée d'abord à une catégorie. Après réflexion, nous avons décidé de créer une catégorie supplémentaire « PARTICIPLE », car attribuer l'étiquette « VERB+Participle » à ces ALU nous a paru inapproprié : en effet, les verbes sont inscrits dans les dictionnaires uniquement à l'infinitif. De plus, certaines de ces ALU sont utilisées aujourd'hui presque exclusivement en tant qu'adjectifs, mais les étiqueter seulement avec la catégorie « ADJECTIVE » nous a paru insuffisant, car cela reviendrait à ignorer leur lien avec le verbe. Par ailleurs, NooJ permet d'ajouter une double catégorie à une entrée : ainsi, pour les ALU utilisées comme des adjectifs, on peut ajouter l'étiquette « ADJECTIVE+PARTICIPLE ».

Ainsi, après avoir ajouté la catégorie « PARTICIPLE » dans le fichier « _properties.def », nous avons préparé un bloc d'étiquettes supplémentaires que nous pouvons voir en Figure 33.

Figure 32. Description des participes en tant que catégorie

```
#####
##### PARTICIPLE #####
#####
# Count of categories = 6

PARTICIPLE_Aspect = Imperfective | Perfective; # 2
PARTICIPLE_Case = Accusative | Dative | Genitive | Instrumental | Nominative | Locative; # 6
PARTICIPLE_Gender = Feminine | Masculine | Neuter; # 3
PARTICIPLE_Number = Singular | Plural; # 2
PARTICIPLE_Tense = Past | Present; # 2
PARTICIPLE_Voice = Active | Passive; # 2
```

En accord avec ces étiquettes, nous avons préparé le dictionnaire secondaire (« Ukr_dictionary_PARTICIPLES_V.1.3 ») qui inclut 13 070 lemmes tels que ceux-ci :

діючий,ADJECTIVE+PARTICIPLE+Imperfective+Active+Present+FLX=ВЕЛИКИЙ
добитий,PARTICIPLE+Perfective+Passive+Past+FLX=ВЕЛИКИЙ

Notons que cette solution reste provisoire, nous comptons y revenir ultérieurement.

4.3.3 Les ALU dans le dictionnaire du module ukrainien

Comme nous l'avons dit dans le chapitre 2, NooJ permet de formaliser les locutions, les mots avec des orthographes alternatives et les formes dérivées directement dans le dictionnaire.

Ainsi, dans notre dictionnaire, nous traitons un grand nombre de locutions grammaticales, comme :

- от тобі й на,INTERJECTION+UNAMB : ici, on décrit l'interjection qui exprime l'étonnement pour un changement de situation ;
- все-таки,PARTICLE+UNAMB : ici, on décrit la particule modale ;
- під час,PREPOSITION+UNAMB : ici, on décrit la préposition *pendant* ;

¹⁹⁷ NooJ offre la possibilité d'associer à chaque ressource linguistique un niveau de priorité ; ainsi, lors de l'annotation du texte, l'outil utilise les ressources en fonction de ces priorités.

- тим часом як, CONJUNCTION+UNAMB : ici, on décrit la conjonction *pendant que*.

Le trait « UNAMB » permet de considérer la locution, écrite avec un espace ou avec un trait d'union, comme une seule unité atomique non-ambiguë.

La plateforme NooJ nous a permis de décrire les ALU qui, selon les normes orthographiques de 1933, de 1990 et de 2019, s'écrivent avec différentes orthographes :

- проект, проєкт, NOUN+Masculine+Common+Inanimate+FLX=ЛИСТ : ici, on traite le nom *projet* qu'on écrit depuis la réforme d'orthographe de 2019 avec « е » ;
- ґрунтувати, ґрунтувати, VERB+FLX=ПИСУВАТИ : ici, on traite le verbe *argumenter*, qu'on écrit depuis la réforme de l'orthographe de 1990 avec « ґ » ;
- віце_прем'єр, NOUN+Masculine+Common+Animate+FLX=АВТОР+UNAMB : ici, on décrit le nom *vice-premier (ministre)* qu'on écrit depuis la réforme de 2019 en un seul mot, alors qu'avant cette ALU s'écrivait avec un trait d'union.

Nous avons décrit aussi quelques variantes de l'ukrainien de la rive gauche et de la rive droite. En effet, aujourd'hui, les premières sont considérées comme des calques du russe, et sont remplacées par les secondes :

- зал, зала, NOUN+Masculine+Common+Inanimate+FLX=ЛИСТ : ici, on décrit le nom *salle* ;
- вилка, виделка, NOUN+Feminine+Common+Inanimate+FLX=КВИТКА : ici, on décrit le nom *fourchette*.

Nous avons également décrit 535 abréviations de la façon suivante :

- ЄС, Європейський Союз, NOUN+Abbreviation+Masculine+Invariable : ici, on décrit l'abréviation invariable *Union européenne (UE)*, dont l'unité principale est un nom masculin ;
- АЕС, атомна електростанція, NOUN+Abbreviation+Feminine+Invariable : ici, on décrit l'abréviation invariable *Centrale nucléaire* dont l'unité principale est un nom féminin.

Cela nous a également permis de formaliser les diminutifs des prénoms et les formes dérivées non productives ou peu productives, exclues des grammaires morphologiques productives :

- Сашко, Олександр, NOUN+Masculine+Animate+Proper+F_Name+Diminutive+FLX=БАТЬКО : ici, on décrit le diminutif non productif du prénom *Sachko* → *Olexendr* ;
- боженька, Бог, NOUN+Masculine+Common+Animate+Diminutive+FLX=ТИТКА : ici, on décrit la forme diminutive non productive du nom masculin *bon dieu* → *Dieu* ;
- гірший, поганий, ADJECTIVE+Comparative+FLX=ВЕЛИКИЙ : ici, on décrit la forme non productive du comparatif de l'adjectif *pire* → *mauvais*.

4.3.4 Les grammaires flexionnelles et dérivationnelles

Comme nous l'avons vu dans les sections 2.5.1 et 2.5.2, dans NooJ, les entrées lexicales variables sont associées aux paradigmes flexionnels et dérivationnels qui sont décrits dans une grammaire rangée dans un fichier avec l'extension « .nof ». Nous avons préparé deux fichiers de ce type : le premier, « VERB_Pardadigms_V.1.3.nof », contient les paradigmes FLX et DRV qui décrivent les verbes, et le second, « OTHER_Pardadigms_V.1.3.nof », contient des paradigmes FLX pour les noms, les adjectifs, les numéraux et les pronoms.

Le lien entre les entrées du dictionnaire et ces paradigmes se fait grâce aux instructions données au début du dictionnaire :

```
#use VERB_Pardadigms_V.1.3.nof
#
#use OTHER_Pardadigms_V.1.3.nof
```

Dans cette section, nous proposons de décrire l'élaboration de nos paradigmes pour chaque classe d'ALU.

Les paradigmes (FLX) pour les noms

Pour décrire les noms, nous avons construit 183 paradigmes. Nous pouvons voir l'exemple d'un paradigme en Figure 33.

Figure 33. Exemple d'un paradigme FLX nominal dans NooJ

```

ГРУПА_МАМА=
<E>/Nominative+Singular|
<B>у/Accusative+Singular|
<B>и/Genitive+Singular|
<B>і/Locative+Singular|
<B>ї/Dative+Singular|
<B>юю/Instrumental+Singular|
<B>о/Vocative+Singular|
<B>и/Nominative+Plural|
<B>и/Accusative+Plural+Inanimate|
<B>/Accusative+Plural+Animate|
<B>/Genitive+Plural|
<B>ам/Dative+Plural|
<B>ами/Instrumental+Plural|
<B>ах/Locative+Plural|
<B>и/Vocative+Plural;

```

Les paradigmes flexionnels prennent en compte la classe flexionnelle¹⁹⁸ et les formes graphiques des formes fléchies. Nous avons aussi décidé de réunir dans un même paradigme les variantes pour les noms animés et les noms inanimés. En effet, la différence dans la déclinaison pour les noms animés et inanimés se fait au niveau de l'accusatif singulier et/ou pluriel : les noms animés qui relèvent de ce registre ont l'accusatif qui est semblable au génitif, tandis que les noms inanimés ont l'accusatif qui est semblable au nominatif.

Ainsi, le paradigme qui est présenté dans la Figure 33 s'applique à la fois aux noms inanimés qui se déclinent à l'accusatif pluriel comme « ГРУПА » et aux noms animés qui se déclinent comme « МАМА ».

Nous avons indiqué pour chaque entrée nominale de notre dictionnaire la valeur de la propriété NooJ « Animation », c'est-à-dire soit « Animate » soit « Inanimate ». Ainsi, lors de la production des formes, notre grammaire vérifiera dans le dictionnaire si l'ALU est animée ou inanimée et la fléchira avec une seule de ces deux formes. Comme une grande partie de nos paradigmes inclut ces deux lignes, cela a permis de réduire le nombre de paradigmes à décrire et à appliquer.

De la même façon, nous traitons les noms qui ne s'utilisent qu'au singulier : nous indiquons qu'une entrée de dictionnaire produit uniquement les formes au singulier, en lui ajoutant l'étiquette « Singular », ainsi les formes du pluriel ne seront pas produites.

Dans le Tableau 33, nous pouvons voir trois entrées qui sont fléchies grâce au même paradigme « ГРУПА_МАМА ». La première de ces entrées ne sera associée qu'à 14 formes fléchies et non 15 qui sont décrites dans le paradigme, dont l'accusatif pluriel semblable au nominatif. La deuxième sera aussi associée à 14 formes fléchies, dont l'accusatif pluriel semblable au génitif. Quant à la troisième, elle ne sera associée qu'à 7 formes fléchies du singulier.

¹⁹⁸ Pour rappel : en ukrainien, les noms sont séparés en quatre classes flexionnelles (voir la section 3.3.1).

Tableau 33. Trois types d'entrées, accordées au paradigme « ГРУПА_МАМА »

N°	Entrées du dictionnaire	Nombre de formes fléchies
1.	адреса, NOUN+Feminine+Common+Inanimate+FLX=ГРУПА_МАМА	14 formes
2.	актриса, NOUN+Feminine+Common+Animate+FLX=ГРУПА_МАМА	14 formes
3.	Аль-Каїда, NOUN+Feminine+Inanimate+Proper+FLX=ГРУПА_МАМА+Singular+UNAMB	7 formes

Dans le Tableau 34, nous pouvons voir tous les paradigmes que nous avons formalisés pour les noms de la première classe flexionnelle. Ils ont des disparités au locatif singulier et/ou au vocatif singulier et/ou au génitif pluriel, que nous avons indiquées dans le tableau.

Tableau 34. Les paradigmes pour les noms de la classe I

Groupe	Paradigme	Locatif sg. (alternance)	Vocatif sg.	Génitif pl. (alternance, + voyelle mobile)	N°
Dur	ГРУПА_МАМА		-o	-∅	1
	РОБОТА_ОСОБА		-o	-∅, o/i	2
	СЛЬОЗА		-o	-∅, bo/i	3
	ЦЕРКВА_ВІДЬМА		-o	-∅, +o	4
	ГРИВНА_КНЯЗІВНА		-o	-∅, +e	5
	ВІЙНА		-o	-∅, iŷ/oε	6
	РУКА_КОНЯКА	к/ц	-o	-∅	7
	ЩОКА	к/ц	-o	-∅, o/i	8
	КВІТКА_ТІТКА	к/ц	-o	-∅, +o	9
	ВИМОГА_ПОДРУГА	з/з	-o	-∅	10
	НОГА	з/з	-o	-∅, o/i	11
	СТРИХА_СВЕКРУХА	х/с	-o	-∅, o/i	12
	ПАНЧОХА_БЛОХА	х/с	-o	-∅	13
Mou	ГҮЛЯ_КРАЛЯ		-e	-∅, b	14
	РУЦЯ_МАТУСЯ		-ю	-∅, b	15
	ВЕЧЕРЯ_ТЕТЕРЯ		-e	-∅	16
	КУХНЯ_ВАХНЯ		-e	-∅, b, +o	17
	ЗЕМЛЯ_ЧАПЛЯ		-e	-∅, b, +e	18
	ВІТАЛЬНЯ		-e	-∅, b, +b/e	19
	ДВІЙНЯ		-e	-∅, b, +ŷ/ε	20
	БІЙНЯ		-e	-∅, b, iŷ/oε	21
	БУКЛЯ_ПЕРЕБЕНДЯ		-e	-iв	22
	СТАТТЯ_ПОПАДЯ		-e	-eŷ	23
Mou (en -[j]я)	АВАРІЯ_МАРІЯ		-ε	-∅, ŷ	24
	СІМ'Я		-ε	-eŷ	25
Mixte	БАРЖА_СВЯТОША		-e	-∅	26
	РАДЖА		-e	-iв	27
	МИША		-e	-eŷ	28

Dans les Tableaux 35 et 36, nous pouvons voir les paradigmes que nous avons formalisés pour les noms de la deuxième classe flexionnelle. Bien que la majorité des noms masculins et des noms neutres soient classés dans cette classe, ils ont beaucoup de différences : par exemple, les noms neutres ont le vocatif singulier semblable au nominatif et n'ont pas de distinction de l'animé/inanimé à l'accusatif singulier, ils ont le locatif II et le datif II uniquement pour les animés en -ко, et ont le nominatif pluriel différent des noms masculins, puis ils ont le génitif pluriel disparate. De ce fait, nous avons dû construire des paradigmes distincts pour les noms masculins et les noms neutres. Notons seulement qu'en tête de ces deux tableaux, nous avons indiqué uniquement les grammèmes qui différencient nos paradigmes.

Tableau 35. Les paradigmes pour les noms masculins de la classe II

Groupe	Paradigme	Gén. -у	Gén. -ю	Gén. -а	Gén. -я	Loc. -і	Loc. -ї	Loc. -у	Loc. -ю	Loc. -ові	Loc. -еві	Loc. -єві	Voc. -е	Voc. -у	Voc. -ю	Changement du radical	N°
Dur (у compris en gutturale sans l'alternance de consonnes et en -р)	ЛИСТ	x		x		x		x		x			x				29
	ПЛАКАТ_АВТОР			x		x		x		x			x				30
	МЕД	x				x		x		x			x				31
	СИН			x		x		x		x				x			32
	СЛОВНИК_АЛХІМІК			x				x		x				x			33
	БАНК	x						x		x				x			34
	ВУЗОЛ_МАЙСТЕР			x		x		x		x			x			voyelle mobile (-1)	35
	СОН	x				x		x		x			x	x		voyelle mobile (-1)	36
	ВІТЕР	x				x		x		x			x			voyelle mobile (-1)	37
	ВІЗОК_БИЧОК			x				x		x				x		voyelle mobile (-1)	38
	БУДИНОК	x						x		x				x		voyelle mobile (-1)	39
	СТІЛ	x		x		x		x		x			x			i/o	40
	СНІП_КІТ			x		x		x		x			x			i/o	41
	ХВІСТ_ШИЛОХВІСТ			x		x		x		x			x			i/o (-2)	42
	ВІСК	x						x		x				x		i/o (-2)	43
	ВЕРТОЛІТ			x		x		x		x			x			i/bo	44
	ВИЛІТ	x				x		x		x			x			i/bo	45
	ПАПІР	x		x		x		x		x			x			i/e	46
	КИЇВ			x		x		x		x			x			i/ε	47
РИМЛЯНИН			x		x		x		x			x			иn (g. pl)	48	
БАТЬКО			x				x		x				x		+1 au nom.sg	49	
Groupe	Paradigme	Gén. -у	Gén. -ю	Gén. -а	Gén. -я	Loc. -і	Loc. -ї	Loc. -у	Loc. -ю	Loc. -ові	Loc. -еві	Loc. -єві	Voc. -е	Voc. -у	Voc. -ю	Changement du radical	N°
Dur (en gutturale + l'alternance de consonnes)	АРХІПЕЛАГ	x				зі		x		x				x			50
	БОГ			x				x		x			же				51
	ПЛУГ			x		зі		x		x			же				52
	ЛУГ	x				зі		x		x			же				53
	РІГ	x		x		зі		x		x			же	x		i/o	54
	НОСОРІГ			x				x		x			же			i/o	55
	КАПЕЛЮХ			x		сі		x		x			ше				56
	СКОМОРОХ			x				x		x			ше				57
	ВІК	x				ці		x		x				x			58
	БАК			x		ці		x		x				x			59
	КОЗАК			x				x		x			че				60
	ПАРУБОК			x				x		x			че			voyelle mobile (-2)	61
РІК	x				ці		x		x			че			i/o	62	
Mixte (en chuintante et en -р)	ГАРАЖ_МОРЖ			x		x		x			x		x				63
	ПРОДАЖ	x				x		x			x		x				64
	НІЖ_НЕБІЖ			x		x		x			x		x			i/o	65
	МЕЧ_ТОВАРИШ			x		x		x			x			x			66
	СМЕРЧ	x				x		x			x			x			67
	КІВШ			x		x		x			x			x		i/o (-2)	68
	ХАБАР_КОМАР			x		x		x		x			x			mou au pl.	69
ЛЕМІШ			x		x		x			x			x		i/e	70	
Mou (en -р)	МОНАСТІР_СЛЮСАР				x	x			x		x				x		71
	ЯНТАР		x			x			x		x				x		72

	ЯКІР				x	x			x		x			x	<i>i/o</i>	73	
Mou (en -ї)	НОСІЙ_ЕНЕРГОНОСІЙ				x		x		x		x			x		74	
	МУЗЕЙ		x				x		x		x			x		75	
	ПРИСТРІЙ		x				x		x		x			x	<i>i/o</i>	76	
	КИТОБІЙ				x		x		x		x			x	<i>i/o</i>	77	
	РІЙ		x		x		x		x		x			x	<i>i/o</i>	78	
Groupe	Paradigme	Gén. -у	Gén. -ю	Gén. -а	Gén. -я	Loc. -і	Loc. -ї	Loc. -у	Loc. -ю	Loc. -ові	Loc. -еві	Loc. -еві	Voc. -е	Voc. -у	Voc. -ю	Changement du radical	N°
Mou (en -ь)	АНСАМБЛЬ		x			x			x		x				x		79
	БАРИЛЬ_ВЧИТЕЛЬ					x	x		x		x				x		80
	КУКІЛЬ		x			x			x		x				x	<i>i/o</i>	81
	ТЕРНОПІЛЬ_КІНЬ					x	x		x		x				x	<i>i/o</i>	82
	ГРЕБІНЬ_ДЖМІЛЬ					x	x		x		x				x	<i>i/e</i>	83
	КИСІЛЬ		x			x			x		x				x	<i>i/e</i>	84
	ДЕНЬ_ЖУРАВЕЛЬ					x	x		x		x				x	voyelle mobile (-2)	85
	ПЕРЕЦЬ		x			x			x		x				x	voyelle mobile (-2)	86
	ВАЛЕЦЬ_ЖИЛЕЦЬ					x	x		x		x				x	<i>e/ь</i>	87
	СМАЛЕЦЬ		x			x			x		x				x	<i>e/ь</i>	88
	ОКРАЄЦЬ_АВСТРАЛ ІЄЦЬ					x	x		x		x				x	<i>ε/ü</i>	89
	БОЄЦЬ					x	x		x		x				x	<i>oe/iü</i>	90
ЧЕПЕЦЬ_ГОНЕЦЬ					x	x		x		x				x	voyelle mob. (-2), <i>e/i, o/i</i>	91	

Tableau 36. Les paradigmes pour les noms neutres de la classe II

Groupe	Paradigme	Gén. -а	Gén. -я	Loc. -і	Loc. -у	Loc. -ю	Ins. -ом	Ins. -ем	Ins. -ям	Ins. pl. -ьми	Gén pl. -ø	Gén pl. -ів	N°
Dur	ТІЛО_БАЗИКАЛО	x		x	x		x				x		92
	КОЛІНО	x		x	x		x			x	x		93
	ВІЙСЬКО_БАБИСЬКО	x			x		x				x		94
	БРОВЕНЯТКО_ДИТЯТКО	x			x		x				x+o		95
	ВІКНО	x		x	x						x+o		96
	ВІДРО	x		x	x		x				x+e		97
	СЛОВО	x		x	x		x				x, o/i		98
	ВЕЛИЧАЙКО	x			x		x					x	99
	ІКЛО	x		x	x		x					x	100
Mou	ПОЛЕ		x	x		x	x					x	101
	МІСЦЕ		x	x		x		x			x+b		102
	ВИНЦЕ		x	x		x		x			x+b, +e		103
	КРИЛЬЦЕ		x	x		x		x			x+b, +b/e		104
	ЦУПАЛЬЦЕ		x	x		x		x			x+b, +b/e	x	105
	ВІДЧУТТЯ		x	x		x			x			x	106
	ЗАНЯТТЯ		x	x		x			x		x+b		107
	ДЕСЯТИРІЧЧЯ		x	x		x			x		x		108
En [j]я	ВЕРХІВ'Я		x	ї		x			x			ів	109
Mixte	ГОРИЩЕ_ДІТИЩЕ	x		x	x			x			x		110

Les paradigmes que nous avons formalisés pour la troisième classe flexionnelle sont décrits dans le Tableau 37. Les ALU qui appartiennent à cette classe se distinguent entre l'accusatif animé et inanimé uniquement au pluriel. Ils ont des disparités au génitif singulier et/ou l'instrumental et peuvent avoir une alternance de voyelles au radical, que nous avons indiquées dans le tableau.

Tableau 37. Les paradigmes pour les noms de la classe III

Groupe	Paradigme	Génitif sg.	Instrumental sg.	Alternance au radical	N°
Mou	БОРТЬ_ЧЕРНЬ	-i			111
	ПОВІСТЬ	-i et -u			112
	ВІДОМІСТЬ_ОСОБИСТІТЬ	-i et -u		i/o	113
	ВІДСТАНЬ_ГАЗЕЛЬ	-i	Consonne double		114
	СІЛЬ_МІЛЬ	-i	Consonne double	i/o	115
	ПОСТІЛЬ	-i	Consonne double	i/e	116
En chuintante	ПРИГОРЩ	-i			117
	ЗУСТРІЧ_ПОКРУЧ	-i	Consonne double		118
	НІЧ	-i	Consonne double	i/o	119
	ПІЧ	-i	Consonne double	i/e	120
En labiale et -p	ВЕРФ_ТВАР	-i	-'ю		121

La quatrième classe regroupe les noms neutres qui ont les terminaisons *-а*, *-я* au nominatif singulier et les suffixes *-ам-*, *-ят-* et *-ен-* lors de la déclinaison. Dans les paradigmes que nous avons formalisés pour cette classe, nous ne faisons pas la distinction entre les ALU qui se terminent par *-а* et celles qui se terminent par *-я*, car techniquement, nous n'avons pas besoin de prendre en compte la voyelle. Par exemple, les ALU *коліща* [petite roue] et *немовля* [bébé] sont associées aux formes fléchies suivantes :

- *коліща* : *коліщ-а*, *коліщ-амі*, *коліщ-амі*, *коліщ-ам*, *коліщ-ата*, *коліщ-ат*, *коліщ-атам*, *коліщ-атами* ;
- *немовля* : *немовл-я*, *немовл-ямі*, *немовл-ямі*, *немовл-ям*, *немовл-ята*, *немовл-ят*, *немовл-ятам*, *немовл-ятами*.

Dans cet exemple, on voit que si l'on ne sépare pas la première lettre du radical, les deux ALU ont la même partie variable. De ce fait, il est possible de préparer un seul paradigme pour ces deux ALU, en établissant comme base de déclinaison non pas le radical, mais le lemme entier.

En conséquence, pour les noms de la classe IV, nous avons formalisé deux paradigmes que nous pouvons voir dans le Tableau 38.

Tableau 38. Les paradigmes pour les noms de la classe IV

Paradigme	Le suffixe	N°
ІМ'Я	-ен-	122
КОЛІЩА	-ам- et -ят-	123

Pour faciliter la lecture, mais aussi l'accord des entrées avec leurs formes fléchies, nous avons décidé de construire des paradigmes distincts pour les ALU au pluriel tantum, d'autant plus que certains noms de ce type ont des terminaisons parallèles. Nous pouvons les voir dans le Tableau 39. Dans notre tableau, nous avons mis les cas par lesquels nos paradigmes se différencient ; les terminaisons parallèles sont séparées par un point-virgule.

Nous avons également formalisé les paradigmes qui traitent les déclinaisons non productives, que nous pouvons voir dans le Tableau 40.

Tableau 39. Les paradigmes pour les noms plurales tantum

Paradigme	Gén.	Loc.	Dat.	Ins.	N°
ОКУЛЯРИ* ¹⁹⁹	-ів	-ах	-ам	-ами	124
НАРДИ*	∅	-ах	-ам	-ами	125
РЕЙТУЗИ	∅, -ів	-ах	-ам	-ами	126
ГОНИ	∅, о/і				127
ВИСІВКИ*	∅, +о	-ах	-ам	-ами	128
ГУСЛА	∅, +е	-ах	-ам	-ами	129
ВОРОТА	∅, і mobile	-ах ; -ях	-ам ; -ям	-ами ; -ьми+і	130
КУРИ	-еї	-ах	-ам	-ми	131
ГРУДИ*	-еї	-ях	-ям	-ьми	132
САНИ	-еї	-ях	-ям	-ьми ; -ями	133
ДВЕРІ	-еї	-ях	-ям	-ума ; -ми	134
ГРОШІ	-еї	-ах	-ам	-ума ; -ми+і	135
БРАТТЯ	-ів	-ях	-ям	-ями	136
ВЕЧОРНИЦІ	∅+ь	-ях	-ям	-ями	137
ГРАБЛІ	∅+ь, +е	-ях	-ям	-ями	138
ПОКОЇ	-ів	-ях	-ям	-ями	139
ВАКАЦІЇ	∅+ї	-ях	-ям	-ями	140
ПОМІЇ	∅+ї ; -ів	-ях	-ям	-ями	141
ДВЕРЦІ	∅+ь, +е	-ях	-ям	-ями	142
ДРІВЦЯ	∅+ь, і/о, +е	-ях	-ям	-ями	143
ГРАБЛІ	∅+ь, +е ; -ів	-ях	-ям	-ями	144

Tableau 40. Les paradigmes particuliers

Paradigme	Description de l'exception	N°	Paradigme	Description de l'exception	N°
І_ГРА	radical ²⁰⁰ ігр- (au pl.)	144			
ДОШКА	дошок (gén. pl.)	145	ЧЕРНЕЦЬ	radical ченц'-	164
ДІВЧИНА	radical дівчата (au pl.)	146	ХАЗЯЇН	хазяї, хазяїв (gén. pl.)	165
ЛЮДИНА	radical люди (au pl.)	147	ЧЕТВЕР	radical четверг-	166
ВІВЦЯ	овець (gén. pl.)	148	ОКО	radical оч'-, (au pl.), вічі (gén. pl.)	167
РІЛЛЯ	ріль (gén. pl.)	149	КОЛЕСО	коліс, колесами, колісьми (inst. pl.)	168
ГАНДЖ	ганджей, ганджів (gén. pl.)	150	СЕРЦЕ	сердець, серць (gén. pl.)	169
ВЕЛИКДЕНЬ	radical великодні (au pl.)	151	ПЛЕЧЕ	плечей, пліч (gén. pl.)	170
ТИЖДЕНЬ	radical тижн-	152	ЖИТІЄ	житій (gén. pl.)	171
ПАН	панове (voc. pl.)	153	ЯЙЦЕ	яєць (gén. pl.)	172
ГІСТЬ	гостями, гістьми (inst. pl.)	154	ВУХО	вух et ушей (gén. pl.)	173
РАЗ	раз et разів (gén. pl.)	155	БЛАГО	блазі (loc. sg.)	175
ОВЕС	radical вісв-	156	ДНО	radical ден- (au pl.)	176
ДРУГ	radical друз'- (au pl.)	157	СКЛО	radical стекл- (au pl.)	177
ГОРЩОК	radical горш-	158	НЕБО	radical небес- (au pl.)	178
МРЕЦЬ	radical мерц'-	159	ВИМ'Я	radical вим'[j]- et вимен- (au sg.)	179
ГОСПОДЬ	господи (voc. pl.)	160	МАТИ	radical матер-	180
ЧОБІТ	чоботами, чобтями, чобітьми (inst. pl.)	161	ВІСЬ	radical ос-	181
СОЛОВЕЙ	radical солов'[j]	162	ШЕВЕЛЬОВ	Pour les noms de familles, s'accorde au m., f. et pl.	182
ШВЕЦЬ	radical шевц'-	163	САВИН	Pour les noms de familles, s'accorde au m., f. et pl.	183

¹⁹⁹ Les paradigmes avec le symbole « * » s'appliquent aux animés et aux inanimés.

²⁰⁰ Cela signifie que le radical des formes fléchies est différent du radical du lemme.

Les paradigmes (FLX) pour les adjectifs, les participes, les numéraux et les pronoms

Pour décrire les formes fléchies des adjectifs, des participes, des numéraux et des pronoms, nous avons formalisé 48 paradigmes supplémentaires.

Pour prendre en compte les formes fléchies des adjectifs, mais aussi des numéraux ordinaux, des pronoms-adjectifs et des participes du dictionnaire secondaires qui ont les mêmes flexions que les adjectifs qualitatifs et relatifs, nous avons formalisé trois paradigmes flexionnels (voir Tableau 41, N°184, 185, 186) qui correspondent aux déclinaisons décrites dans la section 3.3.2 (Tableau 12). Nous pouvons voir le paradigme « ВЕЛИКИЙ » dans la Figure 34. Ce paradigme inclut aussi la description des formes non contractées qui fléchissent uniquement les adjectifs grâce à l'étiquette « ADJECTIVE ». Le paradigme « СИНИЙ » inclut aussi la description des formes non contractées. En revanche ces formes ne sont pas incluses dans le paradigme « БЕЗКРАЙ ».

Figure 34. Extrait du paradigme « ВЕЛИКИЙ »

```

ВЕЛИКИЙ=
<E>/Nominative+Masculine+Singular |
<B2>ого/Accusative+Animate+Masculine+Singular |
<E>/Accusative+Inanimate+Masculine+Singular |
<B2>ого/Genitive+Masculine+Singular |
<B2>ому/Locative+Masculine+Singular |
<B2>им/Locative+Masculine+Singular |
<B2>ому/Dative+Masculine+Singular |
<B2>им/Instrumental+Masculine+Singular |
<B2>е/Nominative+Neuter+Singular |
<B2>ее/Nominative+Uncontracted+ADJECTIVE+Neuter+Singular |
<B2>е/Accusative+Neuter+Singular |
<B2>ее/Accusative+Uncontracted+ADJECTIVE+Neuter+Singular |
    
```

Bien que la plus grande partie des adjectives d'appartenance soit décrite par une grammaire morphologique, nous avons préparé quatre paradigmes flexionnels pour les adjectifs exclus de cette grammaire dont le schéma dérivatif est non productif (voir Tableau 41), comme pour l'entrée *материн* [appartenant à la mère], formalisé par le paradigme « МАМИН » :

материн,ADJECTIVE+Belonging+"мати"+FLX=МАМИН

Tableau 41. Les paradigmes adjectivaux

Groupe	Paradigme	Description	N°
qualitatif et relatif dur	ВЕЛИКИЙ	Radical <i>велик-</i>	184
qualitatif et relatif mou	СИНИЙ	Radical <i>син'-</i>	185
qualitatif et relatif en « yod »	БЕЗКРАЙ	Radical <i>безкра-</i>	186
d'appartenance	МАМИН	Radical est formé à partir des noms f. et n.	187
d'appartenance	БАТЬКІВ	Radical est formé à partir des noms m. du groupe dur	188
d'appartenance	ДІДУСІВ	Radical est formé à partir des noms m. du groupe mou	189
d'appartenance	СЕРГІЇВ	Radical est formé à partir des noms m. du groupe en « yod »	190

À ces paradigmes, nous avons ajouté deux paradigmes supplémentaires, « ОТОЙ » et « ОЦЕЙ », pour formaliser un petit nombre de pronoms-adjectifs démonstratifs, qui ont un génitif et un instrumental féminins différents des adjectifs (voir Tableau 42).

Nous avons formalisé également les paradigmes (voir Tableau 43) qui décrivent les formes fléchies des pronoms que nous ne pouvons pas décrire avec les paradigmes mentionnés plus haut.

Tableau 42. Comparaison des paradigmes des adjectifs et des pronoms-adjectifs démonstratifs

Paradigme « ВЕЛИКИЙ » (groupe dur)	Paradigme « ОТОЙ » (groupe dur)	N° : 191
<B2>ої/Genitive+Feminine+Singular	<B2>ої/Genitive+Feminine+Singular <B2>ієї/Genitive+Feminine+Singular	
<B2>ою/Instrumental+Feminine+Singular	<B2>ою/Instrumental+Feminine+Singular <B2>ією/Instrumental+Feminine+Singular	
Paradigme « СИНІЙ » (groupe mou)	Paradigme « ОЦЕЙ » (groupe mixte²⁰¹)	N° : 192
<B2>ьої/Genitive+Feminine+Singular	<B2>ієї/Genitive+Feminine+Singular	
<B2>ьою/Instrumental+Feminine+Singular	<B2>ією/Instrumental+Feminine+Singular	

Tableau 43. Les paradigmes pronominaux

Groupe	Paradigme	Traduction	Description	N°
Possessif	МІЙ	<i>mon</i>	radical <i>мо[ʃ]</i> -	193
	НАШ	<i>notre</i>	terminaison « zéro » au N. et à l'Acc. in. m.	194
Déterminé	ВЕСЬ	<i>entier</i>	radical <i>yc'</i> -	195
	УВЕСЬ	<i>entier</i>	radical <i>вс'</i> -	196
	КОЖЕН	<i>chaque</i>	le -e- mobile dans le radical	197
Interrogatif	ЧИЙ	<i>à qui ?</i>	radical <i>чуй-</i> pour m. et <i>чу[ʃ]</i> - pour le f., n. et pl.	198
Interrogatif et relatifs	ЩО	<i>que</i>	alternance de <i>щ/ч</i> dans le radical	199
	ХТО	<i>qui</i>	alternance de <i>хт/к</i> dans le radical	201
Personnel	СЕБЕ	<i>réfléchi</i>	alternance de <i>e/o</i> dans le radical	202
	Я	<i>je</i>	radical <i>мен-</i> avec -e- mobil	203
	ТИ	<i>tu</i>	radical <i>теб-</i> avec l'alternance de <i>e/o</i>	204
	ВІН	<i>il (m.)</i>	radicaux <i>ї-, н'-, н-</i>	205
	ВОНА	<i>elle</i>	radicaux <i>ї-, н'-, н-</i>	206
	ВОНО	<i>il (n.)</i>	radicaux <i>ї-, н'-, н-</i>	207
	МИ	<i>nous</i>	trois formes fléchies <i>нас, нам, нами</i>	208
	ВИ ²⁰²	<i>vous</i>	trois formes fléchies <i>вас, вам, вами</i>	209
	ВОНИ	<i>ils/elles</i>	radicaux <i>[ʃ]-, н-</i>	210

Notons que nous n'avons pas créé de paradigmes FLX pour les pronoms indéterminés qui sont formés par la contraction du pronom et de la particule, car ils seront décrits par une grammaire morphologique.

Comme nous l'avons vu dans la section 3.3.3, les numéraux cardinaux ont beaucoup de disparités dans leurs schémas flexionnels. De ce fait, nous avons formalisé plusieurs paradigmes que nous pouvons voir dans le Tableau 44²⁰³. Dans ce tableau, nous avons également ajouté les numéraux indéterminés²⁰⁴.

²⁰¹ Il n'existe pas de groupe mixte pour les adjectifs.

²⁰² Nous avons nommé ce paradigme « ВИ » pour éviter le conflit avec le paradigme DRV « ВИ » (voir Tableau 48).

²⁰³ Par ailleurs, nous aurions pu décrire la flexion des numéraux cardinaux se terminant par *-cmi, -cma, -com* grâce à une grammaire morphologique à restriction, mais nous avons préféré décrire tous les numéraux avec un seul outil formel pour plus de transparence.

²⁰⁴ Par ailleurs, comme c'était déjà évoqué dans le chapitre précédent, certains numéraux cardinaux pourront être fléchis avec les paradigmes nominaux, par exemple :

мільйон, NUMERAL+Cardinal+FLX=ПЛАКАТ_АВТОР+Inanimate.

Tableau 44. Les paradigmes pour les numéraux

Groupe	Paradigme	Chiffre ou traduction	Description	N°
Numéral	ОДИН	1	trois genres (m., n., f.), base dure	211
	ДВА	2	deux genres (m. et n., f.), base dure	212
	ТРИ	3	base molle	213
	ЧОТИРИ	4	base molle, mais <i>чотирма</i> à l'instrumental	214
Numéral	П'ЯТЬ	5	base molle, mais <i>п'яти</i> au génitif	215
	ШІСТЬ	6	comme « П'ЯТЬ », mais alternance de <i>i/e</i>	216
	СІМ	7	base dure, alternance de <i>i/e</i>	217
	ВІСІМ	8	base dure, alternance de <i>i/o</i>	218
	СОРОК	40	<i>сорока</i> dans tous les cas obliques, terminaison « zéro »	219
	ВІСІМДЕСЯТ	80	comme « П'ЯТЬ », sans <i>ь</i>	220
	ДЕВ'ЯНОСТО	90	<i>дев'яноста</i> , dans tous les cas obliques, terminaison <i>-о</i>	221
	ДВІСТІ	200	déclinaison de deux parties <i>дві-</i> et <i>сті-</i>	222
	ТРИСТА	300	déclinaison de deux parties <i>три-</i> et <i>ста-</i>	223
	ШІСТСОТ	600	déclinaison de deux parties <i>шіст-</i> et <i>сот-</i>	224
	СІМСОТ	700	déclinaison de deux parties <i>сім-</i> et <i>сот-</i>	225
	ВІСІМСОТ	800	déclinaison de deux parties <i>вісім-</i> et <i>сот-</i>	226
	ДЕВ'ЯТСОТ	900	déclinaison de deux parties <i>дев'ят-</i> et <i>сот-</i>	227
Indéterminé	КІЛЬКАСОТ	<i>quelques centaines</i>	déclinaison de deux parties <i>кілька-</i> et <i>сот-</i>	228
	БАГАТО	<i>beaucoup</i>	base molle, terminaison <i>-о</i>	229
Collectif	ОБОЄ	<i>tous les deux</i>	base molle, fin de mot variable <i>-є</i>	230
	ДВОЄ	<i>les deux</i>	uniquement le nominatif et l'accusatif	231

Les paradigmes FLX pour les verbes

Pour décrire les verbes, nous avons formalisé deux types de paradigmes : FLX qui formalisent la conjugaison et DRV qui formalisent le lien entre le verbe IPF et PF du couple aspectuel dominant. Tous les verbes ont été associés aux paradigmes FLX ; par ailleurs, seuls les verbes IPF qui entrent en couple aspectuel avec un verbe PF ont été accordés au paradigme DRV. Reprenons l'exemple que nous avons déjà donné précédemment (Tableau 32) :

читати, VERB+Pair="читати/прочитати"+FLX=ЧИТАТИ+DRV=ПРО:ЧИТАТИ_Р

Ici, à l'entrée « читати » est associée l'étiquette « Pair="читати/прочитати" », qui décrit le couple aspectuel dont le verbe *читати* [lire] fait partie. Cette étiquette a été créée pour les besoins pédagogiques : elle affiche le couple aspectuel lors de l'annotation du texte ; ainsi, les étudiants qui utiliseront la plateforme pourront voir rapidement les deux verbes du couple aspectuel. Le premier élément de cette étiquette ("читати") correspond au verbe IPF, écrit à l'entrée du dictionnaire, alors que le deuxième ("прочитати") correspond au verbe PF, lié au premier, grâce au paradigme dérivationnel DRV=ПРО. Le paradigme « ЧИТАТИ » produit les formes fléchies du verbe IPF. Le verbe PF, formalisé grâce au paradigme DRV, est accordé, à son tour, au paradigme flexionnel « ЧИТАТИ _PF » avec l'opérateur « : ».

