UD Treebanks for Esperanto as a Natural Language

Masanori Oya
School of Global Japanese Studies,
Meiji University
masanori_oya2019@meiji.ac.jp

Abstract

This paper describes the details of UD-based morphological and syntactic annotations on Esperanto texts to construct its small-scale UD treebank. Though it was created as an international auxiliary language, Esperanto has increasingly been studied as a natural language both in linguistics and in NLP. This paper introduces the detail of the manual annotation of UD morphological and relational tags and describes how the frequencies of these tags differ across the treebanks and discusses the possibility of future research of Esperanto as a natural language.

1 Introduction

Esperanto is a constructed language created by L.L. Zamenhof in 1887 for the purpose of international communication. Since its creation, it has been spreading across the globe, attracting many language enthusiasts and creating a variety of communities using it as a means of communication. Though it has been criticized for several reasons, it has now developed into one of the natural languages which is worthy of linguistic research and natural language processing (NLP). In this context surrounding Esperanto, this paper introduces small-scale UD treebank for Esperanto. This is the first contribution of Esperanto to UD community (if not the first attempt to construct an Esperanto treebank (Bick 2018)), and it is expected to facilitate not only NLP research of Esperanto but linguistic research in general.

This paper is organized as follows: Section 2 is a summary of previous studies which treat Esperanto not as an artificial language but as one of the natural languages, and those which approach it from the viewpoint of NLP. Section 3 summarizes the process of annotating Esperanto texts with UD POS tags and UD relation tags, along with a brief description of its morphological and syntactic features. Section 4 is a discussion on the issue of how UD POS tags and UD relation tags are used in different genres of Esperanto texts, for the further development of future research on morphological and syntactic properties of Esperanto as a natural language.

2 Previous studies of Esperanto

2.1 Esperanto in linguistics

Goodall (2023) summarizes the history of the relationship between Esperanto and linguistics. He points out decades of hostilities between them due to the overemphasis on the role of native speakers' intuition on naturalness of language. He claims that Esperanto today can be one of the topics of ordinary linguistic research. Moreover, even though Esperanto was constructed as an international auxiliary language, there are researchers who consider it as one of the natural languages and investigate its linguistic characteristics which are similar with those of other natural languages. In Gledhill (2000), which is a comprehensive grammar of Esperanto, he pointed out that it has developed as a natural language, going beyond Zamenhof's original design. Manaris, Pellicoro, Pothering and Hodges (2006) applied Zipf-based metrics (word distribution, word-distance distribution, word-bigram distribution, etc.) to six languages including Esperanto, and found similar statistical proportions between Esperanto and the other five languages. Parkvall (2010)'s typological study based on the World Atlas of Language Structures (Haspelmath 2005, henceforth WALS) revealed that Esperanto has certain features characteristic to European languages, yet it is not a typical one. Stria (2015) assessed Esperanto as a natural language in terms of a variety of criteria. Koutny (2015) conducted a

Proceedings of the Eighth Workshop on Universal Dependencies (UDW, SyntaxFest 2025), pages 22–29

typological description of Esperanto on WALS database and concluded that Esperanto has characteristics which make it accessible for people with different linguistic backgrounds.

As these studies indicate, Esperanto in the 21st century can be considered as a natural language which deserves linguistic investigation just like all the other natural languages, even though its status as an international auxiliary language may not be as realistic as it was intended to be when it was created in the 19th century.

2.2 Esperanto in NLP research

Along with the trend where Esperanto has been one of the topics of linguistic research, several researchers have focused on Esperanto from the point of view of natural language processing (NLP). Minnaja and Paccagnella (2000) proposed a PoS tagger for Esperanto. Bick (2007, 2009) introduced *EspGram*, a Constraint-Grammar-based parser for Esperanto. Bick (2016) dealt with the internal structure of complex Esperanto words (CWs) and reported the construction of a CW dictionary. Bick (2019) presented a method for the automatic generation and semantic evaluation of Esperanto sentences for pedagogical purposes. Poncelas, Buts, Hadley and Way (2020) focused on Esperanto for improvements of the performance of a machine translation system for low-resource languages. Bick (2018, 2020) introduced *Arbobanko*, the first treebank for Esperanto. Arbobanko is a dependency treebank with 52,000 tokens. The texts were taken from *Monato,* an Esperanto news magazine, randomly selected from the year 2000-2010. Words in Arbobanko are annotated with tags for several features, such as lemma, part-of-speech, inflection, syntactic function, dependency links, and verb frames categories. Arbobanko also provides CoNLL-format version, yet it is not annotated with UD Pos tags and UD relation tags.

