The Lexicography of Esperanto

Gobbo, F.

DOI
10.1007/978-3-642-45369-4_88-1

Publication date
2017

Document Version
Final published version

Published in
International Handbook of Modern Lexis and Lexicography

License
Article 25fa Dutch Copyright Act (https://www.openaccess.nl/en/in-the-netherlands/you-share-we-take-care)

Link to publication

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 426, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date: 31 Oct 2023
The lexicography of Esperanto

Federico Gobbo

Abstract

Esperanto is the most successful planned language in the history of humankind. In fact, it succeeded to form a stable community of practice that has now lasted for more than a century, surviving two world wars. The language has been almost naturalized, as it is also used in multilingual families, even though they do not set the linguistic norm. Over time, the actual language use has changed according to the communicative needs and the linguistic repertoire of its speakers, with implications for lexicography, which is the topic of this chapter.

Contents

Introduction ......................................................................................... 1
Description ................................................................................................ 2
Some fundamental traits of the lexicon .................................................. 2
Monolingual dictionaries and the development of the lexicon .................... 3
The role of Esperanto in scientific terminology ...................................... 5
Lexical strategies to modernize the lexicon .......................................... 6
Lexicography in the digital age ............................................................. 8
References ........................................................................................ 9

Introduction

Esperanto is a language planned for international communication by Ludwik Lejzer Zamenhof, an Ashkenazi Lithuanian Jew (litvak), born in Bialystok (Poland), at that time part of the Russian Empire. Esperanto was launched in 1887 through a book...
published in Warsaw, under the pseudonym *Doktoro Esperanto* [the Doctor who hopes], a pen name which eventually led to the name of the planned language.

The first lexicographer of Esperanto is Zamenhof himself, who developed the structure (phonology, morphology, syntax) of the new language and compiled the first dictionary essentially via introspection. The first Esperanto dictionary, reproduced in Itô (1991), is composed of 904 lemmas. This dictionary was addressed to Russians, and therefore it is bilingual, general-purpose, and semasiological. In fact, the aim was to let the readers use the new language quickly and effectively, without too much effort to learn it. All affixes used for word formation were (and are) put in the lemma position, and this choice influenced the lexicography of Esperanto forever. For example, the first entry of the English edition, published in 1888, is “*a*, expresses an adjective; f. in. *hom* – man, *hom,a* – human” (in Itô 1991, p. 188; note the comma used as the word-division marker).

Zamenhof chose agglutination as the main morphological strategy for word formation, according to a principle of min-max optimization: minimization of the number of roots and maximization of productivity, i.e. combining forms using affixes. For example, from the single lemma *frat*, users produce *frat,o* “brother,” *frat,a* “fraternal,” *frat,e* “fraternally” and even *frat,i* “behave fraternally.” Direct verbification of roots emerged with use but is still controversial, because meanings are thereby driven by the root and do not follow any general rule — see the discussion in Jansen (2011). There are four fundamental parts of speech in Esperanto (nouns, verbs, adjectives, adverbs), and accordingly Esperanto dictionaries will show the meaning (by means of a definition) and the use (by means of some examples) of lemmas for every part of speech, wherever possible. Moreover, Zamenhof borrowed the other parts of speech from the Standard Average European Sprachbund: in particular, the definite article, prepositions, pronouns and numerals (Itô 1991).