Ajoutons qu'il existe des couples aspectuels asymétriques :

- un verbe IPF, lié à deux verbes PF synonymiques, comme

наливати, VERB+Pair="наливати/налити, наляти"+FLX=ЧИТАТИ+DRV=ТИ_В4:ЛИТИ_PF
+DRV=ЛЯТИ:ЧУТИ_PF

- deux verbes IPF synonymiques, liés à un verbe PF, comme

висувати, VERB+Pair="висувати,висовувати/висунути"+FLX=ЧИТАТИ+DRV=НУТИ_В4:ПВАТИ_PF

Dans le premier cas, l'entrée a autant de paradigmes DRV que de verbes PF associés au verbe IPF. Dans le deuxième cas, nous indiquons le paradigme DRV pour le premier verbe IPF du couple asymétrique, et le deuxième verbe IPF est inscrit dans le dictionnaire comme entrée indépendante (sans paradigme DRV), le premier verbe figurant après la virgule :

висовувати, висувати, VERB+Pair="висувати,висовувати/висунути"+FLX=РИСУВАТИ

Les deux paradigmes FLX qui formalisent les conjugaisons des verbes IPF *читати* et *висувати* et du verbe PF *прочитати* sont présentés dans le Tableau 45.

Tableau 45. Les paradigmes FLX « ЧИТАТИ » et « ЧИТАТИ_PF »

Paradigme « ЧИТАТИ »	Paradigme « ЧИТАТИ_PF »
<E>/Imperfective+Infinitive	<E>/Perfective+Infinitive
<B2>в/Imperfective+Masculine+Singular+Past	<B2>в/Perfective+Masculine+Singular+Past
<B2>ла/Imperfective+Feminine+Singular+Past	<B2>ла/Perfective+Feminine+Singular+Past
<B2>ло/Imperfective+Neuter+Singular+Past	<B2>ло/Perfective+Neuter+Singular+Past
<B2>ли/Imperfective+Plural+Past	<B2>ли/Perfective+Plural+Past
<B2>ю/Imperfective+1+Singular+Present	<B2>ю/Perfective+1+Singular+Present
<B2>еш/Imperfective+2+Singular+Present	<B2>еш/Perfective+2+Singular+Present
<B2>є/Imperfective+3+Singular+Present	<B2>є/Perfective+3+Singular+Present
<B2>ємо/Imperfective+1+Plural+Present	<B2>ємо/Perfective+1+Plural+Present
<B2>єте/Imperfective+2+Plural+Present	<B2>єте/Perfective+2+Plural+Present
<B2>ють/Imperfective+3+Plural+Present	<B2>ють/Perfective+3+Plural+Present
<E>му/Imperfective+1+Singular+Future	<B2>й/Perfective+2+Singular+Imperative
<E>ме/Imperfective+3+Singular+Future	<B2>ймо/Perfective+1+Plural+Imperative
<E>меш/Imperfective+2+Singular+Future	<B2>йте/Perfective+2+Plural+Imperative
<E>memo/Imperfective+1+Plural+Future	<B2>вши/Perfective+Gerund+Anterior
<E>мете/Imperfective+2+Plural+Future	<B2>но/ Asubjectal+Perfective+Passive;
<E>муть/Imperfective+3+Plural+Future	
<B2>й/Imperfective+2+Singular+Imperative	
<B2>ймо/Imperfective+1+Plural+Imperative	
<B2>йте/Imperfective+2+Plural+Imperative	
<B2>ючи/Imperfective+Gerund+Concomitant	
<B2>вши/Imperfective+Gerund+Anterior	
<B2>но/Asubjectal+Imperfective+Passive;	

Nous pouvons constater que le paradigme qui décrit le verbe IPF est plus complet, car il inclut la déclinaison du futur et le gérondif concomitant, marqués en rouge dans le tableau. Nous aurions pu créer un seul paradigme pour le verbe IPF et le verbe PF qui ont la même conjugaison et faire la distinction entre les valeurs « Imperfective » et « Perfective » en ajoutant :

- l'étiquette « Imperfective » à l'entrée du dictionnaire ;
- l'étiquette « Perfective » au paradigme DRV ;
- l'étiquette « Imperfective » aux lignes, notée en rouge dans notre premier paradigme.

Cette solution, plus simple techniquement, nous aurait évité de dédoubler nos paradigmes. Cependant, pour garder la transparence pour l'utilisateur linguiste, nous avons préféré construire un paradigme pour le verbe IPF, et un autre pour le verbe PF. De plus, le fait d'avoir deux paradigmes

distincts nous permettra, par la suite, de créer plus facilement la grammaire morphologique qui formalise les participes²⁰⁵.

Ainsi, nous avons formalisé 96 paradigmes (voir Tableau 46), dont 89 sont dédoublés pour les verbes IPF et PF, comme le Tableau 45 l'indique. Comme nous l'avons dit dans le chapitre précédent (section 3.3.5), nous avons utilisé les paradigmes proposés par Oleg Chinkarouk. Cependant, nos paradigmes sont légèrement différents, car nous n'avons pas besoin de créer des paradigmes spécifiques pour certains verbes. Par exemple, les verbes du type *любити* [*aimer*] (le suffixe *-u-* à l'infinitif) et du type *conimu* [*respirer fort*] (le suffixe *-i-* à l'infinitif) sont fléchis selon le même paradigme²⁰⁶. Par ailleurs, les verbes du type *клясти* [*jur*, *maudire*] et *п'ясти* [*pousser en avant*] ont la même déclinaison, mais pour les formes fléchies du premier verbe, nous avons besoin d'effacer les trois derniers caractères graphiques, alors que pour le deuxième, nous avons besoin d'en effacer quatre²⁰⁷. De ce fait, nous avons dû construire deux paradigmes différents avec NooJ.

Dans la colonne « Description » du Tableau 46 qui décrit tous les paradigmes flexionnels, nous indiquons la base de l'infinitif, sur laquelle sont construites les formes fléchies du passé, et la base du présent, sur laquelle sont construites les formes fléchies du présent. Par ailleurs, certains paradigmes se différencient uniquement par la forme asubjectale ou par l'impératif que nous indiquons pour ces paradigmes-là.

Tableau 46. Les paradigmes FLX pour les verbes

N° du paradigme verbal	Paradigmes	с. ²⁰⁸	Description	N° NooJ
1	ЧИТАТИ* ²⁰⁹	I	l'infinitif : <i>-a</i> ²¹⁰ , <i>-я-</i> , le présent : <i>-aj-</i> , <i>-яj-</i>	232
2	РИСУВАТИ*	I	l'infinitif : <i>-yva-</i> , le présent : <i>-yj-</i>	233
3	МАЛЮВАТИ*	I	l'infinitif : <i>-yva-</i> , le présent : <i>-yj-</i> , la forme asubjectale <i>-ьовано</i>	234
4	ВОЮВАТИ*	I	l'infinitif : <i>-yva-</i> , le présent : <i>-yj-</i> , la forme asubjectale <i>-йовано</i>	235
5	ВИСМІЮВАТИ*	I	l'infinitif : <i>-yva-</i> , le présent <i>-yuj</i> ²¹¹ , la forme asubjectale <i>-ювано</i>	236
6	ДАВАТИ*	I	l'infinitif : <i>-va-</i> , le présent : <i>-j-</i>	237
7	ВІЯТИ*	I	l'infinitif : <i>-я-</i> , le présent : <i>-j-</i>	238
8	РВАТИ*	I	l'infinitif : <i>-a-</i> , <i>-i-</i> , <i>-ну-</i> , le présent : \emptyset , <i>-u</i> à l'impératif sg.	239
9	ЇХАТИ*	I	l'infinitif : <i>-a-</i> , le présent : \emptyset , le radical du présent <i>їд-</i>	240
10	ГНАТИ*	I	l'infinitif : <i>-a-</i> , le présent : \emptyset , le radical du présent <i>жен-</i>	241
11	СЛАТИ*	I	l'infinitif : <i>-a-</i> , le présent : \emptyset , le radical du présent <i>шл-</i>	242
12	СЛАТИ2*	I	l'infinitif : <i>-a-</i> , le présent : \emptyset , le radical du présent <i>стел-</i>	243
13	ПРАТИ*	I	l'infinitif : <i>-a-</i> , le présent : \emptyset , <i>+e-</i> , la mobile au radical du présent (<i>пер-</i>)	244

²⁰⁵ Nous indiquerons pour la contrainte que le participe imperfectif est créé à partir du verbe IPF tandis que le participe perfectif – à partir du verbe PF.

²⁰⁶ C'est-à-dire qu'ils ont les mêmes variations au niveau du radical et des flexions.

²⁰⁷ Il faut effacer l'apostrophe et les trois dernières lettres du lemme.

²⁰⁸ Conjugaison (I ou II).

²⁰⁹ Les modèles avec l'étoile (*) ont aussi un modèle perfectif, auquel nous avons ajouté « _PF », comme pour ЧИТАТИ_PF.

²¹⁰ Nous notons ainsi soit le suffixe qui fait partie de la base, soit la dernière voyelle du radical, s'il n'y a pas de suffixe, si la base se termine par une consonne, nous noterons cela avec le symbole « \emptyset ». Si la base de l'infinitif se termine par une voyelle, alors nous ajouterons le suffixe *-ø* au passé masculin, et si la base se termine par une consonne, alors nous n'ajouterons rien.

²¹¹ *-j-* signifie la présence des voyelles *-ю-* et *-є-* dans les terminaisons de la 1^{re} conjugaison, et *-ю*, *-ї-*, *-я-*, dans les terminaisons de la 2^e conjugaison.

N° du paradigme verbal	Paradigmes	c. ²⁰⁸	Description	N° NooJ
14	ПИСАТИ*	I	l'infinitif : -а-, le présent : ∅, l'alternance <i>с/ш</i> au radical du présent	245
15	КАЗАТИ*	I	l'infinitif : -а-, le présent : ∅, l'alternance <i>з/ж</i> , -у à l'impératif sg.	246
16	МАЗАТИ*	I	l'infinitif : -а-, le présent : ∅, l'alternance <i>з/ж</i> , ∅ à l'impératif sg.	247
17	ТКАТИ*	I	l'infinitif : -а-, le présent : ∅, l'alternance <i>к/ч</i> , -у à l'impératif sg.	248
18	ПЛАКАТИ*	I	l'infinitif : -а-, le présent : ∅, l'alternance <i>к/ч</i> , ∅ à l'impératif sg.	249
19	ПОЛОСКАТИ*	I	l'infinitif : -а-, le présent : ∅, l'alternance <i>ск/щ</i> , -у à l'impératif sg.	250
20	ЖАТИ*	I	l'infinitif : -а-, le présent : -н- (<B3> ²¹²), le radical du présent <i>жн-</i>	251
21	М'ЯТИ*	I	l'infinitif : -я-, le présent : -н- (<B4>), le radical présent <i>мн-</i>	252
23	ЙНЯТИ*	I	l'infinitif : -я-, le radical du présent <i>йм-</i> (<B4>)	253
24	ВЗЯТИ_РФ	I	l'infinitif : -я-, le radical du présent <i>візьм-</i>	254
25	ПІДНЯТИ_РФ	I	l'infinitif : -я-, le radical du présent <i>піднім-</i>	255
26	МЕТАТИ*	I	l'infinitif : -а-, le présent : 1) -ај- (comme ЧИТАТИ) ; 2) le radical <i>меч-</i>	256
27	ПЛЕСКАТИ*	I	l'infinitif : -а-, le présent : 1) -ај- (comme ЧИТАТИ) ; 2) le radical <i>плещ-</i>	257
28	ДЕРЖАТИ*	II	l'infinitif : -а-, le présent : ∅	258
29	СПАТИ*	II	l'infinitif : -а-, le présent ∅, + л' à la fin du radical au présent après <i>б, п, в</i> , et <i>м</i> à la 1 ^{ère} personne du sg. et la personne du 3 ^{ème} pl.	259
30	СТОЯТИ*	II	l'infinitif : -я-, le présent ∅	260
31	БИТИ*	I	l'infinitif : -у-, le radical du présent <i>б'іј-</i>	261
32	ЛИТИ*	I	l'infinitif : -у-, le radical du présent <i>лл'-</i>	262
33	ЖИТИ*	I	l'infinitif : -у-, le radical du présent <i>жив-</i>	263
34	ГОВОРИТИ*	II	l'infinitif : -у-, -і-, le présent : ∅, la mouillure de la dernière consonne du radical à la 1 ^{ère} personne du sg. et à la 3 ^{ème} personne du pl., -у à l'impératif sg.	264
35	МИЛИТИ*	II	l'infinitif : -у-, le présent : ∅, la mouillure de la dernière consonne du radical à la 1 ^{ère} personne du sg. et à la 3 ^{ème} personne du pl., -ь à l'impératif sg.	265
36	ВІРИТИ*	II	l'infinitif : -у-, le présent : ∅, la mouillure de la dernière consonne du radical à la 1 ^{ère} personne du sg. et à la 3 ^{ème} personne du pl., ∅ à l'impératif sg.	266
37	ЛЮБИТИ*	II	l'infinitif : -у-, -і-, le présent : ∅, -л'- à la fin du radical au présent après <i>б, п, в, м</i> à la 1 ^{ère} personne du sg. et à la 3 ^{ème} personne du pl., -у à l'impératif sg.	267
38	БАРВИТИ*	II	l'infinitif : -у-, le présent : ∅, -л'- à la fin du radical au présent après <i>б, п, в, м</i> à la 1 ^{ère} personne du sg. et la personne du 3 ^{ème} pl., ∅ à l'impératif sg.	268
39	БАЧИТИ*	II	l'infinitif : -у-, le présent : ∅, le radical en chuintante, ∅ à l'impératif sg.	269
40	ВЧИТИ*	II	l'infinitif : -у-, le présent : ∅, le radical en chuintante, -у à l'impératif sg.	270
41	ПРОСИТИ*	II	l'infinitif : -у-, le présent : ∅, l'alternance de <i>с/ш</i> à la 1 ^{ère} personne du singulier, la mouillure de la dernière consonne du radical à la 3 ^{ème} personne du pl., -у à l'impératif sg.	271
42	КВАСИТИ*	II	l'infinitif : -у-, le présent : ∅, l'alternance de <i>с/ш</i> à la 1 ^{ère} personne du singulier, la mouillure de la dernière consonne du radical à la 3 ^{ème} personne du pl., -ь à l'impératif sg.	272
43	ХОДИТИ*	II	l'infinitif : -у-, le présent : ∅, l'alternance de <i>д/дж</i> à la 1 ^{ère} personne du singulier, la mouillure de la dernière consonne du radical à la 3 ^{ème} personne du pl., -у à l'impératif sg.	273
46	ВАДИТИ*	II	l'infinitif : -у-, le présent : ∅, l'alternance de <i>д/дж</i> à la 1 ^{ère} personne du singulier, la mouillure de la dernière consonne du radical à la 3 ^{ème} personne du pl., -ь à l'impératif sg., l'absence de la forme asubjectale	274

²¹² L'opérateur <B3> signifie que nous devons effacer les trois derniers caractères graphiques, avant d'ajouter la partie variable. L'opérateur <B4> signifie que nous devons effacer les quatre derniers caractères graphiques, etc.

N° du paradigme verbal	Paradigmes	c. ²⁰⁸	Description	N° NooJ
47	ГНІЗДИТИ*	II	l'infinitif : -у-, le présent : ∅, l'alternance de зд/ждж à la 1 ^{ère} personne du singulier, la mouillure de la dernière consonne du radical à la 3 ^{ème} personne du pl., -у à l'impératif sg.	275
48	ЇЗДИТИ*	II	l'infinitif : -у-, le présent : ∅, l'alternance de зд/ждж à la 1 ^{ère} personne du singulier, la mouillure de la dernière consonne du radical à la 3 ^{ème} personne du pl., -ь à l'impératif sg.	276
49	ВОЗИТИ*	II	l'infinitif : -у-, le présent : ∅, l'alternance de з/ж à la 1 ^{ère} personne du singulier, la mouillure de la dernière consonne du radical à la 3 ^{ème} personne du pl., -у à l'impératif sg.	277
50	ЛАЗИТИ*	II	l'infinitif : -у-, le présent : ∅, l'alternance de з/ж à la 1 ^{ère} personne du singulier, -ь à l'impératif sg.	278
51	ПЛАТИТИ*	II	l'infinitif : -у-, -і-, le présent : ∅, l'alternance de т/ч à la 1 ^{ère} personne du singulier, la mouillure de la dernière consonne du radical à la 3 ^{ème} personne du pl., -у à l'impératif sg.	279
52	МІТИТИ*	II	l'infinitif : -у-, le présent : ∅, l'alternance de т/ч à la 1 ^{ère} personne du singulier, la mouillure de la dernière consonne du radical à la 3 ^{ème} personne du pl., -ь à l'impératif sg.	280
53	ГУСТИТИ*	II	l'infinitif : -у-, le présent : ∅, l'alternance de см/щ à la 1 ^{ère} personne du singulier, mouillure de la dernière consonne du radical à la 3 ^{ème} personne du pl., -у à l'impératif sg.	281
54	ЧИСТИТИ*	II	l'infinitif : -у-, le présent : ∅, l'alternance de см/щ à la 1 ^{ère} personne du singulier, la mouillure de la dernière consonne du radical à la 3 ^{ème} personne du pl., -ь à l'impératif sg.	282
55	СВИСТИТИ*	II	l'infinitif : -і-, le présent : ∅, l'alternance de см/щ à la 1 ^{ère} personne du singulier, la mouillure de la dernière consonne du radical à la 3 ^{ème} personne du pl., -у à l'impératif sg.	283
56	ЗЕЛЕНІТИ*	I	l'infinitif : le suffixe -і- le présent : le suffixe -іj-, l'absence de la forme asubjectale	284
57	ХОТИТИ*	I	l'infinitif : -і-, le présent : ∅, l'alternance de т/ч au radical du présent, la mouillure de la dernière consonne du radical à la 3 ^{ème} personne du pl.	285
58	ГНОЇТИ*	II	l'infinitif : -ї-, le présent : -j, -ї à l'impératif sg.	286
59	КЛЕЇТИ*	II	l'infinitif : -ї-, le présent : -j, -ї à l'impératif sg.	287
60	ПОЇТИ*	II	l'infinitif : -ї-, le présent : -j, -ї à l'impératif sg.	288
61	РИНУТИ*	I	l'infinitif : -ну-, le présent : ∅, -ь à l'impératif sg., l'absence de la forme asubjectale	289
62	ТВЕРДНУТИ*	I	l'infinitif : -ну-, au passé m. тверд et тверднув, le présent : ∅, -у à l'impératif sg., l'absence de la forme asubjectale	290
63	ЧУТИ*	I	l'infinitif : -у-, -у-, le présent -j-	291
64	ДУТИ*	I	l'infinitif : -у-, le présent : 1) -j- (comme ЧУТИ) ; 2) le radical дм-	292
65	КЛЯСТИ*	I	l'infinitif (en -сму) : -я-, le présent : + -н- (<B3>)	293
66	П'ЯСТИ*	I	l'infinitif (en -сму) : -я-, le présent : + -н- (<B5>), l'absence de la forme asubjectale	273
67	ПРЯСТИ*	I	l'infinitif (en -сму) : -я-, -а-, le radical du présent пряд-	274
68	СІСТИ_RF		l'infinitif (en -сму) : -і-, le radical du présent сяд-	275
69	ГНИСТИ*	I	l'infinitif (en -сму) : -у-, le présent : -j-	276
70	ПЛИСТИ*	I	l'infinitif (en -сму) : -у-, le radical du présent плив-	277
71	ВЕСТИ*	I	l'infinitif (en -сму) : -е-, le radical passé ві-, le radical du présent вед-	278
72	МЕСТИ*	I	l'infinitif (en -сму) : -е-, le radical passé мі-, le radical du présent мет-, la forme asubjectale -мено	279
73	НЕСТИ*	I	l'infinitif : ∅, au passé masculin alternance e/i, le présent : ∅	280
74	РОСТИ*	I	l'infinitif : ∅, au passé masculin alternance o/i, le radical du présent рост-, -	281

N° du paradigme verbal	Paradigmes	c. ²⁰⁸	Description	N° NooJ
			<i>u</i> à l'impératif sg., la forme asubjectale <i>-щено</i>	
75	ПОВЗТИ*	I	l'infinitif : \emptyset , le présent : \emptyset , <i>-u</i> à l'impératif sg., l'absence de la forme asubjectale	282
76	ЛІЗТИ*		l'infinitif : \emptyset , le présent : \emptyset , <i>-ь</i> à l'impératif sg., l'absence de la forme asubjectale	283
77	ТОВКТИ*		l'infinitif : \emptyset , le présent : \emptyset , l'alternance <i>к/ч</i>	284
78	ПЕКТИ*	I	l'infinitif : \emptyset , au passé masculin alternance <i>e/i</i> , le présent : \emptyset , l'alternance <i>к/ч</i>	285
79	РЕВТИ*	I	l'infinitif : \emptyset , au passé masculin <i>ревів</i> , le présent : \emptyset	286
80	ДОВБТИ*		l'infinitif : \emptyset , le présent : \emptyset , <i>-u</i> à l'impératif sg.,	287
81	ГРЕБТИ*	I	l'infinitif : \emptyset , au passé masculin alternance <i>e/i</i> , <i>o/i</i> , le présent : \emptyset , l'absence de la forme asubjectale	288
82	ТЕРТИ*	I	l'infinitif : \emptyset , le présent : \emptyset , <i>тр-</i> (<i>-e-</i> mobile)	289
83	ТЯГТИ*	I	l'infinitif : \emptyset , le présent : \emptyset , <i>+ -н-</i> (<B2>)	290
84	СТРИГТИ*	I	l'infinitif : \emptyset , le présent : \emptyset , l'alternance <i>г/ж</i>	291
85	БЕРЕГТИ*	I	l'infinitif : \emptyset , au passé masculin alternance <i>e/i</i> , \emptyset , l'alternance <i>г/ж</i>	292
86	ОРАТИ*	I	l'infinitif : <i>-а-</i> , le présent : \emptyset , mouillure de la dernière consonne du radical à la 1 ^{ère} personne du sg. et à la 3 ^{ème} personne du pl., la forme asubjectale <i>-но</i>	293
87	КОЛОТИ*	I	l'infinitif : <i>-о-</i> , le présent : \emptyset , mouillure de la dernière consonne du radical à la 1 ^{ère} personne du sg. et à la 3 ^{ème} personne du pl., la forme asubjectale <i>-ено</i> et <i>-мо</i>	294
88	МОЛОТИ*	I	l'infinitif : <i>-о-</i> , le présent : \emptyset , mouillure de la dernière consonne du radical à la 1 ^{ère} personne du sg. et à la 3 ^{ème} personne du pl., alternance <i>o/e</i>	295
89	ВІДПОЧИТИ_РФ	I	l'infinitif : <i>-у-</i> , le présent : <i>+ -н-</i> (<B2>)	296
90	БІГТИ*	II	l'infinitif : \emptyset , le radical du présent <i>біж-</i> , <i>-u</i> à l'impératif sg.,	297
91	ЛЯГТИ_РФ		l'infinitif : \emptyset , le radical du présent <i>ляж-</i> , \emptyset à l'impératif sg.,	298
92	БУТИ*	I	exception pour l'imperfectif et pour le perfectif : l'infinitif : <i>-у-</i> , le radical du présent <i>буд-</i>	299
93	ЙТИ*	I	le radical du passé <i>йш(о)-</i> , et le radical du présent <i>йд-</i>	300
94	РОЗПОВІСТИ_РФ	-	exception	301
95	ЇСТИ*	-	exception	302
96	ДАТИ_РФ	-	exception	303

Les paradigmes DRV pour les verbes

Les 114 paradigmes DRV que nous avons formalisés pour relier les verbes IPF, inscrits dans le dictionnaire aux verbes PF du couple aspectuel dominant correspondants²¹³ sont assez simples. Certains ajoutent un préfixe, comme c'est le cas du paradigme « ПРО » :

ПРО=<LW>про/VERB;

qui ajoute le préfixe *про-* et nous permet de reconnaître le couple *читати* (IPF) / *прочитати* (PF) [lire].

D'autres paradigmes remplacent le suffixe et/ou la terminaison, comme c'est le cas du paradigme « ИТИ_V3 » :

ИТИ_V3=<V3>ити/VERB;

²¹³ Nous avons décidé d'inscrire dans le dictionnaire les verbes IPF par pure convention, et non parce que nous considérons que tous les verbes PF sont créés à partir des verbes IPF.

qui formalise le couple aspectuel *вбачати* (IPF) / *вбачити* (PF) [*percevoir*], en remplaçant le suffixe *-а-* et la terminaison *-ти* par le suffixe *-и-* combiné à la même terminaison.

Les paradigmes du troisième type prennent en compte les alternances de voyelles, comme c'est le cas du paradigme « ALT_10 » :

ALT_10=<B5><L2>о<RW>ати/VERB;

qui traite à la fois l'alternance des voyelles *-i/-o-* dans la racine et le changement du suffixe *-ува-* par le suffixe *-а-* et permet de formaliser le couple aspectuel *виполискувати* (IPF) / *виполювати* (PF) [*rincer*].

Nous avons également décrit le paradigme « 0 » qui ajoute un suffixe vide pour lier au verbe IPF le verbe PF qui a la même composition graphique :

0=<E>/VERB;

Ce paradigme permet de traiter le verbe *акцентувати* [*accentuer*] en tant que le verbe IPF, fléchi grâce au paradigme РИСУВАТИ, et le verbe PF, fléchi grâce au paradigme РИСУВАТИ_PF :

акцентувати,VERB+FLX=РИСУВАТИ+DRV=0:РИСУВАТИ_PF

Nous pouvons voir dans le Tableau 47 tous nos paradigmes DRV avec des exemples d'application.

Tableau 47. Les paradigmes DRV pour les verbes

N° du paradigme DRV	Paradigmes	Exemples
1	0	L'ajout du suffixe vide pour lier au PF qui a la même composition graphique que l'IPF
2	В	IPF : <i>краяти</i> , PF : <i>вкраяти</i> [<i>couper un morceau</i>]
3	ВІД	IPF : <i>дубасити</i> , PF : <i>віддубасити</i> [<i>tabasser</i>]
4	ВИ	IPF : <i>лікувати</i> , PF : <i>вилікувати</i> [<i>soigner</i>]
5	З	IPF : <i>бідніти</i> , PF : <i>збідніти</i> [<i>devenir pauvre</i>]
6	З'	IPF : <i>їсти</i> , PF : <i>з'їсти</i> [<i>manger</i>]
7	ЗА	IPF : <i>барвити</i> , PF : <i>забарвити</i> [<i>colorier</i>]
8	ЗІ	IPF : <i>грати</i> , PF : <i>зіграти</i> [<i>jouer</i>]
9	ЗО	IPF : <i>гнити</i> , PF : <i>зогнити</i> [<i>pourrir</i>]
10	НА	IPF : <i>гріти</i> , PF : <i>нагріти</i> [<i>réchauffer</i>]
11	НАЙ	IPF : <i>іменувати</i> , PF : <i>найменувати</i> [<i>nommer</i>]
12	О	IPF : <i>глухнути</i> , PF : <i>оглухнути</i> [<i>devenir sourd</i>]
13	ОБ	IPF : <i>безсилити</i> , PF : <i>обезсилити</i> [<i>perdre ses forces</i>]
14	ПО	IPF : <i>багрянити</i> , PF : <i>побагрянити</i> [<i>devenir pourpre</i>]
15	ПРИ	IPF : <i>готувати</i> , PF : <i>приготувати</i> [<i>préparer</i>]
16	ПРО	IPF : <i>гнівити</i> , PF : <i>прогнівити</i> IPF : [<i>fâcher</i>]
17	РОЗ	IPF : <i>будити</i> , PF : <i>розбудити</i> [<i>réveiller</i>]
18	С	IPF : <i>казати</i> , PF : <i>сказати</i> [<i>dire</i>]
19	У	IPF : <i>наслідувати</i> , PF : <i>унаслідувати</i> [<i>hériter</i>]
20	ТИ_В3	IPF : <i>вбігати</i> , PF : <i>вбігти</i> [<i>rentrer en courant</i>]
21	ТИ_В4	IPF : <i>забувати</i> , PF : <i>забути</i> [<i>oublier</i>]
22	ИТИ_В3	IPF : <i>вбачати</i> , PF : <i>вбачити</i> [<i>percevoir</i>]
23	ИТИ_В4	IPF : <i>вимовляти</i> , PF : <i>вимовити</i> [<i>prononcer</i>]
24	ИТИ_В5	IPF : <i>вдосконалювати</i> , PF : <i>вдосконалити</i> [<i>perfectionner</i>]
25	ИТИ_В6	IPF : <i>винагороджувати</i> , PF : <i>винагородити</i> [<i>récompenser</i>]
26	ИТИ_В7	IPF : <i>вигранувати</i> , PF : <i>вигранити</i> [<i>graver</i>]
27	СИТИ_В4	IPF : <i>прикрашати</i> , PF : <i>прикрасити</i> [<i>orner</i>]
28	СИТИ_В6	IPF : <i>заквашувати</i> , PF : <i>заквашити</i> [<i>fermenter</i>]

N° du paradigme DRV	Paradigmes	Exemples
29	ТИТИ_В4	IPF : відмічати, PF : відмітити [marquer]
30	ТИТИ_В6	IPF : визолочувати, PF : визолотити [dorer]
31	СТИ_В4	IPF : випадати, PF : випасти [faire tomber]
32	СТИ_В6	IPF : складувати, PF : скласти [entasser]
33	СТИТИ_В4	IPF : заміщати, PF : замістити [remplacer]
34	СТИТИ_В6	IPF : вирощувати, PF : виростити [faire pousser]
35	СЛИТИ_В5	IPF : умишляти, PF : умислити [avoir l'intention]
36	ЗИТИ_В4	IPF : виражати, PF : виразити [exprimer]
37	ЗИТИ_В6	IPF : заморозувати, PF : заморозити [congeler]
38	ЧИТИ_В6	IPF : підскакувати, PF : підскочити [sursauter]
39	ШИТИ_В6	IPF : розполохувати, PF : розполошити [faire peur]
40	ІТИ_В3	IPF : загорати, PF : загоріти [bronzer]
41	ІТИ_В5	IPF : задимлювати, PF : задиміти [enfumer]
42	ІТИ_В6	IPF : перетерплювати, PF : перетерпіти [supporter en attendant]
43	ЇТИ_В5	IPF : заспокоювати, PF : заспокоїти [calmer]
44	АТИ_В5	IPF : вгадувати, PF : вгадати [deviner]
45	АТИ_В7	IPF : використувувати, PF : використати [utiliser]
46	ВАТИ_В5	IPF : виривати, PF : вирвати [arracher]
47	УВАТИ_В7	IPF : арештовувати, PF : арештувати [arrêter]
48	ЮВАТИ_В8	IPF : евакуйовувати, PF : евакуювати [évacuer]
49	ЯТИ_В4	IPF : засівати, PF : засіяти [semer]
50	ЯТИ_В5	IPF : вимірювати, PF : виміряти [mesurer]
51	НЯТИ_В4	IPF : винаймати, PF : винайняти [louer]
52	НЯТИ_В5	IPF : перезнімати, PF : перезняти [filmer encore une fois]
53	НЯТИ_В6	IPF : піднімати, PF : підняти [soulever]
54	ЛЯТИ_В5	IPF : наливати, PF : налляти [verser]
55	ОТИ_В5	IPF : виколювати, PF : виколоти [crever (quelque chose)]
56	НУТИ_В3	IPF : вдягати, PF : вдягнути [tirer (un poids)]
57	НУТИ_В4	IPF : викидати, PF : викинути [jeter]
58	НУТИ_В5	IPF : вигукувати, PF : вигукнути [s'exclamer]
59	НУТИ_В6	IPF : обманювати, PF : обманути [tromper]
60	ЬНУТИ_В3	IPF : тицяти, PF : тицьнути [montrer du doigt]
61	КНУТИ_В4	IPF : кричати, PF : крикнути [crier]
62	КНУТИ_В5	IPF : вимикати, PF : вимкнути [éteindre]
63	ОНУТИ_В2	IPF : гризти, PF : гризнути [gringotter]
64	ОНУТИ_В3	IPF : хапати, PF : хапонути [attraper]
65	ТХНУТИ_В6	IPF : надихати, PF : натхнути [inspirer]
66	УНУТИ_В7	IPF : висовувати, PF : висунути [mettre en avant]
67	ЮНУТИ_В8	IPF : випльовувати, PF : виплюнути [cracher (quelque chose)]
68	ХНУТИ_В3	IPF : запихати, PF : запхнути [pousser (derrière)]
69	АПОСТРОПНЕ_1	IPF : заминати, PF : зам'яти [étouffer]
70	АПОСТРОПНЕ_2	IPF : запинати, PF : зап'ясти [accrocher]
71	АПОСТРОПНЕ_3	IPF : зминати, PF : зім'яти [froisser]
72	ХОДЖУ_1	IPF : винаходить, PF : винайти [inventer]
73	ХОДЖУ_2	IPF : відходити, PF : відійти [reculer]
74	ХОДЖУ_3	IPF : проходжати, PF : пройтися [se promener]
75	ЗІЙТИ	IPF : сходити, PF : зійти [descendre]
76	ЛІЗУ	IPF : облазити, PF : облізти [se délayer]
77	ЕСТИ	IPF : вносити, PF : внести [rentrer quelque chose]
78	ВЕЗУ	IPF : перевозити, PF : перевезти [transporter]
79	ЛЕТІТИ	IPF : долітати, PF : долетіти [voler jusqu'au]
80	СКОЧИТИ	IPF : перескакувати, PF : перескочити [sauter par-dessus]

N° du paradigme DRV	Paradigmes	Exemples
81	ИСТИ	IPF : допливати, PF : доплести [nager jusqu'au]
82	ЖЕНУ_1	IPF : обганяти, PF : обізнати [doubler an avançant]
83	ЖЕНУ_2	IPF : вганяти, PF : ввізнати [faire rentrer quelqu'un]
84	ЖЕНУ_3	IPF : виганяти, PF : вигнати [chasser, faire partir quelqu'un]
85	ІЖДЖ_ІХ	IPF : від'їжджати, PF : від'їхати [s'éloigner autrement qu'à pied]
86	ОЛОТИ	IPF : перемелювати, PF : перемолоти [moudre]
87	СТИЛАТИ_СЛАТИ	IPF : перестилати, PF : переслати [refaire le lit]
88	ЯСТИ	IPF : проклинати, PF : проклясти [maudire]
89	БРАТИ_ВЗЯТИ	IPF : брати, PF : взяти [prendre]
90	ALT	IPF : висихати, PF : висохти [s'assécher]
91	ALT_1	IPF : випікати, PF : випекти [faire cuire la pâtisserie]
92	ALT_2	IPF : вибирати, PF : вибрати [choisir]
93	ALT_3	IPF : допомогати, PF : допомогти [aider]
94	ALT_4	IPF : всихати, PF : всохнути [s'assécher]
95	ALT_5	IPF : вичікувати, PF : вичекати [attendre]
96	ALT_6	IPF : вкорінювати, PF : вкоренити [enraciner]
97	ALT_7	IPF : нашіптувати, PF : нашептати [chouchouter]
98	ALT_8	IPF : відвірчувати, PF : відвертіти [ouvrir en tournant]
99	ALT_9	IPF : розстібати, PF : розстебнути [détacher]
100	ALT_10	IPF : виполіскувати, PF : виполоскати [rincer]
101	ALT_11	IPF : ополіскувати, PF : ополоснути [rincer]
102	ALT_12	IPF : протикати, PF : проткнути [percer]
103	ALT_13	IPF : відстібувати, PF : відстебнути [détacher]
104	ALT_14	IPF : відчіплювати, PF : відчепити [dérocher]
105	ALT_15	IPF : вбирати, PF : вібрати [s'empêigner]
106	ALT_16	IPF : дождити, PF : діждати [attendre jusqu'à la fin]
107	ALT_17	IPF : згинати, PF : зігнути [plier]
108	ALT_18	IPF : обзивати, PF : обізвати [insulter]
109	ALT_19	IPF : відбирати, PF : відібрати [prendre par la force]
110	ALT_20	IPF : відсисати, PF : відіссати [sucrer]
111	ALT_21	IPF : відмикати, PF : відімкнути [déverrouiller]
112	ALT_22	IPF : вмикати, PF : ввімкнути [allumer]
113	ALT_23	IPF : вривати, PF : ввірвати [arracher un morceau]
114	ALT_24	IPF : падати, PF : упасти [tomber]

4.3.5 Évaluation des dictionnaires

Actuellement, le dictionnaire principal que nous avons construit pour le module ukrainien (« Ukr_dictionary_V.1.3.dic ») contient 157 534 entrées et reconnaît 3 184 522 formes fléchies. Il couvre uniquement la langue normative selon les normes orthographiques de 1933, 1990 et de 2019. Le dictionnaire secondaire, comme nous l'avons dit plus tôt, contient 13 070 entrées et reconnaît 444 380 formes fléchies.

Pour évaluer nos dictionnaires, nous avons traité 7 textes²¹⁴ qui représentent les multiples facettes de la langue ukrainienne. Il s'agit de :

²¹⁴ Les cinq premiers textes sont disponibles sur le site de la bibliothèque en ligne UkLib (2020-2025 levhen Vassyliev) <https://www.ukrilib.com.ua>. Ils peuvent être téléchargés au format .txt.

- l'extrait de la nouvelle *Без пумтя* [*Les inutiles*] d'Ivan Netchouy-Levyts'kyy ; l'édition de 2012 (écrite et éditée pour la première fois en 1900 en « jelekhivka »). Bien que la nouvelle ait été réécrite selon les normes orthographiques en vigueur, elle reste imprégnée du vocabulaire de la rive gauche, qui n'est pas considéré comme normatif (7 737 occurrences);
- l'extrait de la nouvelle *В неділю рано зілля копала* [*Elle a récolté des herbes tôt le dimanche*] d'Ol'ha Kobylans'ka; l'édition de 1963 (éditée pour la première fois en 1909 en « jelekhivka »). Ce texte est riche en vocabulaire et expressions bucoviniennes, considérés aujourd'hui comme des dialectismes (18 534 occurrences) ;
- la nouvelle *Зачарована Десна* [*La Desna enchantée*] d'Oleksandr Dovjenko, écrite en 1954. Dans ce texte, il y a des mots familiers et des russismes (16 035 occurrences) ;
- la nouvelle *Казка про калинову сопілку* [*Un conte de fées autour d'une flûte de viorne*], d'Oksana Zaboujko, écrite en 1999. La langue de Zaboujko est connue pour sa richesse, car elle déborde de vocabulaire spécifique (il y a des néologismes, du « sourjyk », des dialectismes de la rive droite) (19 124 occurrences) ;
- l'extrait traduit du *Père Goriot* d'Honoré de Balzac, écrit en 1834, traduit du français par Ielyzaveta Starynkevych (édition 2016). Les œuvres traduites sont souvent des exemples de langue « bien formulée » (21 512 occurrences) ;
- la collection des textes niveau A1-A2 *Читаємо українською* [*Lisons en ukrainien*] (2005)²¹⁵ préparée par Oksana Maliarenko et Alina Chevtsova, les enseignantes de l'Université de Soumy. La langue est très simple, car destinée aux personnes qui commencent l'apprentissage de l'ukrainien (13 339 occurrences) ;
- l'extrait des articles du journal ukrainien *Тиждень.ua* [*La Semaine*]²¹⁶. Il s'agit d'un échantillon de langue de la presse (12 158 occurrences).

L'ensemble de ces textes est nommé ici *Corpus 1*. Nous avons ajouté ce corpus au module ukrainien pour NooJ.

Après avoir appliqué notre dictionnaire principal à ces textes, nous avons constaté qu'il y a encore au total 5,7 % des occurrences qui ne sont pas reconnues. En appliquant le dictionnaire secondaire en plus, nous avons constaté que 5,0 % des occurrences ne sont pas reconnues. Nous pouvons voir les détails de nos tests dans le Tableau 48. Dans les colonnes « Nombre d'occurrences non reconnues » et « Occurrences non reconnues en % », la partie « 1 » contient les statistiques pour le dictionnaire principal « Ukr_dictionary_V.1.3 » et la partie « 2 » contient les statistiques pour le dictionnaire secondaire « Ukr_dictionary_PARTICIPLES_V.1.3 ».

Tableau 48. Évaluation des dictionnaires NooJ pour l'ukrainien

	Nombre d'occurrences dans le texte	Nombre d'occurrences non reconnues		Occurrences non reconnues en %	
		1	2	1	2
<i>Corpus 1</i>	108 137	6 072	5 301	5,6 %	4,9%

Pour évaluer nos dictionnaires, nous les avons ajoutés un par un aux ressources utilisées pour analyser ces textes. Puis nous avons consulté le fichier résultant « Unknowns » qui regroupe les occurrences non reconnues.

