Lemma Selection in the Bilingual Italian-Swedish Dictionary. A Glance at the Past and a Look Ahead

Camilla Bardel
University of Stockholm

Abstract
In this paper I discuss the representation of lemmas of the word family *giocare* in three bilingual dictionaries and one monolingual dictionary. The aim is to come to grips with the criteria for lemma selection in bilingual Italian-Swedish dictionaries, present and future such, and to relate these to Tullio De Mauro's structuring of the Italian lexicon into categories of use. The paper also provides a short presentation of two Italian-Swedish dictionaries.

1 Introduction

The function of a bilingual dictionary is to provide its user with equivalents in the target language (*TL*) for words and expressions of the source language (*SL*). But how many words of the *SL*, and above all, which words can we expect to find in a dictionary? Which are the words that the user needs to understand? In this study, frequency of use and stylistic criteria are considered crucial for the answers to these questions.

The background of the study is a project for developing a new Italian-Swedish dictionary at the Department of French, Italian and Classical Languages, Stockholm University (Bardel in press; Nystedt 2005ms.). In this paper, I will concentrate on the issue of lemma selection, and hereby consider two existing bilingual Italian-Swedish dictionaries from that point of view as well as one Italian-English dictionary, as a comparison. The dictionaries are *Italiensk-svensk ordbok*, from 1973 (henceforth *Esselte*), *Norstedts italiensk-svenska ordbok*, from 1998 (henceforth *Norstedts*) and the concise *Oxford-Paravia Italiano-Inglese*, from 2003 (henceforth *Oxford-Paravia*). I will also compare these three dictionaries with Tullio de Mauro's – much more extensive – monolingual Italian dictionary from 2000 (henceforth *Il monovolume*). In a small pilot study, I will compare how the four dictionaries mentioned deal with words related to the word *giocare* (i.e. the verb 'play') trying to answer the following questions: How many lemmas, and which lemmas belonging to this word family (Nation, 2001:8), are included in the different dictionaries? What are the stylistic/pragmatic values of the lemmas that are included and those that are not? Are there any major differences between the older Italian-Swedish *Esselte* and the more recent Italian-Swedish *Norstedts*?

While monodirectional dictionaries are designed to serve native speakers of one of the languages involved (*TL* or *SL*), bidirectional dictionaries are bilingual dictionaries that can be used with equal success by native speakers of the *TL* or the *SL* (Svensén, 2004:22). As it is designed to serve both groups of users, its structure will be more complicated than that of a

monodirectional dictionary. The two existing Italian-Swedish dictionaries that I will consider in this study, are neither clear-cut monodirectionals, nor are they fully bidirectional. As Svensén points out, because the market for bilingual dictionaries is often restricted, many publishers try to make them appealing to both groups, although with a simplified structure. The Italian-Swedish dictionaries are examples of this: they are not entirely satisfying for either of the groups.

In our dictionary project, we want to address both native speakers of Swedish learning Italian, and native speakers of Italian learning Swedish. However, this implies a complicated and very detailed structure for the bilingual dictionary, and a way to avoid this is to offer two parallel usage possibilities, or versions, as options in an electronic version. We will then start with the Italian as L2 → Swedish as L1 version, since this is the most imminent need in the Swedish society.

2 The two Italian-Swedish dictionaries
2.1 Esselte (1973)

The Italian-Swedish dictionary published by Esselte (1973) is based on Silvia Tomba's *Italiensk-svensk ordbok* (first published 1940), but is presented in *Esselte* as mainly a new work (Italiensk-svensk ordbok, 1973: V). Tomba's work has been completed with a selection of words and phrases, based mainly on the *Garzanti* and the *Zingarelli* monolingual Italian dictionaries from the era (p. V). The preface (pp. V-VI) does not reveal the size of the lemmary, neither does it explicitly mention the mother tongue of the user aimed at. Since the overall metalanguage is Swedish, both in preface and lemmary, it is clear, however, that it serves mainly the Swedish L1 speaker.

