Reversing the Dutch-Estonian Dictionary
to Estonian-Dutch

Anne Tamm
Dutch-Estonian, Estonian-Dutch Dictionary Project
ned_est@yahoo.com

Abstract

This contribution evaluates some aspects of the reversing of the Dutch-Estonian electronic bilingual dictionary database to Estonian-Dutch. The project has linked two monolingual lexical databases and added new lexical and example units with the editor tool OMBI. The links are provided with information about the status of equivalence. The two sources are the Dutch Reference File and an Estonian database of polysemous words. The strategies of deriving correct polysemy representations of the Estonian items in the course of editing the Dutch-Estonian dictionary are be evaluated. Prior to dictionary editing, an Estonian reference file for polysemous words was created. In the course of editing, many missing entries and senses were added. The Estonian reference file consists of three structurally different parts: first, the left side of another bilingual dictionary, second, a database of a monolingual dictionary, third, a part created specially for the database. It is argued that the high quality of the target language database and a correct specification of the equivalence information are crucial for successful reversing. Verbal polysemy and its relation to the Estonian object case have posed a major challenge for the project.

The project, its tools and methods

This paper studies some polysemy-related aspects of a bilingual dictionary reversal. Bilingual dictionary reversal means here the changing the status of the target language to the status of the source language in a bilingual electronic database. I present material from the Dutch-Estonian and Estonian-Dutch project of the Dutch and Belgian governmental commission, the CLVV. Started in 1997, the Dutch-Estonian project aims at creating a database and two bilingual dictionaries, the first Dutch-Estonian and Estonian-Dutch standard modern language dictionaries. Both dictionaries will comprise 25 000 entries. The editors of the database work at different workstations, mainly in Estonia.

Our project has linked the senses and examples of two monolingual lexical databases. The data has been edited and organized with the electronic dictionary database editor tool OMBI. This tool was developed by the CLVV to organize lexicographic material into a database around form, lexical, idiomatic and example units and conceptually typed translations between two languages. While creating one bilingual dictionary, simultaneously a reversibly oriented bilingual dictionary is also being constructed, and the reverse dictionary can be always used for reference in the database. The database tool forms output files in the formats of RTF, HTML and SGML.

Three main sources serve as the basis for creating an Estonian-Dutch dictionary from a Dutch-Estonian one. The two monolingual sources are the Dutch Reference File, created by the CLVV and further referred to as the RBN, and an Estonian database of lexically

polysemous words, created by the project on the basis of several electronic files and manuscripts of the Institute for the Estonian Language in Tallinn. The Dutch-Estonian database serves as a collection of typed and specified conceptual equivalence links between lexically relevant units of the two languages, there are three levels of specification (see [Martin et al 1996] for details). This means that the conceptual, lexical and pragmatic differences between the meaning units of the two different languages are carefully documented and structured in the database. All translation links are provided with information about the status of equivalence between the units.

The strategies of deriving correct polysemy representations of the Estonian items in the course of editing Dutch-Estonian will be evaluated in this article. Here I have concentrated on pointing out the most typical problems and some specific solutions in our project. Why is target language polysemy, i.e. the polysemy of Estonian entries, important for constructing a Dutch-Estonian dictionary? Our aim was to write two dictionaries, the Dutch-Estonian and the Estonian-Dutch. In the case of our language pair, there are more Estonians who know Dutch and have a background in linguistics than the other way round. Since it is easier to translate into one's own native language, we have started first with the Dutch-Estonian direction. Our goal was to derive a basis for the Estonian-Dutch dictionary, which inherits much of the structure and information from the work done for the Dutch-Estonian part. The project has developed a method of linking two monolingual databases (see [Tamm 1997] for details) in order to diminish the amount of work after reversing the direction of the database to Estonian-Dutch. Prior to dictionary editing, an Estonian reference file for polysemous entries was created. In the course of editing, many missing entries and senses were added. The Estonian reference file consists of three structurally different parts: the left side of another bilingual dictionary, a database of a monolingual dictionary and a part that was created specially for the dictionary database. Available sources, not conviction motivated the use of structurally different parts in the project. On the other hand, this necessity allows us to draw conclusions about the pros and cons of various strategies of creating polysemy representations for the target language entries.

