Aspectual Mismatches in Bilingual Dictionaries

Anne Tamm
Research Institute for Linguistics and the University of Szeged
P.O. Box 701/518
1399 BUDAPEST
HUNGARY
annetamm@yahoo.com

Abstract
This article calls the attention to an equivalence problem in bilingual dictionaries. I present the evidence that the equivalence between aspect verbs is systematically incomplete in bilingual dictionaries. Aspectual mismatches emerge between languages that derive sentential aspect by different means. Two Finno-Ugric languages are compared, Hungarian and Estonian. Mismatches emerge despite the typological relatedness and the identical conceptual coverage for the perfective aspectual category. However, Hungarian relies on the lexical means of encoding aspect, whereas Estonian uses also grammatical means. Many problems of lexical aspectual mismatches can be solved by editing the dictionaries on the basis of uniformly designed and comparable monolingual knowledge bases, where the lexically and non-lexically determined aspect is distinguished. Recording the lexical aspect of each verb in a database according to a uniform system enables the database tools to notify the lexicographer when an aspectually incomplete translation is created. The equivalence problems that appear in dictionaries of the two typologically related languages discussed here may be present in bilingual dictionaries containing typologically less related languages. Aspect is a matter of bilingual lexicography if expressing aspect is determined verbal lexically in at least one of the languages.

1. Equivalence of Aspect Verbs

The aim of this article is to call the lexicographers' and bilingual database designers' attention to the following problem: equivalence between aspect verbs is systematically incomplete in bilingual dictionaries. My observation is that equivalence mismatches are an inevitable outcome in bilingual dictionaries containing languages that encode aspectual distinctions by different means: by lexical, grammatical, or compositional. This problem is solved if bilingual dictionaries are edited on the basis of uniformly designed and, therefore, comparable monolingual knowledge bases, where the distinction of lexically and non-lexically determined aspect is clearly drawn for each language.

Equivalence mismatches that are studied here are aspectually unbalanced equivalence relations between the lexical items of source and target languages. Verbs contribute differently to the aspectual composition of a sentence. The aspectual mismatches emerge as the result of underspecification and specification (lexicalization) of aspectual distinctions in verbs. In Slavic languages, there are perfective verbs that fully determine the perfective aspect of the sentence; these verbs do not shift aspectually. In contrast, there are other languages, such as many Germanic languages, that derive aspect compositionally; that is, the verb and the quantification of its arguments determine the aspectual value of a sentence (cf. Verkuyl, 1993). Yet other languages, such as Finnic, have a fairly large class of verbs where the verb and the object—but not the quantificational properties of the object—determine the

aspectual value of the sentence; instead, the object case does. A number of Finnish verbs are underspecified for aspect; they emerge in a perfective ("bounded") sentence when the object case is accusative and in imperfective ("unbounded") sentences when the object case is partitive (Heinämäki, 1984; Kiparsky, 1998). As there are considerable cross-linguistic differences between the specification of aspectual values in a verb, a Russian verb cannot be always precisely translated: the Russian verb specifies lexically "more aspect" than the Finnish or the Germanic equivalent. In this paper, I am primarily concerned with the perfective subcategory of aspect. Perfectivity is a matter of bilingual lexicography if, in a language pair, the means related to expressing it are verbal lexical in at least one of these languages. Before addressing the concrete examples of aspectual mismatches in section 5, I introduce the problematic verb group in section 2, the conditions for equivalence relations in section 3 and the motivation for a study of the Hungarian-Estonian equivalents in section 4. Section 6 sketches some solutions.

2. Problematic Verbs

This article targets the mismatches that are found between verbs that are here referred to as "aspect verbs". These verbs occur in sentences that are described as "perfective", "telic", or "bounded", such as *Mary made a chair,* or *Mary made a speech.* As an example, these verbs with their quantized arguments fail in the progressive-perfect entailment test: *Mary is making a chair/a speech* does not entail *Mary has made a chair/a speech* as opposed to verbs (e.g., sing, run) that do not: *Mary is singing/running* entails *Mary has sung/run.* The term "aspect verbs", as used here, covers a rather wide variety of verbs that are referred to in different languages and theoretical frameworks as verbs that have aspectual preverbs, particles or verbal prefixes, aspect, terminative, transition, boundable, perfective verbs, incremental-theme, change-of-state, change-of location verbs, verbs of (definite) change, accomplishment, achievement, some activity verbs, dynamic, or telic verbs. The activity-accomplishment, prefixed-nonprefixed verb distinctions stand in the focus here.