Since the POS and relation tags in Arbobanko are created only for Esperanto, it cannot be employed for multi-lingual comparison and contrast straightforwardly. We need multi-lingual treebank for that purpose, in which different languages are annotated with the same POS and relation tags, in order to develop Esperanto NLP and linguistic investigations.

3 This work: UD Treebank for Esperanto

3.1 Data

In the context mentioned in the previous section, this study attempts to construct a UD treebank for Esperanto. The first text chosen for this study is *Manifesto de Prago* (Prague Manifesto), drafted at the World Esperanto Congress in Prague in 1996. This manifesto promotes more democratic communication, language rights not only for Esperanto but for all languages on this planet, importance of language diversity and its preservation, and language education facilitated by Esperanto. The original text is available online (https://uea.org/teko/praga_manifesto/pm_esperanto). For this study, the text was annotated according to the UD Guidelines manually by the author of this study; automatic parsing was not conducted because the parsers for Esperanto available at present do not yield parse output in the format of CoNLL-U, which means that we need to convert the output into CoNLL-U, manually or automatically, and this is not the main topic of this study (yet this will be a topic of future study, naturally). The annotated text, *eo_prago-ud-text.conllu* (henceforth *Prago*), contains 1023 lines.

The second text chosen for this study is the Cairo CICLing sentences (https://github.com/UniversalDependencies/cairo) translated from the original English into Esperanto, following the UD guideline for a new treebank (https://universaldependencies.org/release_checklist.html). The annotated text, *eo_cairo-ud-text.conllu* (henceforth *Cairo_eo*), contains 257 lines. Table 1 summarizes the number of tokens, types, and type-token ratios of these texts. Though Prago is more than four times larger than Cairo in terms of the number of tokens, their type-token ratios do not differ substantially.

	Tokens	Types	T/T Ratio
Prago	839	321	.383
Cairo_eo	256	103	.402

Table 1: The numbers of types, tokens, and type-token ratio of *eo_prago_ud-test.conllu* (indicated as Prago) and *eo_cairo_ud-text.conllu* (indicated as Cairo_eo).

	Prago		*Cairo_eo*		*Cairo_en*	
	freq.	%	freq.	%	freq.	%
ADJ	108	12.87	4	2.26	5	2.67
ADP	124	14.78	5	2.82	6	3.21
ADV	35	4.17	13	7.34	7	3.74
AUX	18	2.15	6	3.39	12	6.42
CCONJ	32	3.81	8	4.52	10	5.35
DET	75	8.94	10	5.65	20	10.70
NOUN	196	23.36	23	12.99	24	12.83
NUM	10	1.19	1	0.56	0	0.00
PART	4	0.48	3	1.69	6	3.21
PRON	44	5.24	24	13.56	22	11.76
PROPN	8	0.95	14	7.91	15	8.02
PUNCT	104	12.40	28	15.82	25	13.37
SCONJ	12	1.43	5	2.82	3	1.60
VERB	69	8.22	33	18.64	32	17.11
	839		177		187	

Table 2: The frequencies and the percentages of the UPOS tags in *eo_prago_ud-test.conllu* (indicated as *Prago*), *eo_cairo_ud-text.conllu* (indicated as *Cairo_eo*), and *cairo/en.conllu* (indicated as *Cairo_en*)

3.2 UPOS and morphological annotation of Esperanto

This section summarizes the part-of-speech and morphological features of Esperanto along with brief explanations on how they are annotated in the Esperanto text.

According to Zamenhof (1887), an international auxiliary language should be easy to learn for anybody, and therefore the grammar of Esperanto is characterized by a high degree of regularity. As a result of this, it contains only a few exceptions (of course, the fact that there are only a small number of exceptions in a language does not necessarily mean that the language is simple and easy to learn).