The left side of the bilingual dictionary used in our project originates from an Estonian-Russian dictionary under construction. The left side of entries from this part of the database can have a polysemy structure as in examples (1) and (2), the English translations are added by the author:

(1)   *abielluma (verb)*'to marry'
         [omavahel] 'each other'
         [naise kohta] 'about women'
         [mehe kohta] 'about men'
(2)   *armastama [0]* '(indication of a general sense) to love, like, be fond of'
         [armatsema] 'kiss, hug, make love'.

The left side of an entry from a database of a monolingual dictionary looks typically as the following example (3). The definitions of the monolingual dictionary entries are here presented in a shortened form here and numbered:

(3)   *kinnitama (verb)* 'to fasten, strengthen'

1.[panema (köitma, naelutama, kruvima vms.)] 'to fix, fasten, attach, put, place, tighten with rope, nails, screws etc'
2.[kindlamini, tugevamini ühendama] 'to unite, unify, tighten, stregthen'
3.['peatus'] 'to stop, land (reference to an expression about ships, in a port)'
4.[kinnistama (2. täh.)] 'to register an acquisition (of real estate)'
5.[milleski kindlust, vastupidavust andma] 'to strengthen, ensure, reinforce, consolidate, secure, make resistant'
6. [millegi õigsust, tõelevastavust tõendama] 'to affirm, acknowledge,assure, warrant'

7.[(asjaolude, sündmuste vms. kohta)] 'to confirm, belie (restricted to proofs as circumstances or events)'
8. [tõeks väitma, õigeks tunnistama] 'to certify, verify, corroborate, witness'
9.[lausuma, mainima, toonitama, väitma vms] 'to say, mention, assert, assure, warrant, argue'
10.[otsusele lõplikku kehtivust andma] 'to ratify, validate (a decision)'
11.[(õnnetuse puhuks) kindlustama (3. täh.)] 'to insure'
12.[tähtima (postisaadetise kohta)] 'to register (a letter etc)'
13.[kinnistama (4. täh.)];(a term from specialized language: photography)
14.[kindlustama, tugevdama] 'to fortify'

An entry of the third type belongs to the part that was created specially for the purposes of the database. The polysemy model is minimalistic, the sense groups are rather wide. The description of senses is formulated so that the differentiation of one sense from another is maximally easy for editors creating translation links from Dutch to Estonian. This means also that single synonyms for the adjectives in the given sense are avoided, since they can be ambiguous on their own. Although generally it is not advisable, we applied the labels 'literally' and 'figuratively' with occasional reference to typical modification environments. The following example is characteristic of the left side of the entries created before the bilingual dictionary's editing phase:

(4)   *tuline (adjective)*
        [konkr. kuum, pqlev, hqqguv] 'literally, hot, boiling, burning'
        [piltl. innukas (inimese kohta)] 'figuratively, eager (of humans)'

Some polysemous entries, especially numerous verbs, were created in the course of editing the Dutch-Estonian part, and distributed to other databases. The high quality of the target language database and the correct specification of equivalence information are crucial for successful reversing.

Representations of polysemy of the entries in the target language

After preliminary estimations, comparing the database section with a list of prospective Estonian-Dutch dictionary entries, 95% of the entries, i.e. form units, can be estimated to be derivable from the database after reversal. Almost one half of the Estonian entries created in the course of the Dutch-Estonian dictionary making process are, however, superfluous. In 5% of the cases, an additional lexical unit will have to be added to an existing form unit. Around 70% of the lexical units of lexically polysemous Estonian words will inherit already an equivalent after the reversal.