**Description**

**Some fundamental traits of the lexicon**

Because of its phonemic-oriented alphabet, no Esperanto dictionary shows any phonetic transcription or respelling, and analogously standard dictionary slots, such as those for parts of speech, are considered superfluous. Zamenhof’s intention was to introduce “a complete dismemberment of ideas into independent words, so that the whole tongue is consisting not of words in different grammatical forms, but of unchangeable words” (in Itô 1991, p. 188; emphasis in the original). This choice caused a certain resistance to polysemy, which is tolerated by today’s speakers but definitely not encouraged. Instead, new lemmas – always in the form of roots – are introduced, with the aim of maintaining precision in semantics. For example, the English word “order” is rendered in Wells (2010) with three self-standing lemmas, although etymologically related: *ord,o* (in the sense of the Greek *kosmos*, the opposite of chaos), *ordon,o* (command), and *orden,o* (e.g., Knight Templars).
Furthermore, there is the sense captured by mend,i in the sense of “put an order.” The source languages of the Esperanto lexicon come from Zamenhof’s repertoire; the main part is Romance (in particular French, but also Latin), quite a bit is from the Germanic stock (through German, English, and Yiddish), while the Slavic part (through Russian and Polish) is less relevant in quantitative terms. Ancient Greek is also present. The phonetic rules to import new roots are implicit in the language, and they do not follow a rigorous etymological principle, they are rather according to Zamenhof’s intuition and personal gusto. For example, the Latin aqua becomes akvo in Esperanto, as Zamenhof thought about the pronunciation by Slavic people, and similarly quanquam becomes kvankam. As Esperanto has a well-defined series of affixes for derivational morphology, the derivational morphology inherited from Latin is no longer productive. For example, the borrowing of the Latin syllable ex is idiosyncratic: “exam” is ekzameno, “exile” is ekzilo, “to export” is eksporti, and “to explore” is esplori. Another important implicit phonotactic rule concerns Germanic words, which are borrowed midway between German (for vowels) and English (for consonants). For instance, the English “nightingale” is in German Nachtigall and becomes najtingalo in Esperanto (WELLS, p. 334).

**Monolingual dictionaries and the development of the lexicon**

Despite some implicitness in phonotactics, Zamenhof’s aim of productivity was certainly achieved. According to the corpus-based description by Gledhill (1998), the most frequent 6,000 words of the American Heritage Dictionary are rendered in Esperanto with approximately 850 roots. However, as the language was soon used in practice, the need to expand the lexicon quickly emerged. The Russian to Esperanto dictionary that Zamenhof published in 1888 consisted of 232 pages, with more than 10,000 Russian words, translated with just 3,097 Esperanto roots (Korzhenkov 2009). It is interesting to note that the author considered these extensions of the original dictionary of 1887 only recommendations. In fact, Zamenhof never considered the language its own property; rather, he tried to set the language free, even from himself. Major modifications to the language structure were considered acceptable until 1894, following a referendum among the readers of the first journal, **La Esperantisto** [The Esperantist]. Zamenhof collected all proposals against the then-current structure of the language, but the reformed proposal was rejected by the readers; see Dietterle (1929). Since then, the language has not been subjected to any major changes: any proposals to change the grammar will automatically be rejected by the Esperanto speech community, and therefore innovation mainly focuses on the lexicon. By the time the **Universala Vortaro** (UV) [Universal Dictionary], henceforth UV, was published in 1894, being the first normative dictionary after 1887, the number of lemmas had already doubled.

1905 is a very important year in the history of the language. The first **Universala Kongreso** [Universal Congress] was organized in Boulogne-sur-Mer, France, and the movement presented itself publicly to the world (Forster 1982). Moreover, Zamenhof published a book called **Fundamento de Esperanto** [Foundations of
Esperanto] in order to establish an untouchable base for the language so that Zamenhof himself could renounce his leadership of the language. In that book the UV was included, with 2,768 roots in total (Zamenhof 1905). UV is essentially a word list made of parallel roots in six languages, Esperanto plus French, English, German, Russian and Polish (in this order). The five languages that were paired with Esperanto acquired prestige as the most important substrata of Esperanto: when a new lexeme is proposed, the parallel translation in these languages is often cited in support of the new proposal. Sometimes the equivalence postulated in UV raised semantic problems due to anisomorphism. For example, *offender* is rendered as “*offenser | offend | beleidigen | оби́жать [obrażáć] | krzywdzić*.” In German, *beleidigen* can be used only with persons and feelings, while the French *offenser* can be used in relation to rules or laws. Which semantic domain is more “esperantic,” the German or the French one? No definitive answer can be given. However, since 1905, such questions have been debated in the language committee, an independent organism which eventually became a Language Academy, modelled after similar institutions present in Europe for languages like French, Italian and Spanish. After 1905, a lot of journals in the language, as well as original and translated literature, were published: Esperanto speakers quickly felt the need for a monolingual dictionary. BOIRAC (1909) was the first monolingual dictionary, including an Esperanto-French lexicon, but it was highly criticized by contemporary Esperanto speakers, while KABE (1911) was the first truly monolingual Esperanto dictionary, compiled by Kazimir Bein. KABE became the most influential dictionary until the 1920s. Lemmas are explained by means of definitions and examples. Nowadays, this work is important for the diachronic study of the language, as it registers interesting archaisms. Around the same period, Esperantology, i.e. the branch of linguistics devoted to Esperanto, was founded by René de Saussure, the youngest brother of Ferdinand, the famous linguist. In fact, René de Saussure wrote a rigorous tractate on the morphology of Esperanto, accepted by Zamenhof in 1910 and highly influential for the lexicography of Esperanto afterward (a presentation in French is de Saussure 1918).