2.2 Norstedts (1998)

Norstedts is the most extensive and modern Italian-Swedish/Swedish-Italian dictionary available today. It was first published in two separate volumes in 1994 by the editorial house *Norstedts*. In the second edition, from 1998, the Swedish-Italian part has been revised, and the two parts are published in one volume, including 111 000 words and locutions, according to the information on the book cover. The preface to the Italian-Swedish part of the dictionary (p. XX) reveals that some of the major Italian monolingual dictionaries, with editions ranging from 1987 to 1993, are included in the main sources, e.g. *Il Grande Dizionario Garzanti della lingua italiana* (1987) and *Zingarelli* (1993). According to the Swedish preface on pp. V-VI, the aim of the Italian-Swedish part is to help the user understand written and spoken Italian, whereas the aim of the Swedish-Italian part is to be a tool for communicating with Italian speaking people. In the Italian version of the preface on pp. VII-VIII, the same thing is said about the Italian-Swedish part. As for the Swedish-Italian part, however, the Italian preface says that its aim is to be a means for communication *for* Italian-speaking people. It is not quite clear, thus, which group the dictionary is really addressing. The metalanguage of the lemmary in both the Italian-Swedish and the Swedish-Italian part is Swedish, so it is reasonable to consider Swedish L1 speakers as the main user group of this dictionary.

3 The usage markers of De Mauro (2000)

In the evaluation of the lemma selection of the above-mentioned Italian-Swedish bilingual dictionaries, a comparison will be made with the monolingual Italian dictionary edited by Tullio De Mauro *Il dizionario della lingua italiana* (De Mauro, 2000). This dictionary, also called *Il Monovolume*, is a compact version of the much more extensive *GRADIT* (De Mauro, 1999). *Il Monovolume* contains approximately 130 000 lemmas consisting of one single word (*monorematiche*) and 30 000 lemmas consisting of more than one word with a unitary meaning) (*polirematiche*).

Although a much more extensive work than any of the bilingual dictionaries considered in this study, *Il Monovolume* is a very important source of information. Its major advantage for our project is related to the usage markers (*marche d'uso*), which are based on stylistic values and/or frequency of the lemma. These markings are relevant for the L2 user (students as well as professionals), as they function as a guide on word usage and context. They could also be used as criteria for the selection of words included in a new dictionary. It is a well-known fact that frequency is an important factor in L2 acquisition and use (cf. e.g. Cobb & Horst, 2004 and Vermeer, 2004, but see Nation, 2004 for a critical discussion). Lexical occurrences of the "fundamental" vocabulary, for instance, (*vocabolario fondamentale* – marked *FO*) constitute about 90% of all written and spoken texts (De Mauro, 2000: XVII). The FO category and a further 4000 words, form the so-called *Vocabolario di Base* (*Basic Vocabulary*). The other words in the *Vocabolario di Base* belong to the category *AU* (*alto uso*), i.e., another group of high frequency words, and the category of *AD* (*alta disponibilità*). They will be presented in more detail below.

The lemmas in the *Monovolume* are categorized as follows:

FO (*fondamentale*). These are 2049[1] words of high frequency.

AU (*alto uso*). This category contains 2576 words, less frequent than those belonging to FO, but still very frequent, and they constitute about 6 to 8 % of all oral and written texts.

AD (*alta disponibilità*). 1897 words highly familiar to native speakers of Italian and frequent in spoken discourse and connected to very common situations.

The rest of the lemmas contained in the *Monovolume* (and *GRADIT*), are categorized as follows:

CO (*comune*). About 40 000 one-word lemmas (*monorematiche*) and about 3500 lemmas consisting of more than one word with a unitary meaning (*polirematiche*) are marked with *CO*.

TS (*tecnico-specialistico*) – 53 602 technical terms, *LE* (*letterario*) – 7 325 words that are present in the major works of Italian literature.

RE (*regionale*) – 3 727 regional words, *DI* (*dialettale*) – 171 dialectal words, *ES* (*esotismi*) – 3 762 foreign words, e.g. anglicisms, *BU* (*basso uso*) – 19 323 low frequency lemmas, *OB* (*obsoleto*) – about 14 879 obsolete words.

---

[1] The number of words indicated for the categories of the *Vocabolario di Base* applies to the information given in the *Monovolume* (De Mauro, 2000: VIII-IX). The numbers of lemmas indicated for the following categories applies to data extrapolated from the CD-ROM version of De Mauro (2000).

Consideration of the usage markers will probably be very useful when selecting lemmas for a bilingual dictionary. It seems crucial to include the lemmas from the *Vocabolario di Base*, i.e., the first three categories (Ferreri, in press). Further lemma inclusion will depend on the size of the dictionary, but also on other factors, connected to the kind of users the dictionary will address (students, translators etc.), but for a dictionary of the same size as e.g. *Norstedts* it is reasonable to suggest including the *CO* category as well. *FO, AU, AD* and *CO* would together form around 45-50 000 lemmas.