Nous avons constaté que les ALU qui n'ont pas été reconnues dans tous ces textes sont :

²¹⁵ Le manuel est disponible sur le dépositaire de l'université : <https://essuir.sumdu.edu.ua/handle/123456789/1457>

²¹⁶ Cette édition (2007-2024) est en ligne à <https://tyzhden.ua>

- les verbes avec le postfixe *-ся* ;
- les formes diminutives et négatives des noms, des adjectifs et des adverbes ;
- les adjectifs et les adverbes négatifs ;
- les adjectifs d'appartenance ;
- les adjectifs non contractés et les adjectifs courts ;
- les adjectifs comparatifs, superlatifs et superlatifs renforcés ;
- quelques noms propres ;
- les adverbes et les pronoms indéterminés en « *-ць* » ;
- quelques variantes phonétiques (les ALU qui commencent par « *вид-/од-* », « *в-/у-* ») ;
- les mots en alphabet latin ;
- les mots qui relèvent du registre familier ;
- les néologismes ;
- les dialectismes ;
- le « *sourjyk* ».

Pour reconnaître ces ALU, nous avons construit quelques grammaires morphologiques productives que nous allons présenter dans la section suivante.

4.4 Grammaires morphologiques pour le module ukrainien de NooJ

La langue ukrainienne est très riche, de ce fait, les dictionnaires sont assez conséquents, par exemple VESUM²¹⁷, le plus grand dictionnaire en ligne, contient actuellement plus de 418 000 lemmes. Ce dictionnaire contient des lemmes dérivés, de plus il est constamment mis à jour : on y ajoute des entités nommées, des néologismes, des variantes régionales et diverses formes, retrouvées dans les textes récemment numérisés.

Pour qu'un outil de TAL puisse reconnaître un mot, il faut que ce mot corresponde soit à une entrée de dictionnaire, soit à une forme fléchie ou dérivée associée à l'entrée du dictionnaire.

Les auteurs d'outils empiriques qui utilisent des corpus d'apprentissage annoncent qu'ils sont capables de détecter et classer les mots inconnus. Cependant, comme nous l'avons vu dans le chapitre 2 (section 2.2.1 et 2.2.2), ils ne le font pas correctement.

Les outils linguistiques disponibles actuellement ont accès à des dictionnaires très conséquents, comme c'est le cas de GRAC²¹⁸ qui est basé sur VESUM. Mais si jamais une forme dérivée d'une ALU n'est pas recensée dans le dictionnaire, l'outil n'est pas capable de la reconnaître. Cela implique une mise à jour obligatoire et fréquente du dictionnaire. De plus, cette méthode a un point faible : elle ne permet pas de relier la forme dérivée et la forme de départ²¹⁹ ni les mots avec une orthographe alternative, car ils sont inscrits dans le dictionnaire indépendamment, comme nous l'avons vu dans le chapitre 2 (section 2.4.2).

La plateforme NooJ donne accès à un outil de construction de grammaires morphologiques capables de produire des formes absentes du dictionnaire. Cela permet à la fois de créer un lien entre les formes dérivées et les formes de départ et d'avoir une bonne couverture du vocabulaire.

²¹⁷ Disponible sur <https://vesum.nlp.net.ua/vesum/?w=> et en téléchargement sur https://github.com/brown-uk/dict_uk

²¹⁸ GRAC est une collection de textes ukrainiens, annoté selon l'approche formaliste, disponible sur https://parasol.vmguest.uni-jena.de/grac_crystal/#dashboard?corpname=grac17

²¹⁹ Par exemple, il n'existe pas de lien entre l'adjectif comparatif et l'adjectif qualitatif dont il est issu.

À ce jour, nous avons construit plusieurs grammaires que nous allons exposer dans cette section. Il est important de rappeler que les grammaires NooJ sont réversibles : elles servent à annoter les textes et à produire les formes. Par ailleurs, en préparant nos grammaires, nous avons privilégié l'annotation et non la production. Par exemple, dans la section 4.4.9, nous détaillerons la grammaire formalisant les formes courtes des adjectifs. Si l'on voulait l'utiliser pour produire les formes, elle produirait un adjectif court associé à chaque adjectif long de notre dictionnaire. Or, dans le chapitre 3, nous avons expliqué que tous les adjectifs ne possèdent pas de forme courte. De ce fait, nous n'utiliserons pas nos grammaires pour produire les formes, avec un risque de surproduction, mais uniquement pour annoter les textes.

Pour évaluer nos grammaires, nous utiliserons les dictionnaires décrits dans la section précédente : le dictionnaire principal « Ukr_dictionary_V.1.3 » avec le niveau de priorité élevé et le dictionnaire « Ukr_dictionary_Participles_V.1.3 » avec un niveau de priorité faible. Nous ajouterons par la suite dans nos ressources les grammaires avec un niveau de priorité faible²²⁰ et nous les appliquerons au *Corpus 1* décrit plus haut. Nous inscrirons les données récupérées lors de cette évaluation dans la dernière colonne du tableau. Puis, nous comparerons le nombre d'occurrences non reconnues avant l'application de la grammaire à évaluer et après son application.

Dans chaque section suivante, nous traiterons les données de la deuxième colonne de la section précédente.

4.4.1 Formalisation des participes avec une grammaire morphologique

Description de la grammaire

Comme nous l'avons dit dans la section 4.3.2 et dans le chapitre 3, il est important de relier les participes aux verbes dont ils sont issus. Pour cela, nous avons construit la grammaire morphologique « Participles_V.1.3.nom » que nous pouvons voir dans la Figure 35.

Ici, il s'agit d'une grammaire contextuelle, composée de 26 graphes. Dans le graphe principal, nous pouvons voir 5 nœuds qui décrivent des contraintes :

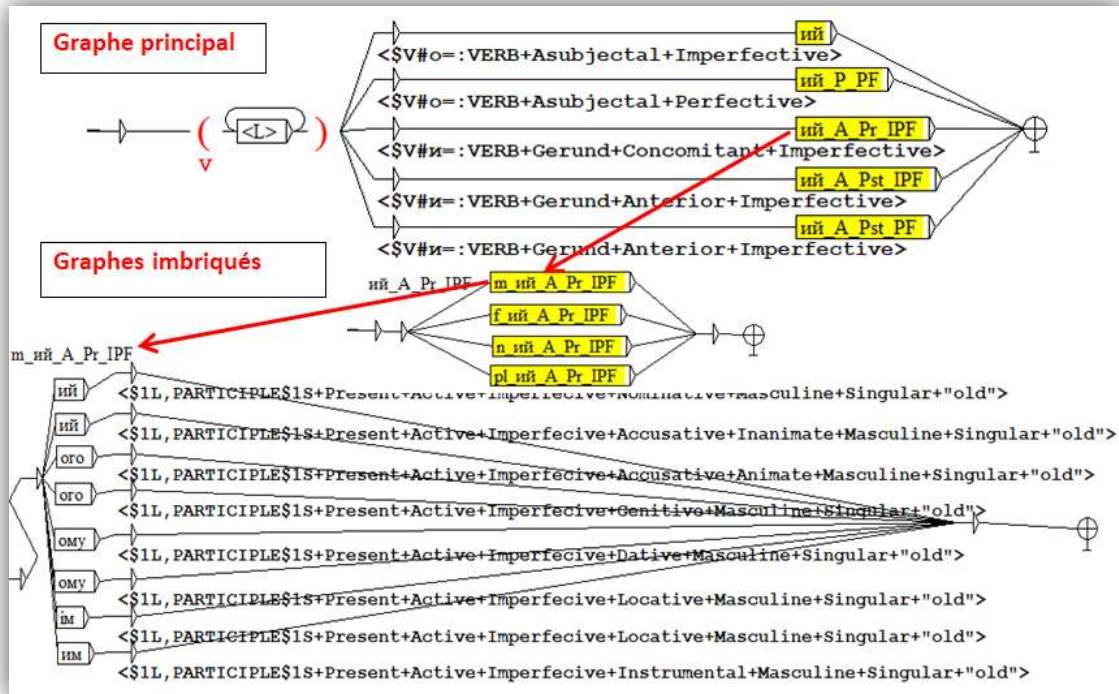
- dans le premier nœud sont imbriqués les schémas qui décrivent les participes passés passifs IPF, créés à partir de la forme asubjectale des verbes IPF (la contrainte <\$V#o=:VERB+Asubjectal+Imperfective>²²¹);
- dans le deuxième nœud sont imbriqués les schémas qui décrivent les participes passés passifs PF, créés à partir de la forme asubjectale des verbes PF (la contrainte <\$V#o=:VERB+Asubjectal+Perfective>);
- dans le troisième nœud sont imbriqués les schémas qui décrivent les participes présents actifs IPF, créés à partir du gérondif concomitant des verbes IPF (la contrainte /<\$V#и=:VERB+Gerund+Concomitant+Imperfective>);
- dans le quatrième nœud sont imbriqués les schémas qui décrivent les participes passés actifs IPF, créés à partir du gérondif antérieur des verbes IPF (la contrainte <\$V#и=:VERB+Gerund+Anterior+Imperfective>);

²²⁰ Nos grammaires auront une priorité inférieure à notre dictionnaire principal pour compléter les formes absentes du dictionnaire ; avoir un niveau plus faible nous évite la double annotation d'une UL, décrite à la fois dans le dictionnaire et dans la grammaire.

²²¹ Cela veut dire que notre lemme du départ doit être reconnu comme « VERB+Asubjectal+Imperfective » auquel l'outil doit enlever la lettre « o » avant de produire la nouvelle forme fléchie.

- dans le cinquième nœud sont imbriqués les schémas qui décrivent les participes passés actifs PF, créés à partir du gérondif antérieur des verbes PF (la contrainte $\langle \$V\#и:VERB+Gerund+Anterior+Imperfective \rangle$).

Figure 35. Grammaire morphologique pour formaliser les participes



Dans chaque graphe imbriqué est décrit le paradigme du participe, c'est-à-dire, la déclinaison qui prend en compte le genre (masculin, féminin et neutre), le nombre – quatre nœuds représentés sur le premier graphe imbriqué, puis le cas et la catégorie de l'animé/inanimé, imbriqués dans les quatre nœuds en question.

Les formes fléchies reconnues ainsi seront étiquetées comme participes. Elles recevront les étiquettes qui indiqueront : le verbe de départ en tant que lemme, le temps, la voix, l'aspect, le cas, le genre, le nombre, puis les participes actifs auront aussi l'annotation supplémentaire « old », car comme nous l'avons dit avant, cette forme n'est plus utilisée actuellement. Nous pouvons voir comment est annoté le participe présent actif IPF *виючий* [hurlant] dans l'exemple, présenté dans la Figure 36.

Figure 36. Annotation du participe *виючий*

виючий
0
ВИТИ,PARTICIPLE+Tense=Present+Voice=Active+Imperfective+Case=Nominative+Gender=Masculine+Number= Singular+old"

Évaluation de la grammaire

Cette grammaire décrit les schémas productifs de formation des participes et ne prend pas en compte les schémas non productifs ou peu productifs, notamment les participes passés actifs, dérivés des verbes PF par ajout du suffixe *-л-*, comme *загаснувший* [éteint]. De ce fait, elle doit être

complétée par le dictionnaire secondaire. Les résultats des tests de cette grammaire sont donnés dans le Tableau 49.

Dans la première colonne de ce tableau, nous avons indiqué les données d'évaluation du dictionnaire principal, dans la deuxième colonne, nous avons exposé les données que nous avons obtenues après l'application de notre grammaire, dans la troisième colonne, nous avons mis les résultats obtenus après l'ajout du dictionnaire secondaire « Ukr_dictionary_Participles_V.1.3 ».

Nous pouvons voir que notre grammaire, malgré le fait qu'elle ne contienne pas tous les schémas de la production, couvre bien les participes trouvés dans notre corpus, vu qu'il en reste 5 496 occurrences après son ajout dans les ressources au lieu de 6 072. Notons qu'en rajoutant le dictionnaire secondaire avec les participes, il nous reste 5 259 occurrences à reconnaître.

Tableau 49. Évaluation de la grammaire « Participles_V.1.3.nom »

	Nombre d'occurrences dans le texte	Nombre d'occurrences non reconnues			Nombre d'occurrences reconnues	Occurrences non reconnues en %		
		1	2	3		1	2	3
<i>Corpus 1</i>	108 137	6 072	5 496	5 259	813	5,6 %	5,1%	4,8%

4.4.2 Traitement des verbes avec le postfixe -ся

Description de la grammaire

En ukrainien, de nombreux verbes se terminent par le postfixe -ся. Certains ne s'utilisent jamais sans ce postfixe, par exemple *сміятися* [rire] dans

Він сміється [il rit],

tandis que d'autres verbes expriment une action qui se réfléchit sur le sujet, par exemple *читатися* [se lire] dans

Книга читається швидко [le livre se lit vite].

En ukrainien, ces verbes sont extrêmement nombreux, car presque tous les verbes ukrainiens peuvent recevoir ce postfixe. Par exemple, le dictionnaire en ligne *Goroh* en contient 50 686²²². Inscrire ces verbes dans le dictionnaire a un gros désavantage, car techniquement cela nous obligerait à dédoubler tous nos paradigmes à cause de deux symboles graphiques supplémentaires qui s'ajoutent au verbe, puis, pour ne pas perdre le lien entre les deux formes, inscrire tous les verbes dans le dictionnaire en ajoutant le verbe sans postfixe, après la virgule, comme ceci :

читатися,читати,VERB/.../

Pour éviter ces descriptions coûteuses et redondantes, nous avons construit une grammaire morphologique qui reconnaît tous les verbes postfixés.

Pour cela, nous avons entré tous les verbes qui n'existent pas sans postfixe -ся dans notre dictionnaire, avec l'étiquette « NW », qui signifie « non-word [ne peut pas apparaître tel quel dans un texte] », comme c'est le cas du verbe *сміятися* [rire] inscrit dans le dictionnaire ainsi :

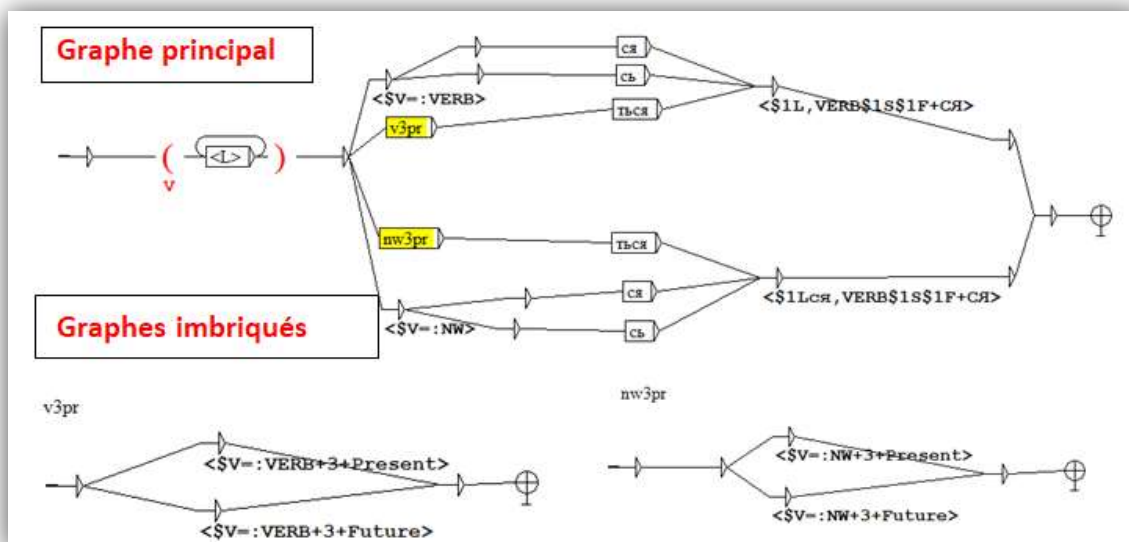
сміяти,NW+FLX=ВІЯТИ

²²² Nous pouvons voir la liste complète ici : [https://goroh.pp.ua/Словозміна/\[рефлексивне\]](https://goroh.pp.ua/Словозміна/[рефлексивне])

En tout, notre dictionnaire contient 358 entrées verbales, étiquetées avec la catégorie « NW » à la place de « VERB ».

Ensuite, nous avons construit la grammaire morphologique « Verbs_Sia_V.1.3 », présentée dans la Figure 37.

Figure 37. Grammaire morphologique pour décrire des verbes en « -ся »

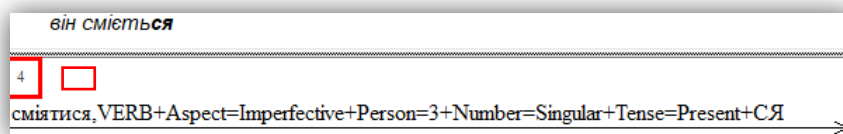


Cette grammaire reconnaît et analyse les verbes postfixés en calculant le lemme à partir de la contrainte $\langle \$V = : VERB \rangle$ ou $\langle \$V = : NW \rangle$. Ici, notre lemme de départ (« V ») est une séquence de lettres qui est reconnue par le système en tant que VERB ou NW. Les graphes imbriqués décrivent les contraintes supplémentaires : la séquence de lettres doit être reconnue comme « VERB » ou « NW » à la troisième personne du présent ou du futur. Ensuite, une des trois combinaisons des formes graphiques (« ся », « сь », « ться ») est ajoutée à la séquence « V ». Lors de l'annotation pour les verbes qui ont été entrés dans le dictionnaire avec l'étiquette « VERB », les informations de la forme fléchie de départ sont récupérées et l'étiquette supplémentaire « +CR » y est ajoutée (voir Figure 38). Par ailleurs, pour les verbes qui ont été entrés dans le dictionnaire avec l'étiquette « NW », on récupère également les informations morphologiques de la forme fléchie initiale et on ajoute l'étiquette supplémentaire « +CR », mais leur lemme serait affiché dans les annotations avec le postfixe -ся (voir Figure 39).

Figure 38. Annotation du verbe *читается*



Figure 39. Annotation du verbe *сміється*



Évaluation de la grammaire

En appliquant cette grammaire à nos textes, nous obtenons les résultats qui sont détaillés dans le Tableau 50²²³. Nous pouvons remarquer que le nombre d'occurrences non reconnues a beaucoup diminué. Au total, nous sommes passés de 5 259 occurrences non reconnues à 2 979.

Tableau 50. Évaluation de la grammaire « Verbs_Sia_V.1.3 »

	Nombre d'occurrences dans le texte	Nombre d'occurrences non reconnues		Nombre d'occurrences reconnues	Occurrences non reconnues en %	
		1	2		1	2
Corpus 1	108 137	5 259	2 979	2 280	4,8%	2,7%

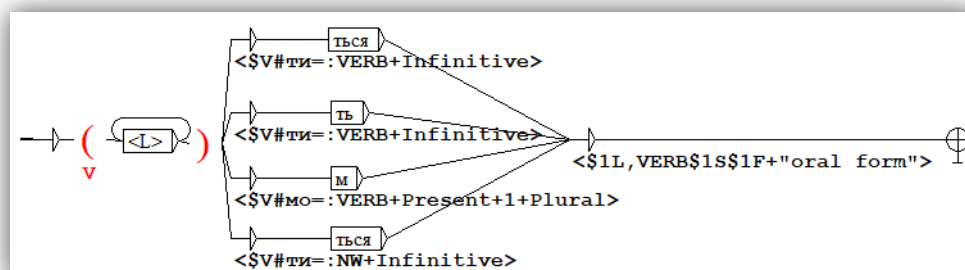
4.4.3 Formalisation des formes verbales orales

Description de la grammaire

En examinant attentivement les formes non reconnues, nous avons remarqué que notre liste contenait quelques formes verbales orales, comme *читат* au lieu de *читати* [lire], *читатся* au lieu de *читатися* [lire], ou encore *чумаем* au lieu de *чумаємо* [(nous) lisons]. En effet, ces formes sont utilisées dans les œuvres littéraires et méritent d'être décrites.

Pour les prendre en compte, nous avons construit la grammaire « Verbs_oral_forms_V.1.3.nom », présentée dans la Figure 40. Nous pouvons voir que notre grammaire remplace à l'infinitif le « ти » par « ть » ou par « ться » pour les verbes avec le postfixe – « ся », et « мо » par « м » à la première personne du pluriel au présent. La nouvelle forme reprend toutes les informations morphologiques de la forme de départ, y compris le lemme, puis une étiquette « oral form » y est ajoutée.

Figure 40. Grammaire formalisant les formes verbales orales



²²³ Pour rappel : dans la première colonne de ce tableau, nous utilisons les données de la colonne 3 du Tableau 53.

Nous pouvons voir l'exemple de l'annotation de la forme orale *читать* [lire] dans la Figure 41.

Figure 41. Annotation de la forme orale *читать*

читать
0
читати, VERB+Pair="читати/прочитати"+Aspect=Imperfective+Mood=Infinitive+"oral form"

Évaluation de la grammaire

Après application de cette grammaire, le nombre d'occurrences à reconnaître passe de 2 979 à 2 941 (voir Tableau 51).

Tableau 51. Évaluation de la grammaire « Adverbs_V.1.3.nom »

	Nombre d'occurrences dans le texte	Nombre d'occurrences non reconnues		Nombre d'occurrences reconnues	Occurrences non reconnues en %	
		1	2		1	2
Corpus 1	108 137	2 979	2 941	38	2,7%	2,7%

4.4.4 Formalisation des noms diminutifs ou affectifs

Description de la grammaire

Comme nous l'avons vu dans le chapitre précédent (section 3.3.1), les noms peuvent exprimer le sens diminutif ou affectif grâce au suffixe qui a été ajouté à la racine. Ces noms sont très nombreux. Dans les dictionnaires académiques, ils sont inscrits avec une mention qui renvoie vers la forme de départ. Voici un exemple tiré du dictionnaire SUM-20 :

« БАНТИК », а, ч. Зменш. до « бант » [« petit nœud », diminutif du « nœud »].

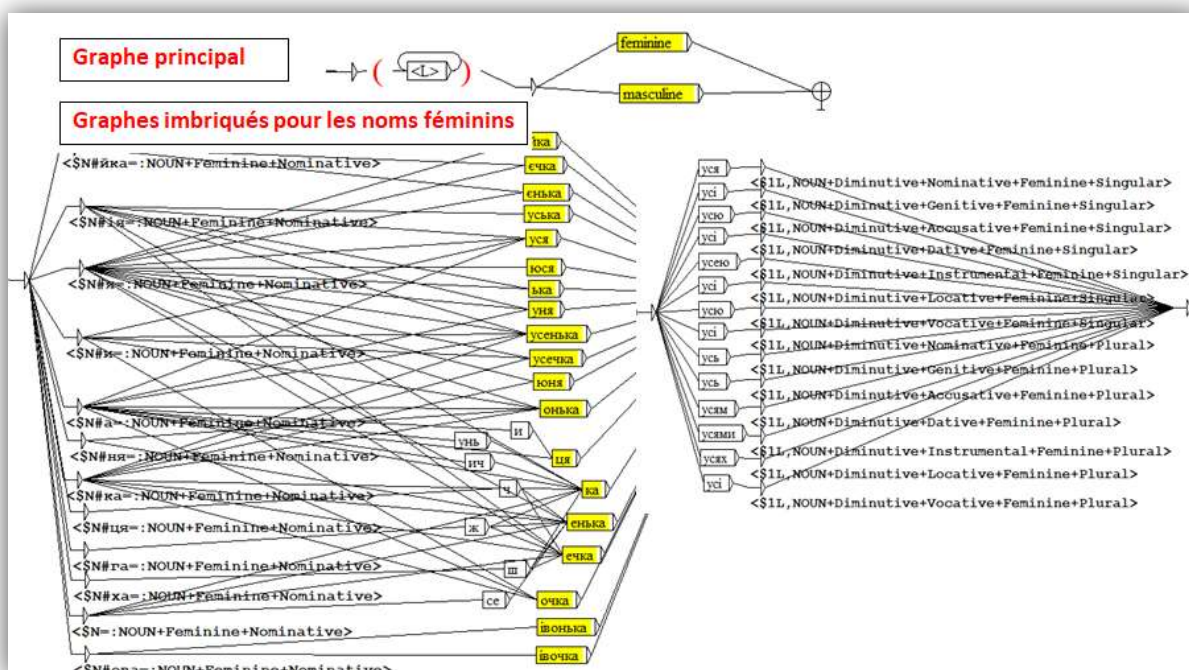
Pourtant, dans les outils que nous avons testés et décrits dans le chapitre 2, ces formes ne sont pas reliées. Nous avons donc construit deux grammaires morphologiques qui reconnaissent ces formes dérivées : une pour les noms masculins et féminins et une pour les noms neutres. Chacune de ces grammaires ajoute le suffixe et décrit la déclinaison. Nous avons séparé le masculin et le neutre, car les noms neutres et masculins ont des terminaisons qui coïncident, mais les étiquettes ne sont pas les mêmes.

Nous pouvons voir la grammaire pour les noms masculins et féminins dans la Figure 43 (Noun_Diminutive_F_M_V.1.3.nom).

Cette grammaire est composée de 27 graphes. Sur le graphe principal, nous pouvons voir le lemme de départ « N » qui est une séquence de lettres (<L>) et deux graphes imbriqués : un pour le féminin et un autre pour le masculin.

À l'intérieur de chaque graphe imbriqué, notre grammaire vérifie la contrainte : la forme de départ est soit le nom féminin au nominatif, soit le nom masculin au nominatif auquel on enlève les fins de mot (« #я », « #я », « #а », etc). Ensuite, ces fins de mots seront remplacées par d'autres, décrites dans les 24 graphes imbriqués (19 pour les noms féminins et 5 pour les noms masculins). Les fins de mots de remplacement correspondent aux suffixes et à la terminaison. Nous pouvons voir dans la Figure 42 le graphe imbriqué « :уся », qui sert à reconnaître les formes fléchies.

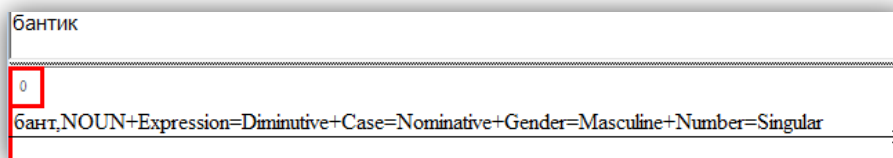
Figure 42. Grammaire morphologique pour formaliser les diminutifs des noms



La grammaire qui prend en compte les noms neutres (Noun_Diminutive_Neuter_V.1.3.nom) est construite de la même façon.

Grâce à ces deux grammaires, tous les diminutifs et leurs formes fléchies sont reconnus, annotés et ramenés au lemme d'origine. Nous pouvons voir le résultat d'annotation pour le mot *бантик* [petit œud] dans la Figure 43.

Figure 43. Annotation de la forme diminutive *бантик*



Évaluation des grammaires

Grâce à ces deux grammaires, le nombre des occurrences non reconnues passe de 2 941 à 2 733. Nous pouvons voir les détails de nos tests dans le Tableau 52.

Tableau 52. Évaluation des grammaires « Noun_Diminutive_F_M_V.1.3.nom » et « Noun_Diminutive_Neuter_V.1.3.nom »

	Nombre d'occurrences dans le texte	Nombre d'occurrences non reconnues		Nombre d'occurrences reconnues	Occurrences non reconnues en %	
		1	2		1	2
Corpus 1	108 137	2 941	2 733	208	2,7%	2,5%

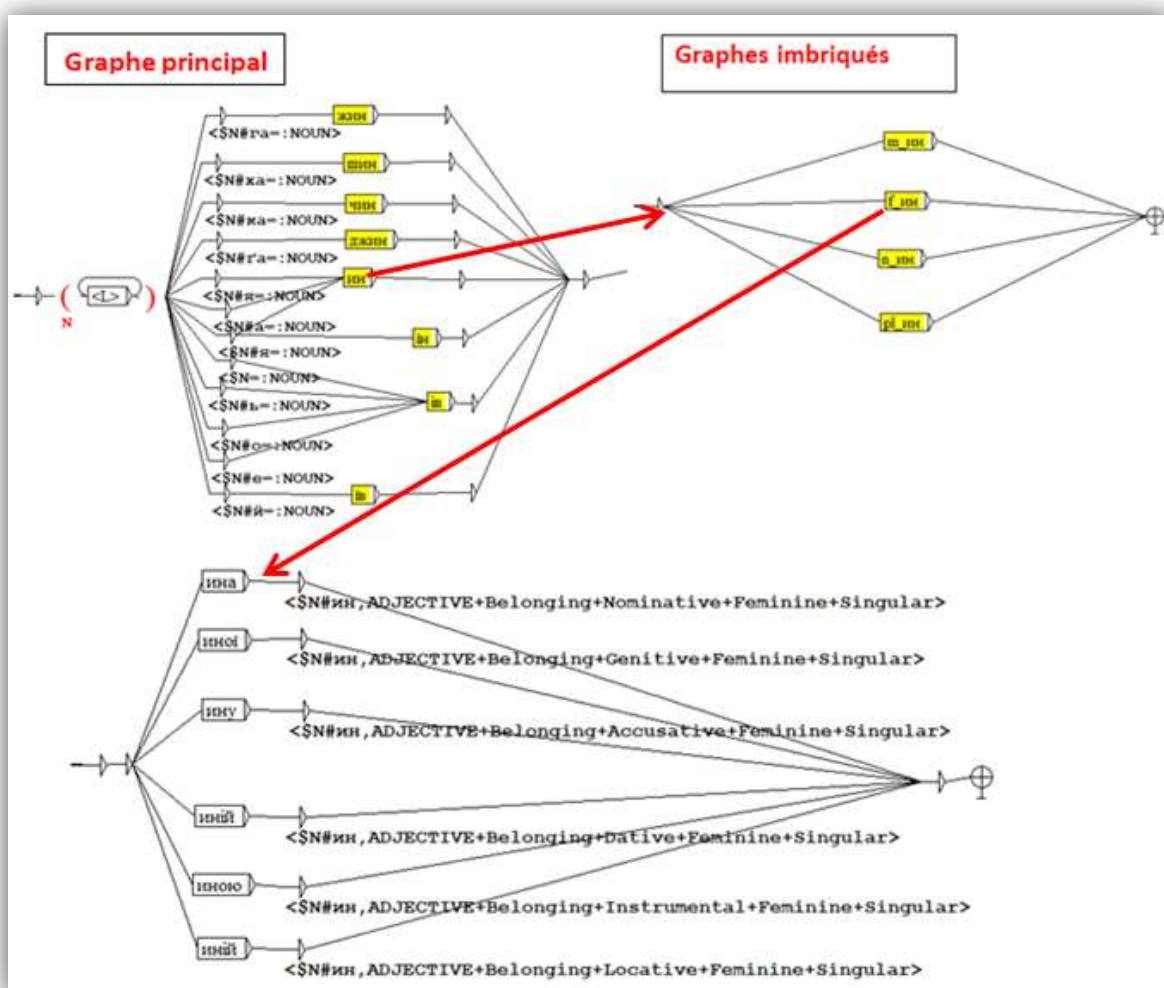
4.4.5 Formalisation des adjectifs d'appartenance

Description de la grammaire

Dans le chapitre 3, nous avons expliqué qu'en ukrainien, on peut créer les adjectifs d'appartenance qui expriment le lien avec un être vivant (par exemple : *Оксанин* [celui qui appartient à Oksana]). Ces adjectifs sont créés à partir de noms désignant une personne, y compris les noms propres ; de ce fait, il est impossible d'avoir un dictionnaire qui recenserait tous les adjectifs d'appartenance.

Pour que ces formes soient reconnues par notre module NooJ, nous avons construit la grammaire « Adjectives_Belonging_V.1.3.nom » représentée dans la Figure 44.

Figure 44. Grammaire formalisant les adjectifs d'appartenance



Cette grammaire est composée de 41 graphes. Le graphe principal montre le mécanisme de la dérivation des adjectifs d'appartenance à partir du nom : nous enlevons à la forme de départ (« :NOUN ») un ou plusieurs caractères graphiques (« #а », « #ra », etc.), et on les remplace par d'autres caractères graphiques (« :ин », « :жин », etc.) décrits dans les huit nœuds. Dans chaque nœud sont imbriqués quatre graphes qui décrivent les formes fléchies au masculin, féminin, neutre et pluriel. Pour chaque forme fléchie, la grammaire calcule le lemme (l'adjectif d'appartenance au

masculin), les informations morphologiques (le cas, le genre, le nombre) et produit l'étiquette spéciale « Belonging».

Nous pouvons voir comment est annoté l'adjectif d'appartenance *Оксанин* [celui qui appartient à *Oksana*] dans la Figure 45.

Figure 45. Annotation de l'adjectif d'appartenance *Оксанин*

Оксанин
0
Оксанин_ADJECTIVE+Type=Belonging+Case=Nominative+Gender=Masculine+Number=Singular

Évaluation de la grammaire

Après application de cette grammaire, le nombre d'occurrences non reconnues passe à 2 646. Nous donnons les détails de nos tests dans le Tableau 53.

Par ailleurs, cette grammaire reconnaît uniquement les adjectifs, créés à partir des noms de notre dictionnaire. Pour mieux faire, nous aimerions trouver plus tard le moyen de reconnaître les adjectifs à partir des noms propres non-inscrits dans notre dictionnaire.

Tableau 53. Évaluation de la grammaire «Adjectives_Belonging_V.1.3.nom»

	Nombre d'occurrences dans le texte	Nombre d'occurrences non reconnues		Nombre d'occurrences reconnues	Occurrences non reconnues en %	
		1	2		1	2
<i>Corpus 1</i>	108 137	2 733	2 646	87	2,5%	2,4%

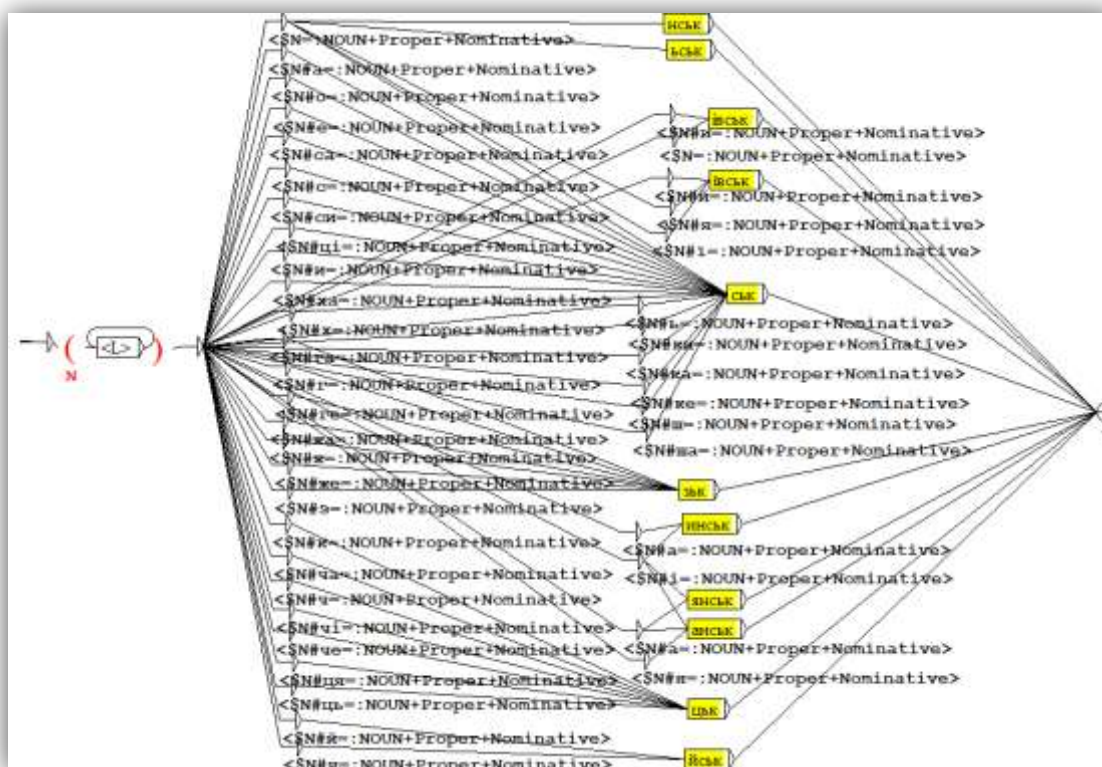
4.4.6 Formalisation des adjectifs relatifs à partir des noms propres

Description de la grammaire

Comme les adjectifs d'appartenance, des adjectifs relatifs peuvent être créés à partir de noms propres, notamment à partir de toponymes ou parfois à partir de prénoms ou de noms de famille. Si les adjectifs relatifs créés à partir des noms des grandes villes, comme c'est le cas pour *київський* [de *Kyiv*] sont inscrits dans les dictionnaires, ceux qui sont créés à partir des noms des petites villes ne le sont pas.

Pour traiter ces formes, nous avons construit la grammaire « Adjectives_Relatives_V.1.3.nom » (voir Figure 46). Cette grammaire est composée de 56 graphes imbriqués et elle est similaire à la grammaire précédente. Sur le graphe principal, nous pouvons voir le mécanisme de remplacement des fins de mots de noms propres (par exemple, <\$N#o=:NOUN+Proper+Nominative>) par des suffixes (comme « :нськ », 11 nœuds en tout), combinés avec la terminaison adjectivale (à l'intérieur de chaque nœud, il y a 4 nœuds qui décrivent les formes fléchies comme pour la grammaire précédente). Chaque adjectif reconnu grâce à cette grammaire aura en plus l'étiquette « Relative ».

Figure 46. Grammaire formalisant les adjectifs relatifs



Après avoir construit cette grammaire, nous nous sommes rendu compte qu'elle fonctionnait à condition que les adjectifs concernés apparaissent dans le texte avec une majuscule, comme le nom propre dont ils sont issus. Or, les adjectifs en question ne s'écrivent avec la majuscule à l'initiale que lorsqu'eux-mêmes sont des noms propres, ce qui n'est pas toujours le cas.

La solution que nous avons trouvée à ce problème est la suivante : nous avons recopié les noms propres (24 721 entrées) de notre dictionnaire, et nous les avons mis dans le dictionnaire secondaire « Ukr_dictionary_Proper_lowercase_V.1.3 ». Dans ce dictionnaire, chaque ALU est écrite en minuscules, puis suit un lemme, écrit après une virgule, commençant par une majuscule, comme pour l'ALU *алтаї* [Altai] :

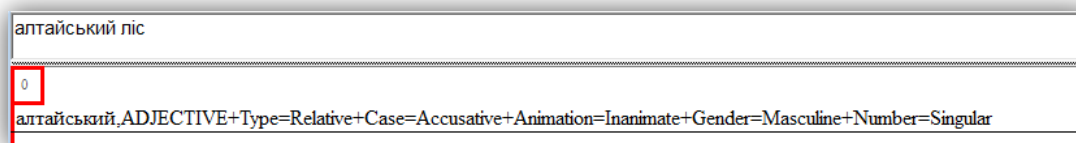
алтаї,Алтай,NOUN+Masculine+Inanimate+Proper+Toponym+FLX=МУЗЕЙ

Nous avons affecté à ce dictionnaire une priorité plus faible que celle du dictionnaire principal, pour ne pas créer d'ambiguïté artificielle (double analyse).

Le dictionnaire « Ukr_dictionary_Proper_lowercase_V.1.3 » nous servira pour reconnaître les adjectifs relatifs en combinaison avec notre grammaire, et aussi pour reconnaître les noms propres écrits par erreur en minuscules. Cela peut être utile pour analyser des textes en provenance des réseaux sociaux, où les noms propres sont écrits fréquemment en minuscules.

Grâce à cette grammaire, les adjectifs relatifs créés à partir des noms propres seront annotés comme dans l'exemple présenté dans la Figure 47.

Figure 47. Annotation de l'adjectif *алтайський*



Évaluation de la grammaire

En appliquant cette grammaire à nos textes de tests nous avons obtenu les résultats présentés dans le Tableau 54.

Tableau 54. Évaluation de la grammaire «Adjectives_Relatives_V.1.3.nom»

	Nombre d'occurrences dans le texte	Nombre d'occurrences non reconnues		Nombre d'occurrences reconnues	Occurrences non reconnues en %	
		1	2		1	2
Corpus 1	108 137	2 646	2 587	59	2,4%	2,4%

Cette grammaire a quelques points faibles : elle prend en compte uniquement les noms propres de notre dictionnaire et elle ne traite donc pas les cas peu productifs (comme - *Москва [Moscou]* → *московський [de Moscou]*). Il faudra recenser ces noms propres dans un dictionnaire, ou alors généraliser notre grammaire.