The First World War signals a stop in the lexicography of the language, and the Esperanto movement had to reorganize itself around the main association, the UEA [Universal Esperanto Association], also because of Zamenhof’s death in 1917 (Forster 1982). The role of Paris and France in general in the Esperanto life became more and more important in the 1920s, where a left-wing Esperanto movement was founded, under the name SAT [World Anational Association]. Its founder, Eugène Adam Lanti, formerly co-founder of the French Communist Party, pursued a radical cosmopolitan ideology, called anationalism, openly critical of the Soviet Union (Lanti 1973). Lanti and SAT gave a lot of importance to the acculturation of workers outside national languages: in order to accomplish this goal, *Plena Vortaro de Esperanto* [Complete Dictionary of Esperanto], henceforth PV, the first complete dictionary of Esperanto, was published in 1930. PV follows the model of the French dictionary *Le Petit Larousse Illustré*; for this reason, the problems of anisomorphism explained above are often solved in favor of French. PV was very successful in the Esperanto community: the revised second edition published in 1934 with corrections
and extensions, had approximately 6,900 lemmas and 5,000 compounds, and became the de facto monolingual standard dictionary of Esperanto for decades. As a comparison, the official dictionary, published by the Academy of Esperanto as a series of extensions to the original Zamenhof’s UV, listed only 4,492 lexemes in 1934. Since the publication of PV, the guiding role of the Academy of Esperanto with regard to the lexicon was taken up by the French Anationalist movement (Gobbo 2010a).

Interestingly, an Esperanto motto arose: “Esperanto words are French with the o-ending.” Even if it is an exaggeration, this motto does attest the big influence of French over the Esperanto lexicon since the 1930s. On the other hand, during the same years, another center of innovation for the Esperanto language emerged: Budapest (Sutton 2007, p. 77). In fact, since 1922 a group of writers led by Kálmán Kalocsay and Julio (Gyula) Baghy published a journal, Literatura Mondo [Literary World], profoundly innovative in poetry and prose. In their Parnasa Gvidlibro [A Guidebook to Parnassus], Kalocsay and Waringhien (1932) proposed almost 600 neologisms, many of which eventually entered the everyday register of the language, such as eseoo “essay,” etoso “atmosphere, ethos, ” sandviĉo “sandwich” (Camacho 1999). However, the leading role of SAT in lexicography never ceased: in 1970 the Plena Ilustrita Vortaro de Esperanto (PIV) is published as the successor to PV, forty years after the first edition. PIV is the first complete illustrated monolingual dictionary of Esperanto: it has approximately 15,250 lexemes for about 45,000 entries including compounds and derivations, while its last revision of 2002, the Nova Plena Ilustrita Vortaro de Esperanto (nPIV), slightly enlarged the content, to about 17,000 lexemes for approximately 47,000 entries. Since 2012, nPIV is freely available on the web (cf. http://vortaro.net/). Compared to major European national languages, Esperanto dictionaries are still small; for instance, the Italian dictionary De Mauro Paravia lists about 129,000 entries with approximately 734,000 derivations (Cicoira 2008).