4 Representation of the *giocare* family in the dictionaries
4.1 The Italian-Swedish dictionaries

I will now look at a very small sample of the lemma selection in the two bilingual dictionaries already mentioned, in order to see which categories are present in these. The sample is too small to be generalized, but it will still give an indication of which categories are represented in the bilingual dictionaries and of the similarities and differences between these, as far as the lemma selection is concerned.

Esselte includes 11 lemmas of the *giocare* family: *giocare, giocata, giocatore, giocattolo, giocherellare, giochevole, gioco, giocoforza, giocoliere, giocosamente, giocoso*. All of these, except *giochevole* and *giocosamente* are also found in the more recent *Norstedts*. In *Norstedts,* another lemma has been added, namely *giochetto*, which makes a total of 10 lemmas of the family. In sum, one obsolete/literary – according to De Mauro (2000) – lemma (*giochevole* – 'playful', 'amusing') has been eliminated, which seems reasonable. On the other hand, one lemma labelled *CO* by De Mauro in the *Monovolume* (*giocosamente* – 'playfully') has also been eliminated and replaced with another of the same category (*giochetto* – 'a child's play', 'a trick'). The remaining lemmas are all part of the *Vocabolario di Base*, or the *CO* category, except for *giocoforza* ('necessarily') which is a low frequency lemma marked *BU* in the *Monovolume*. As we will see in the next section there are several other words in the *Monovolume* that are not present in *Norstedts*, and that seem to be more relevant for an L2 user, than *giocoforza.*

4.2 Representation in the Monovolume

The lemmas of the *giocare* word family that we have found in the *Monovolume* are the following: *giocabile (TS), giocare (FO, CO, OB, TS), giocarsela (CO), giocarsi (CO), giocata (CO), giocato (→ giocare, giocarsi), giocatona (→ giocata, BU), giocatore (AU, BU), giocatrice (→ giocatore), giocattolaia (→ giocattolaio), giocattolaio (CO), giocattolo (AU), giocherellare (CO), giocherellato (→ giocherellare), giocherellona (→ giocherellone), giocherellone (CO), giocheria (CO), giochetto (CO), giochevole (OB, LE), giochevolmente (OB), giochicchiare (CO), giochicchiato (→ giochicchiare), gioco (FO, CO, TS, OB, LE), giocoforza (BU),* [1]*giocolare (s) (LE),* [2]*giocolare (v) (BU), giocolaro (→* [1]*giocolare), giocolato (→*[2]*giocolare), giocolatore (BU), giocoleria (BU), giocoliera (→ giocoliere), giocoliere (CO, TS), giocosamente (CO), giocosità (CO), giocoso (CO), giocoteca (BU), giocucchiare (BU), giocucchiato (→ giocucchiare).*

As is already clear from the preceding section, several of these lemmas are missing in the

Italian-Swedish bilingual dictionaries. Some of the lemmas that are present are not lemmatized as proper lemmas in *Norstedts* and *Esselte*. For instance the past participle *giocato* is not a lemma in *Norstedts*, but is only found in an example, of the infinitive form *giocare*. (In *Esselte, giocato* is not present at all). Neither is *giocarsi* a lemma in its own right in the Italian-Swedish dictionaries, although it is found within *giocare* in both *Esselte* och *Norstedts*.

If one considers the most recent Italian-Swedish dictionary, which is *Norstedts*, none of the lemmas that are missing there belong to the *Vocabolario di Base*. Many of them belong to the *CO* category, or are derived from words belonging to the *CO* category. *CO* lemmas from the *Monovolume* missing in *Norstedts* are: *giocarsela, giocattolaio, giocherellone, giocheria, giochicchiare, giocosamente, giocosità*. The only reason one can think of for inserting a word like *giocoforza* in *Norstedts*, instead of any of the *CO* words from the *Monovolume* just mentioned, is that some of the latter (but only a few) are more transparent, and therefore easier to understand for an L2 speaker than *giocoforza*. This could be said on the basis of the semantics of *gioco/giocare/giocattolo* ('game'/'to play'/'toy') and the derivational morphology of e.g. *giocattolaio, giochicchiare, giocosamente* and *giocosità*, which is quite regular, semantically. On the other hand, it is hardly possible to deduce the meaning of *giocarsela* ('take a chance', 'gamble') or *giocherellone* ('playful person/pet') in the same way. One could also argue that *giocarsela* and *giocherellone* belong to a more modern and colloquial variety of language than *giocoforza*.