Zamenhof's idea of regularity (if not simplicity) is reflected in the fact that the word-ending derivational morpheme of Esperanto indicates the part of speech of the majority of Esperanto words (Zamenhof 1887). Nouns end with *-o* (when nominative singular), infinitive verbs with *-i*, adjectives with *-a*, and adverbs with *-e* (Zamenhof 1887). For example, with the root *parol-*, *parolo* means "speaking", *paroli* means "to speak", *parola* means "oral" or "spoken", and *parole* means "verbally."

Esperanto uses all the UPOS tags (universal part-of-speech tags), and no XPOS tags (optional language-specific part-of-speech/morphological tags). For the details of UPOS and XPOS tags, refer to the Web page of CoNNL-U format (https://universaldependencies.org/format.html). The table below summarizes the frequencies and the percentages of the UPOS tags in *Prago* and *Cairo_eo*, and those in Cairo/en.conllu, for comparison:

The most obvious differences between Prago and Cairo are the percentage of NOUNs, PRONs, and VERBs. The percentage of ADV in Cairo is almost twice as high as that in Cairo_en, and that of DET in Cairo_en is almost twice as high as that in Cairo.

Table 3 summarizes the morphological features used in Esperanto, categorized according to UPOS tags:

UPOS	Features
ADJ	Case=Acc, Nom; Degree=Pos; Number=Plur, Sing
AUX	Mood=Imp, Ind, Sub; Tense=Fut, Past, Pres; VerbForm=Fin, Inf; Voice=Act, Pass
DET	Definite=Def; PronType=Art, Dem, Tot
NOUN	Case=Acc, Nom; Number=Plur, Sing
NUM	NumForm=Word; NumType=Card
PRON	Case=Acc, Nom; Number=Plur, Sing; Number[psor]=Plur, Sing; Person=1, 2, 3; Poss=Yes; PronType=Dem, Int, Prs, Rel, Tot; Reflex=Yes
PROPN	Case=Acc, Nom; Number=Plur, Sing
VERB	Mood=Imp, Ind, Sub; Tense=Fut, Past, Pres; VerbForm=Fin, Inf; Voice=Act, Pass

Table 3: The morphological features of Esperanto

Suffixes on nouns and adjectives indicate their case and number, which are summarized in Table 4 and 5:

	Singular	Plural
Nominative	*-o*	*-oj*
Accusative	*-on*	*-ojn*

Table 4: Esperanto nominal suffixes (based on Zamenhof 1887)

	Singular	Plural
Nominative	*-a*	*-aj*
Accusative	*-an*	*-ajn*

Table 5: Esperanto adjectival suffixes (based on Zamenhof 1887)

For example, in an Esperanto phrase *interesaj artikoloj* "interesting articles", the noun *artikoloj* is annotated with "Case=Nom|Number=Plur," and the adjective *interesaj* is annotated with "Case=Nom|Degree=Pos|Number=Plur."

DET is annotated on the definite article *la* (Esperanto does not have indefinite articles), and some correlatives when preceding a noun, such as

tiu artikolo “that book”, *tiuj artikoloj* “those books”, or *kies artikolo* “whose book”. The definite article *la* is annotated with “Definite=Def|PronType=Art.”

Esperanto pronouns are annotated with POS tag PRON, and with morphology tags, according to their number, person, gender (for 3rd person), and case. For example, the morphology tag on the 3rd-person masculine pronoun *li* is annotated with “Case=Nom|Gender=Masc|Number=Sing|Person=3rd|PronType=Prs.”

Some correlatives used not as determiners but as demonstrative pronouns are annotated with POS tag PRON. For example, the correlative *tiuj* in *Tiuj estas interesaj* “Those are interesting” functions as a pronoun, and it is annotated with “Case=Nom|Number=Plur|PronType=Dem.”