De plus, nous pouvons constater que certains noms peuvent produire deux adjectifs relatifs. Par exemple, les adjectifs *софійський*, *софіївський* sont créés tous les deux à partir du nom propre *Софія [Sophia]* qui se termine par *-ія*. Tandis que le nom propre *Ефіопія [Éthiopie]* qui se termine aussi par *-ія* ne peut produire qu'un seul adjectif *ефіопський* selon la norme orthographique en vigueur. Cela signifie que notre grammaire reconnaitra également les adjectifs non normatifs, par exemple *ефіопійський*, créé selon le même schéma que *софійський*. Nous comptons revenir sur ce problème lorsque nous travaillerons sur la reconnaissance des entités nommées de façon plus approfondie.

4.4.7 Formalisation des adjectifs et des adverbes, partitifs, excessifs, diminutifs

Description de la grammaire

Comme nous l'avons expliqué dans la section 3.3.2, les adjectifs qualitatifs expriment la qualité du substantif auquel ils se rapportent. En ukrainien, cette propriété peut être exprimée complètement, partiellement, excessivement ou encore avec l'affection du locuteur, grâce à des affixes spécifiques.

Ces adjectifs et ces adverbes sont largement utilisés dans les sources écrites, mais seulement une partie d'entre eux est entrée dans les dictionnaires. Pour qu'ils puissent être reconnus dans NooJ, nous avons construit deux grammaires : une pour les adjectifs « Adjectives_Expression_V.1.3.nom » (voir Figure 48) et une autre « Adverbs_Expression_V.1.3.nom » pour les adverbes (voir Figure 50).

La grammaire qui reconnaît les formes adjectivales est composée de 91 graphes. Elle reconnaît à la fois les formes construites par ajout des préfixes (dans la partie basse) et les formes produites par ajout des suffixes (dans la partie haute). En tout, il y a 18 nœuds avec les suffixes, dans lesquels

SUM-20²²⁵ : МЕНШИЙ, а, е. 1. Вищ. ст. до малій і маленький [*« plus petit », le comparatif du « petit »*].

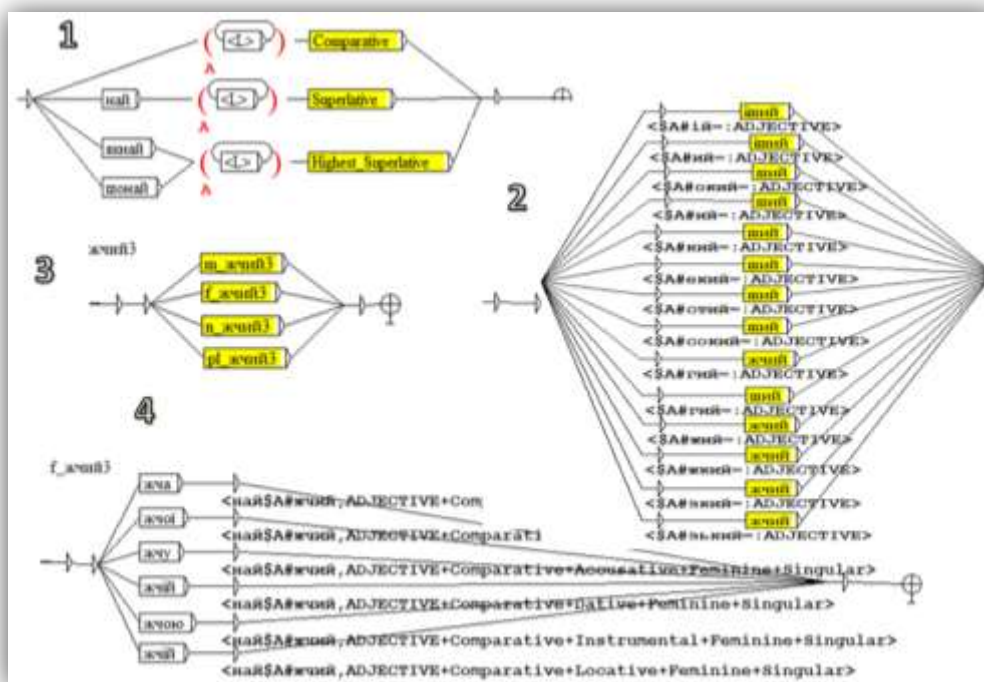
Par ailleurs, les formes qui sont créées grâce au schéma dérivatif productif sont absentes du dictionnaire. Les outils de TAL (par exemple, Sketch Engine, GRAC, VESUM) les reconnaissent, mais sans indiquer le lien avec le lemme de départ, comme c'est le cas du *гарніший* [*plus beau*] :

VESUM²²⁶ : гарніший - прикметник, вища форма [*« plus beau », adjectif comparatif*].

Nous avons décidé de formaliser les formes du comparatif, du superlatif et du superlatif absolu des adjectifs et des adverbes avec deux grammaires morphologiques. Cela permettra de créer le lien entre le lemme de départ et la forme dérivée et de mieux couvrir le vocabulaire. Notons que nous n'avons entré dans notre dictionnaire que les formes créées selon un schéma non productif, ou peu productif.

Ainsi, nous avons d'abord préparé une grammaire pour les adjectifs « Adjectives_Degree_V.1.3 » représenté en Figure 52.

Figure 52. Grammaire formalisant les degrés de comparaison pour les adjectifs



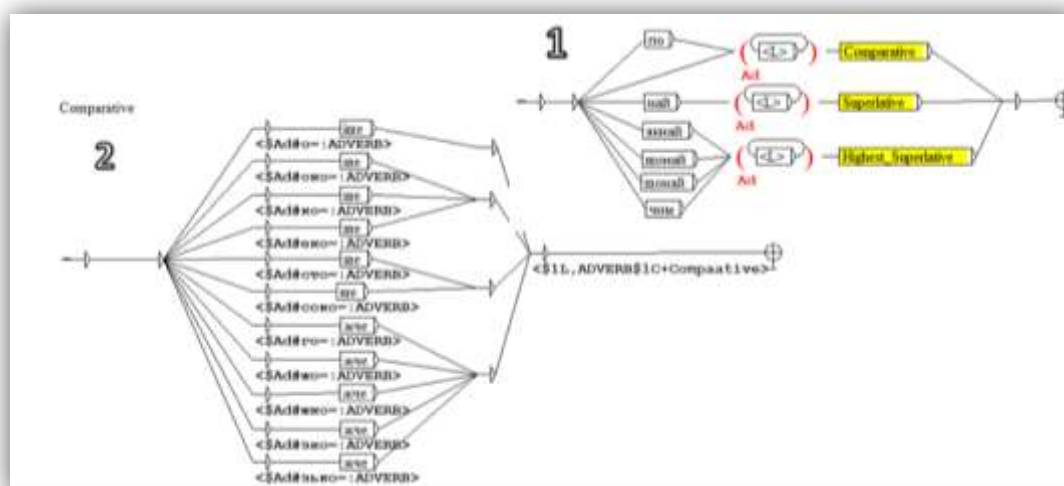
Cette grammaire est composée de 85 graphes. Sur le graphe principal, noté par le numéro 1, il y a un nœud pour chaque degré de comparaison et les préfixes qui permettent de produire le superlatif et le superlatif absolu. Sous le numéro 2, nous pouvons voir le graphe « Comparative » dans lequel sont décrits les suffixes (14 nœuds) qui contiennent le paradigme adjectival (parties 3 et 4). Pour chaque adjectif reconnu ainsi, nous affichons le lemme de départ et nous ajoutons l'étiquette qui indique le degré de comparaison.

²²⁵ <https://sum20ua.com/Entry/index?wordid=50578&page=1601>

²²⁶ <https://vesum.nlp.net.ua/?w=гарніший&format=on>

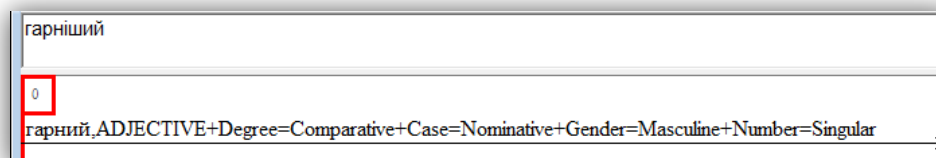
La grammaire « Adverbs_Degree_V.1.3 » qui produit des formes adverbiales est beaucoup plus simple (composée de 4 graphes), car encore une fois, nous n'avons pas besoin d'y inclure de paradigme flexionnel (voir Figure 53). Sur le graphe principal sont décrits les préfixes qui prennent part dans la production des formes qui expriment le degré de comparaison, y compris le préfixe moins fréquent «по-» qui participe à la formation du degré de comparaison avec l'ajout du sens partitif.

Figure 53. Grammaire formalisant les degrés de comparaison pour les adverbes



Nous pouvons voir comment sont annotées les formes de ce type sur l'exemple du comparatif *гарніший* [plus beau] dans la Figure 54.

Figure 54. Annotation de l'adjectif *гарніший*



Évaluation des grammaires

Ces deux grammaires nous permettent de reconnaître quelques dizaines d'occurrences supplémentaires dans nos textes (voir Tableau 56).

Tableau 56. Évaluation des grammaires «Adjectives_Degree_V.1.3.nom» et «Adverbs_Degree_V.1.3.nom»

	Nombre d'occurrences dans le texte	Nombre d'occurrences non reconnues		Nombre d'occurrences reconnues	Occurrences non reconnues en %	
		1	2		1	2
Corpus 1	108 137	2 503	2 258	245	2,3%	2,1%

4.4.9 Formalisation des formes courtes des adjectifs

Description de la grammaire

Comme nous l'avons dit précédemment (voir section 3.3.3) les adjectifs courts s'emploient surtout dans la langue poétique, mais on peut les trouver également dans les textes littéraires. Pour rappel : la forme courte a été créée suite à la contraction de la terminaison du nominatif masculin, comme pour l'adjectif *зеленый* [vert] – forme longue → *зелен* [vert] – forme courte.

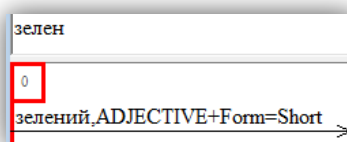
Pour que les adjectifs courts soient reconnus, nous avons construit la grammaire morphologique productive « Adjectives_Short_V.1.3.nom » (voir Figure 55). Cette grammaire est très simple, car les adjectifs courts ne se déclinent pas. Elle nous permet de reconnaître toutes les formes courtes créées à partir des adjectifs de notre dictionnaire par enlèvement de la terminaison adjectivale. Elle prend en compte également l'apparition de la lettre *-e-* mobile avant le suffixe *-н-*.

Figure 55. Grammaire formalisant les adjectifs courts



Elle affiche dans les annotations le lemme de départ et nous permet également d'ajouter l'annotation « Short » à tous les adjectifs courts reconnus (voir Figure 56).

Figure 56. Annotation de l'adjectif court *зелен*



Évaluation de la grammaire

Nos textes n'avaient que peu d'adjectifs de ce type ; nous pouvons voir leur décompte dans le Tableau 57.

Tableau 57. Évaluation de la grammaire « Adjectives_Short_V.1.3 »

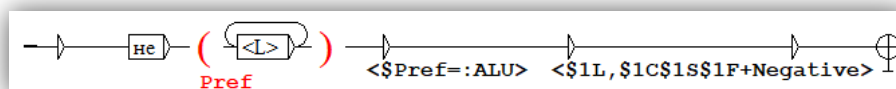
	Nombre d'occurrences dans le texte	Nombre d'occurrences non reconnues		Nombre d'occurrences reconnues	Occurrences non reconnues en %	
		1	2		1	2
Corpus 1	108 137	2 258	2 247	11	2,1%	2,1%

4.4.10 Formalisation des formes négatives

Description de la grammaire

En ukrainien, les noms, les adjectifs, les adverbes peuvent avoir des antonymes qui sont exprimés par le même lexème précédé par « не- »²²⁷. Un certain nombre de ces ALU étaient présentes dans notre source, mais cette liste n'est pas exhaustive. Ainsi, pour les prendre en compte, nous avons construit la grammaire morphologique « Negative_V.1.3 », présentée dans la Figure 57.

Figure 57. Grammaire formalisant les formes négatives



Cette grammaire reconnaît n'importe quelle forme avec préfixe « не- », grâce à la contrainte « <\$Pref=:ALU> ». Pour rappel, le symbole « ALU » décrit soit une entrée lexicale, soit une forme fléchie ou dérivée d'une entrée lexicale, soit une forme reconnue par une grammaire morphologique. Ici, la grammaire attribue aux formes reconnues les mêmes étiquettes qu'aux formes reconnues de départ (le lemme, la catégorie, les informations du dictionnaire et les informations flexionnelles), puis ajoute une étiquette supplémentaire « Négative ». Dans la Figure 58, nous pouvons voir comment est annoté l'adjectif *негарний* [pas beau], reconnu grâce à cette grammaire.

Figure 58. Annotation de l'adjectif négatif *негарний*

негарний
0
гарний,ADJECTIVE+Case=Nominative+Gender=Masculine+Number=Singular+Expression=Negative

Évaluation de la grammaire

Suite à l'application de cette grammaire, le nombre d'occurrences non reconnues passe de 2 247 à 1 967 ; nous pouvons voir le décompte dans le Tableau 58.

Tableau 58. Évaluation de la grammaire « Négative_V.1.3 »

	Nombre d'occurrences dans le texte	Nombre d'occurrences non reconnues		Nombre d'occurrences reconnues	Occurrences non reconnues en %	
		1	2		1	2
Corpus 1	108 137	2 247	1 967	280	2,1%	1,8%

²²⁷ Il s'agit de la particule contractée.

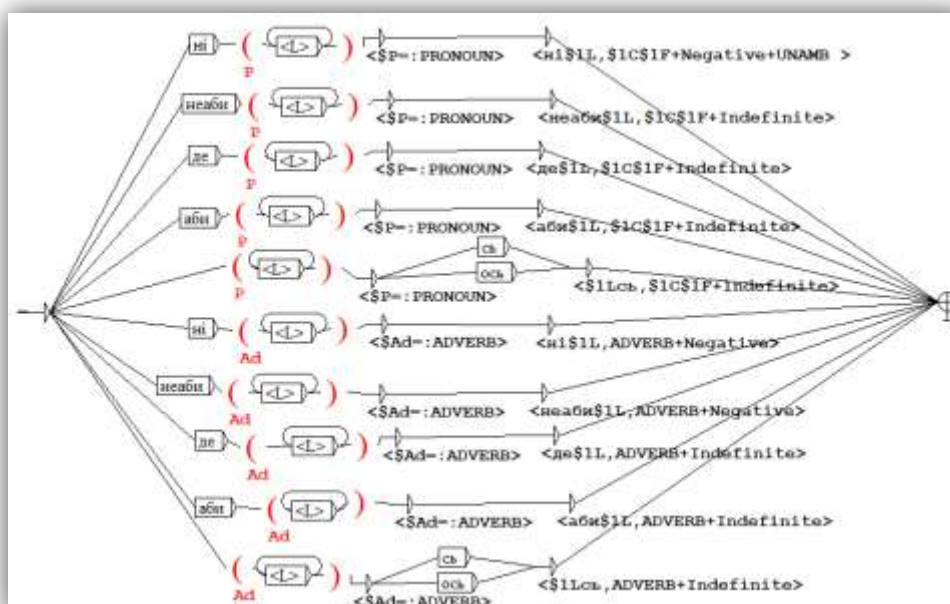
4.4.11 Formalisation des adverbes et des pronoms indéfinis et négatifs

Description de la grammaire

En observant les ALU non reconnues, nous avons remarqué que dans la liste, il y avait des adverbes et des pronoms indéterminés ou indéfinis, formés par la contraction de la particule et un adverbe ou un pronom existant. De plus, nous avons intentionnellement enlevé de notre dictionnaire les pronoms-adjectifs qui sont formés par l'ajout du morphème postposé *-сь*, pour éviter de démultiplier nos paradigmes, comme c'est le cas du pronom indéterminé *хтось* [quelqu'un].

Ainsi, pour compléter notre dictionnaire, nous avons construit la grammaire « Adverbs_Pronouns_Indefinite.V.1.3 » (voir Figure 59) qui prend en compte quatre morphèmes préposés (*ни-, небу-, де-, абу-*) et deux morphèmes postposés (*-сь* et *-ось*).

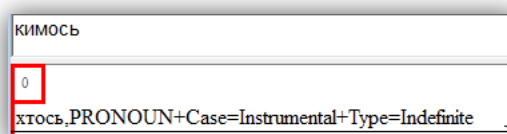
Figure 59. Grammaire formalisant les adverbes et les pronoms indéfinis et négatifs



Dans cette grammaire, nous décrivons le mécanisme d'ajout du morphème au pronom (cinq premières lignes) ou à l'adverbe (cinq lignes suivantes). Dans les étiquettes, nous remplaçons le lemme de départ par le lemme créé grâce à l'ajout de la particule, nous ajoutons également l'étiquette « Indefinite » ou « Negative », puis pour les pronoms, nous récupérons les informations flexionnelles de la forme fléchie de départ.

Nous pouvons voir comment est annotée une forme fléchie de l'ALU *хтось* [quelqu'un] dans la Figure 60.

Figure 60. Annotation du pronom *хтось*



Évaluation de la grammaire

Comme les pronoms et les adjectifs de ce type ne sont pas nombreux, le nombre d'occurrences reconnues par cette grammaire n'est pas très grand (voir Tableau 59).

Tableau 59. Évaluation de la grammaire « Adverbs_Pronouns_Indefinite_V.1.3.nom »

	Nombre d'occurrences dans le texte	Nombre d'occurrences non reconnues		Nombre d'occurrences reconnues	Occurrences non reconnues en %	
		1	2		1	2
Corpus 1	108 137	1 967	1 952	15	1,8%	1,8%

4.4.12 Formalisation des alternances au début des mots

Description de la grammaire

En ukrainien, il existe des prépositions synonymiques et des préfixes synonymiques créés à partir des prépositions en question, comme c'est le cas de *в-/у-* et de *від(i)-/од(i)-* (Dyka et Shvedova 2022 [28]²²⁸). Notons qu'au fil des siècles, ces préfixes ont perdu leur indépendance et sont devenus des parties du radical. Ainsi, il existe des ALU de catégories grammaticales différentes qui commencent par *в-* ou par *у-*, ou bien par *від(i)-* ou *од(i)-*, avec un sens synonymique, par exemple :

- *вчитель / учитель [enseignant], вчений / учений [scientifique], вчиту / учиту [apprendre]* ;
- *від'їзд / од'їзд [départ], відчиняти / одчиняти [ouvrir], віднині / однині [dorénavant]*.

Les variantes en *в-/у-* ont toujours été recensées dans les dictionnaires, comme ceci :

ВЧÉНИЙ див. учений. [ВЧÉНИЙ regarder учений]²²⁹

En outre, les variantes *від(i)-/од(i)-*, à quelques exceptions près, ont été considérées comme régionales et ont fait leur apparition dans les dictionnaires depuis la réforme orthographique de 2019. D'ailleurs, dans le dictionnaire SUM-20, ces variantes sont données, mais pas toutes recensées. Par exemple, nous pouvons trouver en entrée du dictionnaire le début des mots « *од'ї...* », comme ceci :

*« Од'ї... див. від'ї...; напр.: од'їд́а́тися, од'ї́жджа́ти, од'ї́зд і т. ін. див. від'ї́д́а́тися, від'ї́жджа́ти, від'ї́зд і т. ін. [Од'ї... **regarder** від'ї...; **par exemple** : од'ї́д́а́тися, од'ї́жджа́ти, од'ї́зд, etc., **regarder** від'ї́д́а́тися, від'ї́жджа́ти, від'ї́зд, etc.]²³⁰*

Dans le dictionnaire NooJ, nous n'avons entré qu'une seule variante pour ces deux formes. Pour les variantes en *в-/у-* nous avons gardé les variantes trouvées dans notre source, à savoir soit l'une, soit l'autre. Les variantes en *од(i)-* sont d'ailleurs presque toutes absentes. Notons que parfois les deux variantes ont été recensées dans le dictionnaire open source que nous avons utilisé ; nous n'avons pas pris le temps de les trier, car nous avons décidé de construire une grammaire capable de reconnaître les secondes à partir des premières et vice versa, en donnant une priorité différente (soit plus faible, soit plus élevée) à notre dictionnaire, pour éviter de produire une double analyse artificielle. La grammaire « Parallel_forms_V.1.3 » est présentée dans la Figure 62.

²²⁸ Les autrices s'appuient sur plusieurs grammaires, les normes orthographiques et l'*Encyclopédie de la langue ukrainienne*.

²²⁹ <https://sum20ua.com/Entry/index?wordid=214281&page=572>

²³⁰ Nous pouvons trouver cet exemple ici : <https://sum20ua.com/Entry/index?wordid=65024&page=2060>

Notre grammaire est composée de trois graphes : le graphe principal regroupe deux nœuds : le premier pour les variantes *в-/y-* et le deuxième pour les variantes *вид(i)-/од(i)-* ; tous deux ont été construits de la même façon. Dans la Figure 61, nous pouvons voir le graphe imbriqué qui décrit la production des variantes *в-/y-*. Dans la première partie de ce graphe, nous remplaçons « в » par « y » (la commande « #в\$Pref ») et dans la deuxième partie – « y » par « в ». L'ALU peut être un nom, un adjectif, un verbe (y compris un participe), un adverbe. Toutes les annotations de la forme fléchée de départ sont reprises et l'annotation supplémentaire « Parallel_form » est ajoutée (voir Figure 62).

Figure 61. Grammaire formalisant des alternances au début des mots

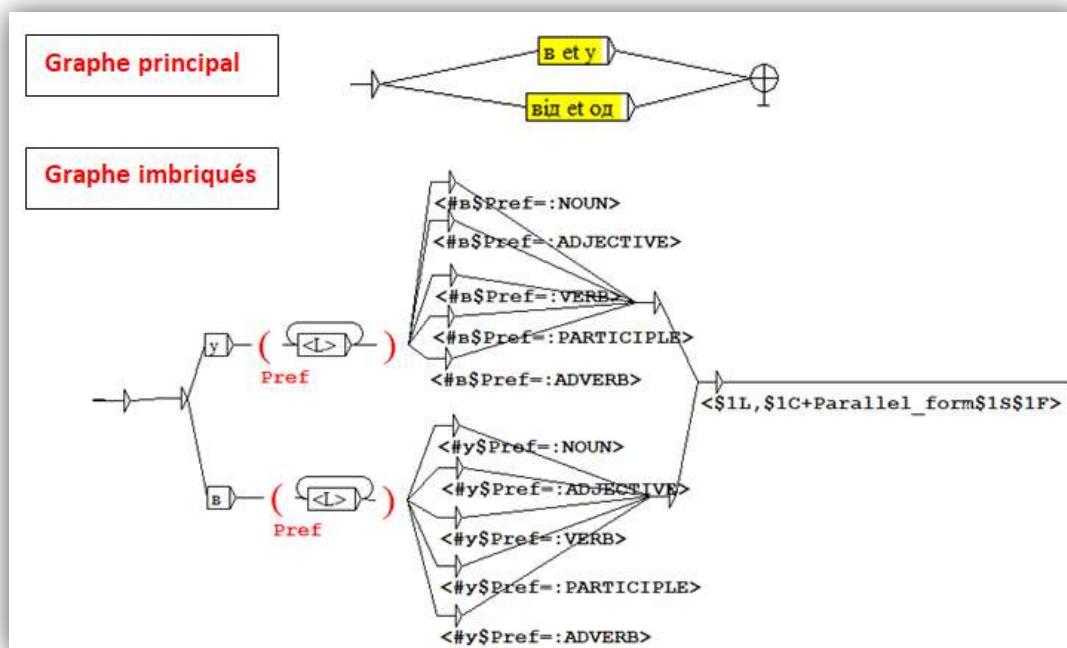


Figure 62. Annotation du nom *учитель*

учитель
0
вчитель,NOUN+Parallel_form+Type=Common+Gender=Masculine+Animation=Animate+Case=Nominative+Number=Singular

Évaluation de la grammaire

Pour que notre grammaire prenne en compte également les formes décrites par les autres grammaires, nous lui donnons une priorité inférieure aux autres grammaires. Nous pouvons voir les détails sur les occurrences reconnues suite à l'application de cette grammaire dans le Tableau 60.

Tableau 60. Évaluation des grammaires « Parallel_forms_V.1.3 »

	Nombre d'occurrences dans le texte	Nombre d'occurrences non reconnues		Nombre d'occurrences reconnues	Occurrences non reconnues en %	
		1	2		1	2
Corpus 1	108 137	1 952	1 887	65	1,8%	1,7%

4.4.13 Formalisation des mots en alphabet latin ou en russe

Description de la grammaire

Dans les textes ukrainiens, aussi bien littéraires que journalistiques, nous pouvons trouver des mots ou des passages entiers, écrits en alphabet latin ou en russe. Pour que NooJ reconnaisse ces mots, nous avons préparé la grammaire « Foreign_words_V1.3.nom » que nous pouvons voir dans la Figure 59.

Cette grammaire est composée de cinq graphes. Sur le graphe principal, il y a trois nœuds : le premier décrit l'alphabet latin, le deuxième décrit les lettres spécifiques du cyrillique russe et le troisième décrit le mélange de lettres alphabétiques cyrilliques et latines.

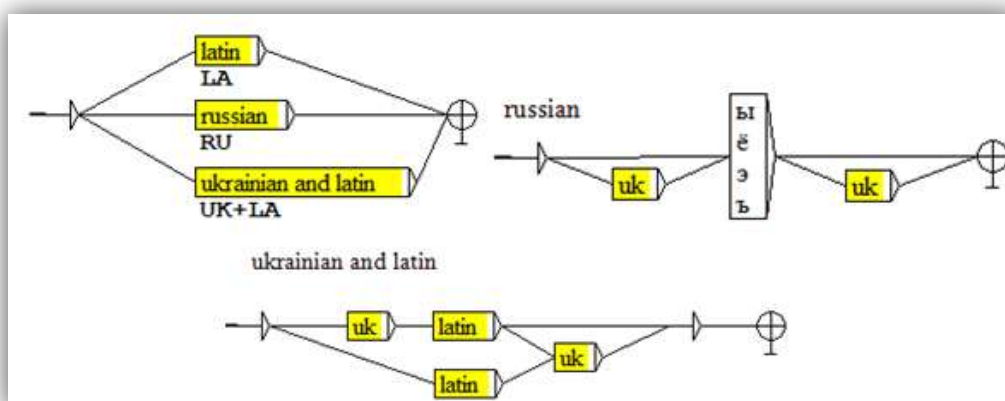
Dans le premier nœud (:latin), sont contenus les caractères issus de la norme ISO/CEI 8859-1 qui couvre, entre autres, l'anglais, le français, l'allemand, le latin liturgique, etc.²³¹, puis les caractères issus de la norme ISO/CEI 8859-2 qui couvre, entre autres, le polonais, le tchèque, etc.²³². Notons que dans les textes écrits en ukrainien figurent le plus souvent des mots en provenance de ces langues.

Dans le deuxième nœud (:russian), sont contenus les caractères spécifiques de l'alphabet russe qu'on peut voir dans la Figure 63. Ces lettres peuvent être combinées à d'autres lettres cyrilliques contenues dans l'alphabet ukrainien (:uk).

Dans le troisième nœud, nous avons formalisé les ALU qui sont orthographiées par erreur avec des lettres de l'alphabet latin au lieu de lettres cyrilliques, par exemple le « a » latin au lieu de « а » cyrillique, qui pour un œil humain est la même lettre, mais pour l'ordinateur, il s'agit de codages différents : « 61 » (ISO 646) pour le premier et « D0 » (ISO 8859-5) pour le deuxième.

Par la suite, tous les mots concernés par cette grammaire seront annotés avec l'étiquette « LA » pour les mots en latin, « RU » pour les mots en russe et « UK+LA » (voir Figure 65). Cette dernière annotation est particulièrement intéressante, car elle permet de détecter très rapidement les fautes de frappe, de les corriger et d'intégrer les mots corrigés dans les statistiques.

Figure 63. Grammaire formalisant des mots orthographiés en alphabet latin et en cyrillique russe



²³¹ La liste complète des langues et des symboles est ici : https://fr.wikipedia.org/wiki/ISO/CEI_8859-1

²³² La liste complète des langues et des symboles est ici : https://fr.wikipedia.org/wiki/ISO/CEI_8859-2

Par ailleurs, pour que les chiffres romains soient analysés non pas en tant que formes « LA », mais en tant que tels (et associés à leur valeur numérique équivalente), nous avons repris la grammaire de Max Silberztein pour le français²³³. Elle produit pour chaque chiffre romain sa valeur numérique correspondante. Nous pouvons voir un extrait de la grammaire « Latin_numerals_V1.1.3.nom » que nous avons repris pour l’ukrainien dans la Figure 64 et l’annotation du chiffre « VII » dans la Figure 64. Nous donnons une priorité supérieure à cette grammaire, pour éviter que les chiffres romains ne soient annotés avec l’étiquette « LA ».

Figure 64. Extrait de la grammaire formalisant des chiffres romains

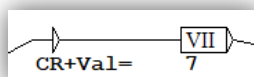
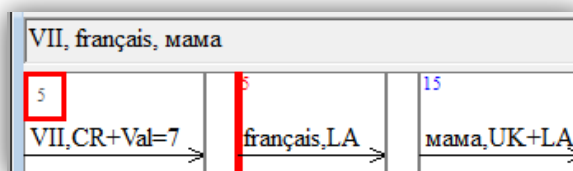


Figure 65. Annotation des ALU VII, français, mama



Dans la Figure 66, nous pouvons voir l’annotation de trois occurrences, la première et la deuxième sont écrites en caractères latins, mais la première est décrite par la grammaire formalisant les chiffres romains. La troisième occurrence, « *mama* » [*maman*], a été écrite en ukrainien pour les trois premiers caractères, et le dernier « *a* » a été entré en français.

Évaluation des grammaires

Les résultats du test, suite à l’application de ces grammaires, sont présentés dans le Tableau 61. Nous pouvons voir que le nombre des ALU formes inconnues a beaucoup diminué.

Notre grammaire « Foreign_words_V1.3.nom » a quelques points faibles, car elle ne fait pas de distinction entre les langues écrites en alphabet latin, elle ne prend pas en compte d’autres langues écrites en alphabet cyrillique, elle différencie le russe de l’ukrainien uniquement lorsque les formes utilisent des lettres distinctives, or cette distinction peut être élargie aux affixes spécifiques pour chaque langue. Mais pour construire une grammaire plus exhaustive, nous aurons besoin de faire une étude comparative de la morphologie de l’ukrainien et du russe.

Tableau 61. Évaluation des grammaires « Foreign_words_V1.3.nom » et « Latin_numerals_V1.1.3.nom »

	Nombre d’occurrences dans le texte	Nombre d’occurrences non reconnues		Nombre d’occurrences reconnues	Occurrences non reconnues en %	
		1	2		1	2
<i>Corpus 1</i>	108 137	1 887	1 738	149	1,7%	1,6%

²³³ Les grammaires « _Chiffres romains.nom » et « _Nom propre.nom » sont disponibles en téléchargement dans le module du français sur le site <https://nooj.univ-fcomte.fr/resources.html>

4.4.14 Formalisation des noms propres

Description de la grammaire

La liste des occurrences non reconnues contenait également un grand nombre de noms propres qui ne sont pas recensés dans notre dictionnaire. Pour les reconnaître, nous avons repris la grammaire construite par Max Silberztein pour le français. Notre grammaire « Proper_V.1.2.nom », tout comme la grammaire du français, reconnaît des mots écrits avec une majuscule et ajoute dans le texte l'annotation « Proper ». Pour que les mots écrits avec une majuscule en début de phrase ne soient pas traités comme noms propres, nous donnons une priorité faible à cette grammaire. Nous pouvons voir cette grammaire dans la Figure 66 et l'annotation du prénom *Анвар* [Anwar], suite à son application dans la Figure 67.

Figure 66. Grammaire formalisant des noms propres

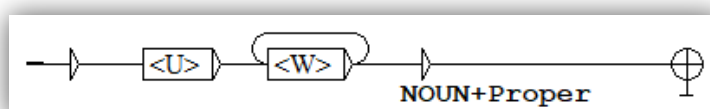
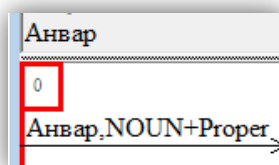


Figure 67. Annotation du nom propre *Анвар*



Évaluation de la grammaire

Cette grammaire a permis d'analyser 350 occurrences écrites avec une majuscule en tant que nom propre (voir Tableau 62), mais elle ne nous permet pas de décrire la morphologie de ces ALU. Nous pouvons supposer qu'il peut exister un contexte dans lequel une occurrence qui n'a pas été formalisée dans notre dictionnaire pourrait être orthographiée en majuscule, bien qu'il ne s'agisse pas d'un nom, mais d'une ALU d'une autre catégorie. Dans ce cas, l'analyse produite sera incorrecte. Nous n'avons pas rencontré ce problème dans notre corpus, mais nous ne pouvons pas pour autant exclure ce cas de figure complètement. Ainsi, nous devons utiliser cette grammaire avec précaution et construire à l'avenir des grammaires qui prennent en compte la morphologie des ALU reconnues ainsi.

Tableau 62. Évaluation de la grammaire «Proper_V1.3.nom»

	Nombre d'occurrences dans le texte	Nombre d'occurrences non reconnues		Nombre d'occurrences reconnues	Occurrences non reconnues en %	
		1	2		1	2
<i>Corpus 1</i>	108 137	1 738	1 393	345	1,7%	1,3%

4.4.15 Évaluation de l'ensemble des grammaires morphologiques

L'ensemble de nos grammaires morphologiques permet une bonne couverture du vocabulaire ukrainien non décrit dans le dictionnaire, puisque même sans compter les trois dernières grammaires « Foreign_words_V1.3.nom », « Latin_numerals_V1.1.3.nom » et « Proper_V1.3.nom »,

98,3% des occurrences de notre corpus sont reconnues et analysées. Ces grammaires nous permettent non seulement de mieux annoter les textes, mais aussi d'analyser des formes dérivées.

Certaines formes dérivées n'ont pas été traitées dans ce travail : par exemple, les noms péjoratifs, les métiers au féminin, les verbes avec le double ou triple préfixe. Nous comptons réaliser ce travail dès cette année.

De plus, certaines formes dérivées peuvent être inscrites aussi bien dans le dictionnaire et être reconnues par une grammaire. Pour éviter d'introduire des ambiguïtés artificielles, il faut donner une priorité différente à notre grammaire : soit plus faible, si l'on veut privilégier le dictionnaire, soit plus élevée, si l'on veut privilégier la grammaire. Pour mieux faire, il faudrait enlever de notre dictionnaire toutes les formes décrites par nos grammaires. Nous pensons le faire, en adaptant notre module à la version NooJ Web²³⁴.

Notons que quelques grammaires pourront être améliorées. Par exemple, nous pourrions compléter notre grammaire « Adverbs_Pronouns_Indefinite.V.1.3 » avec des adverbes et des pronoms qui s'écrivent avec un trait d'union. En effet, il n'est pas possible d'utiliser le trait d'union dans une grammaire morphologique avec la version de NooJ que nous avons utilisée, mais c'est possible avec NooJ Web. Nous pourrions aussi formaliser les adverbes du type *по-турецьки* [à la turque], créé à partir de la préposition *по* et de l'adjectif *турецький* [turque], ou encore des adjectifs composés comme *англо-український* [anglais-ukrainien] dont la première partie a été formée à partir de l'adjectif *англійський* [anglais].

En effet, les ALU que nous avons citées font partie des 1 474 occurrences non reconnues dans notre corpus.

Parmi ces occurrences non reconnues, nous trouvons :

- quelques adjectifs et adverbes composés, y compris sans trait d'union, comme *загальноосвітній* [de l'enseignement général] ;
- quelques ALU rares qui n'étaient pas inscrites dans le dictionnaire source, alors qu'elles devront y être, comme l'ALU *зависть* [jalousie, envie], synonyme de *зздрість*, utilisée plus couramment, ou *бичувати* [taper avec un fouet] ;
- quelques abréviations ;
- les néologismes, ou les emprunts récents, comme *абгрейдити* [mettre à niveau] ou *грант* [bourse, subvention] ;
- un assez grand vocabulaire d'auteur(s), comme le verbe *жбурхнути*, absent des dictionnaires actuels, mais utilisé par Kobylans'ka dans le sens « taper en faisant un bruit fort » ou des variantes régionales, comme l'adverbe *завзятущо* au lieu de *завзятуще* [énergiquement], utilisé par Netchouy-Levyts'kyy ou le verbe *запіє* au lieu de *заспіває* [se mettra à chanter], utilisé par Kobylans'ka ;
- beaucoup de mots du « sourjyk », comme *глупа* qui provient du mot russe *глупая* [idiote] au lieu de l'ukrainien *дурна*, ou encore *воздусі*, le nom au locatif, construit selon les règles ukrainiennes du mot russe *воздух* [aire] au lieu de l'ukrainien *повітря* ;
- les mots tapés ou numérisés avec des erreurs comme *івши*, au lieu de *ївши* [en mangeant], ou *ух* au lieu de *їх* [« eux » à l'accusatif].

Toutes ces formes non reconnues pourront être décrites. Les ALU spécifiques, comme les variantes régionales, le « sourjyk », pourront aussi être reconnues grâce à des grammaires morphologiques, ou alors recensées et reliées aux formes initiales dans un dictionnaire. Nous

²³⁴ La version Web est disponible pour les tests depuis décembre 2023 sur <https://webnooj.univ-fcomte.fr>

pourrons aussi construire un dictionnaire de l'ukrainien de diaspora française, ou des dictionnaires secondaires pour les différents dialectes.

Nous comptons analyser la langue de la presse de la diaspora française et présenter les ressources linguistiques correspondantes à la prochaine conférence NooJ (juin 2024).

4.5 Grammaires syntaxiques de désambiguïsation pour le module ukrainien

Comme nous l'avons déjà dit, il n'existe pas beaucoup d'outils formels en libre accès pour l'ukrainien, et ceux qui existent ne désambigüisent pas les unités textuelles (voir la section 2.4.2 qui décrit GRAC et la section 2.4.3 qui décrit LanguageTool API NLP UK). Or, l'ukrainien est une langue particulièrement nantie d'ambigüités qu'il faut lever pour pouvoir produire des analyses de textes précises.

Sur la plateforme NooJ, les ambigüités peuvent être levées grâce à des grammaires syntaxiques. Dans cette section, nous ferons d'abord le point sur la typologie des ambigüités, puis nous présenterons les grammaires que nous avons construites pour les résoudre.

Pour évaluer nos grammaires, nous utiliserons le *Corpus 1* (108 137 occurrences) présenté dans la section 4.3.5. Avant de construire notre grammaire, nous constatons que notre corpus contient 206 445 annotations, ce qui représente quasiment le double des occurrences du corpus. Le fichier « Ambigüities » calculé par NooJ présente 10 209 ambigüités. Pour évaluer nos grammaires syntaxiques, nous préparerons des tableaux dans lesquels nous mettrons les données récupérées après l'application de la grammaire (ou des grammaires) sur le *Corpus 1*, que nous allons comparer avec les données précédentes.

4.5.1 Typologie des homonymes grammaticaux

L'homonymie grammaticale est un phénomène bien connu en linguistique ukrainienne. Ce sujet a beaucoup été étudié ces quelques dernières décennies (Hriaznoukhina *et alii* 1989 [24], Kouz'menko 2000 [36], Dems'ka 2019 [27], Chypanivs'ka 2008 [68], Bounio 2014 [8], Hlibtchouk 2018 [19]). Par ailleurs, il n'existe pas de classification bien définie des homonymes grammaticaux. Nous en proposons une, adaptée aux outils de TAL. Selon cette classification, les homonymes sont répartis dans les groupes suivants :

- les homonymes flexionnels,
- les homoformes,
- les homonymes lexico-grammaticaux verbaux,
- les homonymes transposés.

Les homonymes flexionnels²³⁵ appartiennent au même paradigme et ont une composition graphique identique, mais sont associés à plusieurs valeurs grammaticales différentes. Par exemple, il existe deux formes fléchies du lemme *cecnpa* [*sœur*] qui ont la même flexion « u » (*cecnpu*). La première désigne le génitif singulier, tandis que la deuxième désigne le nominatif pluriel.

Les homoformes peuvent appartenir à la même classe ou à des classes grammaticales différentes. Leur composition graphique est identique dans une ou quelques formes fléchies (Ponomariv *et alii* 2008 [53]). Par exemple, le nom masculin au nominatif singulier *hic* [*nez*] forme le

²³⁵ Chez certains linguistes, ce type d'homonymie est appelé l'homonymie morphologique (Ponomariv *et alii* 2008 [53]).

couple avec le verbe IPF au passé masculin *ніс [(il) a porté]*. Ou encore, le verbe IPF au passé féminin *спала [(elle) a dormi]* forme le couple avec le verbe PF au passé féminin *спала [(elle) est tombée]*.

