Reversal as Means of Building a New Dictionary

Andrejs Veisbergs
University of Latvia
Visvalža 4a,   Riga, LV-1050, Latvia
anveis@lanet.lv

Abstract
New opportunities offered by information technology make it possible to apply new methods to lexicography. This paper describes a project that will result in compiling a new Latvian-English dictionary of unprecedented size and quality by reversing the definition entries of the many English-Latvian dictionaries currently in print. The paper discusses:
the current lexicographic situation, which forms the backdrop for the reversal project;
the process of reversal itself, conclusions about the reversal project, its pros and cons, problems and findings; and
corollary conclusions about the contents and structure of dictionaries used for reversal. The last area is in some ways the most interesting as such projects might have a future in Latvia and elsewhere.

The Current Situation

Current Latvian lexicography lacks a comprehensive language corpus. This means that bilingual dictionaries with Latvian as the source language (the A part) endemically suffer from a smaller number of entries compared to the second language. The absence of a corpus gives ground to various speculation about the current state of the lexicon and also feeds the purist tendencies still rife in the country. It also entails an absence of information about the lexis as such -- about frequencies of usage, meanings, stylistic registers, diachronic changes, etc. This information is essential for lexicographers working on bilingual dictionaries with Latvian as the source language. Limited reflection on new and not-so-new vocabulary leads in turn to multiple variants of many frequently used terms and notions, because they are not "fixed" in dictionaries. This deepens the chasm between the real language and the reflected one as well. The possible use of Latvian-language internet material as a corpus (Grefenstette 2002) seems to be less advantageous than reversal, because of the variety mentioned above and the small proportion of terminological texts in the corpus. Latvian bilingual dictionaries are in high demand at present as new language contacts are being established and huge amounts of translations (including EU texts) are being undertaken.

Consequently, all possibilities of expanding the volume of the lexicon in dictionaries should be pursued, and one of these is the reversal of existing English-Latvian dictionaries. For historical reasons, it so happens that these dictionaries have always tended to be better and broader in scope compared to Latvian monolingual ones, which were always affected by purism (Latvian and Soviet) and various restrictions of colloquial, borrowed and non-standard lexis. The use of reversing techniques can lead to substantial volume growth of dictionary entries as well as updating and finding "missed" words and translations.

However, it must be mentioned that reversal has a relatively negative reputation in Latvia, because several companies offer sometimes flawed bi-directional English and Latvian online dictionaries (www.tilde.lv/dictionary, www.dictionary.site.lv) based on reversed, monodirectional (i.e. for Latvian users) dictionaries. Unedited reversals into Latvian often lead to serious, and anecdotal, mistakes such as impossible back-translations, strange Latvian entries and long lists of undifferentiated idiographic and stylistic synonyms. Unclear, unmarked and semi-alphabetic enumeration of English definitions when looking for a translation of a Latvian word are most confusing for the uninitiated. The English translations of, for example, **televizors (television set)** are provided in the following order: goggle box, idiot box, television receiver, television set, televisor, telly, television, box, tube box, boob box, teevee;

or **skaists (beautiful)** offers the following enumeration:

beauteous, pulchritudinous, beautiful, lovely, gallant, seraphic, likely, fair, sheen.

Non-experts and people used to printed dictionaries have a tendency to use the first variant in the list. In addition, the amount of lexis offered (with rare labels and no extra information) is too big for a normal user. As a result, lay users (unaware of the existence of reversal) consider the electronic dictionaries dubious, while experienced users think reversal dangerous.

Reversal

There are different views on the efficiency of dictionary reversal. A study of a reversal of an English-Swedish dictionary suggests that the economy of time and money is not particularly great because only about 30% of the reversed material could be retained (Geisler 1999). On the other hand, Estonian experience (Tamm 2002) reports successful solutions in reversing a Dutch-Estonian dictionary. It is clear that serious editing is inevitably required, which may be so time and effort consuming that it cannot be considered worthwhile. Yet one has to consider also the fact that, for well-known reasons and as alluded to above, smaller languages (like Latvian) normally have better bilingual dictionaries *from* than *to* other languages. These are normally based on the use of several large, high-quality, monolingual source-language dictionaries (such as Oxford, Duden, Hachette) worked upon by local language experts with expert technical assistance. As a result the A part is generally <u>comprehensive</u> and the B part fairly <u>accurate and acceptable.</u> It seems reversal is especially good for large dictionaries which contain vast terminological material that is predominantly monosemantic—35-45% of the total number of entries according to Corda (1998: 442)—and consequently undergoes reversal more easily.