Esperanto verbs do not agree with their subject, and they inflect according to their mood and tense, summarized in the table below:

	suffixes
Infinitive	*-i*
Present	*-as*
Past	*-is*
Future	*-os*
Subjunctive	*-us*
Imperative	*-u*

Table 6: Inflectional morphemes of Esperanto verbs (based on Zamenhof 1887)

For example, the verb *verkis* in a sentence *Mi verkis tiujn artikolojn* “I wrote those articles” is annotated with “Mood=Ind|Tense=Pres|VerbForm=Fin.”

Derivational morphemes for participles and gerunds are summarized in the table below:

	Past	Present	Future
Active	*-int-*	*-ant-*	*-ont-*
Passive	*-it-*	*-at-*	*-ot-*

Table 7: Derivational morphemes of Esperanto verbs (based on Zamenhof 1887)

These morphemes are followed by another morpheme for part of speech (nominal *-o*, adjectival *-a*, and adverbial *-e*), and for plural *-j*.

Gerunds and participles are annotated with not only those tags related to verbs but also those related either to nouns or to adjectives. First, gerunds are accompanied by nominal morphemes indicating their case and number, so they are annotated with those tags related to nouns and verbs. For example, the gerund *lernantoj* “learners” is derived from a verb root *lern-*, with the active present *-ant-*, the nominal morpheme *-o*, and the plural morpheme *-j*. Therefore, it is annotated with “Case=Nom|Number=Plur|Tense=Pres|VerbForm=Part|Voice=Act”.

Adjectival participles agree with the noun they modify when used as modifiers or with the subject noun when used as a predicate with the copula *esti*, so they are annotated with those tags related to adjectives and verbs. For example, in *La lernantoj estas legantaj la librojn* “The learners are reading the books”, the present active plural adjectival participle *legantaj* is annotated with “Case=Nom|Number=Plur|Tense=Pres|Verbform=Part|Voice=Act”.

Adverbial participles do not agree with case and number, so they are annotated with those tags related to verbs. For example, in *Legante la libron, la lernantob manĝas* “Reading the book, the learner is eating”, the adverbial participle *Legante* is annotated with “Tense=Pres|VerbForm=Part|Voice=Act”.

The fact that gerunds in Esperanto have often been lexicalized (e.g. *Esperanto* is originally a present active gerund derived from the verb root *esper-* “to hope” plus present active *-ant-* and nominal *-o*, meaning “one who hopes”) while participles are not always lexicalized leads us to categorize gerunds as nouns while participles as verbs, and POS annotation follows this principle.

3.3 Syntactic annotation

Esperanto allows free word-order (e.g., a corpus investigation by Gledhill (2000)). The case morphemes on nouns and adjectives indicate the grammatical relationship between verbs and their direct objects, so changing the order of the verb and its subject and object in a sentence does not change their grammatical relationships. For example, consider an example sentence *Mi verkis tiujn artikolojn* “I wrote those articles”. The object can precede the subject and the verb, such that *Tiujn artikolojn mi verkis*, without changing the grammatical relationship between the verb and its object noun phrase.

Kråkmo (2022) conducted a survey which showed that Esperanto prefers the SVO word order,

though this fact does not mean that other word orders such as SOV are ungrammatical in Esperanto.

The gloss and the UD syntactic annotation on an example sentence *Mi verkis tiujn artikolojn, kiuj ne altiris la atenton de aliaj esploristoj.* "I wrote those articles which did not attract other researchers' attention." are shown below:

(1) *Mi verk-is tiu-j-n artikolo-j-n, kiu-j ne altir-is la atento-n de alia-j esploristo-j.*
I write-*pst* those-*pl*-*acc* article-*pl*-*acc,* which-*pl* not attract-*pst* the attention-*acc* of other-*pl* researcher-*pl*

```
# sent_id = 1
# text = Mi verkis tiujn artikolojn, kiuj altiris la atenton de aliaj esploristoj.
# text_en = I wrote those articles which did not attract other researchers' attention.
1   Mi          mi          PRON    _   Case=Nom|Number=Sing|Person=1|PronType=Prs  2   nsubj   _   _
2   verkis      verki       VERB    _   Mood=Ind|Tense=Past|VerbForm=Fin            0   root    _   _
3   tiujn       tiu         DET     _   Case=Acc|Number=Plur|PronType=Dem           4   det     _   _
4   artikolojn  artikolo    NOUN    _   Case=Acc|Number=Plur                        2   obj     _   _
5   ,           ,           PUNCT   _   _                                         8   punct   _   _
6   kiuj        kiu         PRON    _   Case=Nom|Number=Plur|PronType=Rel         8   nsubj   _   _
7   ne          ne          ADV     _   _                                         8   advmod  _   _
8   altiris     altiri      VERB    _   Mood=Ind|Tense=Past|VerbForm=Fin          4   acl:relcl
9   la          la          DET     _   _                                         10  det     _   _
10  atenton     atento      NOUN    _   Case=Acc|Number=Sing                       8   obj     _   _
11  de          de          ADP     _   _                                         13  case    _   _
12  aliaj       alia        ADJ     _   Case=Nom|Number=Plur|Degree=Pos           13  amod    _   _
13  esploristoj esploristo  NOUN    _   Case=Nom|Number=Plur                       10  nmod    _   _
14  .           .           PUNCT   _   _                                         2   punct   _   _
```

Figure 1: The UD annotation on an example sentence *Mi verkis tiujn artikolojn, kiuj altiris la atenton de aliaj esploristoj.* "I wrote those articles which did not attract other researchers' attention."

In the example above, the pronoun *Mi* depends on the verb *verkis* as its subject. The determiner *tiujn* agrees with the noun *artikolojn* in terms of case (accusative) and number (plural). The relative pronoun *kiuj* refers to *artikolojn*, and it depends on the verb *altiris* as its subject. The negative adverb *ne* depends on the verb *altiris,* which depends on the noun *artikolojn* as a relative clause. This dependency is typed as *acl:relcl*. The noun *atenton* depends on the verb *altiris* as its object. The preposition *de* depends on the noun *esploristoj* with the dependency type *case*, and the noun *esploristoj* depends on the noun *atenton* with the dependency type *nmod*.

The table below summarizes the frequencies and the percentages of the UPOS tags in Prago and Cairo:

	Prago		*Cairo_eo*		*Cairo_en*	
	Freq.	%	Freq.	%	Freq.	%
acl	11	1.31	0	0	1	0.53
acl:relcl	8	0.95	1	0.56	0	0
advcl	4	0.48	2	1.13	1	0.53
advmod	35	4.17	13	7.34	5	2.67
amod	96	11.44	1	0.56	1	0.53
appos	3	0.36	3	1.69	1	0.53
aux	1	0.12	2	1.13	9	4.81
aux:pass	1	0.12	2	1.13	2	1.07
case	121	14.42	6	3.39	8	4.28
cc	29	3.46	8	4.52	10	5.35
cc:preconj	3	0.36	0	0	0	0
ccomp	8	0.95	2	1.13	3	1.60
compound	0	0	1	0.56	2	1.07
conj	34	4.05	8	4.52	7	3.74
cop	16	1.91	2	1.13	4	2.14
det	78	9.30	10	5.65	20	10.7
mark	16	1.91	5	2.82	5	2.67
mwe	0	0	0	0	1	0.53
name	0	0	0	0	2	1.07
neg	3	0.36	2	1.13	2	1.07
nmod	94	11.2	5	2.82	9	4.81
nmod:poss	4	0.48	5	2.82	0	0
nsubj	55	6.56	22	12.43	23	12.30
nsubj:pass	1	0.12	2	1.13	2	1.07
nummod	1	0.12	0	0	0	0
obj	31	3.69	14	7.91	12	6.42

obl	29	3.46	0	0	0	0
orphan	0	0	3	1.69	0	0
parataxis	0	0	1	0.56	0	0
punct	104	12.4	28	15.82	25	13.37
remnant	0	0	0	0	6	13.37
root	46	5.48	20	11.3	20	10.7
vocative	0	0	1	0.56	1	0.53
xcomp	7	0.83	8	4.52	5	2.67
	839		177		187	

Table 8: The frequencies and the percentages of the Universal Dependency Relation tags in *eo_prago_ud-test.conllu* (indicated as *Prago*), *eo_cairo_ud-text.conllu* (indicated as *Cairo_eo*), and *cairo/en.conllu* (indicated as *Cairo_en*)