Il existe plusieurs types d'homofomes. Selon une étude réalisée par Ol'ha Chypaniv's'ka (2008 [68]), les ambiguïtés les plus fréquentes sont entre :

- le nom et le verbe ;
- l'adjectif et le verbe ;
- le verbe et l'adverbe ;
- le nom et l'adverbe ;
- l'adjectif et l'adverbe.

Les homonymes lexico-grammaticaux verbaux ont la même composition graphique dans toutes leurs formes fléchies, mais ils ne sont pas de même aspect. Par exemple, le verbe IPF *заносити [apporter (quelque chose) à l'intérieur]* du couple aspectuel *заносити (IPF) / занести (PF)* et le verbe PF *заносити [user (un vêtement)]* » du couple aspectuel *заношувати (IPF) / заносити (PF)*.

Les homonymes transposés, que nous avons mentionnés dans le chapitre 3 (tableau 6) ont été créés par le biais de la conversion, c'est-à-dire le passage de l'ALU d'une classe à l'autre, ainsi les homonymes de ce type appartiennent à deux (ou plus) classes différentes. Par exemple, *коло* appartient à deux classes : le nom neutre singulier au nominatif *коло [cercle]* et la préposition *коло [à côté de]*²³⁶, tandis que *так* appartient à trois classes : l'adverbe *так [ainsi]*, la particule *так [oui]* et la conjonction *так [mais]*²³⁷. À ce groupe appartiennent aussi les couples d'adjectif-adjectif substantivé, comme *хворий [malade]*, l'adjectif masculin ou *хворий [le malade]*, nom masculin que nous citons dans le chapitre 3, Tableau 6.

Dans cette section, nous présenterons des grammaires syntaxiques qui permettent de traiter certaines ambiguïtés liées aux homonymes flexionnels. Puis, nous présenterons une grammaire qui permet de désambiguïser les couples d'homofomes dont un membre est un verbe et une deuxième qui permet de désambiguïser les couples dont un membre est un adverbe. Nous ne traiterons pas ici les homonymes lexico-grammaticaux verbaux.

4.5.2 Grammaires syntaxiques pour désambiguïser les homonymes flexionnels

En ukrainien, les homonymes flexionnels représentent 98 % de tous les homonymes grammaticaux (Chypaniv's'ka 2008 [68]), donc il convient de les traiter en priorité.

Tous les paradigmes, à l'exception des paradigmes verbaux, incluent deux formes (ou plus) avec des flexions homonymiques. Par exemple, dans le paradigme nominal « СИН » que nous avons préparé pour traiter les noms masculins en consonne dure, il y a cinq flexions homonymiques : « *а* », « *ів* », « *у* », « *ові* » et « *у* ». Nous avons formalisé ce paradigme dans la Figure 68.

Les paradigmes adjectivaux ont encore plus de flexions homonymiques, car les adjectifs masculins et les adjectifs neutres ont les mêmes terminaisons pour tous les cas hormis le nominatif. Par exemple, le paradigme adjectival « ВЕЛИКИЙ » contient dix flexions homonymiques « *уї* », « *ого* », « *ому* », « *ім* », « *ому* », « *ум* », « *іў* », « *і* », « *ії* », « *ух* » (voir Tableau 12).

Il en va de même pour toutes les classes grammaticales qui s'accordent avec le nom ou le pronom personnel, c'est-à-dire les pronoms-adjectifs, les numéraux ordinaux et les participes.

²³⁶ Nous pouvons voir la description des sens de *коло* dans le dictionnaire SUM : <https://sum.in.ua/s/kolo>

²³⁷ Nous pouvons voir la description des sens de *так* dans le dictionnaire SUM : <https://sum.in.ua/s/tak>

Figure 68. Le paradigme nominal « СИН »

СИН=	
<E>/Nominative+Singular	<E>и/Nominative+Plural
<E>а/Accusative+Singular	<E>иb/Accusative+Plural
<E>а/Genitive+Singular	<E>иb/Genitive+Plural
<E>и/Locative+Singular	<E>аx/Locative+Plural
<E>y/Locative+Singular	<E>ам/Dative+Plural
<E>оbi/Locative+Singular	<E>ами/Instrumental+Plural
<E>y/Dative+Singular	<E>и/Vocative+Plural
<E>оbi/Dative+Singular	
<E>ом/Instrumental+Singular	
<E>y/Vocative+Singular	
<E>е/Vocative+Singular	

Par conséquent, sans désambiguïsation, nous obtenons plusieurs étiquettes pour presque toutes les classes variables. Nous pouvons voir un exemple du texte annoté avant que les ambiguïtés soient levées dans la Figure 69.

Figure 69. Annotations ambiguës

дівчи на виді. Наступа була гарна, але трохи блідувата, ніби втомлена, або трохи занеділа, ніби вона дві ночі поспіль танцювала на

218	дівчи, NOUN+Gender=Feminine+Type=Common+Animation=Inanimate+Case=Genitive+Number=Singular
	дівчи, NOUN+Gender=Feminine+Type=Common+Animation=Inanimate+Case=Nominative+Number=Plural
	дівчи, NOUN+Gender=Feminine+Type=Common+Animation=Inanimate+Case=Accusative+Number=Plural+Animation=Inanimate
	дівчи, NOUN+Gender=Feminine+Type=Common+Animation=Inanimate+Case=Vocative+Number=Plural

Désambiguïsation des accords dans le groupe nominal

Pour lever les ambiguïtés liées à l'homonymie flexionnelle, nous avons développé une grammaire syntaxique décrivant l'accord entre les membres d'un syntagme dont le noyau est un nom (GN) ou un pronom personnel (GP), les membres secondaires sont des adjectifs, des pronoms-adjectifs, des numéraux ordinaux, ou des participes qui s'accordent tous en genre, en cas et en nombre avec le noyau du groupe. Quant aux numéraux cardinaux et indéterminés, ils s'accordent tous en cas avec le noyau du groupe ; *два [deux]*²³⁸ s'accorde de plus en genre, *один [un]* s'accorde de plus en genre et nombre.

Avant de construire cette grammaire, nous décrivons comment les membres d'un syntagme peuvent être liés (voir Figure 70). Dans notre schéma les membres secondaires sont décrits par le symbole <ALU> (qui représente n'importe quelle catégorie).

Nous partons de l'hypothèse que l'ordre des membres d'un GN ou d'un GP est libre²³⁹, et que les membres secondaires peuvent se répéter et être liés par une conjonction ou une virgule. La

²³⁸ Et aussi le numéral collectif *обидва [tous les deux]*.

²³⁹ Bien qu'il existe un ordre narratif prédéterminé, cet ordre peut être chamboulé dans diverses situations.

particule *не* [ne pas] ou des adverbes comme *дуже* [très], *надто* [trop] peuvent être aussi insérés dans la construction.

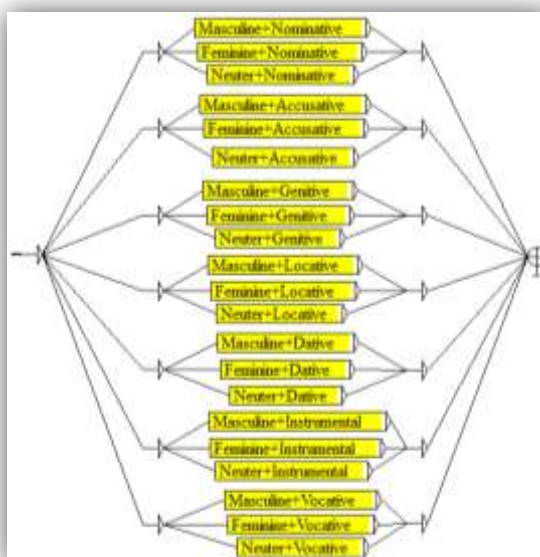
Figure 70. Composition d'un GN ou un GP

1. ALU*²⁴⁰ → NOUN ou PRONOUN+Personal ;
2. PARTICLE « не » et/ou PRONOUN+Possessive+Invariable et/ou ADVERB → ALU* → NOUN ou PRONOUN+Personal ;
3. ALU* → “,” et/ou CONJUNCTION → ALU* → NOUN ou PRONOUN+Personal ;
4. PARTICLE « не » et/ou PRONOUN+Possessive+Invariable et/ou ADVERB → ALU* → “,” et/ou CONJUNCTION → PARTICLE « не » et/ou PRONOUN+Possessive+Invariable et/ou ADVERB → ALU* → NOUN ou PRONOUN+Personal ;
5. NOUN ou PRONOUN+Personal → ALU* ;
6. NOUN ou PRONOUN+Personal → PARTICLE « не » et/ou PRONOUN+Possessive+Invariable et/ou ADVERB → ALU* ;
7. NOUN ou PRONOUN+Personal → ALU* → “,” et/ou CONJUNCTION → ALU* ;
8. NOUN ou PRONOUN+Personal → PARTICLE « не » et/ou PRONOUN+Possessive+Invariable et/ou ADVERB → ALU* → “,” et/ou CONJUNCTION → PARTICLE « не » et/ou PRONOUN+Possessive+Invariable et/ou ADVERB → ALU*.

Ces huit cas sont représentés par la grammaire « GN_GP_V.1.3.nog », présentée en Figure 71. Elle est composée de 23 graphes. Le graphe principal (Figure 71) est structuré en 7 blocs qui décrivent l'accord en cas. Chaque bloc, contient des graphes imbriqués qui décrivent l'accord en genre (masculin, féminin et neutre).

Ces huit cas sont représentés par la grammaire « GN_GP_V.1.3.nog », présentée en Figure 71. Elle est composée de 23 graphes. Le graphe principal (Figure 71) est structuré en 7 blocs qui décrivent l'accord en cas. Chaque bloc, contient des graphes imbriqués qui décrivent l'accord en genre (masculin, féminin et neutre).

Figure 71. Le graphe principal de la grammaire syntaxique « GN_GP_V.1.3.nog »



²⁴⁰ Ici, ALU* désigne un ou plusieurs membres secondaires. La répétition est notée par le symbole « * ».

Figure 72. Le graphe « Feminine+Vocative » de la grammaire « GN_GP_V.1.3.nog »

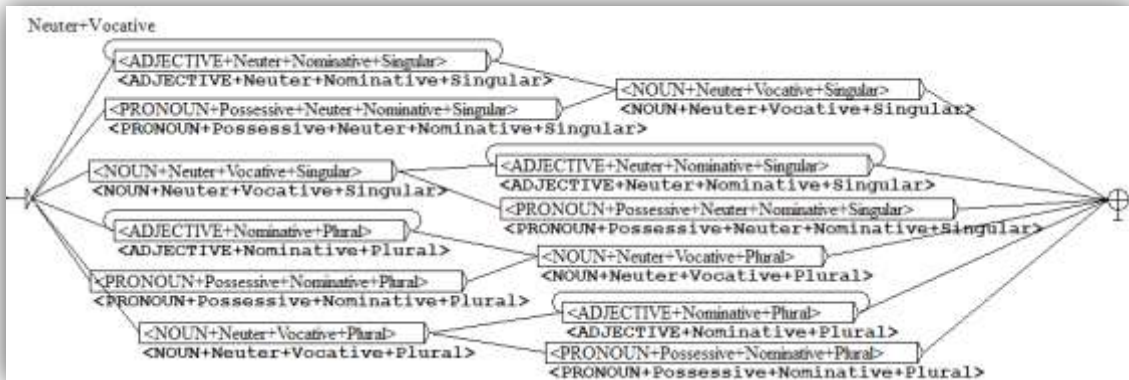
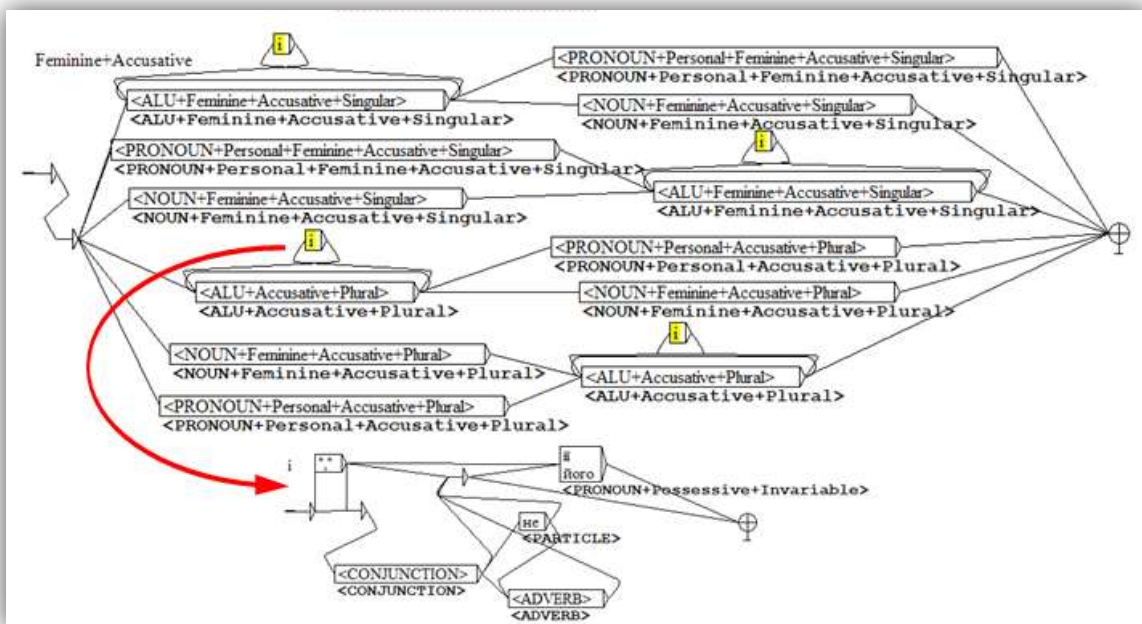


Figure 73. Le graphe « Feminine+Acusative » de la grammaire « GN_GP_V.1.3.nog »



Chaque graphe imbriqué décrit les huit contextes de notre schéma (Figure 70), à l'exception de graphes qui décrivent le syntagme au vocatif. Ces derniers décrivent uniquement le contexte 1 et le contexte 5 et uniquement pour le GN, car les pronoms personnels n'ont pas de vocatif. Nous pouvons voir le graphe qui décrit le vocatif féminin dans la Figure 72, et le graphe qui décrit l'accusatif féminin dans la Figure 73.

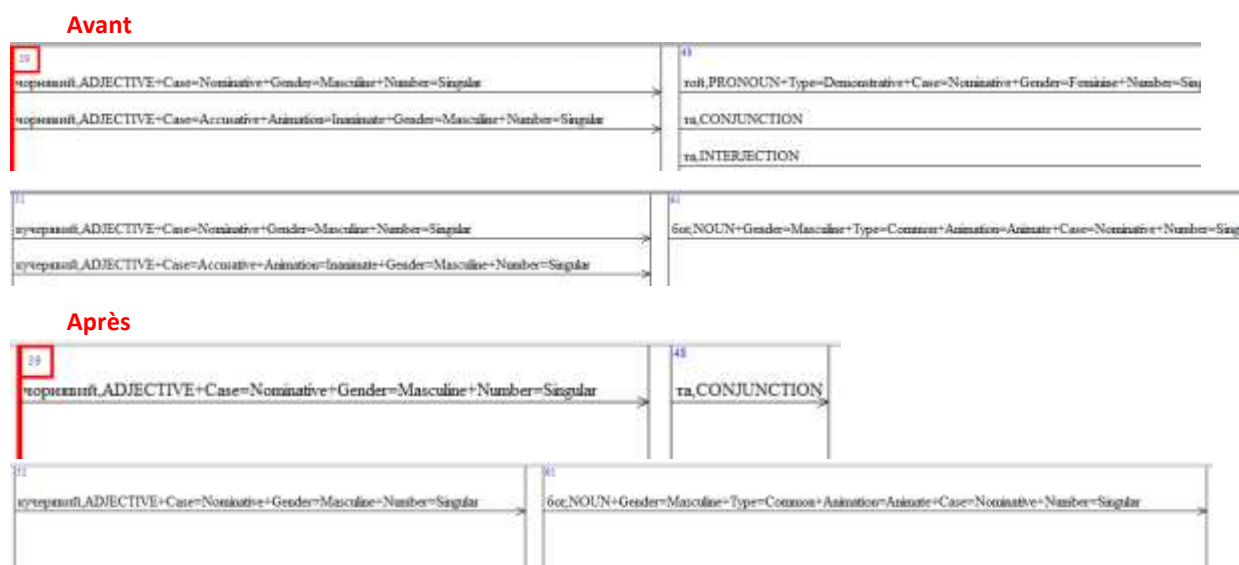
Le graphe « :i », présenté en Figure 74, décrit les membres périphériques et prend en compte aussi la virgule qui peut séparer les membres secondaires.

Évaluation de la grammaire « GN_GP_V.1.3.nog »

Cette grammaire nous a permis de lever de nombreuses ambiguïtés, car si le noyau n'est pas ambigu, alors le membre secondaire est accordé correctement. De plus, elle nous a permis de désambigüiser les membres périphériques²⁴¹ du GN et GP.

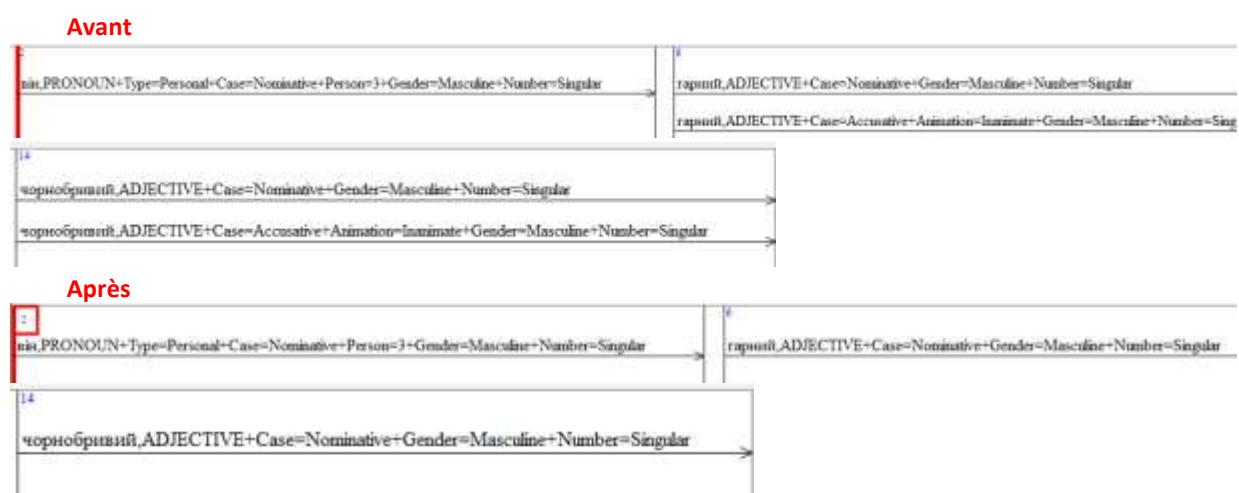
Dans la Figure 74, nous pouvons voir l'accord dans le GN *чорнявий та кучерявий бог* [le dieu aux cheveux bruns et bouclés] avant l'application de la grammaire et après son application : les deux adjectifs et la conjonction ont été annotés correctement.

Figure 74. Désambigüisation du GN *чорнявий та кучерявий бог*



Les résultats pour le GP *він гарний, чорнобривий* [il est beau, il a des sourcils noirs] sont présentés dans la Figure 75.

Figure 75. Désambigüisation du GP *він гарний, чорнобривий*



Avant d'appliquer notre grammaire au *Corpus 1*, celui-ci contenait 206 445 annotations et 10 209 différents types d'ambigüités. Après l'application de cette grammaire, le nombre

²⁴¹ Nous appelons ici « membres périphériques » la conjonction, la particule et l'adverbe.

d'annotations diminue de 27 031 et le nombre d'ambiguïtés différentes de 3 387 (il s'agit du nombre des différents couples homonymiques recensés dans le fichier « Ambiguities ». Nous pouvons voir les détails dans le Tableau 63.

Tableau 63. Évaluation de la grammaire « GN_GP_V.1.3.nog »

Nombre :	d'occurrences	d'annotation			de différents types d'ambiguïtés		
		avant	après	différence	avant	après	différence
Corpus 1	108 137	206 445	179 414	27 031	10 209	6 822	3 387

Désambiguïstation des cas

La première grammaire ne nous permet pas de lever les ambiguïtés si le noyau est ambigu. Pour traiter ce problème, nous avons construit sept grammaires supplémentaires : une pour chaque cas. Pour cela, nous avons préparé pour chaque cas un schéma dans lequel nous décrivons l'utilisation du GN ou GP au cas étudié. Pour chaque cas, nous avons appliqué les principaux emplois, décrits dans le chapitre 3.

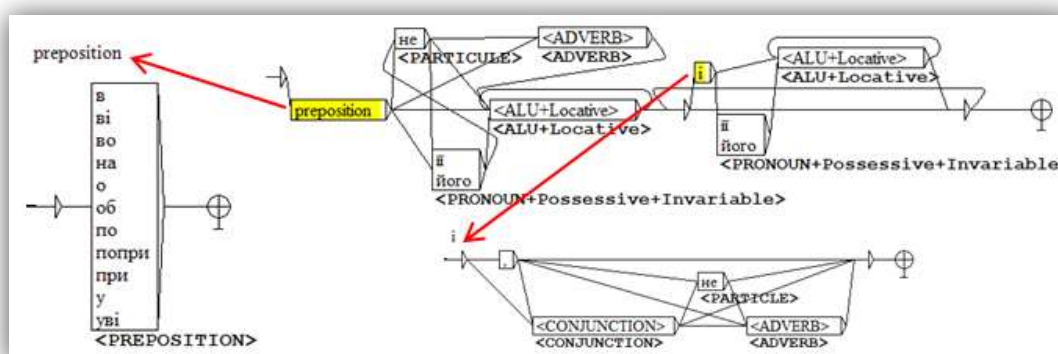
Dans ces grammaires, les séquences constituant le GN et le GP sont désignées par le symbole <ALU> qui se répète (*); le noyau n'est pas distingué des membres secondaires, car ces grammaires complètent la précédente. Elles contiennent également les membres périphériques (conjonction, adverbe, particule). L'élément important de chaque schéma (sauf pour le nominatif et vocatif) est la préposition, parce qu'elle permet de désambiguïser un grand nombre d'homonymes flexionnels.

Dans la Figure 76, nous pouvons voir les différents contextes du locatif. Ce schéma nous a permis de construire la grammaire « Locative_V.1.3.nog » présentée dans la Figure 77.

Figure 76. Description du locatif

1. PREPOSITION → ALU+Locative* ;
2. PREPOSITION → ALU+Locative* → “,” et/ou CONJUNCTION → ALU+ Locative* ;
3. PREPOSITION → PRONOUN+Possessive+Invariable et/ou PARTICLE « не » et/ou ADVERB → ALU+ Locative* ;
4. PREPOSITION → PRONOUN+Possessive+Invariable et/ou PARTICLE « не » et/ou ADVERB → ALU+ Locative* → “,” et/ou CONJUNCTION → PRONOUN+Possessive+Invariable et/ou particle « не » et/ou ADVERB → ALU+ Locative*.

Figure 77. La grammaire « Locative_V.1.3.nog »



La grammaire est composée de trois graphes imbriqués. Dans le graphe principal, nous avons décrit le lien entre l'ALU au locatif et les autres éléments du syntagme. Dans le graphe

« :preposition », sont listées les prépositions qui s'utilisent avec les ALU au locatif ; dans le graphe « :i », sont décrits les membres périphériques du syntagme (et la virgule).

La grammaire « Locative_V.1.3.nog » est assez simple car le locatif s'utilise uniquement après quelques prépositions, tandis que les grammaires qui décrivent les autres cas sont plus complexes puisqu'elles doivent prendre en compte plusieurs types de syntagmes sans préposition. Par exemple, pour les cas du datif présentés dans la Figure 78, nous avons ajouté plusieurs contextes supplémentaires.

Figure 78. Description du datif

1. PREPOSITION → ALU+Dative* = *успереч батькові [en dépit de son père]* ;
2. VERB → ALU+Dative* = *Я телефоную студентці [J'appelle une étudiante]* ;
3. VERB → ALU+Accusative* → ALU+Dative* = *Я дарую подарунок другові [J'offre un cadeau (Accusatif) à un ami]* ;
4. VERB → ALU+Accusative* → ALU+Genitive → ALU+Dative* = *Я даю номер телефону другові [Je donne mon numéro (Accusatif) de téléphone (Génitif) à un ami]* ;
5. ALU+Dative* → VERB+ся (phrase asubjectale) = *Студентці хочеться [L'étudiante a envie]* ;
6. ALU+Dative* → VERB+Asubjectal (phrase asubjectale) = *Мені суджено [Je suis destiné à...]* ;
7. ALU+Dative* → PREDICATIVE = *Студентці треба [L'étudiante a besoin]* ;
8. ALU+Dative* → Age = *Студентці 20 років [L'étudiante a 20 an]s* + question à propos de l'âge: PRONOUN « Скільки » → ALU+Dative* → Age → « ? » = *Скільки студентці років? [Quel âge a l'étudiante]* ;
9. ALU+Dative* → ADVERB (expression de l'état) = *Студентці погано [L'étudiante se sent mal]* ;
10. ALU+Dative* → VERB « бути » ou « стати » → ADVERB (expression de l'état) = *Студентці стало погано [L'étudiante s'est senti mal]* ;
11. ALU+Dative* → ADVERB → VERB+Infinitive = *Мені краще знати [Je sais mieux]*.

Figure 79. La grammaire « Dative_V.1.3.nog » : le graphe principal

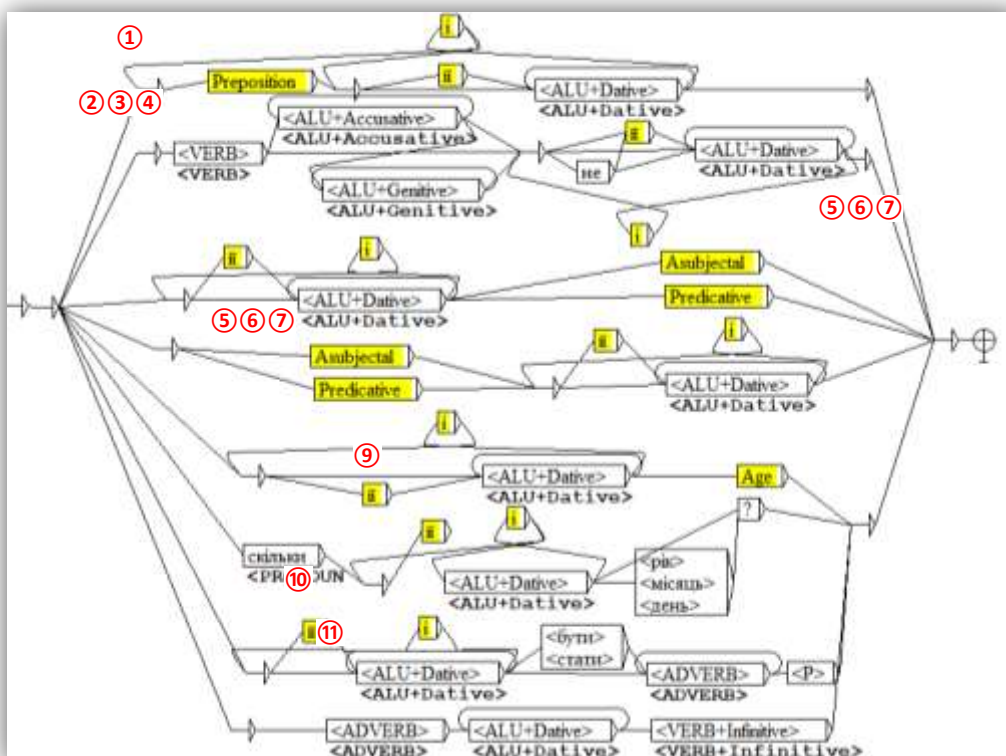
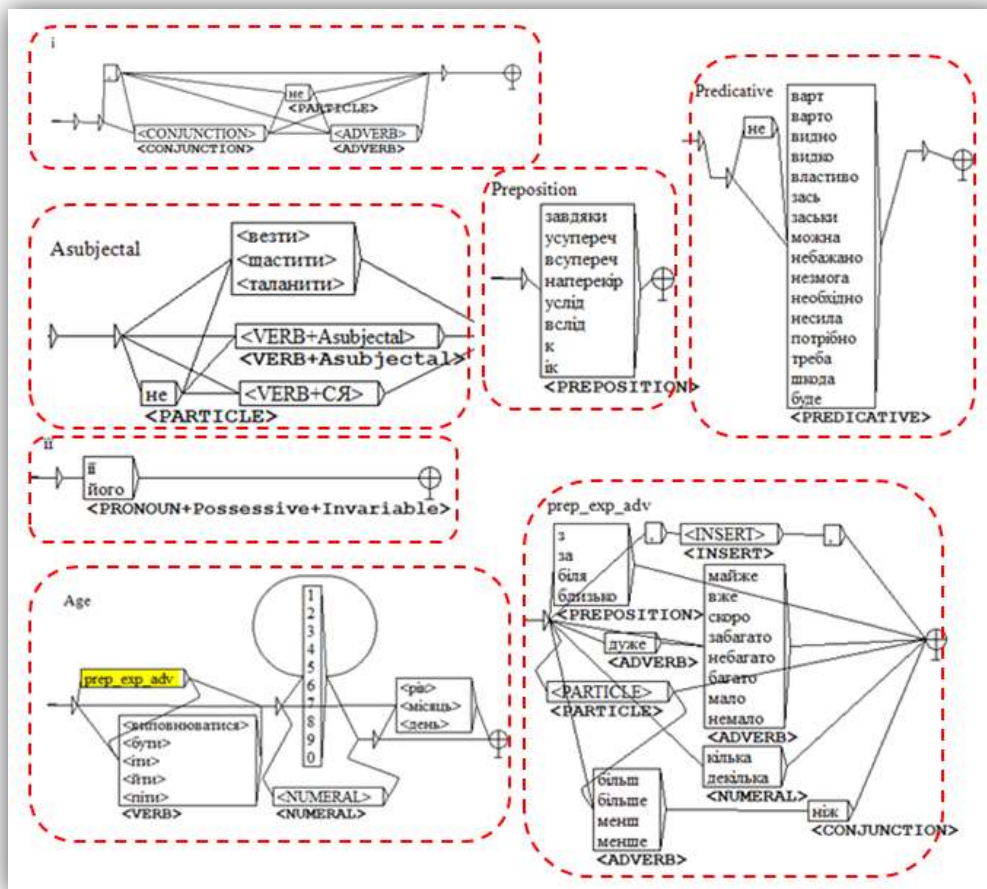


Figure 80. La grammaire « Dative_V.1.3.nog » : les graphes imbriqués



La grammaire « Dative_V.1.3.nog » décrit ces contextes. Le graphe principal avec les contextes numérotés est présenté dans la Figure 79 et les graphes imbriqués dans la Figure 80.

Nous n'avons pas décrit dans cette grammaire les phrases avec l'ordre libre par peur que notre grammaire n'entre en conflit avec les grammaires syntaxiques qui décrivent les ALU à l'accusatif et au génitif.

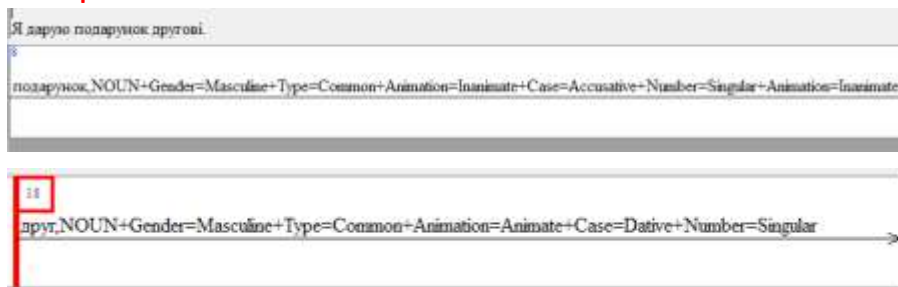
Nous pouvons voir la désambiguïsation des compléments « подарунок » et « друзові » du verbe « дарую » dans la phrase « Я дарую подарунок друзові [J'offre un cadeau à un ami] » dans la Figure 81.

Figure 81. Désambiguïsation des compléments *подарунок* et *друзові*

Avant

Я дарую подарунок друзові.
1 подарунок, NOUN+Gender=Masculine+Type=Common+Animation=Inanimate+Case=Nominative+Number=Singular
18 подарунок, NOUN+Gender=Masculine+Type=Common+Animation=Inanimate+Case=Accusative+Number=Singular+Animation=Inanimate
друз, NOUN+Gender=Masculine+Type=Common+Animation=Animate+Case=Dative+Number=Singular
друз, NOUN+Gender=Masculine+Type=Common+Animation=Animate+Case=Locative+Number=Singular

Après



Nous avons procédé de la même façon pour préparer les grammaires «Nominative_V.1.3.nog », « Acusative_V.1.3.nog », « Genitive_V.1.3.nog », « Instrumental_V.1.3.nog » et « Vocative_V.1.3.nog ».

Évaluation des grammaires de désambiguïsation des cas

Nous appliquons sept grammaires après la grammaire précédente, dans l'ordre suivant²⁴² :

1. « Dative_V.1.3.nog » ;
2. « Acusative_V.1.3.nog » ;
3. « Genitive_V.1.3.nog » ;
4. « Nominative_V.1.3.nog » ;
5. « Instrumental_V.1.3.nog » ;
6. « Locative_V.1.3.nog » ;
7. « Vocative_V.1.3.nog ».

Ces grammaires nous permettent de désambiguïser un grand nombre des homonymes flexionnels ; nous obtenons de bons résultats, puisque nous passons de 179 414 annotations (et 206 445 sans grammaire « GN_GP_V.1.3.nog ») à 133 487 annotations (voir Tableau 64).

Tableau 64. Évaluation des grammaires qui désambiguïsent les cas

Nombre :	d'occurrences	d'annotation			de différents types d'ambiguïtés		
		avant	après	différence	avant	après	différence
Corpus 1	108 137	179 414	133 487	45 927	6 822	4 802	2 020

De plus, ces grammaires permettent de désambiguïser un certain nombre des conjonctions, des adverbes, des pronoms et des numéraux indéterminés décrits dans nos grammaires.

Ces grammaires ne sont pas exhaustives, vu qu'à quelques exceptions près, elles ne prennent pas en compte l'ordre libre des mots. La plateforme NooJ offre les contraintes « ONE » et « ONCE », capables de traiter les séquences de mots avec ordre libre. Il sera donc intéressant d'exploiter ces grammaires et peut-être de les combiner avec les grammaires que nous avons construites, mais pour cela, nous aurons besoin de d'étudier de façon approfondie la syntaxe de la langue ukrainienne.

4.5.3 Désambiguïsation des homofomes et des homonymes transposés

Désambiguïsation des verbes

Comme nous l'avons souligné plus haut, les couples d'homofomes dont un membre est un verbe sont très fréquents. Par exemple, la forme graphique *mamu* peut être reconnue comme

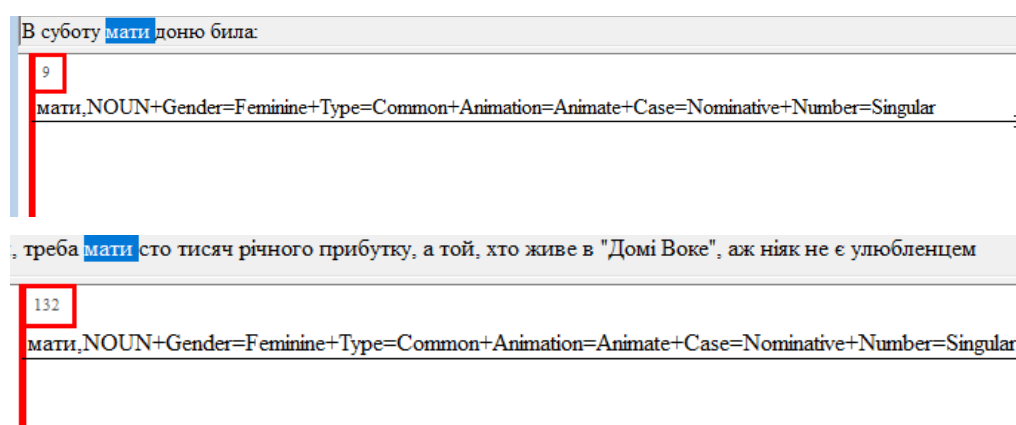
²⁴² L'ordre est égal à la priorité de l'exécution d'une grammaire.

Cette structure est décrite par la grammaire « GV.1.3.nog ». Elle est composée de sept graphes. Dans la Figure 83, nous pouvons voir le graphe principal, le graphe « :Past » et les graphes imbriqués dans ce dernier. Le graphe « :ALU_plus » décrit les compléments sans préposition sans les désambigüiser, car leur désambigüisation sera effectuée par les grammaires qui traitent les cas que nous avons décrites plus haut.

Évaluation de la grammaire « GV.1.3.nog »

Nous appliquons cette grammaire en premier (Nool permet d'automatiser l'application de grammaires en les ordonnant). Après le traitement du *Corpus 1*, nous constatons que la plupart de homoformes verbales ont été désambigüisés. Dans la Figure 84, nous pouvons voir deux exemples de notre corpus de test. Dans la première phrase *В суботу мати доню біла* [samedi, la **mère** battait sa fille], la forme graphique *мати* a bien été reconnue comme nom au nominatif, et dans le syntagme *треба мати сто тисяч річного прибутку*, а той, хто живе в "Домі Воке", аж ніяк не є улюбленцем [il faut **avoir** cent mille], comme verbe.

Figure 84. Annotation de l'ALU *мати*



Par ailleurs, il nous faut trouver des solutions pour y inclure le sujet inversé sans qu'il ne soit confondu avec le complément qui a les mêmes valeurs que le sujet²⁴³. Pour l'instant, ces ALU peuvent être associées à plusieurs annotations concurrentes. De plus, il existe des phrases dépourvues de complément, comme la phrase *Ой мати, мату!*, tirée de notre corpus. Cette phrase peut s'interpréter comme [Et oui ! Il faut avoir ! Il faut avoir !] ou comme [Ma mère, ma mère !]. Notons cependant que cette phrase est ambiguë aussi pour un humain sans contexte.

Cette grammaire nous a permis de supprimer 1 078 annotations dans nos textes. Nous pouvons voir les résultats dans le Tableau 65.

Tableau 65. Évaluation de la grammaire « GV.1.3.nog »

Nombre :	d'occurrences	d'annotation			de différents types d'ambiguïtés		
		avant	Après	différence	avant	après	différence
<i>Corpus 1</i>	108 137	133 487	132 409	1 078	4 802	4 713	89

²⁴³ Pour rappel, certains noms qui s'utilisent en tant que compléments du verbe sans préposition, par exemple, le masculin inanimé à l'accusatif, ont les mêmes terminaisons que le nom au nominatif qui s'utilise en tant que sujet.

Désambiguïisation des adverbes

Bien que certains adverbes soient correctement annotés et désambiguïsés grâce aux grammaires que nous avons décrites plus haut, certains restent ambigus, comme c'est le cas de l'adverbe *дуже* [très] qui peut aussi être analysé comme un adjectif neutre au nominatif ou à l'accusatif inanimé.

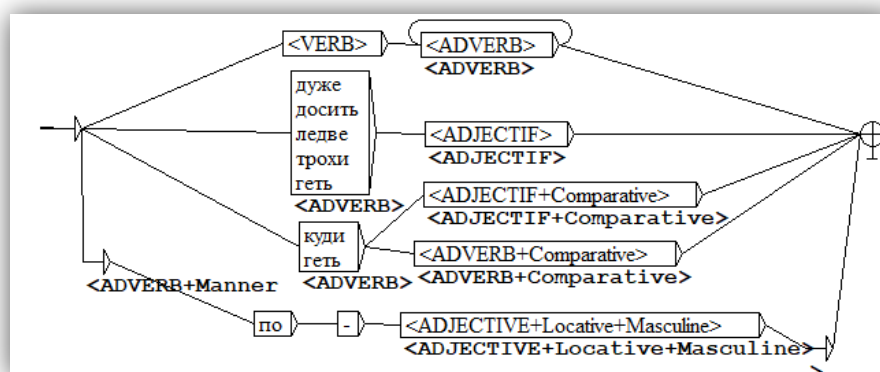
De plus, il existe des adverbes de manière qui sont créés par la contraction de la préposition *по* et d'un adjectif masculin au locatif²⁴⁴ et qui s'écrivent avec le trait d'union. De nombreux adverbes de ce type sont inscrits dans notre dictionnaire avec la propriété « UNAMB » pour interdire à NooJ d'analyser leurs constituants comme des unités indépendantes. Ceux qui ne sont pas décrits dans le dictionnaire sont alors analysés comme des séquences de deux ALU indépendantes et ambiguës, comme c'est le cas de *по-злому* [méchamment]. Nous pouvons voir l'exemple dans la Figure 85.