3. Equivalence Conditions

Equivalence mismatches in a bilingual dictionary pair emerge either (a) if a given equivalent is not lexically insertable in the equivalent target language context since it gives a different value to the sentence, (b) if the equivalence relations are established in the direction of L1-L2 but not in the direction of L2-L1, or (c) if two different L1 items are provided with an identical L2 equivalent (or v. v.) without appropriate meaning discrimination. In order to demonstrate what kinds of mismatches occur with aspect verbs, I present the ideal conditions for equivalence relations in bilingual dictionaries.

A. Insertability and equivalence preserving. The task of bilingual dictionaries is to provide equivalents that are insertable in an equivalent target language text. For instance, if one has to translate *dog* in a sentence *The dog has four legs* and you have a target Hungarian context of *A ___-nak négy lába van*, then in the lexical equivalence pair *dog-kutya*, the equivalent found in the dictionary, that is, in this case, *kutya,* must be lexically a suitable and insertable equivalent in the above context. The inserted item must not change the target sentence in a way that the equivalence of the source and target context disappears.

B. Symmetry. Ideally, an equivalent pair is symmetric; and if it is presented as symmetric in L1-L2, it should be also represented as such in the L2-L1 dictionary. That is, the equivalence pair A→X in a dictionary with the direction of L1-L2 ideally yields an equally valid translation equivalence pair X→ A in the reverse, L2-L1 dictionary. If the item *kutya* is provided as the complete equivalent of the item *dog* in English, then in the reverse, that is, English-Hungarian dictionary, *dog* should also appear as the complete equivalent for *kutya.*

C. Balancing (meaning discrimination). Ideally, if two different entries are provided with an identical equivalent in a bilingual dictionary, there must be some specification for the target user about what is the exact difference between these two source items. For instance, if both *kutya* and *eb* are translated as *dog* in English, then it helps the English user to understand the Hungarian words when there is some extra indication (pragmatic labels, illustrative examples, etc) of the difference between the two Hungarian items (and, also, the English item). Balancing the relation is even more relevant in the opposite case, that is, when one lexical item is provided with two equivalents.

In practice, these ideal equivalence conditions are frequently violated; however, most of them have not been addressed properly for a number of reasons. One reason is that the human users of dictionaries have direct access to several other sources than the given equivalence pair, such as good knowledge of L1 and at least elementary knowledge of L2, synonyms, example sentences, etc. Therefore, the occasional mismatches in various areas have never appeared clearly enough to lead to systematic research. However, in the age of the emergence of production (active) dictionaries and massive electronic databases and dictionaries, where the data must be suitable for automatic "recycling" in new products, it is desirable to articulate equivalence relations systematically. A more serious reason for failing to address mismatches is the existence of less studied semantics-lexicon-syntax interface areas, of which aspect is an example. Not only are subcategories of aspect defined differently across several traditions, rendering the descriptions of lexical material in various languages incomparable but, also, languages lexicalize aspectual distinctions differently across languages. It is the task of lexicographers and lexical database designers to identify the frequent mismatch areas and to find systematic solutions. In the following sections, this contribution makes this effort in the field of aspect verbs.