4.3 The Oxford-Paravia Italian-English dictionary

The lemmas of the *giocare* family found in the concise *Oxford-Paravia* Italian-English dictionary coincide with those found in *Norstedts*, with the addition of *giocherellone*. For the rest, the lemmas are the same: the low frequent *giocoforza* is represented also in the Oxford-Paravia.

5 Conclusion and a look ahead

In this study, I have compared a small group of words, the word family of *giocare*, in two bilingual Italian-Swedish dictionaries, one Italian-English dictionary and one monolingual Italian dictionary. In doing so, I found no major differences between the three bilingual dictionaries. The number of lemmas was about the same in all three bilingual dictionaries and much lower, of course, than that of the *giocare* family in De Mauro's *Monovolume*, which is of another size. The lemmas included in the three bilingual dictionaries were not exactly the same, but the choice of included words differed only slightly. I found that most of the words represented in all the bilingual dictionaries belong to the *Vocabolario di Base* or to the *CO* (*comune*) category from De Mauro's categorization (2000). However, when comparing the material from the bilingual dictionaries to the representation of the same word family in the *Monovolume*, I found that a number of lemmas from the *CO* category were not represented in the three bilingual dictionaries, although one could expect a dictionary of the size of *Norstedts* to include all the CO lemmas if selection is based on frequency. I did not manage to see any obvious criteria, by which lemmas had been selected from the *CO* category. Rather, in some cases, other lemmas from the *CO* category would have seemed more rele-

vant to the Italian L2 user than the ones present. Further, one low frequency lemma, *gioco-forza* (labelled *BU, basso uso*) was found in all three dictionaries, a surprising fact, as it would seem logical to give *CO* words a higher priority for inclusion than *BU* words.

In a future electronic dictionary, it will be possible to include more lemmas, which would certainly be a major improvement of the present Italian-Swedish dictionary situation. First of all, the categories of the *Vocabolario di Base* must be included. These categories plus the CO category would constitute about 45-50 000 lemmas altogether, a number that does not exceed the normal size of a traditional bilingual dictionary. In my view, it is more practical to include all the CO words than to try to make a selection within this category. In this way one would avoid the question about differentiation within the CO category, which would probably be a very complicated matter. Another very important step forward would be to supply the lemmas with the usage markers of De Mauro. This would provide the L2 user with important information and help him or her in choosing the right word in different situations.

References
A. Dictionaries
De Mauro, T. (1999) *Grande dizionario italiano dell'uso*, Torino, Utet.
De Mauro, T. (2000) *Il dizionario della lingua italiana*, Torino, Paravia.
Il grande dizionario Garzanti della lingua italiana (1987), Milano, Garzanti.
Italiensk-svensk ordbok (1940) [1960, 1968,] Stockholm, Svenska bokförlaget.
Italiensk-svensk ordbok (1973) Stockholm, Esselte Studium AB.
Norstedts italienska ordbok (1998) (Second edition.) Stockholm, Norstedts.
Lo Zingarelli. Vocabolario della lingua italiana (1993) Bologna, Zanichelli.

B. Other Literature
Bardel, C. (in press), 'From research to dictionary. Open questions on the pre-dictionary process' to appear in Bardel C., Nystedt J. (eds.), *Progetto dizionario italiano-svedese. Atti del primo colloquio, Stoccolma, 10-12 febbraio 2005*.
Cobb, T. & Horst, M. (2004), 'Is there room for an academic word list in French?', in Bogaards P., Laufer B. (eds.), *Vocabulary in a second language: selection, acquisition, and testing*, Philadelphia, Benjamins.
Ferreri, S. (in press), 'Servono le marche d'uso in un dizionario bilingue? Problemi, compiti, applicazioni' to appear in Bardel & Nystedt (eds.) *Progetto dizionario italiano-svedese. Atti del primo colloquio, Stoccolma, 10-12 febbraio 2005*.
Nation, P. (2001), *Learning vocabulary in another Language*, Cambridge, Cambridge University Press.
Nation, P. (2004), 'A study of the most frequent word families in the British National Corpus', in Bogaards P., Laufer B. (eds.), *Vocabulary in a second language: selection, acquisition, and testing*, Philadelphia, Benjamins.
Nystedt, J. (2005ms.), *A new-generation Italian-Swedish Dictionary.*
Svensén, B. (2004), *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*, Stockholm, Norstedts Akademiska Förlag.
Vermeer, A. (2004), 'The relation between lexical richness and vocabulary size in Dutch L1 and L2 children', in Bogaards P., Laufer B. (eds), *Vocabulary in a second language: selection, acquisition, and testing*, Philadelphia, Benjamins, pp. 173-189.