Figure 85. L'ALU *по-злому* avant l'application de la grammaire « Adverbs_V.1.3.nog »

131	по,PREPOSITION	134	злом,NOUN+Gender=Masculine+Type=Common+Animation=Inanimate+Case=Genitive+Number=Singular
			злом,NOUN+Gender=Masculine+Type=Common+Animation=Inanimate+Case=Dative+Number=Singular
			злом,NOUN+Gender=Masculine+Type=Common+Animation=Inanimate+Case=Locative+Number=Singular
			злий,ADJECTIVE+Case=Dative+Gender=Masculine+Number=Singular
			злий,ADJECTIVE+Case=Locative+Gender=Masculine+Number=Singular
			злий,ADJECTIVE+Case=Dative+Gender=Neuter+Number=Singular
			злий,ADJECTIVE+Case=Locative+Gender=Neuter+Number=Singular

Pour que ces ALU soient reconnues et annotées correctement, nous avons construit la grammaire « Adverbs_V.1.3.nog » présentée dans la Figure 86.

Figure 86. La grammaire syntaxiques « Adverbs_V.1.3.nog »



Nous décrivons ici les adverbes qui sont utilisés après les verbes, les adverbes de manière, mais aussi les adverbes d'intensité, qui sont utilisés avec les adjectifs et peuvent former les couples d'homofomes.

²⁴⁴ Notons que les grammairiens ukrainiens considèrent que le cas qui est utilisé après la préposition « *по* » est le locatif.

Évaluation de la grammaire « Adverbs_V.1.3.nog »

Nous avons appliqué la grammaire « Adverbs_V.1.3.nog » lors de l'analyse linguistique de notre corpus de test. Les résultats sont présentés dans le Tableau 66. Cette grammaire a permis de lever les ambiguïtés de certains adverbes et d'annoter d'autres adverbes qui n'étaient pas décrits comme tels dans nos dictionnaires. Dans la Figure 87, nous pouvons voir comment est annotée l'ALU *по-злому* [méchamment].

Figure 87. L'ALU *по-злому* après l'application de la grammaire « Adverbs_V.1.3.nog »

131	ADVERB+Manner	→	134	
по,PREPOSITION	→		злий,ADJECTIVE+Case=Locative+Gender=Masculine+Number=Singular	→

Notons que le nombre d'annotations supprimées n'est pas très grand (voir Tableau 66), mais cette grammaire est très utile en analyse du discours lorsqu'on a besoin de retrouver tous les syntagmes qui contiennent les adverbes décrits par cette grammaire.

Tableau 66. Évaluation de la grammaire

Nombre :	d'occurrences	d'annotation			de différents types d'ambiguïtés		
		avant	après	différence	avant	après	différence
<i>Corpus 1</i>	108 137	132 409	131 415	994	4 713	4 702	11

4.5.4 Évaluation de l'ensemble des grammaires syntaxiques et perspectives

Nous avons construit 10 grammaires syntaxiques qui permettent de lever 75 030 ambiguïtés, puisque nous passons de 206 445 annotations à 131 415 actuellement, pour un corpus de 108 137 occurrences. Nous avons désambiguïté un grand nombre d'homonymes flexionnels liés à des terminaisons homonymiques dans les déclinaisons des noms, des adjectifs, des participes, des numéraux et des pronoms. Nous avons partiellement levé des ambiguïtés dans les couples d'homofomes, notamment dans les couples dans lesquels un membre est un verbe ou un adverbe. Nos grammaires permettent aussi de désambiguïser les prépositions et quelques cas d'utilisation des particules et des conjonctions.

Parmi les ambiguïtés fréquentes qui n'ont pas été levées, nous comptons des conjonctions, plus particulièrement, celles qui sont composées de quelques unités disjointes (non connexes), les particules, les interjections et les inserts. Avant de nous lancer dans la construction de grammaires pour les désambiguïser, nous devons faire une étude de la syntaxe ukrainienne et préparer les tableaux comparatifs pour classer les différents contextes concernés.

Nous devons également compléter les grammaires qui traitent des cas pour prendre en compte l'ordre libre des mots. Toutefois, l'annotation est déjà satisfaisante pour se lancer dans les études de documents.

4.6 Utilisations pratiques envisagées

Plus de 500 articles sur les diverses applications de la plateforme NooJ sont référencés à la page <https://nooj.univ-fcomte.fr/references.html>. En particulier, les articles suivants²⁴⁵ ont été pertinents pour notre travail :

- l'enseignement assisté par ordinateur (EAO) par Tutin et Silberztein (2004 [148]), Piton (2004 [122], 2008 [123]), Watabe (2018a [151], 2018b [152]), Georgiadou et Papaioannou (2011 [99]), Fehri et Ben Messaoud (2020 [97]), Rodrigo, Méndez et Oliva (2021 [126]), Rodrigo et Harriet (2022 [110]) ;
- les variantes d'une langue, les langues proches : le français oral régional (Chevalier, Kasparian, Silberztein 2004 [81]), (Kasparian 2008 [108]), l'anglais du XVII^e siècle (Pignot et Lardy 2013 [121]) ; les variations lexicographiques en albanais contemporain (Piton et Lagji 2008 [123], Piton et Pěrnaska 2013 [124]), détection des langues proches (Bekavac, Kocijan et Tadić, 2015 [74]), le dictionnaire du romani Watabe (2018a [151]) ;
- les études linguistiques autour des langues slaves : le biélorussien (Hetsevich 2014 [106]), le croate (Kocijan, Janjić, Librenjak 2017 [111]), (Merkler, Berović et Agić 2012 [117]), (Šojat, Kocijan et Filko, 2022 [114]), le polonais (Gwiazdecka 2009 [102], Gwiazdecka et Bogacki 2010, 2012 2013 [103 à 105], Bogacki et Dryjańska 2019 [100]), le russe (Bénet 2013 [94]) ;
- la reconnaissance des entités nommées par Vitas et Pavlović-Lažetić (2007 [149]), Mesfar (2007 [118]), Stoyanova et Lesseva (2008 [142]), Boujelben, Mesfar et Ben Hamadou (2010 [79]), Haddar et Ben Hamadou (2011 [95], 2014 [96]), Khachatryan (2012 [109]), Fehri et Aouini (2015 [72]), Fall Seideh, Fehri et Haddar (2017 [133]), Kocijan, Pia di Buono et Mikić (2019 [112]), Alves, Bekavac et Tadić (2021 [71]) ;
- les études du discours : la représentation des femmes dans les textes de journaux serbes (Gucul-Milojević, Radulović et Krstev 2010 [115]), l'étude du discours de l'équipage de Mars-500 (Ehmann, Balázs, Shved, Bénet et Gushin 2013 [94]), l'analyse des sentiments (Pelosi, Alessandro Maisto 2015 [120], Cirillo 2020 [91]), le discours haineux en ligne (Kocijan, Košković, Bajac 2020 [113]), l'usage de la négation au Parlement croate (Kocijan, Šojat 2022 [114]).

La version actuelle du module ukrainien pour NooJ nous offre plusieurs possibilités d'application pratique, aussi bien en enseignement assisté par ordinateur que dans le domaine des recherches en linguistique ou en TAL.

Nous allons présenter quelques applications pratiques que nous pourrions explorer dans le futur.

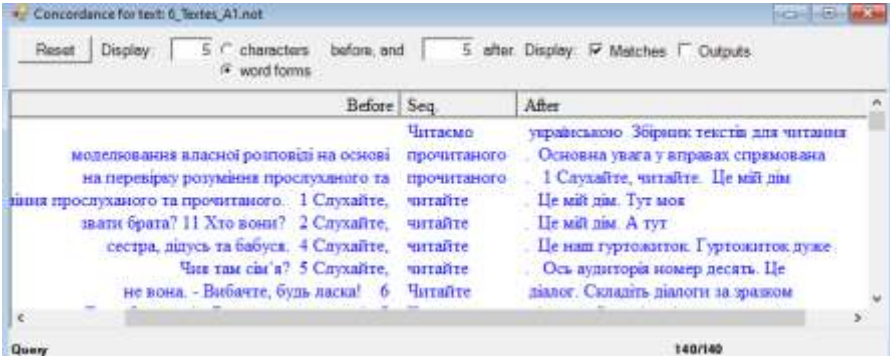
4.6.1 Utilisation pédagogique

Nous pouvons utiliser la plateforme NooJ en cours de grammaire pratique de la deuxième année de la licence d'ukrainien que nous assurons à l'INALCO. Par exemple, nous pouvons proposer à nos étudiants des exercices autour d'un texte dans lequel nous demanderons de trouver certains types d'ALU, puis proposer de faire leur analyse morphologique et de construire des phrases avec les ALU en question. Le Tableau 67 présente un modèle d'exercice qui consiste à faire le travail d'apprentissage des couples aspectuels *читати* (IPF) / *прочитати* (PF) [*lire*], *вивчати* (IPF) / *вивчити* (PF) [*apprendre*], *писати* (IPF) / *написати* (PF) [*écrire*], *повторювати* (IPF) / *повторити* (PF) [*répéter*], *відпочивати* (IPF) / *відпочити* (PF) [*se reposer*]. Ce travail peut être

²⁴⁵ Vu le grand nombre de ces articles, nous ne citons ici que ceux qui ont attiré notre attention en vue de nos futurs travaux.

effectué en utilisant la collection des textes niveau A1-A2 *Читаємо українською [Lisons en ukrainien] (2005)*²⁴⁶ préparée par Oksana Maliarenko et Alina Chevtsova, enseignantes à l'Université de Soumy.

Tableau 67. Exemple d'exercice pédagogique

Instruction	Application
<p>1. Trouver toutes les formes du verbe <i>читати</i>. Combien de fois ce verbe a-t-il été utilisé ?</p>	<p>Effectuer l'analyse linguistique du texte (clic droit, puis commande « Linguistic Analysis », puis rechercher le lemme <читати> (clic droit, puis commande « Locate Pattern », « a NooJ regular expression », cocher « All occurrences »).</p>  <p>Le nombre d'occurrences trouvées est 140.</p>
<p>2. Relever le verbe PF qui forme le couple aspectuel.</p>	<p>A. Afficher le concordancier, puis vérifier les formes qui ont été trouvées.</p>  <p>B. Ensuite, afficher ces formes dans le texte (clic droit, puis commande « Color matching sequence in text »).</p> <p>C. Afficher les annotations de la première forme trouvée (cocher « Show Text Annotation » et cliquer sur la forme coloriée).</p>

²⁴⁶ Le manuel est disponible sur <https://essuir.sumdu.edu.ua/handle/123456789/1457>

Instruction	Application
	 <p>La forme est <i>читаємо</i> [<i>lisons</i>] (« le verbe IPF à la première personne du pluriel au présent). Le couple aspectuel correspondant est <i>читати/прочитати</i>.</p>
<p>3. Combien de fois le verbe IPF a-t-il été utilisé ? Combien de fois le verbe PF a-t-il été utilisé ?</p>	<p>Rechercher le lemme <читати+Perfective> (clic droit, puis commande « Locate Pattern », « a NooJ regular expression », cocher « All occurrences »), puis par lemme <читати+Imperfective>.</p> <p>Trois verbes PF et 137 IPF ont été trouvés.</p>
<p>4. Refaire le même travail pour les verbes <i>вивчати</i> [<i>apprendre</i>], <i>писати</i> [<i>écrire</i>], <i>повторювати</i> [<i>répéter</i>], <i>відпочивати</i> [<i>se reposer</i>]</p>	<p>Mêmes opérations que dans les étapes décrites plus haut.</p>
<p>5. Récrire le texte 27 à partir de la phrase « Студенти вивчають ... » en remplaçant les verbes imperfectifs donnés dans l'étape 4, par les verbes perfectifs.</p>	<p>A. Entrer l'expression : <читати> <вивчати> <писати> <повторювати> <відпочивати>.</p> <p>B. Colorer toutes les formes dans le texte.</p>  <p>C. Réécrire le texte en remplaçant le verbe IPF par PF conjugué ; attention, d'autres verbes, en plus des verbes étudiés, doivent être mis au PF. Comment le sens du texte a-t-il été modifié ?</p> <p>D. Réécrire le texte au passé imperfectif et au passé perfectif. Comment le sens du texte a-t-il été modifié ?</p>

Cet exercice peut être fait en salle informatique ou bien être donné en devoir à la maison.

À notre avis, cet exercice permettra aux étudiants de bien retenir le couple aspectuel des verbes étudiés et d'apprendre à utiliser les verbes en question.

Nous pouvons préparer des exercices similaires pour étudier le comparatif, le superlatif et le superlatif absolu, ou encore pour étudier les formes diminutives des noms, les adjectifs et les adverbes d'expression, etc.

même que la langue de diaspora ou le « sourjyk », comme nous l'avons expliqué dans la section 1.3.8.

NooJ nous permet de décrire ces variantes dans un dictionnaire. Pour cela, nous devons relever les particularités de la variante dans la grammaire de référence source, des grammaires historiques ou des articles récents. Par exemple, Ivan Matviias décrit avec précision les spécificités de « jelekhivka » (Matviias 2011 [41]). En nous basant sur ces recherches, nous pouvons construire un premier dictionnaire, en remplaçant :

- la lettre « i » par « ї », c'est-à-dire *ліс* au lieu de *ліс* ;
- le suffixe *-ськ-* par *-ск-*, c'est-à-dire *сільський* au lieu de *сільський* ;
- et en insérant le signe mou entre les consonnes « c » et « m », c'est-à-dire *сьмішно* au lieu de *смішно* [drôle].

Cette opération peut se faire rapidement et n'exige pas de remplacement du paradigme pour les deux ALU variables. Celles-ci pourront être décrites dans le dictionnaire de cette façon :

ліс,ліс,NOUN+Masculine+Common+Inanimate+FLX=ЛИСТ+"jelekhivka"

сільський,сільський,ADJECTIVE+FLX=ВЕЛИКИЙ+"jelekhivka"

сьмішно,смішно,ADVERB+Manner+"jelekhivka"

Par ailleurs, certaines autres modifications entraîneront une modification des modèles flexionnels. Par exemple, la mouillure des consonnes « ч » et « р » à la fin des noms masculins au nominatif, comme dans *плач* [le pleur] *поштарь* [le facteur]. Par ailleurs, dans ce cas, nous n'avons pas besoin de formaliser de nouveaux paradigmes : nous pouvons réutiliser les paradigmes existants, élaborés pour les noms masculins en consonne molle.

Parfois, nous serons amenée à formaliser un nouveau paradigme (ou à modifier le paradigme existant), comme dans le cas des verbes qui se terminent par une labiale suivie du suffixe *-ити* à l'infinitif. Par exemple, *любити* [aimer] se conjugue selon les normes de « jelekhivka » *люб'ю*, *люб'ять* au lieu de *люблю* [j'aime], *люблять* [ils aiment].

Pour analyser les formes absentes du dictionnaire, nous pouvons également réutiliser les grammaires morphologiques construites pour l'ukrainien standard. Par exemple, dans « jelekhivka », pour former les noms et les adjectifs diminutifs, les suffixes *-ейк-* et *-ойк-* ont été utilisés à la place des suffixes *-еньк-* et *-оньк-*, c'est-à-dire *малейкий* au lieu de *маленький* [petit] et *головоїка* au lieu de *головонька* [tête]. Dans notre module, ces ALU ont été décrites comme des formes dérivées grâce aux grammaires présentées dans les Figures 43 et 49. Nous pouvons adapter alors ces grammaires pour « jelekhivka ». De même, la grammaire présentée dans la Figure 62, qui décrit les alternances au début des mots, pourrait être modifiée pour remplacer les consonnes « хв » par « ф » ou « х », comme dans *фіст* au lieu de *хвіст* [queue] et *хороба* au lieu de *хвороба* [maladie].

Lors de l'analyse linguistique, le nouveau dictionnaire et les nouvelles grammaires auront à une priorité plus élevée que celle affectée aux ressources utilisées pour analyser des textes en ukrainien standard.

Après l'analyse morphologique des textes écrits en « jelekhivka », nous pourrions étudier les ALU non reconnues et repérer d'autres régularités, non décrites dans des articles ou les grammaires publiées aujourd'hui.

Grâce à des grammaires syntaxiques, nous pourrions également étudier certaines structures syntaxiques particulières, comme l'expression de l'âge au locatif, décrite dans la section précédente,

ou l'utilisation du cas après la préposition « *но* » : est-ce le locatif ou le datif ? Cette question n'est pas tranchée actuellement.

De la même façon, nous pourrions construire un module ukrainien de la diaspora ou un module « sourjyk ». Dans ce dernier, il faudra adapter les ressources ukrainiennes et russes, avec la permission de l'auteur, Vincent Bénet. Enfin, ces deux modules seront aussi très intéressants pour étudier les sources en provenance de d'Internet.

4.7 Conclusion du chapitre 4

Nous avons construit un module ukrainien pour NooJ qui est composé d'un dictionnaire principal « Ukr_dictionary_V.1.3 » et de deux dictionnaires secondaires « Ukr_dictionary_Participe_V.1.3 » et « Ukr_dictionary_Proper_lowercase_V.1.3 ». Le dictionnaire principal contient 157 534 entrées et reconnaît 3 184 522 formes fléchies. Il décrit des ALU simples, composées d'une seule forme graphique, mais aussi des locutions composées de deux formes ou plus ; il reconnaît et analyse les ALU avec orthographes alternatives, et explicite les abréviations.

Les formes fléchies des entrées variables sont formalisées grâce à 303 paradigmes flexionnels. Nous avons formalisé également 114 paradigmes dérivationnels qui permettent de lier les verbes perfectifs aux verbes imperfectifs.

Nous avons décrit de nombreuses formes dérivées ou les variantes orthographiques absentes du dictionnaire grâce aux 19 grammaires morphologiques.

Enfin, nous avons recensé certaines formes dans les dictionnaires secondaires, notamment les participes et les noms propres en minuscules. Le dictionnaire « Ukr_dictionary_Participe_V.1.3 » contient 13 070 entrées et complète le dictionnaire principal, quand la grammaire morphologique qui décrit des participes ne permet pas de reconnaître le participe dans le texte. Le dictionnaire « Ukr_dictionary_Proper_lowercase_V.1.3 » contient des noms propres écrits en minuscules. En combinaison avec la grammaire « Adjectives_Relatives_V.1.3.nom », il permet de reconnaître les adjectifs relatifs créés à partir des noms propres.

Grâce à ces ressources, 98,3 % d'occurrences dans le corpus de tests ont été reconnues et annotées avec leurs informations morphologiques correctes.

Nous avons également construit dix grammaires syntaxiques qui permettent de lever un grand nombre d'ambiguïtés, puisque nous passons de 206 445 annotations à 131 415 pour un corpus de 108 137 occurrences.

Nous avons également mis en évidence quelques pistes de travaux pour l'avenir afin d'améliorer notre module, à savoir l'élaboration de nouvelles grammaires morphologiques supplémentaires et de grammaires syntaxiques qui permettront de lever les ambiguïtés restantes.

Nous avons aussi réfléchi sur les cas pratiques d'utilisation du module ukrainien et, notamment en tant qu'outil d'EAO qui permettra aux étudiants d'apprendre certaines notions difficiles, comme l'aspect du verbe ou encore le degré de comparaison. Nous pourrions aussi utiliser ces ressources pour étudier les variantes de la langue ukrainienne qui, à notre connaissance, ne sont décrites par aucun outil de TAL.

Conclusion

Nous avons entrepris ce projet pour décrire la langue ukrainienne d'une part et préparer un outil capable d'analyser correctement les textes ukrainiens. Pour cela, nous avons étudié non seulement la grammaire, mais aussi l'histoire de la langue ukrainienne.

Dans le deuxième chapitre, nous avons dressé un aperçu des outils existants. Nous avons souligné que les outils issus de l'approche empirique sont actuellement majoritaires sur le marché. Certains sont capables d'analyser les textes ukrainiens ; On compte parmi eux Sketch Engine, TreeTagger et RNNTagger. Ils affichent de bons résultats (reconnaissance de 97 % des mots sur l'ensemble des textes disponibles). Mais, comme nous l'avons vu, cette reconnaissance ne se fait pas correctement. Tout cela est dû au fait que la langue ukrainienne est très complexe et que plusieurs phénomènes linguistiques (par exemple l'alternance de voyelles, ou encore les voyelles mobiles, les dérivations ou les orthographes alternatives) ne peuvent pas être couverts par ces outils. L'ensemble de ces erreurs compromet les statistiques et toute autre analyse construite grâce à ces outils.

Pour réaliser une analyse fiable, il vaut mieux avoir recours aux outils formels. Cela a été l'une de nos principales motivations pour nous lancer dans la construction du module ukrainien pour NooJ, qui, en outre, offre une grande souplesse, car cette plateforme permet de construire et de combiner les grammaires des quatre types de la hiérarchie de Chomsky -Schützenberger. De plus, avec NooJ, nous pouvons associer les mots avec leurs variantes, décrire les locutions, les abréviations et les formes dérivées.

Cela nous a permis de construire un module ukrainien pour NooJ composé de trois dictionnaires, 19 grammaires morphologiques et 10 grammaires syntaxiques.

Le dictionnaire principal contient 157 534 entrées et reconnaît 3 184 522 formes fléchies, grâce à 303 paradigmes flexionnels et 114 paradigmes dérivationnels, et contient des locutions, des abréviations et des ALU avec leurs orthographes alternatives.

Nos grammaires morphologiques sont capables d'analyser de nombreuses formes dérivées absentes du dictionnaire. Grâce à l'ensemble des dictionnaires, 98,3 % des occurrences dans le corpus de tests, composé de diverses sources, ont été reconnues et annotées avec leurs informations morphologiques.

Toutefois, nos grammaires ne sont pas complètes, car il manque un certain nombre de formes dérivées, notamment les adjectifs composés. En préparant quelques grammaires supplémentaires, nous pouvons augmenter la couverture du lexique.

Quant aux grammaires syntaxiques, elles permettent de lever de nombreuses ambiguïtés (nous passons de 206 445 annotations à 131 415 pour un corpus de 108 137 occurrences), mais certains problèmes restent non résolus, et notamment, de nombreuses conjonctions, des particules, des interjections, inserts, mais aussi quelques noms, adjectifs et verbes restent ambigus. Étudier la syntaxe de façon plus approfondie et préparer les grammaires supplémentaires sera l'un de nos projets dans l'avenir.

Par ailleurs, dans le premier chapitre, nous avons pu constater que la langue ukrainienne a beaucoup évolué au cours de l'histoire : depuis les deux derniers siècles, les normes orthographiques ont changé à de multiples reprises. Pendant la période soviétique, les textes écrits au XIX^e ont été réécrits selon les normes en vigueur. Mais aujourd'hui, l'intérêt pour ces œuvres est accru : les chercheurs s'intéressent aux textes originaux, leur vocabulaire, leurs particularités morphologiques et syntaxiques. Nous pensons qu'il serait utile de compléter les ressources que nous avons

construites avec un module qui serait capable de reconnaître ces anciennes orthographes, au moins les plus répandues, à savoir « jelekhivka » et « koulichivka ».

Pour cela, nous devons préparer un dictionnaire dans lequel nous pourrions lier ces anciennes formes aux formes actuelles. Nous serons amenée également à préparer un certain nombre de paradigmes spécifiques. Ce module permettra d'étudier le vocabulaire des textes anciens et leur morphologie.

De la même façon, nous pourrions orienter nos recherches pour étudier les variantes contemporaines, c'est-à-dire la langue de la diaspora et le « sourjyk ».

En préparant ces trois modules complémentaires, nous pourrions créer une étiquette spécifique ; ainsi, lorsque nous analyserons un texte qui contient les éléments de tel ou tel module, NooJ sera capable de déterminer de quelle variante il s'agit.

Enfin, nous pourrions aussi utiliser le module pour des études plus concrètes : en tant qu'outil pédagogique ou en tant qu'outil d'analyse du discours.

Liste des figures

Figure 1. Carte allemande de l'Ukraine de 1918	21
Figure 2. Carte postale belge des années 30 représentant une carte de l'Ukraine	25
Figure 3 : Évolution des frontières d'Ukraine entre 1922 et 1954	33
Figure 4 : Erreurs liées à la lemmatisation dans Sketch Engine	49
Figure 5. Texte, pris pour tester les performances des outils	50
Figure 6. Erreurs d'étiquetage morphologique dans Sketch Engine	50
Figure 7. Étiquetage des locutions <i>кінець кінцем</i> et <i>нога в ногу</i> dans Sketch Engine	51
Figure 8. Hiérarchie Chomsky-Schützenberger	58
Figure 9. Recherche du mot <i>музей</i> dans GRAC-11, interface KonText.....	61
Figure 10. Recherches du mot <i>проект</i> (l'orthographe de 1992) dans GRAC-16, interface NoSketch Engine.....	62
Figure 11. Recherches du mot <i>проект</i> (l'orthographe de 2019) dans GRAC-16, interface NoSketch Engine.....	62
Figure 12. Analyseur de texte LanguageTool	63
Figure 13. Annotation de l'abréviation <i>AEC</i> par NooJ	66
Figure 14. Annotation des locutions <i>кінець кінцем</i> et <i>нога в ногу</i> par NooJ	66
Figure 15. Expressions régulières (PERL)	67
Figure 16. Recherche par expression rationnelle (a) dans NooJ et le résultat de recherche (b).....	67
Figure 17. Grammaire flexionnelle (a) dans NooJ et l'annotation (b) dans le texte qui en résulte	68
Figure 18. Graphe à l'état fini dans NooJ.....	68
Figure 19. Graphe morphologique à production (a) dans NooJ et l'annotation (b) dans le texte qui en résulte	69
Figure 20. Graphe syntaxique à production (a) dans NooJ et l'annotation (b) dans le texte qui en résulte.	69
Figure 21. Grammaire avec des graphes imbriqués	69
Figure 22. Grammaires hors-contexte dans NooJ : éditeur de règles (a) et éditeur graphique (b). ...	70
Figure 23. Alternative à la règle principale.....	70
Figure 24. Grammaire hors-contexte représentée graphiquement.....	71
Figure 25. Grammaire contextuelle dans NooJ	71
Figure 26. Exemple de grammaire non restreinte NooJ	72
Figure 28. Exemple du paradigme verbal.....	111
Figure 29. Recherches par étiquettes de NooJ.....	130
Figure 30. Affichage des étiquettes dans les annotations du texte	130
Figure 31. Description des catégories NooJ dans le module ukrainien.....	131
Figure 32. Description de l'adjectif.....	131
Figure 33. Description des participes en tant que catégorie	135
Figure 34. Exemple d'un paradigme FLX nominal dans NooJ	137
Figure 35. Extrait du paradigme « <i>ВЕЛИКИЙ</i> »	143
Figure 36. Grammaire morphologique pour formaliser des participes	157
Figure 37. Annotation du participe <i>виючий</i>	157
Figure 38. Grammaire morphologique pour décrire des verbes en « <i>-ся</i> »	159
Figure 39. Annotation du verbe <i>читається</i>	159
Figure 40. Annotation du verbe <i>сміється</i>	160
Figure 41. Grammaire qui formalise les formes verbales orales.....	160

Figure 42. Annotation de la forme orale <i>читать</i>	161
Figure 43. Grammaire morphologique pour formaliser les diminutifs des noms	162
Figure 44. Annotation de la forme diminutive <i>бантик</i>	162
Figure 45. Grammaire qui formalise les adjectifs d'appartenance	163
Figure 46. Annotation de l'adjectif d'apparence <i>Оксанин</i>	164
Figure 47. Grammaire qui formalise les adjectifs relatifs	165
Figure 48. Annotation de l'adjectif <i>алтайський</i>	166
Figure 49. Grammaire qui formalise les adjectifs partitifs, excessifs, diminutifs.....	167
Figure 50. Annotation de l'adjectif <i>малуватий</i>	167
Figure 51. Grammaire qui formalise les adverbes qualitatifs nuancés	168
Figure 52. Annotation de l'adjectif <i>малувано</i>	168
Figure 53. Grammaire qui formalise les degrés de comparaison pour les adjectifs.....	169
Figure 54. Grammaire qui formalise les degrés de comparaison pour les adverbes.....	170
Figure 55. Annotation de l'adjectif <i>гарніший</i>	170
Figure 56. Grammaire qui formalise les adjectifs courts	171
Figure 57. Annotation de l'adjectif court <i>зелен</i>	171
Figure 58. Grammaire qui formalise les formes négatives	172
Figure 59. Annotation de l'adjectif négatif <i>негарний</i>	172
Figure 60. Grammaire qui formalise les adverbes et les pronoms indéfinis et négatifs	173
Figure 61. Annotation du pronom <i>хтось</i>	173
Figure 62. La grammaire qui formalise des alternances au début des mots.....	175
Figure 63. Annotation du nom <i>учитель</i>	175
Figure 64. Grammaire formalisent des mots orthographiés en alphabet latin et en cyrillique russe	176
Figure 65. Extrait de la grammaire qui formalise des chiffres romains.....	177
Figure 66. Annotation des ALU <i>VII, français, мама</i>	177
Figure 67. Grammaire qui formalise des noms propres	178
Figure 68. Annotation du nom propre <i>Анвар</i>	178
Figure 69. Le paradigme nominal « СИН »	182
Figure 70. Annotations ambiguës	182
Figure 71. Composition d'un GN ou un GP.....	183
Figure 72. Grammaire principal de la grammaire syntaxique « GN_GP_V.1.3.nog »	183
Figure 73. Le graphe « Feminine+Vocative » de la grammaire « GN_GP_V.1.3.nog »	184
Figure 74. Le graphe « Feminine+Acusative » de la grammaire « GN_GP_V.1.3.nog ».....	184
Figure 75. Désambiguïsation du GN <i>чорнявий та кучерявий бог</i>	185
Figure 76. Désambiguïsation du GP <i>він гарний, чорнобривий</i>	185
Figure 77. Description du locatif	186
Figure 78. La grammaire « Locative_V.1.3.nog »	186
Figure 79. Description du datif.....	187
Figure 80. La grammaire « Dative_V.1.3.nog » : le graphe principal.....	187
Figure 81. La grammaire « Dative_V.1.3.nog » : les graphes imbriqués.....	188
Figure 82. Désambiguïsation des compléments <i>подарунок et друзів</i>	188
Figure 83. Description du groupe verbal	190
Figure 84. La grammaire syntaxique « GV.1.3.nog ».....	190
Figure 85. Annotation de l'ALU <i>мати</i>	191

Figure 86. L'ALU <i>по-злому</i> avant l'application de la grammaire « Adverbs_V.1.3.nog »	192
Figure 87. La grammaire syntaxiques « Adverbs_V.1.3.nog ».....	192
Figure 87. L'ALU <i>по-злому</i> après l'application de la grammaire « Adverbs_V.1.3.nog »	193
Figure 87. La grammaire syntaxique « Age_V.1.3 ».....	197
Figure 88. Annotation de l'entité nommée « AGE ».....	197

Liste des tableaux

Tableau 1. Étiquettes POS pour <i>Настуся</i> dans TreeTagger	53
Tableau 2. Lemmatisation de <i>Київщині</i> et <i>Київщини</i> dans Sketch Engine	53
Tableau 3. Lemmatisation de <i>Настусина</i> et <i>Настусині</i> dans Sketch Engine	53
Tableau 4. Erreurs de lemmatisation de dans RNNTagger	54
Tableau 5. Erreurs d'étiquetage dans RNNTagger	55
Tableau 6. Conversion	76
Tableau 7. Le GN à l'accusatif	87
Tableau 8. Les classes flexionnelles des noms	89
Tableau 9. Exemple d'une épithète et d'un attribut	90
Tableau 10. Exemple d'un adjectif court et d'un adjectif long dans le contexte	93
Tableau 11. Exemple des adjectifs contractés et non contractés dans le contexte	93
Tableau 12. Les trois classes flexionnelles des adjectifs	94
Tableau 13. Changements dans le radical des adjectifs d'appartenance	94
Tableau 14. Les modèles de flexion « МАМИН » et « БАТЬКІВ »	95
Tableau 15. Le comparatif, le superlatif et le superlatif absolu de <i>веселий</i>	96
Tableau 16. Exemples des adjectifs partitifs, excessifs et diminutifs dans le contexte	98
Tableau 17. Le paradigme de 1	99
Tableau 18. Les paradigmes de 2, 3, 4	100
Tableau 19. Les paradigmes de 5, 6, 7, 8, 50	100
Tableau 20. Les paradigmes de 40 et 90	100
Tableau 21. Les paradigmes de 200 et 300	100
Tableau 22. Les paradigmes des pronoms-adjectifs	102
Tableau 23. Les paradigmes des pronoms personnels	103
Tableau 24. Les paradigmes des pronoms <i>хто</i> et <i>що</i>	103
Tableau 25. Les classes morphologiques des verbes	110
Tableau 26. Le gérondif <i>читаючи</i> dans des outils existants	111
Tableau 27. Le participe <i>читаний</i> dans des outils existants	113
Tableau 28. Le comparatif, le superlatif et le superlatif absolu des adverbes	118
Tableau 29. Exemples des adverbes partitifs, excessifs et diminutifs dans le contexte	118
Tableau 30. Les couples homonymiques qui contiennent la conjonction	122
Tableau 31. Les couples homonymiques qui contiennent une particule	124
Tableau 32. Exemples des entrées du dictionnaire NooJ par catégorie	134
Tableau 33. Trois types d'entrées, accordées au paradigme « ГРУПА_МАМА »	138
Tableau 34. Les paradigmes pour les noms de la classe I	138
Tableau 35. Les paradigmes pour les noms masculins de la classe II	139
Tableau 36. Les paradigmes pour les noms neutres de la classe II	140
Tableau 37. Les paradigmes pour les noms de la classe III	141
Tableau 38. Les paradigmes pour les noms de la classe IV	141
Tableau 39. Les paradigmes pour les noms plurales tantum	142
Tableau 40. Les paradigmes particuliers	142
Tableau 41. Les paradigmes adjectivaux	143
Tableau 42. La comparaison des paradigmes des adjectifs et des pronoms-adjectifs démonstratifs	144
Tableau 43. Les paradigmes pronominaux	144

Tableau 44. Les paradigmes pour les numéraux	145
Tableau 45. Les paradigmes FLX « ЧИТАТИ » et « ЧИТАТИ_PF »	146
Tableau 46. Les paradigmes FLX pour les verbes.....	147
Tableau 47. Les paradigmes DRV pour les verbes	151
Tableau 48. Évaluation des dictionnaires NooJ pour l'ukrainien.....	154
Tableau 49. Évaluation de la grammaire « Participles_V.1.3.nom »	158
Tableau 50. Évaluation de la grammaire « Verbs_Sia_V.1.3 ».....	160
Tableau 51. Évaluation de la grammaire « Adverbs_V.1.3.nom ».....	161
Tableau 52. Évaluation des grammaires « Noun_Diminutive_F_M_V.1.3.nom » et « Noun_Diminutive_Neuter_V.1.3.nom ».....	162
Tableau 53. Évaluation de la grammaire «Adjectives_Belonging_V.1.3.nom».....	164
Tableau 54. Évaluation de la grammaire «Adjectives_Relatives_V.1.3.nom»	166
Tableau 55. Évaluation des grammaires « Adjectives_Expression_V.1.3.nom » et « Adverbs_Expression_V.1.3.nom.....	168
Tableau 56. Évaluation des grammaires «Adjectives_Degree_V.1.3.nom» et «Adverbs_Degree _V.1.3.nom»	170
Tableau 57. Évaluation de la grammaire « Adjectives_Short_V.1.3 »	171
Tableau 58. Évaluation de la grammaire « Négative_V.1.3 »	172
Tableau 59. L'évaluation de la grammaire « Adverbs_Pronouns_Indefinite_V.1.3.nom »	174
Tableau 60. Évaluation des grammaires « Parallel_forms_V.1.3 ».....	175
Tableau 61. Évaluation des grammaires « Foreign_words_V1.3.nom » et « Latin_numerals_V1.1.3.nom ».....	177
Tableau 62. Évaluation de la grammaire «Proper_V1.3.nom»	178
Tableau 63. Évaluation de la grammaire « GN_GP_V.1.3.nog ».....	186
Tableau 64. Évaluation des grammaires qui désambigüisent les cas.....	189
Tableau 65. Évaluation de la grammaire « GV.1.3.nog ».....	191
Tableau 66. Évaluation de la grammaire	193
Tableau 67. Exemple d'exercice pédagogique	195

Annexes

Le fichier « *_properties.def* »

```
# Count of categories = 12
#
# ADJECTIVE
# ADVERB
# CONJUNCTION
# INSERT
# INTERJECTION
# NOUN
# NUMERAL
# PARTICLE
# PREDICATIVE
# PREPOSITION
# PRONOUN
# VERB
```

```
INFLECTION = Accusative | Dative | Genitive | Instrumental | Nominative | Locative | Vocative |
Singular | Plural | Animate | Inanimate | Uncontracted | Feminine | Masculine | Neuter |
1 | 2 | 3 | Future | Past | Present | Imperfective | Perfective | Anterior | Concomitant |
Asubjectal | Gerund | Indicative | Infinitive | Passive;
```

```
#####
##### ADJECTIVE #####
#####
# Count of categories = 8
```

```
ADJECTIVE_Animation = Animate | Inanimate; # 2
ADJECTIVE_Case = Accusative | Dative | Genitive | Instrumental | Nominative | Locative; # 6
ADJECTIVE_Degree = Comparative | Superlative | Highest_Superlative; # 3
ADJECTIVE_Expression = Diminutive | Excessive | Partitive | Negative; # 4
ADJECTIVE_Form = Short | Uncontracted; # 2
ADJECTIVE_Gender = Feminine | Masculine | Neuter; # 3
ADJECTIVE_Number = Singular | Plural; # 2
ADJECTIVE_Type = Belonging | Qualitative | Relative; # 3
```

```
#####
##### ADVERB #####
#####
# Count of categories = 3
```

```
ADVERB_Degree = Comparative | Superlative | Highest_Superlative; # 3
ADJECTIVE_Expression = Diminutive | Excessive | Partitive; # 3
ADVERB_Type = Aim | Cause | Comparison | Direction | Indefinite | Intensity |
Interrogative_Relative
| Limit | Modal | Negative | Place | Quality | Quantity | State | Time; # 15
```

```
#####
##### CONJUNCTION #####
```


#####

Count of categories = 1

CONJUNCTION_Type = Explanatory | Causal | Conditional | Consequetial | Contrastive |
Comparative
| Coordinative | Correlative | Relative | Separative | Spatial | Subordinating | Temporal;# 13

#####

INTERJECTION

#####

Count of categories = 3

INTERJECTION_Type = Emotive | Onomatopoeia | Politeness | Swearing | Volitive; # 5
INTERJECTION_Origin = Adverb_transposed | Conjunction_transposed | Noun_transposed |
Particue_transposed | Pronoun_transposed
| Verbal_derivative | Verb_transposed; # 7
INTERJECTION_Structure = Elongate | Composite | Contracted | Pairs | Simple; # 5

#####

NOUN

#####

Count of categories = 8

NOUN_Animation = Animate | Inanimate; # 2
NOUN_Case = Accusative | Dative | Genitive | Instrumental | Nominative | Locative | Vocative; # 7
NOUN_Form = Abbreviation | Invariable | Negative | SubAdjectif; # 4
NOUN_Gender = Feminine | Masculine | Neuter; # 3
NOUN_Type = Common | Proper; # 2
NOUN_Number = Singular | Plural; # 2
NOUN_Expression = Abstract | Diminutive | Pejorative; # 3
NOUN_Semantic = Animal | Arts | Body | City | Color | Country | Ethnicity
| F_Name | Food | Humain | L_Name | Lake | Language | Little
| Mountain | Music | Profession | Proper | River | Slang | Sports | Time | Toponym; # 23

#####

NUMERAL

#####

Count of categories = 5

NUMERAL_Animation = Animate | Inanimate; # 2
NUMERAL_Case = Accusative | Dative | Genitive | Instrumental | Nominative | Locative; # 6
NUMERAL_Gender = Feminine | Masculine | Neuter; # 3
NUMERAL_Type = Cardinal | Ordinal | Collective; # 3
NUMERAL_Number = Singular | Plural; # 2

