Lexicographic Approaches to Sense Disambiguation in Monolingual Dictionaries and Equivalent Differentiation in Bilingual Dictionaries

Summary

The article discusses methods of sense disambiguation in monolingual dictionaries and equivalent differentiation in bilingual dictionaries. In current dictionaries, sense disambiguation and equivalent differentiation is presented in the form of specifiers or glosses, collocators or indications of context, (domain) labels, metalinguistic and encyclopaedic information. Each method is presented and illustrated by actual samples of dictionary articles taken from monolingual and bilingual dictionaries. The last part of the article is devoted to equivalent differentiation in bilingual decoding dictionaries. In bilingual dictionaries, equivalent differentiation is often needed to describe the lack of agreement between the source language (SL) and target language (TL). The article concludes by stating that equivalent differentiation should be written in the native language of the target audience and sense indicators in a monolingual learner's dictionary should be words that the users are most familiar with.

Key words: sense disambiguation, equivalent differentiation, monolingual dictionaries, bilingual dictionaries

Leksikografski pristopi k razločevanju pomena v enojezičnih slovarjih in razločevanju slovarskih ustreznic v dvojezičnih slovarjih

Povzetek


Ključne besede: razločevanje pomena, razločevanje slovarskih ustreznic, enojezični slovarji, dvojezični slovarji
Lexicographic Approaches to Sense Disambiguation in Monolingual Dictionaries and Equivalent Differentiation in Bilingual Dictionaries

1. Introduction

One of the problems encountered in (general) lexicography is the problem of polysemy. The question posed by lexicographers is how to tackle polysemy in a way that is most user-friendly. Is it enough to split different senses without giving any more detailed information on the meanings themselves? Does information supplied in brackets help dictionary users to find the sense they are looking for? Is it necessary to make additional information typographically visible? Should the same criteria be observed in short and in long entries? Are the same principles equally suitable for both monolingual and bilingual dictionaries? These are some of the questions that need to be answered before starting any lexicographic work.

The problem of polysemy has been addressed by numerous scholars who deal with lexicography and it is dealt with in all important lexicographic works (cf. Landau 2001; Béjoint 2000; Cowie 1999; Hartmann 2001; Jackson 2002; Svensén 2009). Ascertaining how many senses a lexeme has and in what order to arrange them are difficult decisions for a lexicographer to make, and dictionaries may differ quite markedly in their respective policies. The area of sense disambiguation (also referred to as demarcation of meaning) has an immediate impact on lexicography. If we take, for example, the context surrounding the noun hand as the part of the body at the end of the arm, we can see that it differs from the context surrounding the hand as a part of a clock or watch that points to the numbers. It seems sensible to try to devise principles to distinguish between them.

In a bilingual context, the problem of equivalent differentiation is even more complex, since we are dealing with two different languages and only in rare cases does the distribution of meaning coincide exactly. For that reason, it is especially important to describe the lack of agreement between the SL and TL (cf. Svensén 2009, 261–2). This can be done by adding certain distinctive features on the side where the meaning range is wider or the degree of stylistic or other marking is lower. Such comments have an important role in equivalent differentiation (also called equivalent discrimination) – that is, they provide a more detailed specification of usage and shades of meaning in order to guide the user towards the correct equivalent.

This article is aimed at discussing ways of sense disambiguation in monolingual dictionaries and equivalent differentiation in bilingual dictionaries – a piece of information present in all modern dictionaries and an element that helps the user to locate the information he/she is looking for.

2. Sense Disambiguation and Equivalent Differentiation in Existing Dictionaries

Modern dictionaries use several means of sense disambiguation and equivalent differentiation. A more implicit way is the definition, since the definition itself often suggests the context (e.g.
hand – a person who does physical work on a farm or in a factory; underlined by the author of the contribution). Besides this implicit way of including semantic information, more explicit ways can be found in different dictionaries. In mono- as well as bilingual dictionaries, sense indicators (also called sense discriminations by some scholars) are used. Sense indicators can also be referred to as equivalent differentiation or equivalent discrimination in bilingual dictionaries. They are realized by specifiers or glosses, collocators or indication of context, (domain) labels and metalinguistic and encyclopaedic information.

In sections 2.1, 2.2, 2.3 and 2.4, the above-mentioned techniques used for sense disambiguation and equivalent differentiation are explained in greater detail.

