Challenges in the Semi-automatic Reversion
of a Latvian-English Dictionary

Daiga Deksne[1], Andrejs Veisbergs[2]

[1] Tilde, Vienības gatve 75a, Riga, Latvia, LV-1004
[2] University of Latvia, Visvalža iela 4a, Riga, Latvia, LV-1050
E-mail: daiga.deksne@tilde.lv, andrejs.veisbergs@lu.lv

Abstract

The electronic version of the Latvian-English dictionary has been significantly supplemented over the last year with new linguistic material from corpora, databases and other sources. In contrast, the English-Latvian dictionary can be considered outdated as its electronic version was updated 10 years ago. This motivated us to create a semi-automatic process for reversion of the Latvian-English dictionary in order to supplement the English-Latvian dictionary with missing entries. Some of the major challenges for automatic reversion were as follows: grouping translations by part of speech, deciding to which entry the example should be attached, and ordering translations with similar meaning. By using automatic scripts it was possible to create reversed entries of quite good quality within a short time. Three groups of entries were prepared for manual post-editing: new entries with a single translation, new entries with a more complex structure, and existing entries with additional new content. The tasks for post-editing are: to check the suitability of the chosen headword, part of speech and translation order, to group the translations having the same meaning, and to move examples after appropriate translations.

Keywords: electronic dictionaries; bilingual dictionary reversing; phraseology

1. Introduction

The electronic version of the Latvian-English dictionary has undergone a series of significant revisions and has been supplemented with a significant number of entries that users previously lacked. There are words that have recently entered the language (both Latvian and English neologisms), words that have spelling variants, and words that are frequently found in the corpus but not found in dictionaries. It is common practice not to include regular derivatives in a dictionary. But as not every user of electronic dictionaries is a grammar expert and able to derive the needed word on his or her own, it is helpful to have some regular derivative forms included as well, such as deverbalized nouns, participles, and feminine forms of nouns. At present the Latvian-English dictionary comprises 54,465 entries, 139,796 translations and 23,617 usage examples. It can be considered to be the most up-to-date Latvian bilingual dictionary.

The English-Latvian dictionary was published in 1995, its electronic version was slightly updated in 2009. It comprises 52,202 entries, 118,723 translations and 32,510 usage examples. This dictionary can be considered outdated, and this motivated us to create an automatic process for reversion of the Latvian-English dictionary in order to supplement the English-Latvian dictionary with missing entries.

2. Studies of Reversion

Numerous reports describe attempts at compiling dictionaries by semi-automatic reversion of the opposite direction dictionaries. The language pairs of target dictionaries involve languages from the same language group, for example Estonian-Finnish (Langemets et al., 2017), as well as languages of different groups, such as English-Albanian (Newmark, 1999), Estonian-Dutch (Tamm, 2002), Latvian-English (Veisbergs, 2004), and Slovenian-English (Krek et al., 2008). The main motivation is to save time and the very valuable human lexicographers' resources, and to get the maximum benefit from abundance of examples and translation equivalents in the source dictionary. However, it is also noted that the process does not always go smoothly, and some, often unexpected, manual post-editing is required (Veldi, 2010).

3. Dictionary Structure

Both dictionaries are monodirectional, aimed at the Latvian user, but their microstructures differ. Usually every entry of both dictionaries starts with a single headword. There can be several headwords as well if they are absolute synonyms or phonetic variations. Generally the headword is in its canonical form. For the English-Latvian dictionary, the headword is followed by the pronunciation written using phonetic alphabet. Such information is not included in the Latvian-English dictionary. Within a given part of speech, translation equivalents are grouped into senses in the English-Latvian dictionary. Translation equivalents are grouped into senses in the Latvian-English dictionary as well, but part of speech information is absent as normally the Latvian ending clearly signals the part of speech (except in some minor cases). The most frequent senses are placed first. The sense may contain information about the usage domain (e.g., 'biol.'), register (e.g., 'slang') or some comment about usage context. Examples and their translation equivalents are included at the end of the particular sense. Idiomatic expressions and their equivalents are given at the end of an entry.