However, some reorganization in the microstructure and adding more equivalents and meaning descriptions is inevitable. Reorganization means in most cases the conflation of lexical units (see example (5)). Only 5 senses seem to be really relevant: 1, 6, 9, 10 and the not translated 5. See also example (3) for the English translations for the senses, the lexical units. Any information in Dutch is presented in italics or in italics and bold, the material in Estonian is set in regular or bold. The lexical units that are not frequent or not standard language and therefore must to be deleted in the course of post-reversal editing in the Estonian-Dutch dictionary carry a mark '=0' (added now by the author). The lexical units that will be placed under another lexical unit, are marked with '=' and the number of the lexical unit where the unit will be subsumed. Note that except lexical unit 5, which has no translation yet, the lexical units are translated or irrelevant for the Estonian-Dutch dictionary.

(5)   **kinnitama** verb 1.[panema (köitma, naelutama, kruvima vms.)] *bevestigen, vastmaken, aanzetten, vastzetten (alleen van concreta),* (pandla, lõksuga) *vastgespen,* (konksuga, haagiga) *vasthaken,* (tihedalt, kõvasti) *vasthechten,* (rõhknaela, nööpnõelaga, väikeste naeltega) *vastpinnen,* (ahela, ketiga) *vastketenen,* (kokku pigistades v. üksteise külge/vahele surudes, hrl. klambri v. kruviga) *vastklemmen,* (nööpnõeltega, väikeste varrastega) *vastprikken,* (köiega ümberringi köites) *vastsjorren,* (millegi järele) *aanhangen,* (pistetega) *aanhechten,* (niidi- või lõngaotsa) *afhechten,* (rihma, nööri, köiega) *aanbinden,* (fikseerides, hrl. köie v. ketiga) *vastleggen,* (klambriga, neetidega), *nieten (twee voorwerpen met elkaar)* **kinnitasin ratta/sadula oma kohale** *ik heb het wiel/zadel vastgezet;* **turvavöö/püksirihma kinnitama** *de autogordel/broeksriem vastgespen;* [mille] [mille] **külge/peale kinnitama** *iets aan/op iets vasthechten;* **nad ei kinnitanud laadungit köitega korralikult** *ze hadden de lading niet goed vastgesjord;* **riive uksele kinnitama** *grendels aanbrengen op de deuren;* **külgvaatepeeglit autole kinnitama** *een zijspiegel aan de auto bevestigen;* 2.[kindlamini, tugevamini ühendama]=1., 3.['peatus']=1. 4.[kinnistama (2. täh.)]=1., 5.[milleski kindlust, vastupidavust andma]; 6. [millegi õigust, tõelevastavust tõendama] *beamen,* (millegi toimumisest kinnitava tõendi andma) *constateren* **kahtlust kinnitama** *een vermoeden bevestigen;* **sadam pakub foto järgi vähe huvitavat ja ka kirjeldus kinnitab sama** *de haven ziet er op de foto's weinig interessant uit en de beschrijving is navenant;* 7.[(asjaolude, sündmuste vms. kohta)]=6.; 8. [tõeks väitma, õigeks tunnistama] = 6. *bevestigen* [mida] **täielikult kinnitama** *iets volmondig beamen;* 9.[lausuma, mainima, toonitama, väitma vms] (järjekindlalt) *volhouden, de verzekering geven, verzekeren* **ei kinnita ega lükka ümber** *bevestigen noch ontkennen;* 10.[otsusele lõplikku kehtivust andma] *vaststellen,* (jur.) *sanctioneren,* (ametisse) *bevestigen,* (tugevalt, kindlalt) *verankeren* **programmi kinnitama** *een programma vaststellen;* **põhiseadusega kinnitatud olema** *in de grondwet verankerd liggen;* 11.[(õnnetuse puhuks) kindlustama (3. täh.)]=0; 12.[tähtima (postisaadetise kohta)]=0; 13.[kinnistama (4. täh.)]=0; 14.[kindlustama, tugevdama] *bevestigen.*=0

The percentage of already translated lexical units can be higher after the conflation of senses than the preliminary estimations would suggest. The necessity for conflation is particularly evident in the case of verbs, adjectives and function words: their polysemy is treated with maximal precision in our Estonian monolingual reference source. The example section is the main area where deeper quantitative changes are necessary. This means mainly the deletion of the irrelevant illustrations, i.e. the Estonian lexically free example sentences that are translations of lexically restricted Dutch combinations. The addition of new Estonian lexical

combinations is occasionally necessary. This means also the adding of the respective Dutch translations.

