

# Syntactic and Lexical Changes in Esperanto: A Quantitative and Corpus-Based Survey

**Joshua Herring**

Department of Linguistics  
Indiana University Bloomington, IN, 47403  
jwherrin@indiana.edu

## Abstract

Esperanto is unique among the world's communicative languages in that, rather than being the culmination of hundreds of years of natural language change and evolution, it arose as the result of a conscious human plan. From a certain point of view, this makes it also uniquely interesting for the historical linguist. Despite its relative youth (it was developed only 130 years ago), Esperanto already demonstrates many of the patterns of historical changes that are typical among natural languages. Indeed, *because* of its youth, Esperanto may offer an unparalleled opportunity to watch the early stages of language development in process. This paper details a corpus study of historical changes in Esperanto. Journal articles freely available for use on the internet from 1922, 1957, and 2001 were processed using a Python script. Specific changes predicted are those consistent with a code/signal processing view of language. In particular, it was expected that the so-called "unofficial" suffixes would drop from the language, that derived wordforms for highly frequent words would give way to loanwords, that adverbial forms would become less frequent, and that relative clauses would begin to give way to Japanese-style adjectival modifier phrases. Excepting the pre-

diction that adverbial forms would become less frequent, all of these changes were, in fact, observed.

## 1 Introduction: The Structure of Esperanto

Esperanto is designed to be flexible so as to accommodate as many natural language patterns as possible, the idea being that all learners from all linguistic backgrounds be on "equal footing." As a result, Esperanto grammar is largely dependant on a rich and highly regular agglutinative morphology for syntactic encoding. In theory, all words are made up of a basic stem which takes on the functioning of any available part of speech depending on which ending is attached to it. Thus, all words from the same grammatical categories end in the same characters, and any stem can be made to fill any grammatical role. To give a common example, *amo*, which has a noun ending, means *love*, where *amas*, which has a verb ending, means *loves*. This scheme makes Esperanto ideal for online corpus processing studies since individual items can be easily and accurately tagged.

In theory, Esperanto sentences can be scrambled in any possible way. In reality, however, there are any number of minor rules which specify which items should precede or follow given other types of items. Prepositions and determiners, for example, *always* come before words they specify, and the same is true in the majority of cases for adverbs as well.

Also, complex meanings in Esperanto are meant

to be built up from a stem plus available concatenative atomic concept markers. The original word for *expensive*, for example, is *malmultekosta*, which is formed by *mal* (“not”) + *multe* + *e* (“great/many” – “adverbial”) + *kosta* (“cost” – “adjectival”). In practice, however, such forms are often replaced by foreign borrowings – in this case *ĉipa*, from the English “cheap.”

## 2 Methods

Three corpora were gathered from freely-available sources on the internet. They are:

1. Literatrua Mondo (1922-1923, 47611 words)
2. La Nica Literatura Revuo (1955-56, 35968 words)
3. La Fasko (2000-2001, 24554 words)

These corpora were parsed using a homemade Esperanto parsing application. Syntactic patterns were simply counted and frequencies were compared across the different corpora.

To look for predicted lexical changes, the corpora were fed to a simple regular-expression pattern-matching application which separated out known prefixed and suffixes and counted out their frequencies. Where specific specific replacements have been documented in other literature (e.g. the *ĉipa* example from above), the replacement loanwords were also included in the search.

## 3 Results

Results at this time are still preliminary given the limited nature of the corpora. I am, however, currently searching for more extensive corpora for the late 1950s and early 1920s period.

Based on the corpora available, several actual changes were found and can be crudely summarized as follows:

1. As predicted, while prefixes have been added to the language, “unofficial” suffixes are being dropped – albeit more slowly than might have been expected. Interestingly, however, the addition of prefixes was more evident in the 1955 period than in the 2001 period, indicating that this might have been a tendency which was,

for lack of a better word, “fashionable” in the Esperanto community for a time but has since been replaced by a more practical tendency to either use the available form or simply borrow from foreign languages.

2. Syntactic patterns show a tendency toward simplification in two ways:

- When possible, the number of syntactic patterns available is reduced. For example, the compression of progressive forms from *estas + part.* to simply *part. + verb ending*.

*Ŝi estas scribando leteron*

She is writing a letter.