#####

PARTICLE

#####

Count of categories = 1

PARTICLE_Type = Negative | Interrogative | Conditional; # 3

#####

PRONOUN

#####

Count of categories = 6

PRONOUN_Animation = Animate | Inanimate; # 2

PRONOUN_Case = Accusative | Dative | Genitive | Instrumental | Nominative | Locative; # 6

PRONOUN_Gender = Feminine | Masculine | Neuter; # 3

PRONOUN_Type = Definite | Demonstrative | Indefinite | Interrogative_Relative | Negative | Personal | Possessive | Reflexive; # 8

PRONOUN_Number = Plural | Singular; # 2

PRONOUN_Person = 1 | 2 | 3; # 3

#####

VERB

#####

Count of categories = 15

VERB_Aspect = Imperfective | Perfective | Pair; # 2

VERB_Gender = Feminine | Masculine | Neuter; # 3

VERB_Gerund_Form = Anterior | Concomitant; # 2

VERB_Mood = Asubjectal | Conditional | Gerund | Indicative | Infinitive | Imperative | Participle ; # 7

VERB_Motion_Type = Determinate | Indeterminate; # 2

VERB_Number = Plural | Singular; # 2

VERB_Part participle_Case = Accusative | Dative | Genitive | Instrumental | Nominative | Locative; # 6

VERB_Person = 1 | 2 | 3; # 3

VERB_Reflexivity = Reflexive; # 1

VERB_Tense = Future | Past | Present; # 3

VERB_Transitivity = Transitive | Intransitive; # 2

VERB_Type = Motion | Position | Modal; # 3

VERB_Voice = Active | Passive; # 2

#####

PARTICIPLE

#####

Count of categories = 6

PARTICIPLE_Aspect = Imperfective | Perfective; # 2

PARTICIPLE_Case = Accusative | Dative | Genitive | Instrumental | Nominative | Locative; # 6

PARTICIPLE_Gender = Feminine | Masculine | Neuter; # 3

PARTICIPLE_Number = Singular | Plural; # 2

PARTICIPLE_Tense = Past | Present; # 2

PARTICIPLE_Voice = Active | Passive; # 2

Exemples de paradigmes

Nom

КВІТКА_ТІТКА=

<E>/Nominative+Singular |
и/Genitive+Singular |
у/Accusative+Singular |
<B2>ці/Dative+Singular |
ою/Instrumental+Singular |
<B2>ці/Locative+Singular |
о/Vocative+Singular |
и/Nominative+Plural |
<B2>ок/Genitive+Plural |
<B2>ок/Accusative+Plural+Animate |
и/Accusative+Plural+Inanimate |
ам/Dative+Plural |
ами/Instrumental+Plural |
ах/Locative+Plural |
и/Vocative+Plural;

Adjectif

ВЕЛИКИЙ=

<E>/Nominative+Masculine+Singular |
<B2>ого/Genitive+Masculine+Singular |
<B2>ого/Accusative+Animate+Masculine+Singular |
<E>/Accusative+Inanimate+Masculine+Singular |
<B2>ому/Dative+Masculine+Singular |
<B2>им/Instrumental+Masculine+Singular |
<B2>ому/Locative+Masculine+Singular |
<B2>ім/Locative+Masculine+Singular |
<B2>е/Nominative+Neuter+Singular |
<B2>ее/Nominative+Uncontracted+ADJECTIVE+Neuter+Singular |
<B2>ого/Genitive+Neuter+Singular |
<B2>е/Accusative+Neuter+Singular |
<B2>ее/Accusative+Uncontracted+ADJECTIVE+Neuter+Singular |
<B2>ому/Dative+Neuter+Singular |
<B2>им/Instrumental+Neuter+Singular |
<B2>ому/Locative+Neuter+Singular |
<B2>ім/Locative+Neuter+Singular |
<B2>а/Nominative+Feminine+Singular |
<B2>ая/Nominative+Uncontracted+ADJECTIVE+Feminine+Singular |
<B2>ої/Genitive+Feminine+Singular |
<B2>у/Accusative+Feminine+Singular |
<B2>ую/Accusative+Uncontracted+ADJECTIVE+Feminine+Singular |
<B2>ій/Dative+Feminine+Singular |
<B2>ою/Instrumental+Feminine+Singular |
<B2>ій/Locative+Feminine+Singular |
<B2>и/Nominative+Plural |

<B2>ii/Nominative+Uncontracted+ADJECTIVE+Plural |
<B2>их/Genitive+Plural |
<B2>их/Accusative+Animate+Plural |
<B2>i/Accusative+Inanimate+Plural |
<B2>ii/Accusative+Inanimate+Plural+Uncontracted+ADJECTIVE |
<B2>им/Dative+Plural |
<B2>ими/Instrumental+Plural |
<B2>их/Locative+Plural;

Pronom-adjectif interrogatif

ЧИЙ=

<E>/Nominative+Masculine+Singular |
<E>ого/Genitive+Masculine+Singular |
<E>ого/Accusative+Animate+Masculine+Singular |
<E>/Accusative+Inanimate+Masculine+Singular |
<E>ому/Dative+Masculine+Singular |
ему/Dative+Masculine+Singular |
ім/Instrumental+Masculine+Singular |
<E>ому/Locative+Masculine+Singular |
ему/Locative+Masculine+Singular |
ім/Locative+Masculine+Singular |
i/Nominative+Plural |
ix/Genitive+Plural |
ix/Accusative+Animate+Plural |
i/Accusative+Inanimate+Plural |
ім/Dative+Plural |
ими/Instrumental+Plural |
ix/Locative+Plural |
е/Nominative+Neuter+Singular |
<E>ого/Genitive+Neuter+Singular |
<E>/Accusative+Neuter+Singular |
<E>ому/Dative+Neuter+Singular |
ему/Dative+Neuter+Singular |
ім/Instrumental+Neuter+Singular |
<E>ому/Locative+Neuter+Singular |
ему/Locative+Neuter+Singular |
ім/Locative+Neuter+Singular |
я/Nominative+Feminine+Singular |
єї/Genitive+Feminine+Singular |
ю/Accusative+Feminine+Singular |
ій/Dative+Feminine+Singular |
єю/Instrumental+Feminine+Singular |
ій/Locative+Feminine+Singular;

Pronom personnel

Я=

<E>/Nominative+1+Singular |
<BW>мене/Genitive+1+Singular |

<BW>мене/Accusative+1+Singular |
<BW>мені/Dative+1+Singular |
<BW>мною/Instrumental+1+Singular |
<BW>мені/Locative+1+Singular;

Numéral 1

ОДИН=
<E>/Nominative+Masculine+Singular |
<B2>ного/Genitive+Masculine+Singular |
<B2>ного/Accusative+Animate+Masculine+Singular |
<E>/Accusative+Inanimate+Masculine+Singular |
<B2>ному/Dative+Masculine+Singular |
<B2>ним/Instrumental+Masculine+Singular |
<B2>ному/Locative+Neuter+Singular |
<B2>нім/Locative+Masculine+Singular |
<B2>но/Nominative+Neuter+Singular |
<B2>но/Accusative+Neuter+Singular |
<B2>не/Nominative+Neuter+Singular |
<B2>не/Accusative+Neuter+Singular |
<B2>ному/Dative+Neuter+Singular |
<B2>ним/Instrumental+Neuter+Singular |
<B2>ному/Locative+Neuter+Singular |
<B2>нім/Locative+Neuter+Singular |
<B2>на/Nominative+Feminine+Singular |
<B2>ної/Genitive+Feminine+Singular |
<B2>нієї/Genitive+Feminine+Singular |
<B2>ну/Accusative+Feminine+Singular |
<B2>ній/Dative+Feminine+Singular |
<B2>ною/Instrumental+Feminine+Singular |
<B2>нією/Instrumental+Feminine+Singular |
<B2>ній/Locative+Feminine+Singular |
<B2>ні/Nominative+Plural |
<B2>них/Genitive+Plural |
<B2>них/Accusative+Animate+Plural |
<B2>ні/Accusative+Inanimate+Plural |
<B2>ним/Dative+Plural |
<B2>ними/Instrumental+Plural |
<B2>них/Locative+Plural;

Numéral 2

ДВА=
<E>/Nominative |
ох/Genitive |
ох/Accusative+Animate |
<E>/Accusative+Inanimate |
ом/Dative |
ома/Instrumental |
ох/Locative |

i/Nominative+Feminine;

Verbe (FLX)

БАРВИТИ=

<E>/Imperfective+Infinitive |
<B2>в/Imperfective+Masculine+Singular+Past |
<B2>ла/Imperfective+Feminine+Singular+Past |
<B2>ло/Imperfective+Neuter+Singular+Past |
<B2>ли/Imperfective+Plural+Past |
<B3>лю/Imperfective+1+Singular+Present |
<B3>иш/Imperfective+2+Singular+Present |
<B3>ить/Imperfective+3+Singular+Present |
<B3>имо/Imperfective+1+Plural+Present |
<B3>ите/Imperfective+2+Plural+Present |
<B3>лять/Imperfective+3+Plural+Present |
<E>му/Imperfective+1+Singular+Future |
<E>ме/Imperfective+3+Singular+Future |
<E>меш/Imperfective+2+Singular+Future |
<E>мемо/Imperfective+1+Plural+Future |
<E>мете/Imperfective+2+Plural+Future |
<E>муть/Imperfective+3+Plural+Future |
<B3>/Imperfective+2+Singular+Imperative |
<B3>мо/Imperfective+1+Plural+Imperative |
<B3>те/Imperfective+2+Plural+Imperative |
<B3>лячи/Imperfective+Gerund+Concomitant |
<B2>вши/Imperfective+Gerund+Anterior |
<B3>лено/Imperfective+Passive+Asubjectal;

БАРВИТИ_PF=

<E>/Perfective+Infinitive |
<B2>в/Perfective+Masculine+Singular+Past |
<B2>ла/Perfective+Feminine+Singular+Past |
<B2>ло/Perfective+Neuter+Singular+Past |
<B2>ли/Perfective+Plural+Past |
<B3>лю/Perfective+1+Singular+Future |
<B3>иш/Perfective+2+Singular+Future |
<B3>ить/Perfective+3+Singular+Future |
<B3>имо/Perfective+1+Plural+Future |
<B3>ите/Perfective+2+Plural+Future |
<B3>лять/Perfective+3+Plural+Future |
<B3>/Perfective+2+Singular+Imperative |
<B3>мо/Perfective+1+Plural+Imperative |
<B3>те/Perfective+2+Plural+Imperative |
<B2>вши/Perfective+Gerund+Anterior |
<B3>лено/Perfective+Passive+Asubjectal;

Verbe (DRV)

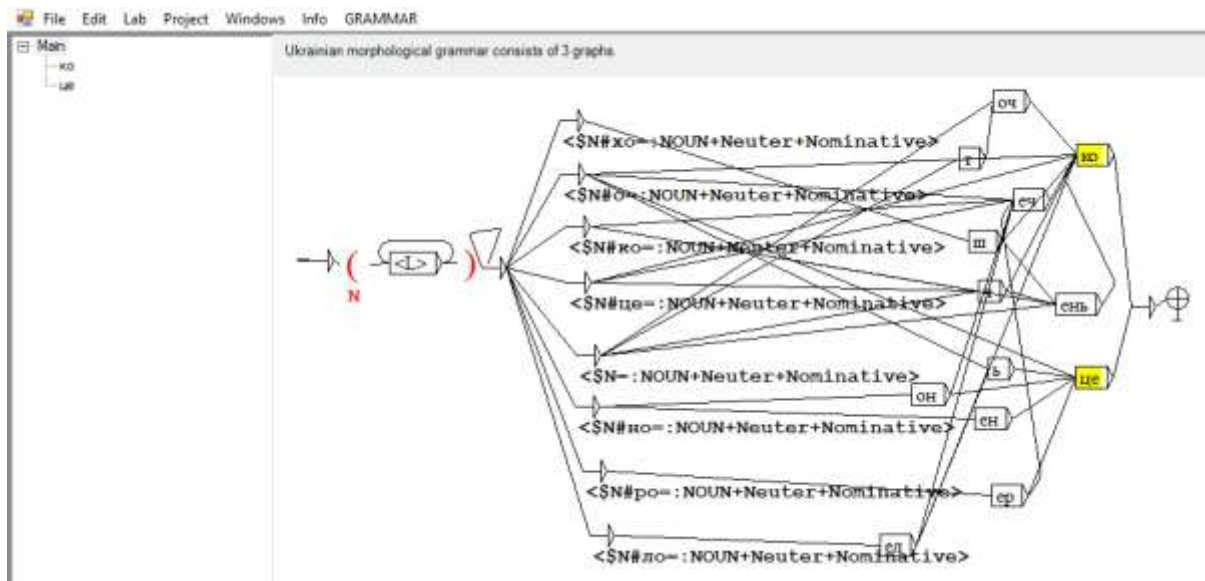
O=<E>/VERB;

B=<LW>в/VERB;

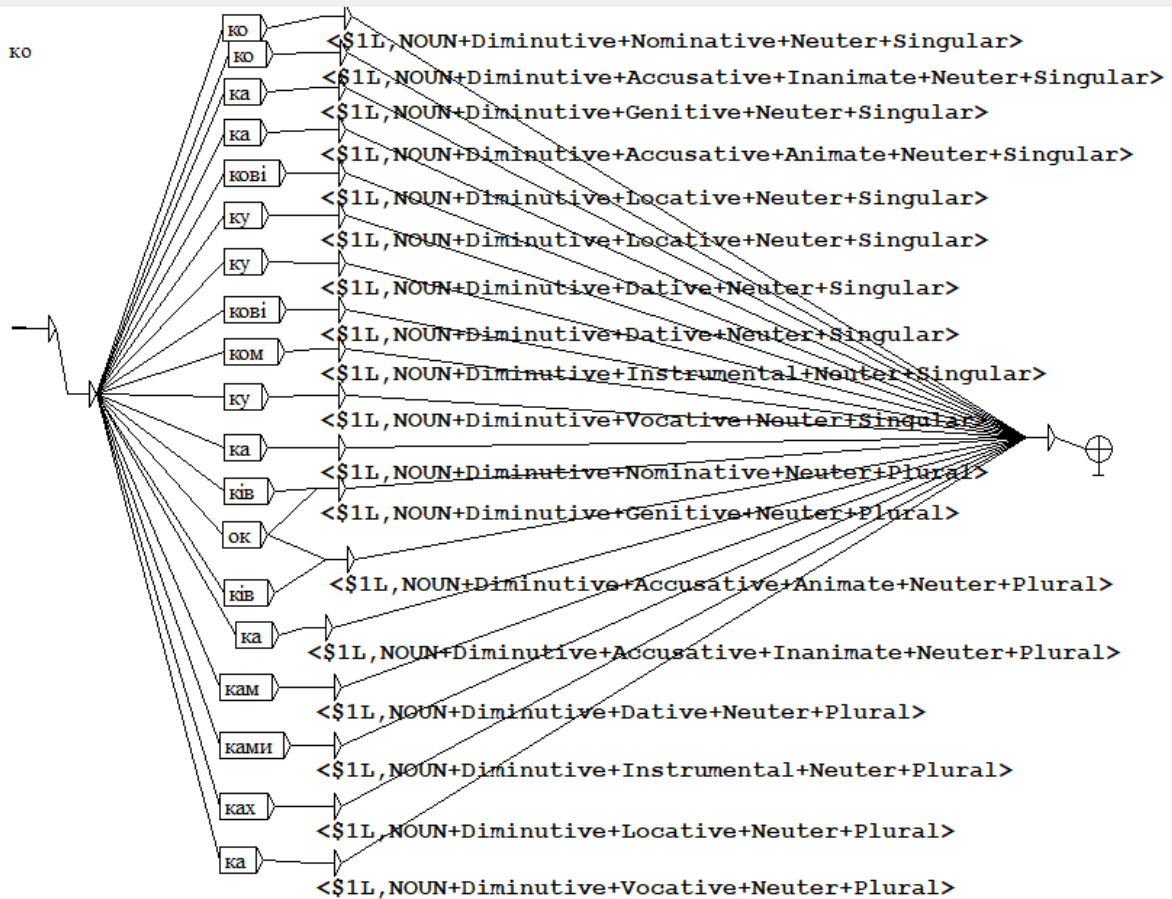
ТИ_ВЗ=<ВЗ>ти/VERB;

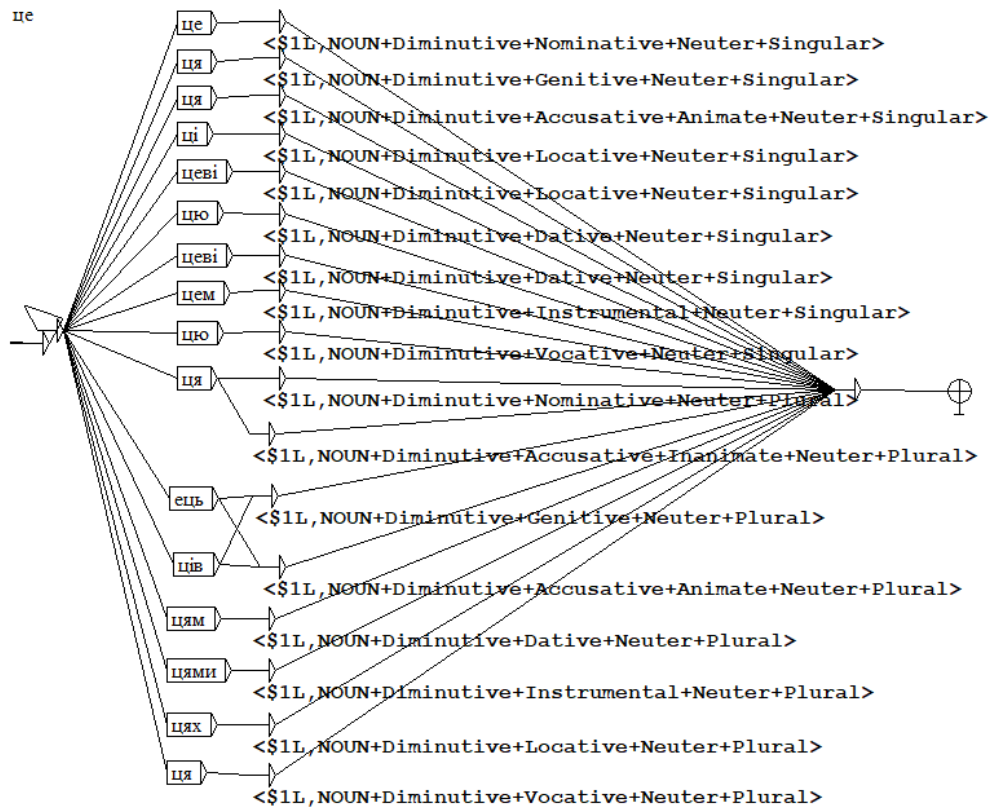
ALT=<L4><В>о<RW><ВЗ>ти/VERB;

La grammaire morphologique « Noun_Diminutive_Neuter_V.1.3.nom » en détail

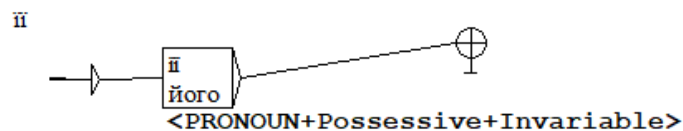
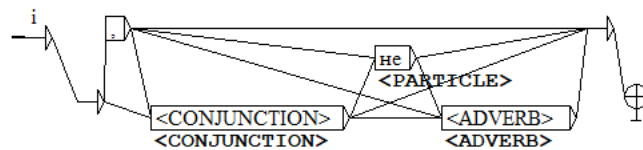
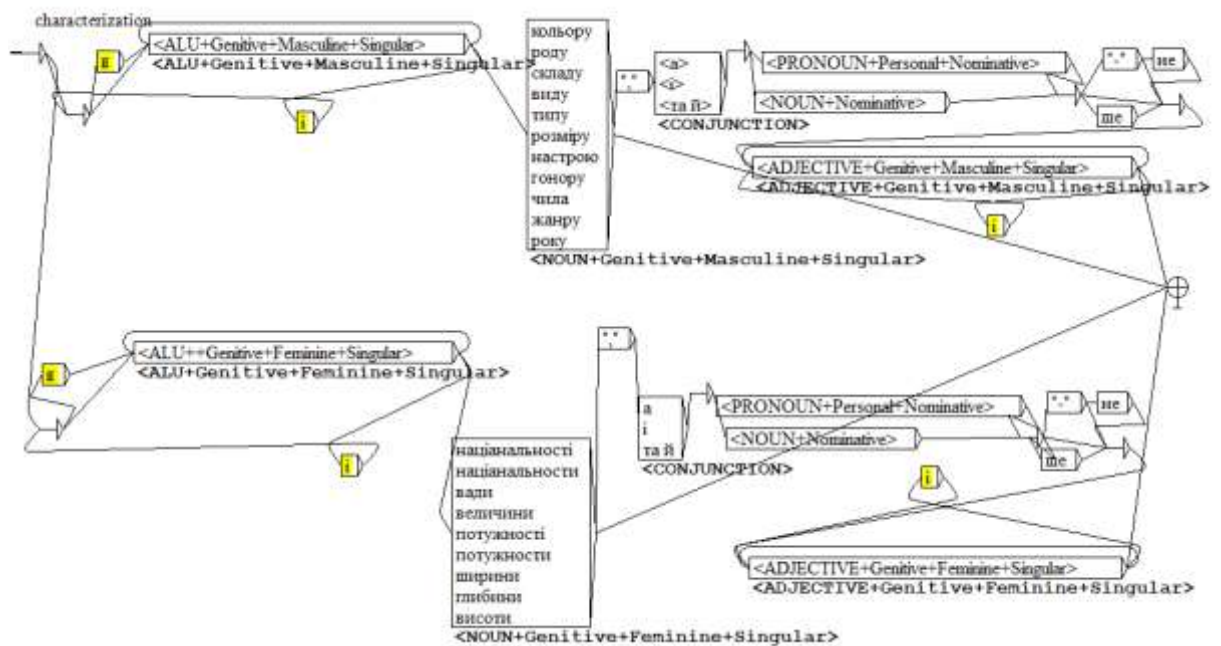
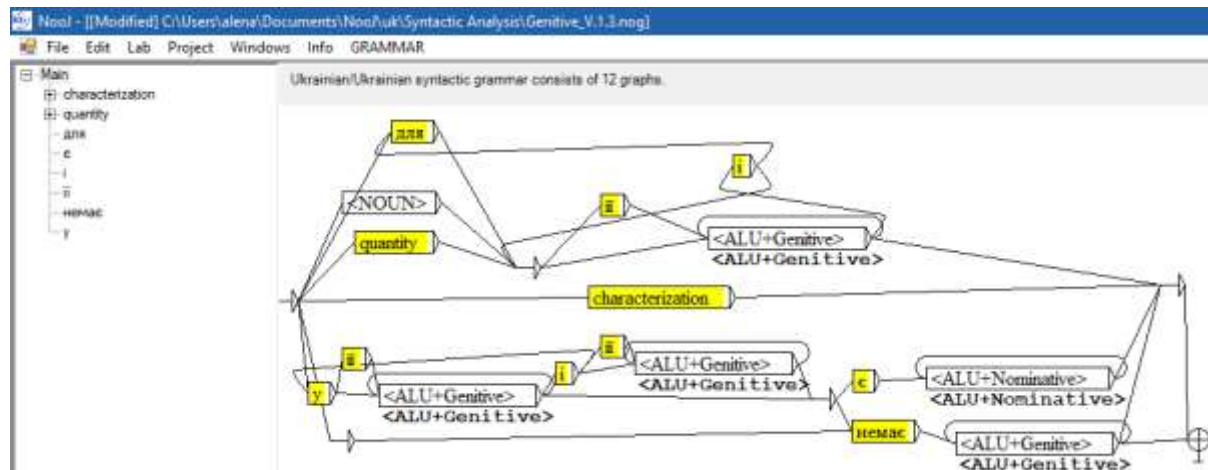


Ukrainian morphological grammar consists of 3 graphs.

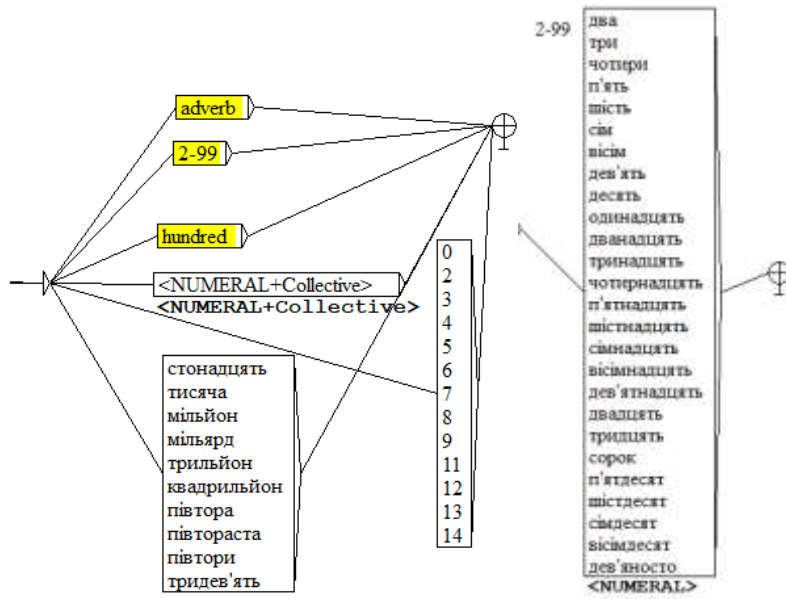




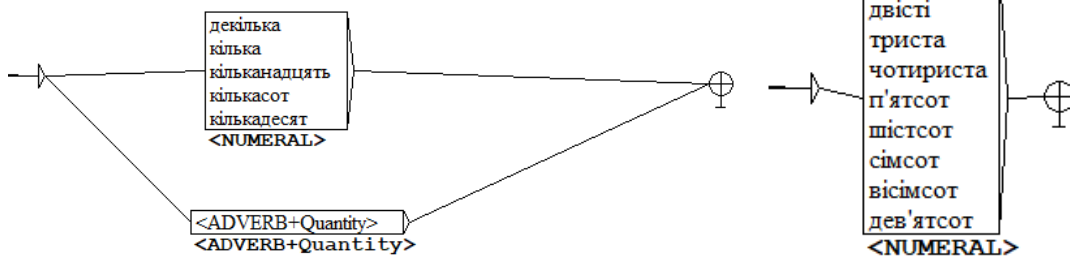
La grammaire locale « Genitive_V.1.3.nog » en détail



quantity

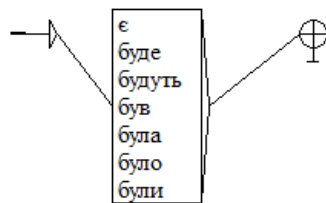


adverb

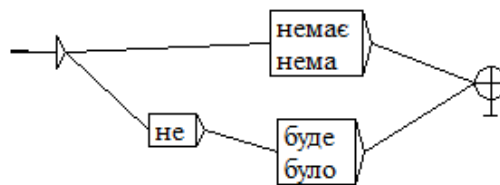


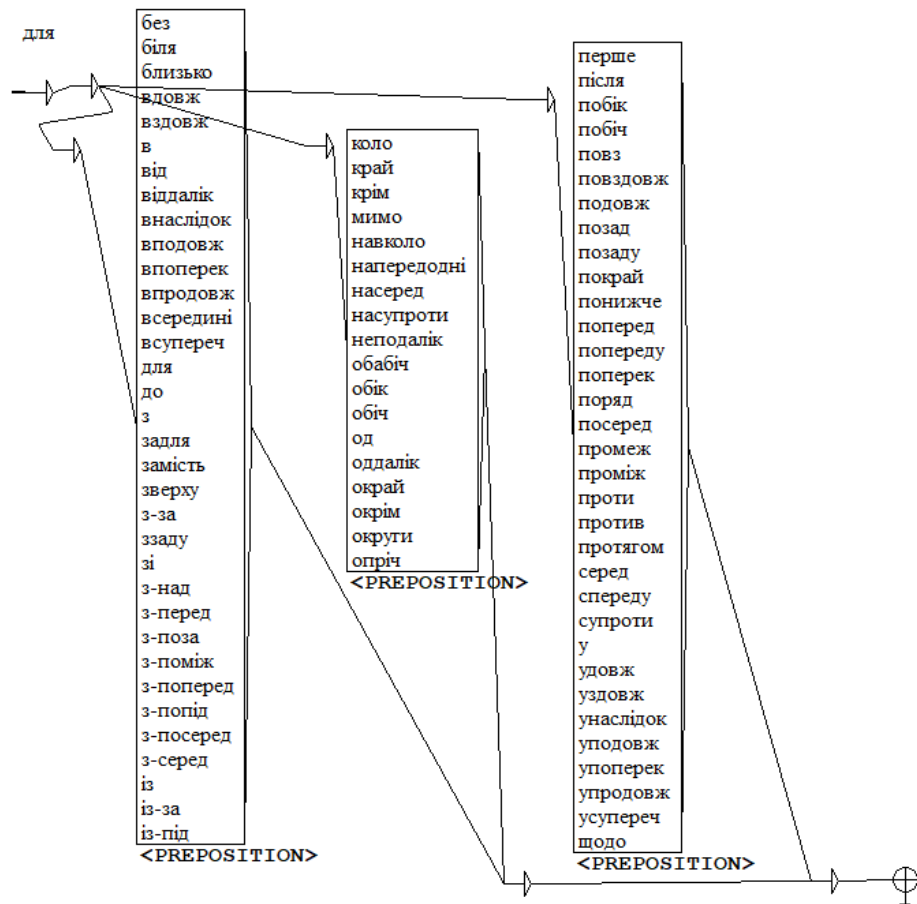
hundred

є



немає





Bibliographie

Bibliographie en ukrainien et en russe

1. Баденкова, Вікторія (2018). *Історія української літературної мови* [Badenkova. Histoire de la langue littéraire ukrainienne]. Миколаївський національний університет імені В. О. Сухомлинського, Миколаїв. Consulté en ligne : https://chtyvo.org.ua/authors/Badenkova_Viktoriia/Istoriia_ukrainskoi_literaturnoi_movy/
2. Бевзенко, Степан Пилипович (1960). *Історична морфологія української мови* [Bevzenko. Morphologie historique de la langue ukrainienne]. Закарпатське обласне видавництво, Ужгород. Consulté en ligne : http://irbis-nbuv.gov.ua/cgi-bin/ua/elib.exe?Z21ID=&I21DBN=UKRLIB&P21DBN=UKRLIB&S21STN=1&S21REF=10&S21FMT=online_book&C21COM=S&S21CNR=20&S21P01=0&S21P02=0&S21P03=FF=&S21STR=ukr0001113
3. Безпалько О.П., Бойчук М.К., Жовтобрюх М.А. (1962). *Історична граматики української мови* [Bezpal'ko et alii. Grammaire historique de la langue ukrainienne]. Радянська Школа, Київ.
4. Безпояско О. К., Городенська К. Г., Русанівський В. М. (1993). *Граматики української мови. Морфологія* [Bezpoïasko et alii. Grammaire de la langue ukrainienne. Morphologie]. Либідь, Київ.
5. Брус, Марія (2016). *Історична граматики* [Brous. Grammaire historique]. ПП Голіней, Івано-Франківськ. Consulté en ligne : https://shron1.chtyvo.org.ua/Brus_Mariia/Istorychna_hramatyka_ukrainskoi_movy_Chastyna_1_Teoretychnyi_material_Tematychni_tablytsi.pdf
6. Булаховський Л.А. (1956). *Питання походження української мови*. [Boulakhovs'kyu. L'origine de la langue ukrainienne]. Вид-во АН УРСР, Ін-т мовознавства імені О. О. Потебні, Київ.
7. Булик-Верхола С. З., Верхола А. О. (2018). *Теорія та історія держави і права* [Boulyk-Verkhola et Verkhola. Théorie et histoire de l'État et du droit]. Dans *Серія: право*, n°1 (59). Consulté en ligne : http://www.law.stateandregions.zp.ua/archive/1_2018/3.pdf
8. Буньо Г. Б. (2014). *Сучасні методи вирішення проблеми граматичної омонімії в тексті* [Bounio. Méthodes modernes de résolution du problème de l'homonymie grammaticale dans le texte]. Dans *Наукові записки національного університету «Острозька академія»*, n° 49, pp. 12-16.
9. Вихованець, Іван (1988). *Частини мови в семантико-граматичному аспекті* [Vykhovanets'. Catégorie grammaticale sous aspect sémantique et grammatical]. Наукова думка, Київ.
10. Вихованець, Іван, Городенська, Катерина. (2004). *Теоретична морфологія української мови* [Vykhovanets' et Horodens'ka. Morphologie théorique de la langue ukrainienne]. Пульсари, Київ.

11. Вихованець, Іван (2012). *Дієслівно-іменниковий граматичний тип української мови* [Vykhovanets'. Type grammatical «verbe-nom» de la langue ukrainienne]. Dans *Українська мова*, n° 2, pp. 3-10.
Consulté en ligne : http://nbuv.gov.ua/UJRN/Ukrm_2012_2_3
12. Вихованець І. Р., Городенська К. Г., Загнітко А. П., Соколова С. О. (2017). *Граматика сучасної української літературної мови. Морфологія* [Vykhovanets' et alii. La grammaire de la langue ukrainienne. Morphologie]. Видавничий дім Дмитра Бурого, Київ.
13. Вірченко, Ніна (2018). *Документи про заборону української мови (XVII–XX ст.)* [Virtchenko. Documents sur l'interdiction de la langue ukrainienne (XVIIe-XXe siècles)]. Dans *Голота Л., Букет Є., Без мови — немає нації* [Sans langue il n'y a pas de nation]. Український пріоритет, Київ. Consulté en ligne : <http://slovoprosvity.org/pdf/bsp/bsp2.pdf>
14. Волкова І.В., Масло О.В. (2012). *Українська мова. Морфологія* [Volkova et Maslo. Langue ukrainienne. Morphologie]. Харківська гуманітарно-педагогічна академія, Харків.
15. Голоюх Л. В. (2021). *Історія української літературної мови* [Holoïoukh. Histoire de la langue littéraire ukrainienne]. Надстир'я, Луцьк.
Consulté en ligne : <https://evnuir.vnu.edu.ua/bitstream/123456789/20032/1/035.pdf>
16. Гриценко П.Е. (1993). *Некоторые замечания о диалектной основе украинского литературного языка* [Hrytsenko. Quelques remarques à propos de la base dialectale de la langue littéraire ukrainienne]. Dans *Philologia slavica: K 70-letiu akademika N.I. Tolstogo*. Consulté en ligne : <http://istoriya.ru/forum/?showtopic=3943&page=17>
17. Гаценко І. О. (2009а). *Синтаксична функція звуконаслідувальних слів української мови* [Hatsenko. La fonction syntaxique des onomatopées dans la langue ukrainienne]. Dans *Сіверянський літопис*, n° 6, pp. 97-100.
18. Гаценко І. О. (2009b). *Особливості структури звуконаслідувальних слів української мови* [Hatsenko. Caractéristiques de la structure des onomatopées dans la langue ukrainienne]. Dans *Актуальні проблеми української лінгвістики: теорія і практика*, n° 8, pp. 50-56.
19. Глібчук Н. (2018). *Препозиціоналізація у площині міжчастиномовної омонімії* [Hlibtchouk. La prépositionnalisation de l'homonymie intralinguistique]. Dans *Вісник Львівського університету. Серія філологічна*, n° 68, pp. 58-67.
20. Горбач О. (1993). *Засади періодизації історії української мови й етапи її розвитку*. [Horbatch. Les principes de périodisation de l'histoire de la langue ukrainienne et étapes de son développement]. Dans *Міжнародний конгрес українців: доповіді і повідомлення*. – Львів, – pp. 7-12.
21. Горпинич, Володимир. (2004). *Морфологія української мови*. [Horpunytch. Morphologie de la langue ukrainienne]. Академія, Київ.
22. Грабовський, Сергій (2020). *Українська мова в часи СРСР. Друга хвиля «великої» українізації* [Hrabovs'kyu. La langue ukrainienne à l'époque de l'URSS. La deuxième vague de la "grande" ukrainisation]. Dans *Radio Svoboda (en ligne uniquement)* : <https://www.radiosvoboda.org/a/30709133.html>

23. Гриценко С. П. (2017). *Динаміка лексики української мови XVI– XVII ст.* [Hrytsenko. La dynamique du lexique de la langue ukrainienne aux XIVe et XVIIIe siècles]. КММ, Київ. Consulté en ligne : https://chtyvo.org.ua/authors/Hrytsenko_SP/Dynamika_leksykonu_ukrainskoi_movy_XVIXVII_st
24. Грязнухіна Т. О., Братченко, Л. Н., Дарчук, Н. П., Критська, В. І., Пузирева, Т. К., Орлова, Л. В. (1989). *Шляхи уникнення омонімії в системі автоматичного морфологічного аналізу* [Hriaznoukhina et alii. Méthodes pour éliminer l'homonymie dans un système d'analyse morphologique automatique]. Dans *Мовознавство*, n° 5, pp. 3–12.
25. Гуменюк І. М. (2003). *Семантичні особливості інтер'єктивів* [Houmeniouk. Caractéristiques sémantiques des interjections]. Dans *Studia metodologica*, pp.26-33.
26. Гуменюк, Тетяна (2014). *Внесок львівських мовознавців у розвиток української філології в УРСР у другій половині ХХ ст.* [Houmeniouk. La contribution des linguistes de Lviv au développement de la linguistique ukrainienne dans la République socialiste soviétique d'Ukraine au cours de la seconde moitié du XX^e siècle]. Dans *Наукові записки Тернопільського національного педагогічного університету імені Володимира Гнатюка. Сер. Історія*, n° 2, Тернопіль. Consulté en ligne : <http://dSPACE.tnpu.edu.ua/bitstream/123456789/7664/1/Humeniuk.pdf>
27. Демська О. М. (2019). *10 історій про омонімію* [Dems'ka. 10 histoires sur l'homonymie]. Ранок, Харків.
28. Дика Л., Шведова М. (2022). *Історія і нормативний статус прийменника / префікса од- в українській мові нового часу.* [Dyka et Shvedova. L'histoire et statut normatif de la préposition/préfixe од- dans la langue ukrainienne moderne]. Dans *Slavia Orientalis*, vol. LXXI, No 4, pp. 797-818. Consulté en ligne : <https://journals.pan.pl/Content/125858/PDF/2022-04-SOR-06.pdf>
29. Жовтобрюх М. А., Кулик Б. М. (1965). *Курс сучасної української літературної мови* [Jovtobrioukh. Cours de langue littéraire ukrainienne moderne]. Радянська школа, Київ. Consulté en ligne : <http://irbis-nbuv.gov.ua/ulib/item/UKR0000852>
30. Жовтобрюх М.А., М. А., Волох, О. Т., Самійленко, С. П., Слинко, І. І. (1980). *Історична граматики української мови* [Jovtobrioukh et alii. Grammaire Histoire de la langue littéraire ukrainienne]. Вища школа, Київ.
31. Кавера Н. В. (2008). *Семантична типологія предикатів стану : автореф. дис. на здобуття наук. ступеня канд. філол. наук : 10.02.01.* [Kavera. La typologie sémantique des prédicats d'état : une thèse de doctorat pour le grade de candidat en philologie, soutenue le 10/02/01]. Київ.
32. Козачевська Л., Сидоренко О. (2013). *Мова української діаспори як віддзеркалення мовної свідомості та мовної діяльності носіїв.* [Kozatchevs'ka et Sydorenko. La langue de la diaspora ukrainienne en tant que reflet de la conscience linguistique et de l'activité linguistique de ses locuteurs]. Dans *Мовні і концептуальні картини світу*, vol. 46(2), pp. 142-151. Consulté en ligne :: http://nbuv.gov.ua/UJRN/Mikks_2013_46%282%29__20