The noun phrase in the nominative case dependent on a verb functions as nominal subject and its dependency is annotated with *nsubj*. An infinitive verb can function as the subject of another verb, and its dependency is also annotated with *nsubj*. A finite clause introduced by *ke* can function as the subject of another verb (e.g., *Ke la studento legos ĉi tiun libron, ne surprizas min* "That the student will read this book does not surprise me"), and its dependency will be annotated with *csubj*, yet it is missing both in Prago and Cairo. The passive voice of Esperanto is expressed by a participle accompanied by the auxiliary *esti* in its finite form, such as *Tiu ĉi libro estas legata de studento* "This book is read by a student" and the subject of a participle in a passive sentence will be labeled with *nsubj:pass* (or *csubj:pass* if the subject is a finite clause), yet it is missing both in Prago and Cairo.

A noun with the accusative case morpheme *-on* in singular or *-ojn* in plural dependent on a verb is its object, hence the dependency between them is annotated with *obj*. The dependency between a verb and a prepositional phrase is annotated with either *obl* or *nmod*. Since the distinction between them remains uncertain in Esperanto and at present based on the annotator's intuition, this issue must be addressed in future study. According to the guideline of Universal Dependency Relations, the verb *esti* "be" is either treated as the auxiliary when it depends on a verbal predicate with the dependency type *aux* or *aux:pass*, or as the copula when it depends on a non-verbal predicate with the type *cop*. The verbs *devi* "must" and *povi* "can" are treated not as auxiliaries but as full verbs in Prago and Cairo.

4 Discussion

4.1 Comparison and contrast between Prago and Cairo_eo

Due to the large difference in the token size between Prago and Cairo_eo, we need to be cautious about drawing any conclusion from the comparisons of the frequencies of UD POSs and UD relations. Provided that, based on the findings in Prago and Cairo_eo, we can point out that some POSs are worthy of focus, such as those for content words (ADJ, NOUN, VERB) and function words (PRON). The same logic applies to UD relations such as *amod*, *nmod*, *nsubj*, and *obj*. Further studies using UD-annotated text data will reveal differences in style of Esperanto texts, which will contribute to the understanding of Esperanto as a natural language.

4.2 Comparison and contrast between Cairo_eo and Cairo_en

The same size of tokens (however small it is) and the same sentential meaning between Cairo_eo and Cairo_en will provide us with more reliable insight into the difference between English and Esperanto. The most obvious difference between them in terms of UD POS is ADV; Cairo_eo has twice as large number of ADVs as Cairo_en. Also, UD relations *advcl* and *advmod* are more frequent in Cairo_eo than in Cairo_en. This may suggest that Esperanto prefers adverbial expressions than English does. In contrast, Cairo_en has twice as large number of AUXs as Cairo_eo. This must be the result of the annotation policy of Cairo_eo where the verb *devi* "must" and *povi* "can" are annotated as VERB. This policy might be revised in future annotation processes.

4.3 Research in the future

The insights in the preceding subsections lead us to realize the necessity of much larger-scale Esperanto corpus with UD annotation which is also in parallel with other languages. Currently, manual annotation of *Deklaratio pri Homaraniso* "Declaration of Homaranism" (Zamenhof 1913) with UD POS tags and UD relation tags is under way; Comparing and contrasting this new UD-annotated text written by Zamenhof in the early 20th century with existing UD treebank based on *Manifesto de Prago* written in the end of 20th century are expected to give us discovery in diachronic change of Esperanto: Through more

than 100 years of its history, Esperanto must have gone through historical changes since its creation, just like other natural languages, and articulating how it has actually changed in terms of usage of morphemes and dependency relations is of linguistic interest. In order to extend this line of research, large-scale UD-annotated Esperanto treebank will function as the ample data for such studies. In addition to this, adding Esperanto version into Parallel Universal Dependencies (PUD) will give us opportunities for multi-lingual contrastive research of Esperanto. However, it is time-consuming, and therefor unrealistic, to manually translate the 1,000 English original sentences in PUD into Esperanto by one researcher and to annotate them with UD POS tags and UD relation tags all manually. We need to conduct these tasks on a team basis, developing an Esperanto UD parser and evaluating its performance with a UD-annotated gold-standard Esperanto texts. These issues must be addressed in future studies, which will contribute to the development not only of Esperanto NLP, but also of understanding of Esperanto as a genuine natural language.