Becomes, without ambiguity,

*Ŝi skribandas leteron.*

- Though not as obvious, evidence was found that the language is tending away from complicated relative-clausal forms and in favor of Japanese-style descriptive adjective phrases (which are syntactically similar to *wh*-relative clauses, of course, but require the use of fewer wordforms – specifically, rich, morphologically inflected relative pronouns are no longer needed). This kind of pattern is easy to detect as it gives rise to a string involving the definite article *la* before any preposition, something that otherwise never occurs. An example:

*la de vi solvita problemo* → *the by*

*you solved problem* vs. *la prob-*

*lemo kion vi solvis* → *the problem*

*you solved*

3. The agglutination of freestanding function words as derivational prefixes:

Most notable among these changes are four words cited by Gledhill as highly frequent in his corpus:

*tiumomente*: “this + moment” → “currently”

*tiutempe*: “this + time” → “now”

*tiukaze*: “this + case” → “in this case”

*tiuscene*: (meaning not given)

4. Derived forms give way to foreign borrowings where possible. There is some evidence (though admittedly nothing conclusive from a corpus-based study of this scope) that when this happens, the foreign word takes on a more specific meaning, while the native word fills a generic function. In the example cited earlier, the English loanword *ĉipa* has a negative connotation (i.e. “cheap”), whereas the native *mal-multekosta* comes to mean something relatively non-judgemental like “inexpensive.”

More precise numerical results will be given in graphical format with the presentation. As stated earlier, there is reason to expect that better corpora – especially for the 1922 period – can be found.

## 4 Conclusion

Although Esperanto is unlike natural language in many important ways, it is still spoke by people like any other language and exhibits many of the same historical effects noted by linguists for “natural” languages. In particular, it fits Hock’s contention that while people will tolerate a great deal of meaning overlap between phonetically dissimilar forms (synonymy) and will even tolerate different meanings from phonetically identical forms (homonymy), they tend *not* to tolerate *identical* meanings from phonetically distinct forms. When phonetically distinct forms have or come to have nearly identical meanings, the individual forms will take on subtle distinctions which can grow, with time, into more significant splits. In this respect, Esperanto is no different from any other natural languages.

Esperanto is *unlike* other languages, however, in that its syntactic transformations, such as they are, seem to be more gradual and based either on second-language learner preferences or compression of existing forms rather than any real categorical reanalysis. This is to be expected given that it is, in many ways, a “work-in-progress,” currently under construction by many people from diverse linguistic backgrounds. Though it would be prudent to wait to draw any firm conclusions until more corpora data is available, and a closer analysis can be performed, one can tentatively say that changes in Esperanto seem motivated by two things:

1. “Voting Effects” in which members of the community “pull” the general grammar of the language toward that of their native language’s preferences
2. A desire to simplify whereby atomically stored items and grammatically simple forms are preferred over complex forms and derived forms.

It should be stressed, however, that these conclusions are tentative and will need to be further verified by closer analysis prior to presentation.

## References

- Lightfoot, David. 1999. *The Development of Language: Acquisition, Change, and Evolution* Oxford: Blackwell
- Gledhill, Christopher. 2000. *The Grammar of Esperanto: A Corpus-Based Description*. Munich: Lincom Europa.
- Das Kausativ und das Antikausativ in der gemischten Plansprache Esperanto. 1988. Schaffhausen: Weidemanns MondoServo.
- Kellerman, Ivy 2003. *A Complete Grammar of Esperanto* Ebook: <http://ggd.tuwien.ac.at/books/gutenberg/etext05/esper10h.htm>
- Association for Computing Machinery. ????. *Nativization Processes in L1 Esperanto*. Berkely: internet pdf
- Harlow, Don. 1995. *Comments on the Sixteen Rules of Esperanto Grammar*. Internet Resource: <http://www.webcom.com/donh/Esperanto/rules.html>
- Hock, Hans Henrick 1991. *Principles of Historical Linguistics* Berlin: Mouten de Gruyter
- Literatura Mondo 1922-1923 Internet Resource: <http://www.tekstoj.nl/lm/index.html>
- La nica literatura Revuo 1955-1956. Internet Resource: <http://donh.best.vwh.net/Esperanto/Literatura/Revuoj/nlr/>
- La Fasko 2000-2001. Internet Resource: <http://www.medialabinc.net/temp/fasko/x/index.html>
- Wells, J.C. 1969. *Concise Esperanto and English Dictionary* London: NTC Publishing Group
- Verdadrako Internet Resource: <http://www.verdadrako/50g.com/asap/25index.htm>