4. The Process of Reversion

4.1 Retrieval of data from the Latvian-English dictionary

The first step of reversion consists in the retrieval of words, translation equivalents and examples from the Latvian-English dictionary. The dictionary is internally stored in an XML format (Deksne et al., 2013). The XML tag names describe all pieces of information found in the microstructure of the dictionary entry. The following example (see Figure 1) shows an entry with three senses, each having a single translation, and two examples for the first sense as well as comments clarifying the second and third senses and the usage information for the second.

```
<entry title="leģenda">
  <title>leģenda</title>
  <mean digits="1" symbol="." />
    <transl>legend</transl>
    <from_sample>l. vēsta</from_sample><to_sample>legend goes</to_sample>
    <from_sample>leģendu krājums</from_sample><to_sample>legendary</to_sample>
  <mean digits="2" symbol="." /><comment>(spiega)</comment><usage>pārn.</usage>
    <transl>cover story</transl>
  <mean digits="3" symbol="." /><comment>(skaidrojums)</comment>
    <transl>caption</transl>
</entry>
```

Figure 1: Sample xml entry for the entry *leģenda* 'legend'.

To ease the reversion process, the data is transformed into a tabular format. Separate files are created for translations and for translation examples. Each line of the first file contains the title of an entry and the translation. The same entry title is on several lines if the entry contains several translations (see Figure 2). It is not important to preserve the division in senses, as translations of the Latvian-English dictionary will be the headwords of different entries in the English-Latvian dictionary.

leģenda	legend
leģenda	cover story
leģenda	caption

Figure 2: Translations from the entry *leģenda* 'legend' in a tabular format.

Each line of the second file contains the title of an entry, the example and the translation of the example with some optional comment (see Figure 3). There are several lines with the same entry title and the same example if the particular entry contains several examples or the example contains several translations. In a dictionary a word in an example may be abbreviated to the first letter of an entry title. In the further process, it will be expanded to a full word.

leģenda	l. vēsta	legend goes
leģenda	leģendu krājums	legendary

Figure 3: Samples from the entry *leģenda* 'legend' in a tabular format.

The entries of the English-Latvian dictionary are prepared in a similar way. An automatic process will be used to ignore translations and examples that are already in the dictionary.

4.2 Determining part of speech

Latvian words from the Latvian-English dictionary are morphologically analysed using the morphological analyser developed by Tilde (Deksne, 2013), by which their part of speech is determined. This is where the problems start: often a word is not in its basic form and it is attributed to various parts of speech or a part of speech which the

corresponding English word will not have, e.g. *izglītības* in Latvian is a noun in genitive. The English counterpart/equivalent 'educational' is (and should be labelled) as an adjective. Several parts of speech are attributed to the Latvian word *ātri*, but one has to choose adverb, since the English equivalent 'quickly' is an adverb. For many Latvian words the part of speech is undetermined, as they have not been included in the morphological analyser's dictionary. Among them are non-traditional compounds, foreign words, abbreviations, non-literary vocabulary, and so on. For 3,631 words out of 55,920 the morphological analyser does not return part of speech information.

The algorithm for choosing the most appropriate part of speech is the following:

- if the part of speech is unknown but a word ends with *-ot*, *-ēt* or *-ties*, it is a verb;

- if the part of speech is unknown but a word ends with *-ošs*, *-īgs* or *-isks*, it is an adjective;

- if the part of speech is unknown but a word ends with *-ējs*, *-tājs*, *-isms*, *-ists*, *-ums*, *-iņš* or some other common noun 'suffix + ending' pattern (35 in total), it is a noun;

- if a word is the past active participle in masculine singular nominative form with a definite ending it receives the part of speech 'noun', as such words have completed the process of nominalization; for example, the participle *pieaugušais* ('the grownup') in the dictionary is included as a noun;

- if a word is a noun in genitive it receives the part of speech 'adjective'; for example, *vietniekvārda* ('pronominal');

- if a word is the adjective in masculine plural nominative form with an indefinite ending it receives the part of speech 'adverb'; for example, *viesmīlīgi* ('hospitably');

- if a word is the adjective in masculine nominative form with a definite ending it receives the part of speech 'noun'; for example, *ļaunais* ('the evil one');

- other words keep the part of speech assigned by the morphological analyser if the current word form coincides with the basic form.