More important changes pertain to the quality of the representation of the data once the direction of the dictionary is reversed. We will identify below the changes in the microstructure, more particularly, the sense structures. More than deletion and addition, the alteration of the existing material poses problems for the post-reversal editor. This means reorganization within the sense and example material, changing the order of units, merging some and splitting other units.

After first estimations, the conceptual coverage of the equivalence is relatively high as demonstrated in the example (5) presenting an entry from the reversed Dutch-Estonian dictionary. Not only one-to-one translations are found back after reversal, but synonym translations and conceptual variants are also found, together with information about conceptual (or pragmatic, etc) differences in the translation pairs (see the first sense of example (5)).

The alterations thus are mainly concerned with improvements to the representation of polysemy and especially, verbal polysemy in relation to cross-linguistic differences in approaches to argument structure.

Which polysemy model is the best?

A polysemy model is also dependent on the strategy of the dictionary user. A decoding user has little access to the meaning structure of the source language, so he would not profit as much as the encoding user from the careful sense structure within the entry. For him, rather formal anchors to rely on are much more relevant. For an encoding user, a transparent polysemy structure of the entry is an important anchor for starting a search. So in our case the Estonians who would profit most from a sense structure in an Estonian-Dutch dictionary. If our aim is to create a database that serves as a basis for both language users, we have to find a simple and transparent structure for polysemy. Therefore the type of structure displayed in the previous example (5), presenting polysemy as in a monolingual dictionary, must be reduced in quantity. The opposite strategy of one single lexical unit that has to be split if necessary, however, has proved to be time-consuming in the post-reversal editing phase.

One-by-one splitting into senses must be carried out with the words that are polysemous, but not included in the database of polysemous words. They had to be entered in the database by the editors, each in their own database and only if there emerged the need for it, i.e. if there was a Dutch word that had to be translated with that particular word. This problem is encountered mostly with verbs and adjectives with general or vague meaning and several instantiations of a central, unique lexical sense, e.g. *sööma* 'to eat'. Contrasting the marginal uses of those verbs and adjectives by means of ambiguity tests would yield polysemy. However, since all uses listed (see example (6)) are intuitively derivable from the central sense, none of the editors has felt compelled to distinguish a separate sense, therefore the identifier for 'monosemous' and general, central meaning, the mark '[0]', has been

unanimously added by each editor. As our final database comprises the five databases of editors working in different geographical locations, we derived the following data:

(6)   **sööma** verb [0] *eten* (loomade kohta) *vreten,*(palju) *verstouwen; spijzigen* '(hea) isuga'/isukalt (sööma)(eten) *met graagte (eten)*; magusat/maiustusi sööma *snoepen*; õhtust sööma *souperen* ta ei söö liha ega kala *hij eet geen vlees of vis*.rooste sööb terase ära/läbi *de roest vreet het staal uit*; kui sa oma taldrikut tühjaks ei söö, siis... *als je je bord niet leeg eet...*; väljas sööma *buitenshuis eten*; siin tuleb süüa, mis ette pannakse *je moet hier eten wat de pot schaft*; neil ei ole midagi süüa *ze hebben geen eten*; (lapsele/loomale) süüa andma *(een kind/dier) eten geven*; süüa tegema *eten klaarmaken*; meil pole midagi süüa *we hebben niets te eten*; meil on (midagi) süüa *we hebben iets te eten* ; me pole veel söömist lõpetanud *we zijn nog niet klaar met eten*; hiina/türgi/prantsuse toitu sööma *Chinees/Turks/Frans eten*; isuga praekala sööma *zich tegoed doen aan gebakken vis*; edasi sööma *dooreten*; isuga/kiiresti sööma *dooreten*; ära lobise ja söö parem edasi *stop met praten en eet eens door*; ära/läbi sööma (söövitama) *uitvreten*; poisid, sööma! *jongens, etenstijd!*; nad kutsuvad sõbrad (enda juurde/poole) sööma *ze geven een etentje voor vrienden*; see ei kõlba süüa *dat is niet te eten* ; ennast ümmarguseks sööma *zich een tonnetje rond eten;* end [kuhu] sisse sööma *zijn kont ergens indraaien.*