4. Motivation for the Data Set

The best sources for examining equivalence mismatches are dictionary pairs of L1-L2 and L2-L1 if they are similar in volume, editorial style and user orientation. These characteristics meet in the Estonian-Hungarian, Hungarian-Estonian dictionaries. The second reason for the study of the Hungarian-Estonian language pair is that comparing two languages with an identical conceptual coverage for a category, perfectivity in this case, is an asset in a systematic research into mismatches. From the languages known well by the author of this article, Dahl's perfectivity tests (Dahl, 1984:20-22) single out two that yield an identical conceptual coverage pattern: Hungarian and Estonian (Dahl, 1984:21). The third reason that makes the study of these two languages interesting is that from the viewpoint of the lexicon, perfectivity is expressed by different means by these two languages. For Hungarian, Dahl's questionnaires test out perfectivity that is associated with verbal prefixation. Therefore, what

Dahl tests as perfectivity is at least partly a lexical matter, a matter of lexicography. On the other hand, in Estonian it is the grammatical means of what is frequently referred to as accusative (or total, the morphological genitive or nominative) object case that is associated with the expression of perfectivity. However, only certain verbs—the aspectual ones—appear in perfective sentences and have accusative (total) objects. Therefore, the distinction of perfectivity is based on lexical features. Conceptually, thus, the perfective category in Hungarian and Estonian is typologically identical, but the means of expressing perfectivity are different, but in both cases dependent on the lexicon. The following section shows examples how the differences lead to aspectual mismatches in equivalence pairs of these two languages; section 5 is not meant as dictionary criticism of a specific dictionary.

5. Estonian-Hungarian and Hungarian-Estonian Examples

The aspectual nature of lexical items in the Hungarian-Estonian dictionary pair yields aspect-based mismatches, cases where the equivalence conditions are systematically violated.

1) Insertability and equivalence preserving is problematic. Given the following translation equivalence pair from the Estonian-Hungarian dictionary: *pidama* → *tart,* translating from Estonian to Hungarian, there is an aspectual mismatch. The Hungarian equivalent is not always the correct lexical insertion. Consider the following context:

Homme pean loengu → *Holnap előadast x.*
tomorrow give.1sg lecture.acc tomorrow lecture.acc x
'I'll give a lecture tomorrow – I'll be giving a lecture tomorrow.'

Here, instead of x, the item *tart* from the equivalence representation or, more precisely, the grammatically suitable form *tartok* should be inserted according to the information from the given equivalence pair. This would yield a pair where the Estonian sentence is clearly perfective, but the Hungarian equivalent is imperfective. The better, insertable lexical item that would lead to an unambiguously perfective sentence would be the prefixed version of the verb, *megtart,* which is, however, missing from the equivalence list. The translation equivalent of the context would be the following: *Homme pean loengu* → *Holnap megtartom az előadást* 'I'll give the lecture tomorrow.' The contribution of the lexicalized verbal prefix *meg-* is, thus, comparable here with that of the accusative object case in Estonian; there are also issues of definiteness (specificity) and Estonian particles that I do not address in this article. The point here is that an evident insertable translation equivalence problem emerges that calls for a solution.

2) Symmetry. The equivalent pairs in the Hungarian-Estonian dictionary are not presented as being symmetric either. The perfective forms with the verbal prefix *meg-* do not occur as equivalents for Estonian verbs that typically occur in sentences with the accusative object. For instance, while the Hungarian verb *megtesz* is translated into Estonian with an equivalent *tegema,* the reverse dictionary does not record this equivalence relation. In the bilingual complement dictionary, Estonian-Hungarian, *megtesz* is missing from the equivalents of the entry word *tegema* (Nurk & Pusztay, 1993:202). This is not an isolated, accidental case, since *megtart* is also missing from the equivalent set of the verb *pidama* (Nurk & Pusztay, 1993:140-141). The equivalent pairs in Hungarian-Estonian are systematically not represented as symmetric.

3) Balance. In addition to the violations of the conditions A and B, condition C is also violated, since two different entries are provided with an identical equivalent in the Hungarian-Estonian dictionary. The Hungarian *megtart* or *megtesz* are translated with identical equivalents that one also finds under *tart* or *tesz,* that is, *pidama* and *tegema* respectively (Pusztay & Rüütma, 1995:313, 465, 314, 477). Since these translation pairs occur without any further specification for the target user, the Estonian user has no clue to the correct aspectual reading of these verbs.