5 Conclusion

This paper described the details of UD-based morphological and syntactic annotations on Esperanto texts to construct its small-scale UD treebank. After reviewing the previous research on Esperanto as a natural language, the detail of manual annotation of UD morphological and relational tags is described. Then the frequencies of these tags are shown to differ across different genres of Esperanto texts, and between English sentences and their Esperanto translations. It is suggested that larger-scale Esperanto UD treebank will open the possibility of future research which will contribute to the idea that Esperanto is not an artificial language but a natural language.

References

Eckhard Bick. 2007. Tagging and Parsing an Artificial Language: an annotated web-corpus of Esperanto. *Proceedings of Corpus Linguistics* 2007.

Eckhard Bick. 2009. A Dependency Constraint Grammar for Esperanto. *Constraint Grammar and robust parsing: Proceedings of the NODALIDA 2009 workshop*. 8-12.

Eckhard Bick. 2016. A Morphological Lexicon of Esperanto with Morpheme Frequencies. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. 1075-1078.

Eckhard Bick. 2018. Arbobanko-A treebank for Esperanto. *International Conference on Computational Linguistics and Intelligent Text Processing*. 248-260. Cham: Springer Nature Switzerland.

Eckhard Bick. 2019. Automatic Generation and Semantic Grading of Esperanto Sentences in a Teaching Context. *Proceedings of the 8th Workshop on NLP for Computer Assisted Language Learning*. 10-19.

Eckhard Bick. 2020. Syntax and Semantics in a Treebank for Esperanto. *Proceedings of the Twelfth Language Resources and Evaluation Conference*. 5120-5127.

Chris Gledhill. 2000. *The Grammar of Esperanto: A Corpus-based description*. München: Lincom Europa.

Grant Goodall. 2023. Esperanto kaj lingvistiko: cent jaroj da (mal)amikeco. "Esperanto and linguistics: A hundred years of friendship and hostility." *Esperantologio / Esperanto Studies*, 4(12). DOI: 10.59718/ees43692

Martin Haspelmath, Matthew Dryer, David Gil and Bernard Comrie, ed. 2005. *The World Atlas of Language Structures*. Oxford: Oxford U. P.

Ilona Koutny. 2015. A typological description of Esperanto as a natural language. Język. *Komunikacja. Informacja "Language. Communication. Information"*. vol.10, 43-62.

Marte Djupvik Kråkmo. 2022. *English vs. Esperanto: A comparative study of clausal word order in a Minimalist framework*. Master's thesis, Department of Foreign Languages and Translation University of Agder.

Bill Manaris, Luca Pellicoro, George Pothering, and Harland Hodges. 2006. Investigating Esperanto's statistical proportions relative to other languages using neural networks and Zipf's law. *Proceedings of the 24th IASTED International Conference on Artificial Intelligence and Applications*. Innsbruck, Austria: Acta press. 102–108.

Carlo Minnaja and Laura Paccagnella. (2000). A Part-of-Speech Tagger for Esperanto oriented to MT. *Proceedings of the International Conference on Machine Translation and Multilingual Applications in the New Millennium: MT 2000*.

Mikael Parkvall. 2010. How European is Esperanto?: A typological study. *Language Problems and Language Planning*, 34(1). 63-79.

Alberto Poncelas, Jan Buts, James Hadley, and Andy Way. (2020) Using multiple subwords to improve English-Esperanto automated literary translation quality. *Proceedings of the 3rd Workshop on Technologies for MT of Low Resource Languages*, 108–117.

Ida Stria. 2015. Esperanto as a natural language. *Język. Komunikacja. Informacja "Language. Communication. Information"*. vol.10, 32-42.

Ludoviko Lazaro Zamenhof. 1887. *Dr. Esperanto's International Tongue (La Unua Libro)*.