Perhaps, strictly speaking, the verbs of the above kind should not be regarded polysemous lexical-semantically, but it is a common practice in dictionaries to split them up for practical reasons. It is much more difficult to sort examples and lexical units if they are all added in the order of the translations from the Dutch-Estonian editing phase; all under one single structural unit. This problem is also relatively common with the polysemous entries that were derived from an earlier bilingual dictionary. Consider the entry *armastama* 'to love' (example 7, translations of senses are in example 2) as represented in the different sub-databases:

(7)   **armastama** [0] *beminnen* **oma laeva/perekonda/vabadust üle kõige armastama** *je schip/gezin/vrijheid boven alles beminnen* [armatsema].
        armastama [0][armatsema] [söögi kohta] **ma ei armasta krokette** *ik houd niet van kroketten*
        armastama [0] **heeringat/Amsterdami/Pavarottit kohutavalt armastama** *dol op haring/Amsterdam/Pavarotti* [armatsema].
        armastama [0] *liefhebben* **oma ligimest armastama** *uw naaste liefhebben*; **teineteist armastama** *een verhouding hebben*; **nad tunnistasid, et nad armastavad üksteist** *ze gaven toe dat ze een verhouding hadden met elkaar* [armatsema].
        armastama [0] (millegi lembene olema) *minnen* (sümpaatne olema, hoolima) *mogen* (meeleldi nägema, sööma vms.) *lusten* **vaikust armastama** *de stilte minnen*; **spinatit mitte armastama** *geen spinazie lusten* [armatsema].

This example shows that in the entry 'kiss, hug, make love', on the one hand it is necessary to distinguish more lexical units and on the other hand less lexical units than in the original. The sense worded as *armatsema* 'make love' has been relevant for the Estoinan-Russian dictionary, but it is obsolete in present-day Estonian language. Therefore it must be deleted for the purposes of our dictionary. This is a typical problem, also illustrated by example (8), in the database part that used the left side of an earlier bilingual dictionary. The general sense should be split or its equivalents and example units must be at least rearranged. One group of equivalents centers around the concept of loving, the special love relation between

people. The other group of equivalents expresses more generally the concept of liking or preferring, especially about food. In too wide senses there is always one editor who has evidence that the sense is too wide, and who adds a new sense. In this case, one editor has added a lexical unit 'about food'. This specification is narrower than reasonable, but in its context, it might have seemed quite justifiable. In order to get any solution, many items have to be regrouped in the course of post-reversal editing. A more suitable pre-constructed polysemy structure would have yielded a less time-consuming solution.

Many lexical units must be eliminated since they have been included for their relevance in the Estonian Russian contrastive perspective. Occasionally editors have placed examples under the too narrow lexical unit specifications and the link must be relinked under a more general lexical unit. Editors working on Dutch-Estonian do not always activate the meaning structure of the whole target word, and they rely on the information on the target word's polysemy as they see it on the screen, picking out the most specific lexical unit that suits their source language information. In the following example, *abielluma* 'to marry', the last two lexical units 'about women' and 'about men' should be in fact deleted. Before deleting the example sentence under 'about men', 'He announced that he was going to have a wedding in december' must be regrouped under the first lexical unit 'each other'. The lexically irrelevant reference to gender in the sentence and in the polysemy structure of the database have mistakenly 'found each other'. See an example of the structure of two editors' work in example (8):

(8)   **abielluma** [omavahel] *trouwen, trouwen* noorelt abielluma *jong trouwen*; [kellega] abielluma *trouwen met iemand*; [naise kohta] [mehe kohta]
        abielluma [omavahel] [naise kohta] [mehe kohta] **ta teatas, et abiellub detsembris**

        hij deelde mee dat hij in december ging trouwen

However, it turns out that the native speaker's intuition about words that they are lexically polysemous is often amazingly correct. Many senses that are identified as distinct by one editor are identified as distinct by other editors as well. This is especially true if there are two quite distinct meaning clusters associated in a word, or the usage area of the each lexical unit is pragmatically distinct. But formally it is a matter of just one letter that the merger of the lexical units of the two editors yields us two items instead of one, as it should. This problem is illustrated with the following example (9) 'to eat' *söötma* 1= 'to feed, to give food or fodder to someone', 2 = 'to throw the ball to someone in a ball game', 3 = serve or give as input for processing. I have numbered the conceptually identical, but differently formulated senses identically:

(9)   [**söötma** [süüa, söögiks andma] = 1 *geven (van het eten)*
        [söötu andma] = 2 *doorspelen* (mbt balspelen) **palli (kaasmängijale) söötma** *de bal (aan een medespeler) doorspelen.*
        söötma [toitma] = 1 (loomi) *voeren,*(loomi) *voederen,*(loomi) *voer geven* **loomade söötmine**
        keelatud! *verboden te voeren!*;
        [palli viskama] = 2
        [sisse andma] = 3 *voeden* **arvutisse andmeid söötma** *een computer voeden (met informatie)*

However, it is easier to merge two already identified structural units with all their translations and examples and the examples' translations than to consider each item

individually for regrouping. Here again we must conclude that the most straightforward way to derive a suitable polysemy structure is to have it pre-constructed before the dictionary's editing phase. A lexical sense structure, which is specially created for the purposes of the dictionary's user profile and where the identifying material of the lexical units is worded in a compact, sketchy way means a huge bonus for deriving the correct grouping of translations and examples under lexical units. Consider the following succint representation for the pre-constructed polysemy of the adjective *tuline* 'hot' in example (10) (see the English translations in 4). This representation is shared by the editors, and the three editors contributed the following translation links to the lexical units [konkr. kuum, põlev, hõõguv] 'literally, hot, boiling, burning' and [piltl. innukas (inimese kohta)] 'figuratively, eager (of humans)':

(10) **tuline** [konkr. kuum, põlev, hõõguv] *vurig, heet* **tuline!** *je bent warm!*; **(nagu) tulistel sütel istuma** *op hete kolen zitten*
[piltl. innukas (inimese kohta)] *vurig, heet, heftig,* **tuline pooldaja** *een vurig voorstander.*
tuline [konkr. kuum, põlev, hõõguv]
[piltl. innukas (inimese kohta)] *ziedend,*(vurig, fervent) *ijverig,*(vurig, heftig) *driftig.*
tuline [konkr. kuum, põlev, hõõguv] [piltl. innukas (inimese kohta)] *fel, fervent***tuline pooldaja/vastane** *een fervent voorstander/tegenstander (van iets)*

Example (11) displays the merger of the work of three editors:

(11) **tuline** [konkr. kuum, põlev, hõõguv] *vurig, heet;* **tuline!** *je bent warm!*; **(nagu) tulistel sütel istuma** *op hete kolen zitten* [piltl. innukas (inimese kohta)] *vurig, heet, heftig* (vurig, fervent) *ijverig,*(vurig, heftig) *driftig, fel, fervent, ziedend,* **tuline pooldaja/vastane** *een fervent voorstander/tegenstander (van iets);* **tuline pooldaja** *een vurig voorstander.*
Even if later in the course of editing, a polysemous item was discovered, it was never too late to make a unified polysemy structure for it and distribute it to the editors working with other databases.

Estonian verbs and their objects' case

Verbal polysemy and its relation to the Estonian object case have posed major challenges for the project. Let us put aside numerous specific function words, and give an example of a more pervading representational problem from the given language pair. There are many languages where the tense, mood or aspect is grammatically encoded on the object (or another dependent, or the subject), contrary to the pattern of the lexicographically well-described languages where these categories are encoded morphologically on the verb, or compositionally. How they are exactly encoded is partly lexically dependent and therefore it is the task of a dictionary to specify the details (see [Tamm to appear]). Linguistic or theoretical changes must bridge the gaps in the structures of any two typologically different languages and reflect a uniform theoretical background in the description of those languages.