The project with TILDE, the largest Latvian software company, was carried out in two stages. The first stage included reversing an English-Latvian dictionary (the notorious internet version) as well as many English-Latvian terminological dictionaries and imposing this material on <u>the existing entries </u>of the Latvian-English dictionary (Veisbergs 2001), thus enlarging the material. In the first stage of editing, the imposed part was easily recognisable (in a different font) bearing the tag for the source dictionaries and field labels. This was followed by thorough editing (performed by linguists), consisting mainly of deleting endless equivalent meanings from various thematic fields, expunging redundancies, reordering senses and removing unnecessary working labels and tags, e.g. **dog** *customs,* which signifies

that a dictionary of customs terms as the source. The result was a 25-30% increase in the volume of material. The second stage (now underway) consists of <u>adding new entries</u> from reversed, mainly terminological sources, which is being carried out and checked by experts.

Conclusions about the Reversed Dictionary

Apart from purely technical problems, some of which took a lot of time to solve, such as difficulties matching the various systems of the dictionaries like differing methods of variant inclusion (slashes versus brackets), the editing work offered interesting insights not only into the reversal process but also concerning the original entry systems and lexical material (see below). Some of the problem issues interfering with smooth reversal and editing were comprehensive enough to be considered endemic in reversal cases. These fall into three main groups:

- A multitude of synonyms, much repetition, redundancies (Newmark 1998) that have to be automatically or manually deleted.
- The reversed material contained numerous definition-style translations (useless for productive reversal) instead of equivalents.
- Connotative inaccuracies (lack of equivalence), the result of doubtful strategies in the source dictionaries.

One of the main problems in reversal is the long lists of entries with endless synonyms, only a few of which are adequate. These long lists come as a result of:

- Overly generalised translations in source dictionaries;
- Overly neutralised translations (loss of connotation);
- Imprecise translations; or
- The situation when Latvian has no precise counterpart, or the compilers have not been aware of it and have used a somewhat similar synonym.

What often emerge are huge groups of entries containing unimportant auxiliary components—again useless for reversal: *tāds, kas..; neliels...; cilvēks, kas...; bez....; būt....; izdarīt...;ar...;*

that which...; small...; A person who...; without...; to be...; to do...; with...
tāds, kas izdala mitrumu (that which produces moisture);
tāds, kas iznāk divreiz mēnesī (that which is published twice a month)
tāds, kas lauž starus (such that breaks rays);
būt par partneri (to be a partner);
būt par pierādījumu (to serve as proof);
būt par rakstnieķeli (to be a poor writer).

This suggests that bilingual dictionaries, instead of translation equivalents, often operate with the definitions characteristic of monolingual explanatory dictionaries. Similarly, some idioms and set expressions tend to have explanatory translations, some of which are useless for reversal:

- vientulības (izolācijas) trūkums (lack of solitude, isolation) -- goldfish bowl;
- Vinčesteras universitātes studentu izteiciens, paraža vai tradīcija (an expression, saying or tradition of Winchester University students) – notion.

A similar, yet also different problem is presented by slang: first, because it is ephemeral and subject to rapid diachronic changes in both languages and consequently difficult to match; secondly, because in many bilingual (including English-Latvian) dictionaries, English slang words tend to have rather neutral or mildly colloquial Latvian equivalents. As a result, for the reversed dictionary entry ļoti (very), 95% of the translations were English slang words of the intensifier type. The Latvian entry nauda (money) was followed by about 50 slang counterparts, and so on, not to mention the enormous volume and scope of the usual derogatory terms.

The problem of stylistic non-equivalence can also be seen on a broader scale --numerous translations distort the connotative meaning of the entry. Most bilingual dictionaries tend to deviate from connotatively strong items in the source language towards a more neutral/standard register in the target language. When reversed, connotatively-neutral words tend to have numerous bright and expressive translations (the Latvian equivalent is provided by the author in brackets):

šūšanas rūpniecība (textile industry) -- rag trade;
entomologs (entomologist) -- beetle sticker;
izdarīt plaģiātu (to plagiarise) -- to crib;
nieres (kidneys) – waterworks;
čiks (nothing) – fuck all.

The connotative difference between denotative equivalents requires very thorough editing. While in the original English-Latvian dictionaries the neutralising approach is not disastrous and in some cases even acceptable, in a reversed dictionary it becomes a serious problem, distorting any equivalence. For example, **rag trade** in serious context, **waterworks** at a medical conference, or **fuck all** at the end of a fairy tale, speak for themselves.

Similar deviation can also be observed on the denotative level—in the case of non-equivalence, the translated meaning of the entry usually errs by becoming generalised. In reversal, accordingly, semantically broad entries are given narrower, more concrete counterparts (the Latvian equivalent is provided by the author in brackets):

platforma (platform) – catwalk;
iemīļots dzīvnieks (favourite animal) -- pet;
transporta tunelis zem jūras šauruma (transport tunnel under a sea straight) – Chunnel.