4.3 Adding examples and translations

Supplementing a dictionary based on the principle of nesting is complicated. For digital purposes a bilingual dictionary based on the alphabetic principle may be more convenient, although that would mean changing the whole pattern, which is not feasible in the short term.

Automatic joining of examples to the entry is the hardest task in the process. Should common phrases and multiword units (MWU) (Fellbaum, 2016) be included as separate entries, or as examples in the existing entry or as examples of contextual use? Which component of the MWU should we choose as the headword for joining? Which is the dominant word, e.g. in the collocation 'the language of the proceedings'? The problem is similar to that of deciding on keywords in the treatment of idioms in lexicography (Yong & Peng, 2007; Mulhall, 2010), and it is well known that users are not sure where to find idioms (Atkins & Varantola, 1998: 30).

English pattern	Latvian pattern	% of all examples	English example	Latvian equivalent
adj. + noun	noun	12.48%	'folic acid'	*folijskābe*
noun + noun	noun	7.93%	'savings account'	*krājkonts*
adj. + noun	adj. + noun	6.37%	'jolly crowd'	*jautra kompānija*
noun + noun	noun + noun	5.94%	'sports hall'	*sporta halle*
adj. + noun	noun + noun	5.31%	'normative act'	*tiesību akts*
'to' + verb + adv./particle	verb	2.58%	'to pay off'	*atpirkties*
adj. + noun	participle + noun	2.06%	'decisive battle'	*izšķiroša kauja*
'to' + verb + adj.	verb	1.94%	'to get fat'	*aptaukoties*
'to' + verb + det. + noun	verb + noun	1.68%	'to call the police'	*izsaukt policiju*
'to' + verb + adv.	verb	1.63%	'to beat back'	*atsist*
'to' + verb + prep.	verb	1.44%	'to blend in'	*iederēties*
noun + prep. + noun	noun + noun	1.26%	'field of action'	*darbalauks*
'to' + verb + 'a/the' + noun	verb	1.24%	'to get a fright'	*izbīties*
'to' + verb + noun	verb	1.03%	'to shed light'	*izgaismot*

Table 1: The most popular structural correspondences of English examples and their Latvian equivalents.

There are 28,155 examples in the Latvian-English dictionary. The English and Latvian examples often present different syntactical patterns (see Table 1 for the most popular structural correspondences). The most popular correspondences are as follows: 12.48% MWUs that have the construction 'adjective + noun' and 7.93% MWUs with the structure 'noun + noun' are translated as 'noun' in Latvian; 2.58% of examples with the structure 'verb + adverb/particle' are translated as 'verb'. English phrasal verbs are translated into Latvian predominantly as prefix-verbs (Veisbergs, 2013: 110-112). We generally see Latvian compounds as corresponding to English MWUs.

The issue of compounds is complicated both theoretically (Burger, 2007; Scalise, 2010) and practically, and increased due to the possibility of hyphenation (The Chicago, 2010; Vrbinc & Vrbinc, 2011: 256). First, while Latvian compounds by definition are written together (which ensures their separate entry status), this is not the case in English. Second, in both languages compound spelling often fluctuates both diachronically and synchronically, with a general tendency for two-component phrases to merge into a compound. In both languages normativizing tendencies (Levin-Steinmann, 2007: 37) exist but are hard to follow. This uncertainty and asymmetry in contrastive aspect has been noted by Čermak (2007: 20).

Thus we have to decide which compounds can be considered full entry words. In the existing English-Latvian dictionary many compounds frequently appear only as contextual examples, while the first component does not have a Latvian translation, for example, 'citric acid' (in Latvian *citronskābe*) is included within the entry with a headword 'citric'. It seems worth avoiding the "categorial bias" (Granger et al., 2012) and leaving some decisions as to where to place the word, compound or MWU for post-editorial work.