Like Finnish, Estonian is one of those languages where aspect is encoded by the object case, by means of the morphological partitive versus genitive/nominative case: both can be regarded as a rough equivalent of the accusative object case in many other languages. The morphological partitive case on objects yields activities, genitive and nominative object

cases yield accomplishments and achievements. This parallels in Dutch, Swedish, German, Hungarian and the Slavic languages partly with the function of the perfective verbal prefix or verbal particle, as in Estonian *Ta sõi melonit*(.part) 'He was eating the melon' and the Dutch equivalent *Hij was een meloen aan het eten* and *Ta sõi meloni*(.gen) 'He ate up a melon', and the Dutch equivalent *Hij heeft een meloen opgegeten.*

This is how both object cases are possible within one sense or lexical unit in case of verbs that can give expression to the opposition between activities and accomplishments. Moreover and crucially for the Dutch user, the verb-object combinations with different case forms must often be translated with different Dutch verbs, as the previous examples show. We would not like to split the Estonian lexical unit in two according to the two possible object cases, since the strictly lexical meaning is identical in both cases, even in the following example where distinct syntactic patterns typically co-occur with either object case:

(12)   *Mehed*   *veeretasid*   *vaadi*   *õue.*
        men.nom   roll.past.3pl   barrel.gen   yard.illat
        'The men rolled a/the barrel into the yard.'
(13)   *Mehed*   *veeretasid*   *vaati.*
        men.nom   roll.past.3pl   barrel.part
        'The men rolled/were rolling a/the barrel.'

The grammars of the Estonian language have not been able to explain the phenomenon of object case alternation adequately, while the choice for the object case forms a real problem for non-Estonian language users. An additional complication of the lexicographer's work is the canonical form of the verb-object collocations, which allows the emergence of the partitive case only. The object-infinitive verb collocations in Estonian dictionaries, e.g. in the English-Estonian dictionary, never specify an object's case other than partitive: *to break a record* – *rekordit* (record.part) *purustama* (break.mainf)[Silvet 1990:170]. This restriction pertains to objects only. Lexically determined inherent case of complements, however, does not emerge in partitive in dictionaries: *to break with (somebody)(kellegagi)*(somebody.**comit**.clitic) *läbikäimist* (communication.part) *või* (or) *suhteid* (relationship.partpl) *katkestama* (cut.mainf) [Silvet 1990:171]. In Estonian, this canonical form does not adequately represent the aspectual meaning potential of verbs and the verbs' object case possibilities in sentences. The object case is revealed only in full sentences:

(14)   *Ta*   *purustas*   *rekordi/*rekordit.*
        s/he.nom   break.3sgpast   record.gen/recort.part
     'He broke the record.'

The problem has not been addressed due to the historical tradition and isolation. This particular drawback of the canonical form has not disturbed anybody, since the bilingual dictionaries containing Estonian were written by Estonians and for Estonians. Non-Estonian speakers, who encounter problems identifying the lexical items that contain the information in their lexical structures that restricts the possibilities in object case, did not practically exist until recent years.

Therefore, it is lexicographically difficult to represent the meaning of an Estonian transitive verb adequately in Estonian without any object-verb collocations. It is in a certain sense comparable to excluding all verbal prefixation and verb-particle combinations from a Germanic or Slavic dictionary. So in a dictionary all transitive verbs should be provided with at least one example unit with a typical, conventionalized object-verb combination, and this creates additional work in the example section after the reversal. Also, an occasional activity and accomplishment distinction between lexical units must be removed.

Our project has experienced that a theoretical framework that takes only the general, object-case and syntactic alternation independent meaning as the basis for lexical polysemy would unify the approach to syntactic alternations. So we do not have to double the lexical units, where sense 1 corresponds to the verb's combination with partitive object and sense 2 to the verb's combination with genitive/nominative object (see examples (12) and (13)). Two lexical units are assumned if the difference in object case is accompanied with a real difference in the lexical meaning, as demonstrated in the following sentence pair:

(15)   *Ta*   *saatis*   *lauljat*   *klaveril.*
         s/he.nom   accompany.3sgpast   singer.part   on the piano
         'He accompanied the singer (on the piano).'