The role of Esperanto in scientific terminology

The movement for an international auxiliary language attracted the interest of numerous scientists and scholars, among others the Italian mathematician Giuseppe Peano (Gobbo 2010b); as such a language could be used as a neutral medium for scientific and technical communication. Already before the First World War, some specialized Esperanto associations were founded, devoted to science in general, medicine, and the railways. Afterward, others were to be added, for instance the post and telegraphy (1966), radio amateurs (1970), and cybernetics, systemics, and informatics (1983) — see Blanke (2008) for details. In the field of terminology, Esperanto plays a special role. In fact, one of the most important results is the first edition of the International Electrotechnical Vocabulary [IEV], published in 1938 by the International Electrotechnical Commission (IEC), where the definitions of 2,000 entries were given in English and French, while translations in German, Italian, Spanish and also Esperanto were provided. The Austrian engineer Eugen Wüster, pioneer in terminology standardization, was a leading figure in this field.
In particular, he wrote an encyclopedic dictionary Esperanto-German [WUST] in eight volumes devoted to science and technology (Blanke 2008). The International Academy of Sciences San Marino, established in 1983, works with Esperanto as the main scientific language, both for education and research at university level (Gobbo and Fößmeier 2012). Finally, since 1987 the Terminology Esperanto Centre, founded within the UEA, works with Infoterm and TermNet, the most important terminology organization in the world.

An interesting case is the domain of fauna and flora, which has been covered through a long-standing project called Lexicon Silvestre (SIMON 2010). It is a multilingual project with Esperanto at its core. One of its outcomes, ULLRICH (2009), is a nomenclatura in English, Esperanto, French, German and of course Latin as the reference (with the acronym “sc,” meaning “scientific name”). From the Preface (version in English): “this work was created as a contribution to the communication among experts in the field of forestry and its related fields speaking different mother tongues but having a good knowledge of Esperanto.” It is worth noting that there are two versions of the entries in Esperanto, the colloquial form versus the scientific one, and also parallel versions in the natural languages are given, when relevant. Variants of the plants are carefully listed. The scholars who classified the species and the essential home territories of the plants are also mentioned, with additional remarks at the end of each entry. For example, the entry for “walnut” states:

“Juglans L. 1753 / Juglanso / Walnuss / Walnut / Noyer. 149 [being an index, author’s note].

In brackets, readers can find the different uses in the various references used, with “ul” indicating the recommendation by the author. In italics there are indications of grammatical gender, in the case of German. This dictionary illustrates that there is a general tendency in the Esperanto movement toward the implementation of strict normative habits in managing the lexical needs of the ordinary language users, by the Esperanto élite – i.e., presidents of the associations, journal editors, reckoned writers, and university-level scholars.

**Lexical strategies to modernize the lexicon**

There are two basic strategies to import a new concept into the language: either a new word combination can be established using existing lexemes and affixes, or a new lexeme can be introduced. A well-known example is “hospital,” where both strategies are attested and used: the first possibility is malsanulejo (lit. “place for ill people,” compare with the German compound Krankenhaus), while the second one
is *hospitalo* (WELLS, p. 285). According to nPIV, *hospitalo* is a hypernym of *malsanulejo*, also indicating, for instance, clinics and sanatoria. While the first strategy exploits the agglutinative morphology of the language, the second strategy is straightforward for anyone educated in English or in European languages such as French, Italian, Spanish or Portuguese, but also in languages influenced by them, for example Malay. However, Esperanto speakers often consider word combinations more respectful of non-European speakers because they do not request an extra investment in learning new lexical material, and in general they are somehow less prone to import the Western way of thinking in the *Weltanschauung* of Esperanto. On the other hand, Esperanto is part of Standard Average European, and therefore lexicalization is driven by the substrata of Esperanto (such as Greek, Latin, and French but also German and English). For example, even if *falakvo* for “waterfall” is attested and considered valid by Zamenhof, as *akvofalo* (cf. Dietterle 1929, p. 526), the first form is now an archaism, while the second is still used and felt as more “natural.” Analogously, *hom-manĝi* does not mean “to eat as a man,” but “to act as a cannibal,” as in the Greek-derived word “anthropophagy” (anthropo- “human,” phagy “to eat”). Quite often substrata influence lexicalization in combinations. For example, *servojo* is not “a way made of iron,” but a railway, and *terpomo* is not “an apple found in the ground,” but a potato. These two examples show that, while the word compounding rules are Germanic, the lexical material is borrowed from French: compare *chemin de fer* for “railway” and *pomme de terre* for “potato.”

The Esperanto phraseology was started in 1910 by Zamenhof himself, through the publication of a collection of proverbs, the *Proverbaro Esperanta*, based on Zamenhof’s father’s interlinguistic work on proverbs in Russian, Polish, French and German. As in the case of the lexicon, the phraseology of Esperanto is fully part of Standard Average European: “Egg of Columbus” is *ovo de Kolumbo*, “castle in the air” is *kastel’en aero*, and “flesh and blood” is *sango kaj karno* – for a dissertation on this topic, see Fiedler (1999).