The automatic process starts with putting the content of the tab separated files of both dictionaries into hash tables. The data from the Latvian-English dictionary is treated in a reversed way, i.e. the key of a hash table is an English word/phrase and the value is a concatenation of the corresponding Latvian words/phrases. Only word/phrase pairs not existing in the English-Latvian dictionary are considered.

The phrases containing all content words with an initial capital letter are considered to be headwords. Phrases with the capital letters usually are some named entities like 'Little Red Riding Hood', 'the Atlantic Ocean', and 'the Book of Psalms'. The single words are considered to be headwords as well. We accept Latvian phrases consisting of one or two words as translations.

The most complex part of the process is to sort out examples. We ignore phrases containing more than five words. It is too risky to decide automatically which word of a phrase should be taken as a headword of an entry. We avoid full sentence-like examples. They frequently have almost word-for-word translation and do not provide any new information. For example, we do not process the example 'this accusation is unfounded' (in Latvian, *šis apvainojums nav dibināts*). We avoid such phrases by

looking for words 'is', 'am', 'are', 'were', 'was', 'has', 'have', and 'had' in the middle of a phrase or by checking if a phrase starts with 'I', 'you', 'he', 'she', 'we', 'they', 'are', and 'is'. Of course, in such a manner we could filter out some valuable examples as well, but with our abundance of examples the potential loss is far smaller than the benefit of quality assurance.

For some popular entry headwords the automatic process assigns up to 80 examples, and these examples are not found in the existing English-Latvian dictionary. Of course, this is too many for a single entry. We thus set a maximum limit and print out only the first ten examples per entry. The most example-rich headwords are 'time', 'work', 'right', 'way', 'call', 'cut', 'stand', 'covered', 'place', 'cover', 'pay', 'side', 'plant', 'hand', 'look', 'hold', 'throw', and 'day'. In our first experiments the process assigned numerous examples to both the stop words and common verbs. To avoid this, we compiled the lists of the stop words and common verbs which we do not choose as entry headwords for particular examples. The stop word list contains pronouns, prepositions, numerals, and some adverbs. The common verb list contains such verbs as 'make', 'be', 'give', 'get', 'put', 'push', 'pull', 'take', 'become', 'come', 'grow', 'turn', 'set', 'run', 'keep', 'bring', 'fall', 'let', 'make', 'break', 'play', 'draw', and 'use'.

In our final version, the algorithm for processing examples is the following:

- in a two-word phrase starting with a capital letter the first word is considered as a headword (e.g., for the examples 'Devonian era' and 'Devonian period' the headword 'Devonian' is chosen);

- we delete stop words from the beginning and end of the example and common verbs from the beginning if followed by a stop word, then we look for an appropriate headword in the remaining text string:

  o if a single word is left we consider it as a headword for the entry in which the current example is included (e.g., for the example 'to accustom oneself to' the headword 'accustom' is chosen);

  o if a text string starts with a common verb we take the word after the verb as a headword (e.g., for the example 'to make suffer' the headword 'suffer' is chosen);

  o if in the middle of a text string one finds the words 'into' , 'to', 'of', 'on', 'with', 'by', 'from', 'in', 'for', or 'a' and there are two words before one of them having an attributive ending, the word without an attributive ending is considered as a headword (e.g., for the example 'additional edition of copies' the headword 'edition' is chosen), otherwise all words before are taken as a head phrase (e.g., for the example 'a hard nut to crack' the head phrase 'hard nut' is chosen);

- if the phrase starts with the word 'to' we take the next word as a headword (e.g., for the example 'to adjust the fire' the headword 'adjust' is chosen);

- if the first word of a two-word phrase is a hyphenated compound and it is a headword in the existing dictionary we take it as a headword for the current example otherwise the second word is chosen (e.g., for the example 'colour-blind person' the headword 'colour-blind' is chosen, but for the example 'computer-composed music' the headword 'music' is chosen);

- for the other two-word examples we take the last word as a headword unless the number of examples for that headword has exceeded ten; then we take the first word as a headword (e.g., we choose the headword 'limit' for the example 'credit limit', but the headword 'credit' for the example 'credit line' as the word 'line' has too many examples);

- for the remaining examples, we take the last word of the example as a headword.