(16)   *Ta*   *saatis*   *laulja*   *ukseni.*
         s/he.nom   accompany.3sgpast   singer.part   to the door
         'He saw the singer to the door.'

Differently from traditional sources that offer no reliable guide to the Estonian verb meaning and object case for non-Estonian dictionary users, we assume the existence of a constructional argument structure, separated from the Frame Semantic participant structure. This argument structure with for example its specific case assignment properties is conventionally associated with a given verb in the given lexical meaning. This association must be represented in the example section illustrating the lexical unit, with an explicit genitive or partitive case on the object. For the cases of the two lexical units from examples (15) and (16), we create the Estonian examples with an explicit genitive or partitive case on the object even in the canonical form (17):

(17)   *lauljat(.part) (klaveril) saatma* 'to accompany the singer on the piano'

        laulja(.gen) (ukseni) saatma 'to see the singer to the door'

Sometimes it was necessary to include the canonical form, as a translation from Dutch, more succintly in the database by means of pronouns and reference to substitution classes:

(18)   *[keda] [millel] saatma* 'to accompany someone on something'

        [kelle] [kuhu] saatma 'to see someone somewhere'

By means of indicating the object case possibilities, our project has made a major contribution to help non- Estonian users. This dictionary project is unique in this aspiration, but hopefully not unique for years to come.

Conclusion

The most important insight gained in this project is the fact that the tool and the method have proven successful in the structural organization of the material in the main subdivisions, but the lexical unit and example section have to undergo qualitative changes, deletions and additions. This is partly motivated by the problem of object case alterations in Estonian verb-object collocations. High quality of the target language database and a correct documentation of the equivalence information are of prime relevance for the successful reversing of the database.

The best strategy for deriving correct target-language polysemy representations is to construct a preliminary database of lexically polysemous target language entries. The pre-constructed database of polysemous entries has also the advantage of relatively little post-reversal editing, since the polysemy model is already tailored for the L2-L1 dictionary user profile. The left sides of other bilingual dictionaries often present their material according to the profile of the given dictionary. Additionally, there is a tendency to shape the polysemy structure of the source language according to the structure of the target language: many senses are invented even if in the monolingual context they do not exist as lexical senses, other important lexical senses are omitted. The left side of a monolingual dictionary, if it contains senses, is a reasonable alternative. However, post-reversal editing of the polysemy structure will be in this case time consuming, since many senses must be merged and formulated more concisely. There is also a tendency in monolingual dictionaries to present the central senses with analytic definitions, and differentiate the less central senses from the central senses with other types of definition, e.g. synonym definitions. Neither full analytic definitions nor synonym definitions are fortunate for editors who approach those definitions as simple anchors or reference points in the course of linking lexical units form source language. The fact that this is a problem is also reflected in the number of mislinked senses. It is therefore crucial that the descriptions of the senses do not simply describe. They must provide first and foremost an anchor for identification, an unambiguous formulation of the main difference between the lexical units of an entry.

Distinguishing a new sense of a verb on the basis of a different syntactic pattern it occurs in, or its Vendlerian class is only reasonable if it is accompanied by a change in the verb's lexical meaning. Our project has changed the canonical form of Estonian object-verb collocations.

References

[Martin et al 1996] Martin, W., Tamm, A., 1996. OMBI: An Editor for Constructing Reversible Lexical Databases, In: eds. Martin Gellerstam et al (eds.) *Proceedings of the Seventh Euralex International Congress, Euralex 1996, volume I*, pp. 675-688. University Gothenburg, Sweden.

[Silvet 1992] Silvet, J., 1992. *Inglise-eesti sõnaraamat, Vol I*. Tallinn, Valgus.

[Tamm to appear] Tamm, A., 2002. Estonian object case, progressive and transitive verb classes. To appear in: *Proceedings of the 19th Scandinavian Conference of Linguistics*. Tromsø, Norway.

[Tamm 1997] Tamm, A., 1997. *Bilingual lexicography of Estonian and Dutch, reversible database building of these languages*, M.A.-thesis, University of Tartu, Tartu, Estonia.