In general, there is no clear sociolinguistic distinction between the two strategies (exploit word compounding vs. import of a new lexeme): some Esperanto speakers consider the first strategy the best, while some others prefer the second one. However, there is some tendency in opting for word combinations among language teachers, a preference which is reflected in language learning books. This is hardly surprisingly, as one of the most important features of Esperanto to be learned by beginners is morphology. In Esperanto publications — such as journals, magazines, and novels — it is common to read non-official lexemes explained through a brief definition in footnotes or on the last page as an addendum, sometimes along with etymological information. In many cases, one strategy wins over the other one, the loser becoming an archaism. For example, nowadays “computer” is *komputilo* (lit. “tool for computation,” according to a very productive paradigm of compounds) for almost every Esperanto speaker, while until the 1980s the borrowings *komputoro* and *komputero* were still used. On the other hand, in the late 1990s the form *blogo* for “blog” quickly won over *retjurnalɔ* (lit. “journal of the net”) or similar compounds. Today, we are witnessing the competition between *fejsbuko* and *vizaĝlibro* as possible designations for Facebook, with *vizaĝaro* (lit. “set of faces”) as the third
contender, probably doomed to be the first loser. Another recent example is Twitter, already designated as *Pepejo* (lit. “place for tweeting,” personal observation). In fact, the act of tweeting is *pepi*, which seems to be more frequent than *tviteri*, perhaps because semantically it is less precise (lit. “using Twitter”). These competing forms attest the language vitality of Esperanto.

**Lexicography in the digital age**

The Internet gave a new impulse to the Esperanto language. Analogously to other lesser-used languages, Esperanto speakers started to design websites early on, namely, in the 1990s, so as to get visibility and to stay in touch with one another almost for free. In particular, there is an analogy between the Esperanto movement and the movement for free software (Gobbo 2004): Esperanto is a free language just as Linux is a free operating system. For this reason, the interface of many free software programs is fully translated into Esperanto, while the Wikipedia in Esperanto is one of the most active in the world in terms of quantity of articles and the frequency of contributions. Sometimes neologisms are added. For example, the Esperanto translation of the wiki page *Bracco Italiano* (Italian Pointing Dog, a breed developed in Italy as a versatile gun dog) needed the introduction of the lexeme *braco*, still absent in nPIV.

The list of bilingual dictionaries available on the web — sometimes with a mobile version encapsulated in an application, so as to be used with smartphones and tables — is quite rich (cf. [http://eo.wikipedia.org/wiki/Esperantaj_vortaroj_en_la_reto](http://eo.wikipedia.org/wiki/Esperantaj_vortaroj_en_la_reto)). However, in most cases they are compiled without serious lexicographic reflection.

The publication on the web of Esperanto corpora is a relatively recent phenomenon, still to be exploited by lexicographers. Apart from newspapers and journals (for original articles, *Monato* is one of the most important, while *Le Monde Diplomatique* has a translation service in Esperanto as well), there is still only one proper monolingual corpus of the language, called *Tekstaro de Esperanto* (cf. [http://tekstaro.com/](http://tekstaro.com/)), encoded following the guidelines of the Text Encoding Initiative (TEI). It contains approximately ten million words, from the Bible to recent novels. The project Visual Interactive Syntax Learning (VISL) promoted by the University of Southern Denmark has annotated part of this corpus following the linguistic framework of Constraint Grammar, using a general tagset valid for different natural languages (Bick 2007). A Danish-Esperanto machine translation interface is also provided. Machine translation research has a long history for Esperanto (Gobbo 2006). Finally, since February 23, 2012, Google has added Esperanto in its statistical machine translation system, available for free worldwide.

The development of electronic lexicography in Esperanto is a key factor in maintaining and developing the language. Recently, mobile apps for smartphones came out with word plays, multilingual dictionaries and self-education courses. The challenge faced by Esperanto is to form a new generation of élite speakers with the necessary linguistic and computing skills in order to modernise the available linguistic resources and eventually the Esperanto movement.
References


de Saussure, R. (1918). La structure logique des mots dans les langues naturelles, considérée au point de vue de son application aux langues artificielles. Büchler, Bern.


Dictionaries