As headwords are chosen from examples, they are frequently not in their base form, like most of the headwords in the English-Latvian dictionary are. In order not to create too many separate entries unnecessarily, small adjustments are performed to the chosen headwords. If the headword is a verb in the simple past or present participle form and the corresponding root form is a headword in the English-Latvian dictionary, we take the root form for a headword ('praised' → 'praise', 'praying' → 'pray'). If the headword has the plural ending and the corresponding singular form is a headword in the English-Latvian dictionary, we take the singular form for a headword ('activities' → 'activity'). If the headword has a comparative or superlative ending and the corresponding base form is a headword in the English-Latvian dictionary, we take the base form for a headword ('smallest' → 'small').

Entries in the XML format are generated from the processed data. Translations with the same part of speech are grouped together in an entry. Groups are sorted alphabetically by part of speech abbreviation, e.g., the first are adjective translations and the last are verb translations. All examples are at the end of an entry, as it is impossible to determine after which translation a particular example should be. There is a single exception with the non-verb phrases. If an example does not start with the particle 'to' it is moved to the previous part of speech translation group.

4.4 Merging the new entries with the entries from the English-Latvian

dictionary

We store the dictionary data in the Microsoft SQL Server database on a permanent basis. For editing purposes, the data is exported to a plain text file. The new content and the existing dictionary are in the same XML format. The unique identifier of the

entry is its headword. The entry is left unchanged if the existing dictionary does not contain the entry with a specific headword or an example that equals the new content's headword. Otherwise we try to merge the new content with the existing entry, although there are some restrictions. We merge the existing dictionary entry and the entry with new content if both entries have translations with a single part of speech grouped in a single meaning only, and if the existing entry does not have examples. It would require too much manual work to merge entries with a more complex structure. For the tags containing a new content the colour attribute is added. This allows users to keep track of the part that is automatically included in an entry. When the XML format is transformed to the HTML format tags with a colour attribute provide a good visual indicator of the new content (see Figure 4). These entries still require post-editing, possibly with regard to changing the order of translations or grouping some translations in the separate senses.

agnomen [æg'nəʊmən] *n* palama, iesauka; pavārds
agriculturist ['ægrɪ'kʌltʃərɪst] *n* agronoms; agronome, lauksaimniece, lauksaimnieks
agrimony ['ægrɪmənɪ] *n* dadzis; ancītis; hemp agrimony – krastkaņepe
agronomist [ə'grɒnəmɪst] *n* agronoms; agronome

Figure 4: Existing entries automatically updated with new content.

Depending on the outcome of the merging process regarding the structure of an entry, we define three different post-editing tasks of various complexity:

1) for the new entries with a single translation and an optional example, the suitability of the chosen headword and part of speech should be checked (11,500 such entries);

2) for the new entries with several translations and/or examples, the suitability of the chosen headword, part of speech and translation order should be checked; translations should be grouped in meanings; every example should be moved after the appropriate translation (2,992 such entries);

3) for the existing entries with some additional content, new translations and examples should be moved to the appropriate position (6,368 such entries).

4.5 Separating senses of translations

Automatic separation of translations into senses is impossible. Thus, a convenient editing format is defined for manual processing. A special script is developed for transforming entries of the first and second tasks to a tabular format with six columns. The first column is reserved for the headword. If an entry has several headwords they are separated by a vertical line '|'. The second column contains the part of speech abbreviation. The third column contains the sense number for grouping of translations. Numeration of senses is organized within the framework of the part of speech. By

default, this column contains sense number '1'. The fourth column contains the translation. The fifth column contains one or several English examples separated by '|'. The sixth column contains one or several Latvian translations of examples separated by '|'. Not all columns are filled. Each line contains either the fourth column with the translation or the fifth and sixth columns with an example(s) and its translation(s) (see Figure 5).

agitated	a	1	uzbudināts		
agitated	a	1	uztraukts		
agitated	a	1		to be agitated|get agitated	uztraukties
agonizing	a	1	mokošs		
agonizing	a	1	mokpilns		
aiding	n	1	palīdzēšana		
aiding	n	1	veicināšana		
Aids	abbr	1	AIDS		
Aids	abbr	1	Aids		
Aids	abbr	1		contract Aids|get Aids	saslimt ar Aidu

Figure 5: New entries in tabular format prepared for post-editing.