This example shows how the origin of the mismatches is the aspectual content of verbs. Intuitively, Estonian verbs contain aspectually "more possibilities" than the Hungarian verbs, but the average lexicographer bases his equivalence decisions systematically on the "minimum" of the Estonian verbs' aspectual potential. Estonian verbs that occur in perfective sentences are not felt to be perfective when a target equivalent is being searched for. Therefore, no perfective Hungarian verbs are listed as equivalents. The access to the lexically perfective verbs is systematically blocked. This is a problem for encoding target sentences, since the prefix-verb combinations are generally opaque, not predictable combinations in Hungarian. The object case, which would provide fuller equivalence coverage, does not belong to the Estonian canonical form of the verb.

5. Aspectual Distinctions as a Problem to be Solved

Apart from several representational problems of objects and their case in Finnic-like languages, there are more serious questions to be asked about the exact aspectual nature of the verbs discussed above. It is desirable to give a cross-linguistically reliable model for verb representation that takes the (prefixed or particle) verbs' aspectual nature into consideration, and to enrich the lexicographer's toolkit with a cross-linguistically valid battery of tests to establish the aspectual distinctions. To give just one example in addition to the entailment test in Section 2, it is clear that the discussed sentences containing the Estonian verb *pidama* and *tegema* yield a positive result with two aspectually opposite standard tests: one with the Estonian equivalent of the time frame adverbial test *in an hour* and the other, with the durative adverbial test *for an hour.* The Hungarian *megtart* is compatible only with the Hungarian equivalent of the time frame adverbial test, whereas the Hungarian *tart* is rather compatible with the durative adverbial test. Space considerations prevent me from discussing the exact implications of this test and those of many other tests. The point to be made here is that a suitable a test set enables one to describe verbs so that the verbs are rendered comparable with each other. Uniformly described data paves the way to automatic detection of equivalence mismatches in a lexical knowledge base. A lexical knowledge base is understood here as a lexical database where certain generalizations are stated and, therefore, systematic inferences can be performed (cf. Boguraev & Levin, 1993). Knowledge bases contain more lexical semantic information than strictly needed for printed bilingual dictionaries. This knowledge becomes relevant in the cases described in section 4, where the equivalence mismatch data shows that the intuition systematically leads the lexicographer to creating incomplete equivalence relations. The lexicographer can be warned automatically if a translation pair is created that has equivalence mismatch with regard to a defined value.

6. Conclusion

On the basis of aspectual equivalence mismatches between the verbs of Estonian and Hungarian, this article identifies a regularly occurring problem in bilingual dictionaries. The equivalence problems of these two typologically related languages may emerge in many bilingual dictionaries containing typologically less related languages.

References

Boguraev, B., B. Levin. 1993. 'Models for lexical data bases' in J. Pustejovsky (ed.), *Semantics and the lexicon.* Kluwer Academic Publishers. 325-340.

Dahl, Ö. 1984. 'Perfectivity in Slavonic and other languages' in C. de Groot, and H. Tommola. (eds.) *Aspect Bound. A voyage into the realm of Germanic, Slavonic and Finno-Ugrian aspectology.* Dordrecht: Foris. 3-22.

Heinämäki, O. 1984. 'Aspect in Finnish' in C. de Groot, and H. Tommola. (eds.) *Aspect Bound. A voyage into the realm of Germanic, Slavonic and Finno-Ugrian aspectology.* Dordrecht: Foris. 153-177.

Kiparsky, P. 1998. 'Partitive Case and Aspect' in M. Butt and W. Geuder (eds.), *The Projection of Arguments.* Stanford: CSLI Publications. 265–307.

Nurk, A., and Pusztay, J. 1993. *Észt-magyar kisszótár.* Szombathely: Savaria Press.

Pusztay, J., Rüütma, T. 1995. *Magyar-észt kisszótár.* Szombathely: Savaria Press. Szombathely.

Verkuyl, H. 1993. *A Theory of Aspectuality: The Interaction between Temporal and Atemporal Structure.* Cambridge: Cambridge University Press.