2.1 Specifiers or Glosses

Specifiers or glosses (cf. Atkins and Rundell 2008, 214–6; Svensén 2009, 262–3) may consist of expressions having a certain (content-)paradigmatic relationship to the headword whose meaning is to be specified. They can contain many different types of information, such as superordinates (e.g. *wolverine* below),1 synonyms, co-hyponyms, typical modifiers and paraphrases (e.g. *corridor* below).

| **wolverine** sam. ZOOL. (sesalec) rosomah, žeruh |
| **corridor** sam. 1. hodnik, koridor 2. (pas ozemlja) koridor |

Figure 1. Specifiers in a bilingual dictionary.

In monolingual learners’ dictionaries, specifiers are now used in the form of signposts, guidewords, short cuts or items in a menu (as they are variously called in different dictionaries) and are dealt with in detail in section 3. This is the reason for the absence of an example from a bilingual dictionary.

2.2 Collocators or Indications of Context

Collocators2 or indications of context (cf. Atkins and Rundell 2008, 217–8; Svensén 2009, 263–4) are an entry component and are thought up by lexicographers to help a user choose the appropriate sense of the headword or the appropriate translation equivalent. They are words that represent a lexical set, i.e. a group of words which belong to the same wordclass and which are similar in meaning. They show typical textual surroundings of a certain lemma. The grammatical relationship of collocator to headword depends on the wordclass of the lexical unit. For example, collocators of adjectives are usually nouns typically modified by the headword (e.g. *stiff* below).

---

1 The English-Slovene sample entries are taken from an ongoing project aimed at the compilation of a general English-Slovene dictionary.

2 Collocators must not be confused with collocates (= words with significant co-occurrence frequencies in corpora).
### Domain Labels

Domain labels are one of the most common methods used for sense disambiguation.

<table>
<thead>
<tr>
<th>head ... noun ...</th>
<th>head ... noun ...</th>
</tr>
</thead>
<tbody>
<tr>
<td>OF RIVER 10 [sing.] the ~ of the river the place where a river begins</td>
<td>OF RIVER 10 [sing.] the ~ of the river the place where a river begins</td>
</tr>
<tr>
<td>SYN source</td>
<td>SYN source</td>
</tr>
<tr>
<td>OF TABLE 11 [sing.] the ~ of the table the most important seat at a table: The President sat at the head of the table.</td>
<td>OF TABLE 11 [sing.] the ~ of the table the most important seat at a table: The President sat at the head of the table.</td>
</tr>
</tbody>
</table>

OALD7: 716–17

| massive ... adj 1 (of objects) large in mass; bulky, heavy, and usually solid ... 5 geology a (of igneous rocks) having no stratification, cleavage, etc; homogeneous b (of sedimentary rocks) arranged in thick poorly defined strata ... | massive ... adj 1 (of objects) large in mass; bulky, heavy, and usually solid ... 5 geology a (of igneous rocks) having no stratification, cleavage, etc; homogeneous b (of sedimentary rocks) arranged in thick poorly defined strata ... |

CED9: 1002

### Metalinguistic and Encyclopaedic Information

Metalinguistic information is provided by part-of-speech labelling which has an evident differentiating function (cf. Svensén 2009, 265).
**round** ... adjective ...
1 shaped like a circle or a ball: a round plate ◊ These glasses suit people with round faces. ◊ The fruit are small and round. ◊ Rugby isn't played with a round ball. ... 2 having a curved shape: the round green hills of Donegal ◊ round brackets (= in writing) ◊ She had a small mouth and round pink cheeks. ...

**adverb** (especially BrE) (**NAmE** usually **around**) 1 moving in a circle: Everybody joins hands and dances round. ◊ How do you make the wheels go round? ... 2 measuring or marking the edge or outside of sth: a young tree measuring only 18 inches round ◊ They've built a high fence all round to keep intruders out.

**preposition** (especially BrE) (**NAmE** usually **around**) 1 in a circle: the first woman to sail round the world ◊ The earth moves round the sun. 2 on, to or from the other side of sth: Our house is round the next bend. ◊ There she is, coming round the corner. ...

**noun** ...
STAGE IN PROCESS 1 a set of events which form part of a longer process: the next round of peace talks ◊ the final round of voting in the election
IN SPORT 2 a stage in a sports competition: the qualifying rounds of the National Championships ◊ Hewitt was knocked out of the tournament in the third round.