The editor is asked 1) to correct the headword if it is not appropriate; 2) to check the part of speech; 3) to correct the sense number if an entry has translations with a different sense; 4) to move lines with examples directly after the appropriate translation. If some translations seem very distant from the headword or are used in a very narrow context the line should be deleted. Any spreadsheet application can be used for editing, and we use Microsoft Excel for this task.

5. Results and Discussion

The manual post-editing has not been completed yet. But the automatic part of the dictionary reversion process has prepared the rough material of quite good quality in a short time. The creation of the scripts for the reversion process took less than a month. As a result, 11,500 new entries have been created containing one translation equivalent or usage example and 2,992 new entries with more complex structure. These entries will help fill the gaps in the English-Latvian dictionary. The addition of new translation equivalents or examples to the existing 6,368 entries is not vital but enriches the dictionary, making its content more up to date, allowing the user to choose from a wider range of translation equivalents or to better understand the meaning of some unknown English word by exploring the newly added usage examples. Usage examples, multi-word terms or idiomatic expression meanings that have translations with different structures (frequently a single word) are especially valuable, for example: 'to appear publicly for the first time' (*debitēt* in Latvian), 'employee buy-out' (*uzņēmuma pārdošana darbiniekiem* in Latvian), 'to lie like a trooper' (*šausmīgi melot* in Latvian). Though some researchers have spoken in favour of omitting idioms when encoding

dictionaries (Hausmann, 2004), it seems they can contribute to a better overall reflection of the linguistic system of the language as well as improve users' choice and production capability.

The first results of the post-editing process reflect the quality of automatically generated entries. Of the first 610 post-edited entries containing one translation equivalent or usage example 64% did not require any editing, while 2% contained the wrong part of speech; for 16% of entries the part of speech tag was added as it was unknown before; 8% of entries were deleted as inappropriate; headwords of 4% of entries were corrected; the translations of 2% of entries were corrected; and the examples of 2% of entries were corrected.

The existing version of the English-Latvian dictionary is available online at https://www.letonika.lv/groups/default.aspx?g=2&r=10331062&f=1. After the post-editing process is completed the new version will be available at the same address.

6. Acknowledgements

7. References

Atkins, B.T.S. & Varantola, K. (1998). Language Learners Using Dictionaries: The Final Report on the EURALEX/AILA Research Project on Dictionary Use. In B.T.S. Atkins (ed.) *Using Dictionaries: Studies of Dictionary Use by Language Learners and Translators.* Lexicographica Series Maior, Number 88, Tübingen: Max Niemeyer Verlag, pp. 21–81.

Burger, H. (2007). Semantic aspects of phrasemes. In H. Burger., D. Dobrovol'skij, P. Kuehn, N.R. Norrick (eds.) *Phraseologie.* Vol.1. Berlin, New York: Walter de Gruyter, pp. 90–109.

Čermak, F. (2007). Idioms and morphology. In H. Burger., D. Dobrovol'skij, P. Kuehn, N.R. Norrick (eds.) *Phraseologie.* Vol.1. Berlin, New York: Walter de Gruyter, pp. 20–26.

Deksne, D. (2013). Finite State Morphology Tool for Latvian. In Mark-Jan Nederhof (ed.) *Proceedings of the 11th International Conference on Finite State Methods and Natural Language Processing.* St. Andrews: Scotland, pp. 49–53.

Deksne, D., Skadina, I., & Vasiljevs, A. (2013). The modern electronic dictionary that always provides an answer. In I. Kosem, J. Kallas, P. Gantar, S. Krek, M. Langemets, M. Tuulik (eds.) *Electronic lexicography in the 21st century: thinking outside the paper: proceedings of the eLex 2013 conference.* Ljubljana: Trojina, Institute for Applied Slovene Studies; Tallinn: Eesti Keele Instituut, pp. 421-434.