33. Костусьяк, Наталія (2012). *Структура міжрівневих категорій сучасної української мови* [Koustoussiak. La structure des catégories intralinguistique de la langue ukrainienne]. Волинський національний університет імені Лесі Українки, Луцьк.
34. Кримський, Агатангел Юхимович (1922). *Українська мова, звідкіля вона взялася і як розвивалася* [Kryms'kyu. La langue ukrainienne, son origine et son évolution]. Dans *Кримський, Агатангел Юхимович, Т.3* (1973). [Kryms'kyu Ahatanhel loukhymovytych, vol. 3], Наукова думка, Київ. Consulté en ligne : <http://irbis-nbuv.gov.ua/ulib/item/UKR0002817>
35. Критська В.І., Недозим Т.І., Орлова Л.В., Пуздишева Т.К., Романюк Ю.В. (2011) *Граматичний словник української літературної мови. Словозміна*. [Kryts'ka .I., Nedozym T.I., Orlova L.V., Pouzdyrieva T.K., Romaniouk Y.V. Dictionnaire grammatical de la langue littéraire ukrainienne. Les paradigmes flexionnels]. Видавничий Дім Дмитра Бурого, Київ.
36. Кузьменко Г. І. (2000). *Омонімія: аспектологія, проблематика* [Kouz'menko. Homonymie : aspectologie, enjeux]. Вид-во Київськ. ун-ту., Київ.
37. Курило О. Й. (2000). *Відвигукова лексика сучасної української літературної мови: склад і структура* [Kourylo. Interjection de la langue littéraire ukrainienne : composition et structure]. Dans *Збірник наукових праць Ужгородського державного університету «Сучасні проблеми мовознавства та літературознавства»*, n° 2, pp. 130-135.
38. Курило О. Й. (2009). *Семантичні та словотвірні явища у відвигуковій лексиці* [Kourylo. Interjection : phénomènes sémantiques et dérivatifs]. Dans *Матеріали Міжнар. наук. конф. «Семантика мови і тексту»*, pp. 295 - 298.
39. Мариненко, Ірина (2018). *Шляхи реалізації пуристичних тенденцій у сучасній українській мові* [Marynenko. Tendances puristes dans la langue ukrainienne moderne]. Dans *Стиль і текст*, n° 2, ВПЦ «Київський університет», Київ. Consulté en ligne : http://www.library.univ.kiev.ua/ukr/host/viking/db/ftp/univ/sit/sit_2018_18.pdf#page=41
40. Масенко, Лариса (2011). *Суржик: між мовою і языком* [Massenko. Le « sourjyk » : entre « мова » et « іязук »]. Видавничий дім «Києво-Могилянська академія», Київ. Consulté en ligne : https://chtyvo.org.ua/authors/Masenko_Larysa/Surzhyk_mizh_movoiu_i_iazukom
41. Матвіяс І. Г. (2011). *Особливості фонетичної системи в західноукраїнському варіанті літературної мови* [Matviias. Caractéristiques du système phonétique dans la variante ukrainienne occidentale de la langue littéraire]. Dans *Мовознавство [Linguistique]*, n° 4, pp. 16-21. Consulté en ligne : http://nbuv.gov.ua/UJRN/MoZn_2011_4_3
42. Німчук, Василь (1985). *Мовознавство на Україні в XIV—XVII ст.* [Nimtchouk. La linguistique en Ukraine aux XVe-XVIIe siècles]. Наукова думка, Київ.
43. Німчук В., Пуряева Н. (2004). *Історія українського правопису: XVI—XX століття: хрестоматія* [Nimtchouk et Pouriaieva. Histoire de l'orthographe ukrainienne : XVI-XX siècles]. Наукова думка, Київ. Consulté en ligne : https://chtyvo.org.ua/authors/Nimchuk_Vasyl/Istoriia_ukrainskoho_pravopysu_XVI-XX_stolittia_Khrestomatiia/

44. Німчук, Василь (2015). *Історія української мови. Хрестоматія X-XIII ст.* [Nimtchouk. Histoire de la langue ukrainienne. X-XIII^e siècles. Manuel]. Полісся, Житомир. Consulté en ligne : https://chtyvo.org.ua/authors/Nimchuk_Vasyl/Istoriia_ukrainskoi_movy_Khrestomatiia_X-XIII_st/
45. Огієнко, Іван (1927). *Нариси з історії української мови. Система українського правопису. Популярно-науковий курс з історичним освітленням* [Ohiienko. Essais sur l'histoire de la langue ukrainienne. Le système orthographique ukrainien. Un cours de vulgarisation scientifique avec une dimension historique]. Варшава. Nous avons consulté la deuxième édition de 1990, d'édition « Волинь », Winnipeg. Consulté en ligne : <https://diasporiana.org.ua/wp-content/uploads/books/3606/file.pdf>
46. Огієнко, Іван (Митрополит Іларіон) (1949). *Історія української літературної мови* [Ohiienko. Histoire de la langue littéraire ukrainienne]. Наша культура, Вінніпег. Consulté en ligne : <https://library.kr.ua/wp-content/elib/ogienko/ogienko.pdf>
47. Огієнко, Іван (Митрополит Іларіон) (1961). *Граматично-стилістичний словник Шевченкової мови* [Ohiienko. Dictionnaire grammatical et stylistique de la langue de Chevtchenko]. Видавництво « Волинь », Вінніпег. Consulté en ligne : <http://irbis-nbuv.gov.ua/ulib/item/UKR0001725>
48. Остапчук О. А. (2013). *Український літературний язык в свете социолінгвістической типології слов'янських мов* [Ostapchouk. La langue littéraire ukrainienne à la lumière de la typologie sociolinguistique des langues slaves]. Dans Молдован, А. М. *Слов'янське мовознавство: XV Міжнародний конгрес мовознавців (Мінськ, 2013)*. Индрик, 633-655. Consulté en ligne : <https://cyberleninka.ru/article/n/kulturno-etnicheskoe-dvuyazychie-na-ukraine-kak-rezultat-dlitelnogo-kontaktirovaniya-ukrainskogo-i-russkogo-yazykov>
49. Павленко, Лариса. (2010). *Історична граматика української мови* [Pavlenko. Grammaire historique de la langue ukrainienne]. Волинський нац. ун-т ім. Лесі Українки, Луцьк. Consulté en ligne : https://chtyvo.org.ua/authors/Pavlenko_Larysa/Istorychna_hramatyka_ukrainskoi_movy/
50. Півторак, Григорій (1993). *Україні: звідки ми і наша мова* [Pivtorak. Les Ukrainiens : nos origines et notre langue]. Наукова думка, Київ. Consulté en ligne : <http://irbis-nbuv.gov.ua/ulib/item/UKR0004644>
51. Півторак, Григорій (2001). *Походження Українців, Росіян, Білорусів та їхніх мов. Міфи і правда про трьох братів слов'янських зі «спільної колиски»* [Pivtorak. L'origine des Ukrainiens, des Russes, des Bélarussiens et de leurs langues. Mythes et vérités sur trois frères slaves issus d'un « berceau commun »]. Академія, Київ. Consulté en ligne : <http://litopys.org.ua/pivtorak/pivt.htm>
52. Плющ М. Я. (2010). *Граматика української мови. Морфеміка. Словотвір. Морфологія* [Pliouchtch. Grammaire de la langue ukrainienne. Morphèmes. Formation des mots. Morphologie]. Видавничий Дім «Слово», Київ.
53. Пономарів О. Д., Різун В. В., Шевченко Л. Ю. та ін. (2008). *Сучасна українська мова* [Ponomariv et alii. La langue ukrainienne]. Либідь, Київ.

54. Рак О. М. (2000). *Про питання походження вигуків* [Rak. À propos de l'origine des interjections]. Dans *Науковий вісник Чернівецького університету: Збірник наукових праць*, n° 72, p. 11, Чернівці.
55. Русанівський, Віталій (1985). *Джерела розвитку східнослов'янських літературних мов* [Roussaniv's'kyu. Les origines des langues littéraires slaves orientales]. Наукова думка, Київ. Consulté en ligne : <http://irbis-nbuv.gov.ua/ulib/item/UKR0002759>
56. Русанівський, Віталій (2001). *Історія української літературної мови* [Roussaniv's'kyu. Histoire de la langue ukrainienne]. Артек, Київ. Consulté en ligne : https://shron1.chtyvo.org.ua/Rusanivskyy_VM/Istoriia_ukrainskoi_literaturnoi_movy.pdf
57. Сербенська, Олександра (1994). *Антисуржик. Вчимося ввічливо поводитись і правильно говорити* [Serbens'ka. Anti-sourjyk. Apprendre à se comporter avec courtoisie et à parler correctement]. Нац. ун-т ім. І. Франка, Львів. Consulté en ligne : https://chtyvo.org.ua/authors/Serbenska_Oleksandra/Antysurzhyk_Vchymosia_vvichlyvo_povydytys_i_pravylno_hovoryty/
58. Смаль-Стоцький С. (1923). *Розвиток поглядів на сем'ю слов'янських мов і їх взаємне споріднення* [Smal'-Stots'kyu. L'évolution des opinions sur la famille des langues slaves et leur parenté mutuelle]. Прага. Consulté en ligne : https://chtyvo.org.ua/authors/Smal-Stotskyi_Stepan/Rozvytok_pohliadiv_pro_simiu_slovianskykh_mov_i_ikh_vzaiemne_sporidnennia/
59. Смаль-Стоцький С. (1933). *Українська мова, її початки, розвиток та характеристичні прикмети* [Smal'-Stots'kyu. La langue ukrainienne, ses origines, son développement et ses caractéristiques]. Відбитка з «Дзвонів», Бібліос, Львів. Consulté en ligne : https://chtyvo.org.ua/authors/Smal-Stotskyi_Stepan/Ukrainska_mova_ii_pochatky_rozvytok_ta_kharakterystychni_ii_prykmety/
60. Старко, Василь, Ризін, Андрій (2020). *Великий електронний словник української мови (ВЕСУМ) як засіб NLP для української мови* [Starko et Rysin. Le grand dictionnaire électronique de la langue ukrainienne (VESUM) en tant qu'outil NLP pour la langue ukrainienne]. Dans *Галактика Слова. Галіні Макарівні Гнатюк*, pp. 135–141. Consulté en ligne : https://www.researchgate.net/publication/344842033_Velikij_elektronnij_slovník_ukrainskoi_movi_VESUM_ak_zasib_NLP_dla_ukrainskoi_movi_Galaktika_Slova_Galini_Makarivni_Gnatuk
61. Тараненко О. О. (2008). *Сучасні тенденції до перегляду нормативних засад української літературної мови і явище пуризму (на загальнослов'янському тлі)* [Tatarenko. Tendances modernes à la révision des fondements normatifs de la langue littéraire ukrainienne et phénomène du purisme (dans un contexte panslave)]. Dans *Мовознавство*, n° 2-30. Consulté en ligne : <http://dspace.nbuv.gov.ua/bitstream/handle/123456789/183048/15-Taranenko.pdf?sequence=1>
62. Тараненко О. О. (2013). *Мова української західної діаспори і сучасна мовна ситуація в Україні* [Tatarenko. La langue de la diaspora ukrainienne occidentale et la situation linguistique actuelle en Ukraine]. Dans *Мовознавство*, n°2-3. Consulté en ligne : http://nbuv.gov.ua/UJRN/MoZn_2013_2-3_7

63. Шахматов А. А. (1916). *Введение в курс истории русского языка. Ч. 1. Исторический процесс образования русских племен и наречий* [Chakhmatov. Introduction au cours d'histoire de la langue russe. Volume 1. Le processus historique de formation des tribus et dialectes russes]. Студентческая издательская книга, Петроград. Consulté en ligne : <http://elib.shpl.ru/ru/nodes/16637-shahmatov-a-a-vvedenie-v-kurs-istorii-russkogo-yazyka-ch-1-istoricheskiy-protsess-obrazovaniya-russkih-plemen-i-narechii-pg-1916#mode/inspect/page/5/zoom/4>
64. Шевельов, Юрій (1962). *Покоління двадцятих років в українському мовознавстві* [Cheveliov. La génération des années 20 dans la linguistique ukrainienne]. Dans *Збірник на пошану українських учених знищених большевицькою Москвою* [Une collection en l'honneur des scientifiques ukrainiens tués par la Moscou bolchevique]. Записки НТШ, Чикаго, pp. 309-332. Consulté en ligne : https://shron1.chtyvo.org.ua/Sheveliov_Yurii/Pokolinnia_dvadtsiatykh_rokiv_v_ukrainskomu_movoznavstvi.pdf
65. Шевельов, Юрій (1987). *Українська мова в першій половині двадцятого століття (1900-1941): стан і статус* [Cheveliov. La langue ukrainienne dans la première moitié du XXe siècle (1900-1941) : état et statut]. Сучасність, Нью-Йорк. Consulté en ligne : https://chtyvo.org.ua/authors/Sheveliov_Yurii/Ukrainska_mova_v_pershii_polovyni_dvadtsiatioho_stolittia_1900-1941_stan_i_status/
66. Шевельов, Юрій (1994). *Чому общерусский язык, а не вітчоруська мова? З проблем східнослов'янської глотогонії : статті про постання української мови* [Cheveliov. Pourquoi la langue russe commune et non la langue vitchoroussienne ? À propos des problèmes de la polysémie slave orientale : articles sur l'émergence de la langue ukrainienne]. КМ Academia, Київ. Consulté en ligne : <https://zbruc.eu/node/64725>
67. Шевельов, Юрій (2002). *Історична фонологія української мови* [Cheveliov. Phonologie historique de la langue ukrainienne]. Акта, Харків. Consulté en ligne : <http://litopys.org.ua/shevelov/shev.htm>
68. Шипанівська О. О. (2008). *Проблема виділення міжчастиномовних морфологічних омонімів* [Shypaniv's'ka. Le problème de l'identification des homonymes morphologiques intralinguistique]. Dans *Лексикографічний бюлетень*, n°17, pp. 83-89.
69. Чаварга, Анна. (2017). *Екстралінгвальні чинники розвитку української літературної мови на теренах історичного Закарпаття у 20-40-х рр. ХХ ст.* [Tchavarha. Facteurs extralinguistiques dans le développement de la langue littéraire ukrainienne dans la Transcarpatie historique dans les années 20-40 du XXe siècle]. Dans *Сучасні проблеми мовознавства та літературознавства*, n°22. Consulté en ligne : <https://dspace.uzhnu.edu.ua/jspui/handle/lib/21999>
70. Юшук І. П. (2000). *Українська мова* [Iouchtchouk. La langue ukrainienne]. Либідь, Київ.

Bibliographie en anglais et français

71. Alves D., Bekavac B., Tadić M. (2021). The Optimization of Portuguese Named-Entity Recognition and Classification by combining Local Grammars and Conditional Random Fields trained with a Parsed Corpus. In *Formalizing Natural Languages: Applications to Natural Language Processing and Digital Humanities. CCIS Series*, Springer-Verlag, Berlin Heidelberg. Consulté en ligne : <https://zenodo.org/records/7947846>

72. Aouini M. (2015). A NooJ Module for Named Entity Recognition in Middle French Texts. In *Formalising Natural Languages with NooJ 2014*. Monti J., Silberztein M., Monteleone M., Di Buono M., Cambridge Scholars Publishing, pp.99-112.
73. Olga, Artyushkina. (2022). Le phénomène du suržyk en Ukraine : deux langues en dialogue... ou en dispute ? In *Slovo. Langues en mouvement : changements linguistiques dans l'espace postcommuniste au tournant du XXIe siècle*, 52, pp.35-60. Consulté en ligne : <https://hal.science/hal-03685992/>
74. Bekavac B., Kocijan, Kristina, Tadić, Marko (2015). Near Language Identification Using NooJ. In *Formalising Natural Languages with NooJ 2014*. Monti J., Silberztein M., Monteleone M., Di Buono M., Cambridge Scholars Publishing, pp.152-165.
75. Bilaniuk, Laada (1997). Speaking of suzhyk: Ideologies and Mixed Languages. In *Harvard Ukrainian Studies*, n°21 (1/2), pp. 93-117.
76. Bilaniuk, Laada (2004). A typology of surzhyk: Mixed Ukrainian-Russian language. In *International Journal of Bilingualism*, vol. 8 (4), pp. 409-425.
77. Bilaniuk, Laada (2010). Language in the balance: the politics of non-accommodation on bilingual Ukrainian–Russian television shows. In *Walter de Gruyter GmbH & Co. KG*, pp. 105-133.
78. Boitet, Christian (2024). Preface. Linguistic Resources for Natural Language Processing - On the Necessity of Using Linguistic Methods to Develop NLP Software, by Max Silberztein, Springer, Switzerland.
79. Boujelben I., Mesfar S., Ben Hamadou A. (2010). Lexicalization of Arabic named entities: a methodological approach. In *Finite-State Language Engineering with NooJ 2009 International Conference (Tozeur, Tunisia)*. Edited by Abdelmajid Ben Hamadou, Slim Mesfar and Max Silberztein, Centre de publication Universitaire, Sfax, Tunisia, pp. 129-142.
80. Boyko, Olena (2023). *Comment l'Ukraine s'est battue pour la Crimée il y a un siècle*. En ligne uniquement : <https://tyzhden.fr/comment-l-ukraine-s-est-battue-pour-la-crim-ee-il-y-a-un-si-ecle/>
81. Chevalier G., Kasparian S., Silberztein. M. (2004). Éléments de solution pour le traitement automatique d'un français oral régional. Dans *TAL [Traitement automatique des langues]*, vol. 45, n°2, pp. 41-62.
82. Chomsky, Noam (1956). Three Models for the Description of Language, I.R.E. Transactions on Information Theory. In *IRE Transactions on Information Theory*, vol. 2, n° 3, pp. 113-124. Consulté en ligne : <https://ieeexplore.ieee.org/abstract/document/1056813>
83. Chomsky, Noam (1957). Syntactic Structures. In *Januar Linguarum, Series Minor 4*, The Hague, Mouton. Consulté en ligne : <https://www.degruyter.com/document/doi/10.1515/9783110218329.bm/html>
84. Chomsky, Noam, Miller, George A (1958). Finite State Languages. In *Information and Control*, n°1, pp. 91-112. Consulté en ligne : <https://www.sciencedirect.com/science/article/pii/S0019995858900822>

85. Chomsky, Noam, (1959a). On certain formal properties of grammars, *Information and Control*, n° 2/2, pp. 137-167. Consulté en ligne : <https://www.sciencedirect.com/science/article/pii/S0019995859903626>
86. Chomsky, Noam (1959b). A note on phrase structure grammars. In *Information and Control*, vol. 2, n° 4, pp. 393-395. Consulté en ligne : <https://www.sciencedirect.com/science/article/pii/S0019995859800176?via%3Dihub>
87. Chomsky, Noam (1963). Formal properties of grammars. In *Handbook of math. Psychology*, n° 2, pp. 328-418.
88. Chomsky, Noam, Schutzenberger, Marcel-Paul (1963). The algebraic theory of context-free languages. In *Computer Programming and Formal Dy stems, Studies in logic series*, Amsterdam, North-Holland, pp. 119-161.
89. Chomsky Noam, Schützenberger Marcel Paul (1968). Théorie algébrique des langages "context-free". In *Langages*, 3^e année, n°9, *Les modèles en linguistique*, sous la direction de Maurice Gross, pp. 77-118. Consulté en ligne : www.persee.fr/doc/lgge_0458-726x_1968_num_3_9_2362
90. Chomsky, Noam (1984). La connaissance du langage. In *Communications, Vol 40 : Grammaire générative et sémantique*, pp. 7-24. Consulté en ligne : https://www.persee.fr/doc/comm_0588-8018_1984_num_40_1_1596
91. Cirillo N. (2020). Dealing with Producing and Consuming Expressions in Italian Sentiment Analysis. In *Formalising Natural Languages with NooJ 2019 and its Natural Language Processing Applications*. Fehri, Mesfar, Silberztein, Springer CCIS Series, pp. 198-208.
92. Dickinson M. (2015). Detection of annotation errors in corpora. In *Language & Linguistics Compass*, vol. 9 (3), pp. 119-138.
93. Dmytrychyn, Iryna (2018). Fin des empires et naissance d'un État : le rêve ukrainien de Guillaume de Habsbourg. Dans *Austriaca*, n°87. Consulté en ligne : <http://journals.openedition.org/austriaca/413>
94. Ehmann B., Balázs L., Shved D., Bénet V., Gushin V. (2013). The Russian Linguistic Resources in Space Psychological Research. In *Formalising Natural Languages with NooJ 2012 International Conference (Paris, France)*. Edited by Anaïd Donabédian, Victoria Khurshudian and Max Silberztein, Cambridge Scholars Publishing, Newcastle, pp. 150-161.
95. Fehri H., Haddar K., Ben Hamadou A. (2011). Recognition and Translation of Arabic Named Entities with NooJ Using a New Representation Model. In *FSMNLP 2011*, France, pp. 134-142.
96. Fehri H., Haddar K., Ben Hamadou A. (2014). Co-reference Resolution Using NooJ Recognition Process of Arabic Named Entities. In *Formalising Natural Languages with NooJ 2013 : Selected Papers from the NooJ 2013 International Conference (Saarbrücken, Germany)*. Edited by Svetla Koeva, Slim Mesfar and Max Silberztein, Cambridge Scholars Publishing, Newcastle, pp. 129-140.
97. Fehri, H., Ben Messaoud, Ines (2020). Construction of Educational Games with NooJ. In *Formalising Natural Languages with NooJ 2019 and its Natural Language Processing Applications*. Fehri, Mesfar, Silberztein, Springer CCIS Series, pp. 173-184.

98. Flier, Michael S. (1998). Surzhyk: The Rules of Engagement. In *Harvard Ukrainian Studies*, 22, pp. 113–136. Consulté en ligne : <http://www.jstor.org/stable/41036734>
99. Georgiadou A., Karakos A., Papaioannou J. (2011). Learning the Greek language via Greeklish. In *Formalising Natural Languages with NooJ 2013 : Selected Papers from the NooJ 2013 International Conference (Saarbrücken, Germany)*. Edited by Svetla Koeva, Slim Mesfar and Max Silberstein, Cambridge Scholars Publishing, Newcastle, pp. 31-39.
100. Gucul-Milojević S., Radulović V., Krstev C. (2010). A View on the Representation of Women in Serbian Newspaper Texts. In *Applications of Finite-State Language Processing : Selected Papers from the NooJ 2008 International Conference (Budapest, Hungaria)*. Edited by Kuti Judit, Silberstein Max, Varadi Tamas, Cambridge Scholars Publishing, Newcastle, pp. 166-176.
101. Guiraud-Weber, Marguerite (1987). Oppositions aspectuelles et sémantisme verbal en russe. Dans *Revue des études slaves*, tome 59, fascicule 3, pp. 585-596.
102. Gwiazdecka, Ewa. (2009). Quelle description pour le préverbe polonais ? Dans *Cognitive Studies*, n° 9. pp. 243-255. Consulté en ligne : https://www.researchgate.net/publication/264277445_QUELLE_DESCRIPTION_POUR_LE_PREVERBE_POLONAIS
103. Gwiazdecka, Ewa, Bogacki, Krzysztof (2010). Grammatical categories for Polish resources in NooJ. Consulté en ligne : https://www.researchgate.net/publication/264288649_Grammatical_categories_for_Polish_resources_in_NooJ
104. Gwiazdecka, Ewa, Bogacki, Krzysztof (2012). Derivational structure of Polish verbs and the expansion of the dictionary. Consulté en ligne : https://www.researchgate.net/publication/264277367_Derivational_structure_of_Polish_verbs_and_the_expansion_of_the_dictionary
105. Gwiazdecka, Ewa, Bogacki, Krzysztof (2013). Disambiguating Polish Verbs of Motion, Cambridge Scholars Publishing, pp. 102-112. Consulté en ligne : https://www.researchgate.net/publication/264286856_Disambiguating_Polish_Verbs_of_Motion
106. Hetsevich Y., Skopinava A. (2014). Processing of Quantitative Expressions with Measurement Units in the Nominative, Genitive, and Accusative Cases for Belarusian and Russian. In *Proceedings of the 17th International Conference (Brno, Czech Republic, September 8-12, 2014)*. Eds. Petr Sojka, Aleš Horák, Ivan Kopeček, Karel Pala, Springer International Publishing, pp. 101-107.
107. Karcevski, Serge (1927). *Système Du Verbe Russe*. Essai de linguistique synchronique. Imprimerie Legiografie, Geneva.
108. Kasparian S. (2008). The Acadian NooJ module: Automatic Processing of a regional oral. In *Linguistica Atlantica*, n° 29, pp. 117-136.
109. Khachatryan L. (2012). Formalization of Proper Names in the Western Armenian Press. In *Formalising Natural Languages with NooJ : Selected Papers from the NooJ 2011 International Conference (Dubrovnik, Croatia)*. Edited by Kristina Vučković, Božo Bekavac and Max Silberstein, Cambridge Scholars Publishing, Newcastle, pp. 75-85.

110. Kilgarriff, Adam, Rychly, Pavel, Smrz, Pavel, Tugwellv David (2004). The Sketch Engine. In *Proceedings of the Eleventh EURALEX International Congress*, Lorient, France, pp. 105-116.
111. Kocijan K., Janjić, Marijana, Librenjak, Sara (2017). Recognizing Diminutive and Augmentative Croatian Nouns. In *Automatic Processing of Natural Language Electronic Texts with NooJ: Selected Papers from the International Conference NooJ 2016*, Springer CCIS, Series #667, pp. 23-36.
112. Kocijan K., Maria Pia di Buono, Mikić, Linda (2019). Detecting Latin-Based Medical Terminology in Croatian Texts. In *Formalising Natural Languages with NooJ 2018 and its Natural Language Processing Applications. I. Mauro Mirto, M. Monteleone, M. Silberstein*, Springer CCIS Series, pp. 38-49.
113. Kocijan K., Košković, Lucija, Bajac, Petra (2020). Detecting Hate Speech Online: A Case of Croatian. In *Formalising Natural Languages with NooJ 2019 and its Natural Language Processing Applications. Fehri, Mesfar, Silberstein*, Springer CCIS Series, pp. 185-197.
114. Kocijan K., Šojat K. (2022). Negation Usage in the Croatian Parliament. In *Formalizing Natural Languages: Applications to Natural Language Processing and Digital Humanities. Bigey M., Richeton A., Silberstein M. Thomas I.*, Springer-Verlag, Berlin Heidelberg, pp. 101-113
115. Krzysztof Bogacki, Agnieszka Dryjańska (2019). Les prénoms et les patronymes dans les ressources dictionnaires pour le traitement automatique du polonais par NooJ, In *Strona domowa*, n° 19. Consulté en ligne : czasopisma.filologia.uwb.edu.pl
116. Mazon, André. (1913) La notion morphologique de l'aspect des verbes chez les grammairiens russes. Dans *Mélanges offerts à M.Emelie Picot*, Paris, E.RAkir, pp.343-367.
117. Merkler D., Berović D., Agić Ž. (2012). Disambiguation of Homographic Adjective and Adverb Forms in Croatian Texts. In *Formalising Natural Languages with NooJ : Selected Papers from the NooJ 2011 International Conference (Dubrovnik, Croatia). Edited by Kristina Vučković, Božo Bekavac and Max Silberstein*, Cambridge Scholars Publishing, Newcastle, pp. 142-151.
118. Mesfar S. (2007). Named Entity Recognition for Arabic Using Syntactic Grammars. In *Proceedings 12th International Conference NLDB 2007*, LNCS Series, Springer Verlag, France, pp. 305-316.
119. Naber, Daniel. (2003). A Rule-Based Style and Grammar Checker. Memoire de fin d'étude, Bielefeld University, Bielefeld. Consulté en ligne : https://danielnaber.de/language/tool/download/style_and_grammar_checker.pdf
120. Pelosi S., Alessandro Maisto (2015). A Lexicon-based Approach to Sentiment Analysis. The Italian Module for NooJ. In *Formalising Natural Languages with NooJ 2014. Monti J., Silberstein M., Monteleone M., Di Buono M.*, Cambridge Scholars Publishing, pp. 226-231.
121. Pignot H., Lardy M. (2013). Formalising a Dictionary of 17th Century English with NooJ. In *Formalising Natural Languages with NooJ : Selected Papers from the NooJ 2012 International Conference (Paris, France). Edited by Anaïd Donabédian, Victoria Khurshudian and Max Silberstein*, Cambridge Scholars Publishing, Newcastle, pp. 37-47.
122. Piton O. (2004). À propos d'une expérience d'enseignement avec Intex. In *INTEX pour la Linguistique et le Traitement Automatique des Langues. Edited by Claude Muller, Jean*

Royauté and Max Silberztein. *Cahiers de la MSH Ledoux*, Presses Universitaires de Franche-Comté, France, pp. 331-358.

123. Piton O., Lagji K. (2008). Morphological Study of Albanian Words, and Processing with NooJ. In *Proceedings of the 2007 International NooJ Conference (Barcelona, Spain)*. Edited by Xavier Blanco and Max Silberztein, Cambridge Scholars Publishing, Newcastle, pp. 189-205.
124. Piton O., Pěrnaska R. (2013). Variations Lexicographiques en Albanais Contemporain, à l'Épreuve du Tal. Proceedings of 7th International Corpus Linguistics Conference, Lorient JLC 2013. Dans *Texte et corpus*, n°5, France. Consulté en ligne : <https://hal.science/hal-01088895/document>
125. Poibeau, Thierry (2014). La linguistique est-elle soluble dans la statistique ? Dans *Revue Sciences/Lettres [En ligne]*, n° 2. Consulté en ligne : <http://journals.openedition.org/rsi/402>
126. Rodrigo, Andrea, Méndez, Bárbara, Oliva, Ivan (2021). Impacto de las TIC en la enseñanza: hacia un cambio de paradigma en la clase de lengua y literatura, COVID-19. In *Aprendo con NooJ*. Consulté en ligne : <https://aprendoconnooj.unr.edu.ar/index.php/revista/article/view/2>
127. Rodrigo, Andrea, Harriet, Soledad (2022). Avances en la morfología del adjetivo en español y en francés: aplicación pedagógica de la plataforma NooJ. *Aprendo con NooJ*. Consulté en ligne : https://www.researchgate.net/publication/360666475_Avances_en_la_morfologia_del_adjetivo_en_espanol_y_en_frances_aplicacion_pedagogica_de_la_plataforma_NooJ
128. Saïdi, Victoriya (2009). Le problème de la nomination de la langue ukrainienne. Dans *Études de lettres*, n°4. Consulté en ligne : <http://journals.openedition.org/edl/315>
129. Saint-Joanis, Olena, Silberztein, Max (2021). Formalisation de la relation entre les verbes imperfectifs et perfectifs en ukrainien. Dans *Traitement Automatique des Langues Naturelles*, Lille, France. pp.171-178.
130. Schmid, Helmut (1994). [Probabilistic Part-of-Speech Tagging Using Decision Trees](#). In *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK.
131. Schmid, Helmut (1995). [Improvements in Part-of-Speech Tagging with an Application to German](#). In *Proceedings of the ACL SIGDAT-Workshop*, Dublin, Ireland.
132. Schmid, Helmut (2019). [Deep Learning-Based Morphological Taggers and Lemmatizers for Annotating Historical Texts](#), In *DATECH, May*, Brussels, Belgium.
133. Seideh Aly Fall M., Fehri, Hela, Haddar Kais (2017). Recognition and Extraction of Latin Names of Plants for Matching Common Plant Named Entities. In *Automatic Processing of Natural Language Electronic Texts with NooJ: Selected Papers from the International Conference NooJ 2016*, Springer CCIS Series #667, pp. 132–144.
134. Sériot, Patrick. (2005). Diglossia, bilingualism or mixture of languages: The case of surzyk in the Ukraine. In *La linguistique*, n°41., pp. 37-52.
135. Silberztein, Max (2005). NooJ: a Linguistic Annotation System for Corpus Processing. In *Proceedings of HLT/EMNLP 2005 Interactive Demonstrations*, Association for Computational Linguistics, Vancouver, British Columbia, Canada, pp. 10–11.

136. Silberztein, Max (2015). Formalisation des langues. L'approche de NooJ. Université de Franche-Comté, ISTE Edition.
137. Silberztein, Max (2019). Les outils informatiques au service des linguistes : présentation. Dans *Langue française*, n° 203. Consulté en ligne : <https://www.cairn.info/revue-langue-francaise-2019-3-page-7.htm>
138. Silberztein, Max (2024). Linguistic Resources for Natural Language Processing - On the Necessity of Using Linguistic Methods to Develop NLP Software. Springer, Switzerland.
139. Starko, Vasyl (2020). Semantic Annotation for Ukrainian: Categorization Scheme, Principles, and Tools. In : *COLINS 2020*, pp. 239-248.
140. Starko V., Rysin A., Chvedova M. (2021a). Ukrainian Text Preprocessing in GRAC, In *2021 IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT)*, LVIV, Ukraine, 2021, pp. 101-104. Consulté en ligne : <https://ieeexplore.ieee.org/document/9648705>
141. Starko, Vasyl, Rysin Andriy (2022). VESUM: A Large Morphological Dictionary of Ukrainian As a Dynamic Tool. In *Vasyl Lytvyn, Natalia Sharonova, Izabela Jonek-Kowalska, Agnieszka Kowalska-Styczen, Victoria Vysotska, Yevhen Kupriianov, Olga Kanishcheva, Olga Cherednichenko, Thierry Hamon, Natalia Grabar, editors, Proceedings of the 6th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2022)*, vol. I: Main Conference, Gliwice, Poland, May 12-13, 2022, Vol. 3171 of CEUR Workshop Proceedings, pp. 61-70. CEUR-WS.org, 2022. Consulté en ligne : <https://ceur-ws.org/Vol-3171/paper8.pdf>
142. Stoyanova, I., Lesseva, S. (2008). The Treatment of Named Entities in Machine Translation. In *Proceedings of the 2007 International NooJ Conference (Barcelona, Spain)*. Edited by Xavier Blanco and Max Silberztein, Cambridge Scholars Publishing, Newcastle, pp. 254-272.
143. Shvedova M. (2020). The General Regionally Annotated Corpus of Ukrainian (GRAC, uacorus.org): Architecture and Functionality. In *CEUR Workshop Proceedings. Proceedings of the 4th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2020)*, vol. I: Main Conference. Lviv, Ukraine, April 23-24, 2020. pp. 489-506. Consulté en ligne : <https://ceur-ws.org/Vol-2604/paper36.pdf>
144. Starko V., Rysin A., Shvedova M. (2021b). Handling of Nonstandard Spelling in GRAC. In *IEEE 16th International Conference on Computer Sciences and Information Technologies, CSIT 2021*, Lviv, Ukraine, September 22-25, 2021. pp. 105-108. Consulté en ligne : <https://ieeexplore.ieee.org/document/9648834>
145. Shvedova M., von Waldenfels R. (2021). Regional Annotation within GRAC, a Large Reference Corpus of Ukrainian: Issues and Challenges In *CEUR Workshop Proceedings. Proceedings of the 5th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2021)*, vol. I: Main Conference. Kharkiv, Ukraine, April 22-23, 2021. pp. 32-45. Consulté en ligne : <https://ceur-ws.org/Vol-2870/paper4.pdf>
146. Shvedova, Maria, Prydvorova, Nataliia, Prydvorova, Ilona. (2022) Normalization of Early Modern Ukrainian in GRAC: the Case of Lesia Ukrainka's Works. In *CEUR Workshop Proceedings 2022*. pp. 71-80. Consulté en ligne : <https://ceur-ws.org/Vol-3171/paper9.pdf>

147. Sychak, Olena, Starko, Vasyl. (2022). Ukrainian Feminine Personal Nouns in Online Dictionaries and Corpora. User Modeling and User-Adapted Interaction. In *COLINS-2022: 6th International Conference on Computational Linguistics and Intelligent Systems, May 12–13, 2022*. Gliwice, Poland. Consulté en ligne : https://www.researchgate.net/publication/362345324_Ukrainian_Feminine_Personal_Nouns_in_Online_Dictionaries_and_Corpora
148. Tutin A., Silberztein M. (2004). NooJ : un outil de TAL de corpus pour l'enseignement des langues et de la linguistique. In *Actes de la journée ATALA TAL et apprentissage des langues*, Grenoble. Consulté en ligne : <https://journals.openedition.org/alsic/336?file=1>
149. Vitas D., Pavlović-Lažetić G. (2007). Extraction of named entities in Serbian using INTEX. In *Formaliser les langues avec l'ordinateur : De INTEX à NooJ*. Edited by Svetla Koeva, Denis Maurel and Max Silberztein. Cahiers de la MSH Ledoux, Presses Universitaires de Franche-Comté, France, pp. 281-304.
150. Volokh A., Neumann G. (2011). Automatic detection and correction of errors in dependency tree-banks. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, vol. 2, Stroudsburg, Association for Computational Linguistics, pp. 346-350.
151. Watabe M. (2018a). A NooJ dictionary for the romani language: toward a NooJ-relevant sorting of morphosyntactic tags. In *Formalising Natural Languages with NooJ and Its Natural Language Processing Applications*. Edited by Mbarki S., Mourchid M., Silberztein M., Springer CCIS Series #811, pp. 3-15.
152. Watabe M. (2018b). The NooJ approach of automatic language processing as a tool for systematization of romani grammar in both description and formal teaching. In *Preservation, Safeguarding and Prospects of the Romani Language in Serbia*. Edited by Varady T., Sikimić B. Serbian Academy of Sciences and Arts, Belgrade, Serbia, pp. 175-201.
153. Włodarczyk, Hélène (1984). Les catégories grammaticales et lexicales dans l'énoncé : l'aspect et les modalités d'action des langues slaves. Dans *Revue des études slaves*, tome 56, fascicule 2, 1984. Grammaire et lexicque. pp. 271-278. Consulté en ligne : https://www.persee.fr/doc/slave_0080-2557_1984_num_56_2_5408
154. Włodarczyk, André, Włodarczyk, Hélène (2001). La Préfixation verbale en polonais. Dans *Études Cognitives / Studia Kognitywne*, n° 4, SOW, Warszawa, pp. 93-109.

Ressources en ligne

- Encyclopédie de l'Ukraine (2022) (version en ligne) // Dziuba, A. I. et alii, L'Académie nationale des sciences d'Ukraine. © 2024 : <https://esu.com.ua/>
- Encyclopédie d'histoire d'Ukraine (2019) (version en ligne) // L'Académie nationale des sciences d'Ukraine. © 2016 -2024 : <http://resource.history.org.ua/>
- GRAC, General Regionally Annotated Corpus of Ukrainian // M. Shvedova, R. von Waldenfels, S. Yaryhin, A. Rysin, V. Starko, T. Nikolayenko, et alli.
© 2017-2024 : <https://uacorporus.org/Kyiv/en>. Interfaces pour GRAC :

- KonTex (GRAC 10 et 11) : <https://parasol.vmguest.uni-jena.de/grac/corpora/corplist>
- NoSketch Engine (GRAC 10 à 17) : https://parasol.vmguest.uni-jena.de/grac_crystal/#dashboard?corpname=grac17
- Interface graphique (CQL) pour GRAC // Nikolajenko, Tymofiy : https://parasol.vmguest.uni-jena.de/grac_batch/NGyears.html
- Goroh, portail linguistique (plusieurs dictionnaires). © 2024 : <http://goroh.pp.ua>
- MOVA.info, portail linguistique (dictionnaires, corpus, normes orthographiques, manuels) // Dartchouk, Nataliia et alii. © 2003-2024 : <http://www.mova.info>
- LanguageTool, module ukrainien // Rysin, Andriy et alii. © 2024 : <https://community.languagetool.org/analysis?lang=uk>
- NooJ, environnement linguistique multilingue // Silberztein, Max : <https://nooj.univ-fcomte.fr/>
- Normes orthographiques 2019 : <https://mon.gov.ua/storage/app/media/zagalna%20serednya/%202019.pdf>
- RNNagger // Schmid, Helmut: <https://www.cis.uni-muenchen.de/~schmid/tools/RNNagger/>
- Sketch Engine // Adam Kilgarriff, Pavel Rychly, Pavel Smrz et David Tugwell : <https://auth.sketchengine.eu/>
- SUM, le grand dictionnaire académique ukrainien en ligne (1970-1980). © 2023 : <https://sum.in.ua>
- SUM-20, le grand dictionnaire ukrainien en ligne (volume 1 à 14) // L'Institut linguistique de l'Académie nationale des sciences d'Ukraine. © 2014 – 2024 : <https://sum20ua.com/>
- TreeTagger // Schmid, Helmut : <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>
- Powers and Hungary And Protocol and Declaration, Signed at Trianon June 4, 1920 : https://wwi.lib.byu.edu/index.php/Treaty_of_Trianon
- VESUM, le grand dictionnaire électronique ukrainien // Rysin, Andryi, Starko, Vasyl. ©2005-2022 : <https://vesum.nlp.net.ua/>
module à télécharger : https://github.com/brown-uk/dict_uk
- Ukrainian Dictionary Open Source, version 2.9.1 // Rysin, Andryi. © 2016 : http://extensions.services.openoffice.org/en/project/dict-uk_UA
- ULIF, portail linguistique // L'Institut linguistique de l'Académie nationale des sciences d'Ukraine. © 2001 – 2023 : <https://lcorp.ulif.org.ua>