One has also to consider the time frame when using reversal of many dictionaries. As language is in a state of flux, some words and terms may have dated variants and definition-type equivalents which have been replaced by new terms today, such as (in brackets Latvian equivalent provided by the author):

kapāti dārzeņi ar majonēzi (cut vegetables with mayonnaise) - **tartar sauce** (today **tatāru mērce**);

remonta punkts autosacīkstēs (repairs point in car races) – **pit** (today **pitbokss**)

Finally, an interesting problem occurs when simultaneously reversing material of many dictionaries. When creating the many terminological dictionaries, the compilers obviously made use of existing bilingual dictionaries along with the inevitable errors which are repeated again and again in the reversed material. For the editors, it can created false

assumptions as all sources point at one and the same equivalent, which was sometimes wrong.

Corollary Conclusions about the Source Dictionaries

Regarding input material, one of course has to consider that the dictionaries used for reversal were not planned or designed for such a future operation and that the inadequacies should not be viewed as a reproach to the original compilers and editors. However, reversal makes it possible to see clearly some tendencies and faults.

The most evident problem is connotative neutralisation. One can state that the Latvian part of the English-Latvian dictionaries tends to be very simple, bland, simplistic, avoiding nuanced equivalents. This is not obvious when looking at separate words, but it is evident in a larger reversed corpus: Here are some examples from the English-Latvian dictionary and possible more precise counterparts:

weird dīvains, savāds   ērmīgs, ērmots, spocīgs
quaint dīvains, savāds   īpatns, vecmodīgs
odd savāds, dīvains   jocīgs, "īpatnējs", savāds.

One is left with the impression that English is so much more varied and Latvian has a rather primitive and connotatively limited set of words, which is not true. It seems that bilingual dictionaries should not simply provide denotative equivalents but also stress the connotative component. This refers both to the lower and higher layers of lexis.

Similar deviation can be also observed on the denotative level. The meaning of the entry is usually more generalised (discussed above), often with awkward definitions that focus not on the essential functions but rather on a visual description:

disks uz kārts (disk on a pole) – lollipop;
dzeltenbrūngans ābols (yellowish brown apple) – russet;
liels apelsīns (a large orange) -- jaffa.

There are alsi some non-existent English neologisms created on the basis of Latvian words, mostly from one dictionary (ELDO 1995):

konkurētspēja -- competitive capacity (correct: competitiveness);
kravnesība -- load carrying capacity (correct: cargo (carrying) capacity).

In general we conclude that this way of augmenting existing bilingual dictionaries may be worth the effort. A serious increase has been observed in the number of entries and in the number of meanings and collocations (missed, new, terminological). For smaller languages whose lexicographical projects are often not commercially interesting, it allows a fast and considerable "beefing up" of existing resources and augmenting the list of existing words. The latter can also involve dictionaries of other language combinations (it is no secret that the source language tends to affect the Latvian part in a dictionary). On the other hand there are inherent limitations, most of which were discussed above which in many language combinations (having good monolingual and bilingual corpora) might render reversal impracticable and too time-consuming.

If this method is used in the future one of the consequences of this project might be that the compilers would practice avoiding explanatory translations in bilingual dictionaries. If this is not possible, the target component should with start the nearest synonym and then provide the differences. Such an approach would provide semantically more co-ordinated

material. Also more attention should be paid to the stylistic register and its retention in the translated part.

The new dictionary would be available in both printed and electronic versions.

References

Austermuehl, F. 2001. *Electronic Tools for Translators*. Manchester: St Jerome Publishing.

Cabre, T.M. *1999. Terminology. Theory, Methods and Application.* Amsterdam, Philadelphia: John Benjamins.

Corda, A. et al. 1998. 'Automatic Reversal of a Bilingual Dictionary: Implications for Lexicographic Work' in Th. Fontanelle et al (eds.), *EURALEX 1998 Proceedings*.ELDO 1995. Riga: LVI

Geisler, C. 1999. *Reversing a Swedish-English dictionary for the Internet.* www.stp.ling.uu.se/pkd99/synopsis/Geidlrt.html

Grefenstette, G. 2002. 'The WWW as a resource for lexicography' in M.H.Correard (Ed.), *Lexicography and Natural Language Processing*. Euralex.

Newmark, L. 1998. 'Reversing a One-Way Bilingual Dictionary' in Th. Fontanelle et al. (Eds.), *EURALEX 1998 Proceedings*.

Nesi, H. 2000. 'Electronic Dictionaries in Second Language Vocabulary Comprehension and Acquisition: the State of the Art' in *Proceedings of the EURALEX International Congress 2000*. University of Stuttgart.

Tamm, A. 2002. 'Reversing the Dutch-Estonian dictionary to Estonian-Dutch' in Proceedings of the Tenth EURALEX International Congress. Copenhagen: CST.

Veisbergs, A. 2001. *The New Latvian English Dictionary*. Riga: EAI.

Veisbergs, A. 2003. 'Computer Use in Translator Work, Terminology Management and Training' in *Terminology and Technology Transfer in the Multilingual Information Society*. Vienna-Riga: LLI.