Fellbaum, C. (2016). Treatment of Multi-Word Units. In P. Durkin (ed.) *The Oxford*

Handbook of Lexicography. Oxford: Oxford University Press, pp. 411-424.

Granger, S., & Lefer, M. A. (2012). Towards more and better phrasal entries in bilingual dictionaries. In R. Vatvedt Fjeld & J. M. Torjusen (eds.) *Proceedings of the 15th EURALEX International Congress.* Oslo: University of Oslo, pp. 682–692.

Hausmann, F. J. (2004). Was sind eigentlich Kollokationen? In K. Steyer (ed.) *Wortverbindungen – mehr oder weniger fest.* Institut für deutsche Sprache. Jahrbuch 2003. Berlin, New York: de Gruyter, pp. 309–334.

Krek, S., Šorli, M., & Kocjančič, P. (2008). The Funny Mirror of Language: The Process of Reversing the English-Slovenian Dictionary to Build the Framework for Compiling the New Slovenian-English Dictionary. In E. Bernal & J. DeCesaris (eds.) *Proceedings of the XIII EURALEX International Congress.* Barcelona: Universitat Pompeu Fabra, pp. 535–542.

Langemets, M., Hein, I., Heinonen, T., Koppel, K., & Viks, Ü. (2017). From Monolingual to Bilingual Dictionary: The Case of Semi-automated Lexicography on the Example of Estonian–Finnish Dictionary. In I. Kosem et al. (eds.) *Electronic Lexicography in the 21st Century: proceedings of eLex 2017 Conference, 19-21 September 2017, Leiden, Netherlands,* Brno: Lexical Computing, pp. 155–171.

Levin-Steinmann, A. (2007). Orthographie und Phraseologie. In H. Burger., D. Dobrovol'skij, P. Kuehn, N.R. Norrick (eds.) *Phraseologie.* Vol.1. Berlin, New York: Walter de Gruyter, pp. 36–41.

Mulhall, C. (2010). A Semantic and Lexical-Based Approach to the Lemmatisation of Idioms in Bilingual Italian-English Dictionaries. In A. Dykstra & T. Schoonheim (eds.) *Proceedings of the XIV EURALEX International Congress,* Ljouwert: Fryske Akademy, pp. 1355–1369.

Newmark, L. (1999). Reversing a One-Way Bilingual Dictionary. Dictionaries. *Journal of The Dictionary Society of North America,* 20(1), pp. 37–48.

Scalise, S. & Vogel, I. (eds.) (2010). *Cross-Disciplinary Issues in Compounding,* Amsterdam, Benjamins.

Tamm, A. (2002). Reversing the Dutch-Estonian Dictionary to Estonian-Dutch. In A. Braasch & C. Povlsen (eds.) *Proceedings of the Tenth EURALEX International Congress.* Vol. 1, Copenhagen: Center for Sprogteknologi, pp. 389–399.

The Chicago Manual of Style. (2010). 16th ed. University of Chicago Press.

Veldi, E. (2010). Reversing a Bilingual Dictionary: a mixed blessing? In A. Dykstra & T. Schoonheim (eds.) *Proceedings of the XIV EURALEX International Congress,* Ljouwert: Fryske Akademy, pp. 861–865.

Vrbinc, A. & Vrbinc, M. (2011). Treatment of multi-word lexical items in the dictionary: the current situation and the potential problems facing dictionary users. *Eesti Rakenduslingvistika Ühingu aastaraamat 7,* pp. 249–263.

Veisbergs, A. (2004). Reversal as Means of Building a New Dictionary. In G. Williams & S. Vessier (eds.) *Proceedings of the Eleventh EURALEX International Congress.* Vol. 1, Lorient: UBS, pp. 327–332.

Veisbergs, A. (2013). *English and Latvian Word Formation Compared.* Rīga: Latvijas

Universitātes Akadēmiskais apgāds.

Yong, H. & Peng, J. (2007). *Bilingual Lexicography from a Communicative Perspective.* John Benjamins.