**verb**
1 [vn] to go around a corner of a building, a bend in the road, etc.: The boat rounded the tip of the island. ◊ We rounded the bend at high speed. 2 to make sth into a round shape; to form into a round shape: [vn] She rounded her lips and whistled. ◊ [v] His eyes rounded with horror.

**OALD7**: 1324–25

**Figure 4.** Shaded examples show part-of-speech labelling.

Apart from part-of-speech labelling, metalinguistic equivalent differentiation consists of explicit constructional information or notes. For example:
**government** ... *noun 1* [C+ sing./pl. v.] (often *the Government*) (abbr. *govt*) the group of people who are responsible for controlling a country or a state: to *lead/ form a government* ◊ the last Conservative government ◊ the government of the day ◊ Foreign governments have been consulted about this decision. ◊ She has resigned from the Government. ◊ The Government has/have been considering further tax cuts. ◊ government policies/officials/ministers ◊ a government department/agency/grant ◊ government expenditure/intervention

OALD7: 672

---

**government** *sam. 1.* vlada: government forces vladne sile 2. *samo edn.* vlada, vladanje: *be in government* vladi 3. *Jez. vezava, rekcija RABA* Edininsko obliko samostalnika *government* lahko v prvem pomenu uporabljamo z glagolom v ednini ali v množini. Glagol v ednini uporabimo, če imamo v mislih vladu kot celoto (npr. *The new government does not have popular support*.), glagol v množini pa uporabimo v primeru, ko imamo v mislih posamezne člane vlade (npr. *The government are planning further cuts in public spending*.). Ta razlika velja samo v britanski angleščini. V ameriški angleščini edninški obliki samostalnika *government* lahko sledi samo glagol v ednini (npr. *The government says it is committed to tax reform*.). Kadar z edninško obliko samostalnika *government* uporabljamo glagol v ednini, mora biti tudi zaimek, ki se na nj nanaša, v ednini (npr. *The government says it is committed to tax reform*.), kadar pa uporabljamo glagol v množini, mora biti tudi zaimek v množini (npr. *The government have made up their minds that they're going to win*.).

---

**Figure 5. Metalinguistic equivalent differentiation in the form of explicit constructional information.**

Encyclopaedic information is useful when dealing with culture-specific items. For example:

**the FTSE Index** ... the Financial Times Stock Exchange Index: an average of the prices of shares from the 100 most important companies on the London stock exchange

MED2: 605

**FT-SE 100 Index** *sam.* krajšava za *Financial Times Stock Exchange 100 Index* EKON. borzni indeks londonske borze (ki ga Financial Times izračuna na podlagi delnic 100 velikih podjetij)

**gallon** *sam.* galona (= 4.5 litra v Veliki Britaniji, 3.8 litra v ZDA): *My car does 35 miles to the gallon.* Moj avto porabi 8 litrov bencina na 100 km.

---

**Figure 6. Encyclopaedic information in culture-specific items.**

### 3. Special Feature Used for Sense Disambiguation in Monolingual Learners’ Dictionaries

A feature relatively newly introduced into monolingual learners’ dictionaries to aid users with the disambiguation of polysemous items is called signposts (LDOCE5), guidewords (CALD3), short cuts (OALD7) or items in a menu (MED2, COBUILD5). The difference between signposts, guidewords and short cuts on the one hand and menus on the other is that signposts, guidewords and short cuts are placed at the beginning of definitions within entries, whereas menus are placed at the top of entries. They all give the core meanings of highly polysemous words. None of the five dictionaries under scrutiny use the two devices conjointly. A comparison between

---

3 Encyclopaedic information can take various forms, such as encyclopaedic notes, encyclopaedic labelling, an encyclopaedic section or illustrations. Frequently, it is realized by field or subfield labels or by means of short phrases that perform the function of sense disambiguation in monolingual dictionaries or equivalent differentiation in bilingual dictionaries (in cases of polysemy, of course).
the previous editions of LDOCE and COBUILD and the current editions shows that the two
dictionaries have inverted their strategies: LDOCE has abandoned the combined use of menus
and signposts in favour of signposts alone, while COBUILD has replaced signposts by menus.

A signpost\(^4\) is a particular type of specifier which is increasingly used in monolingual learners’
dictionaries. Signposts help the users to make mental connections with the word in the context
in which they have encountered it. Signposts do not replace the full definition, but rather form
a quick menu for the user’s eye to scan. They guide the user quickly to the meaning he/she
wants and are often realized by a synonym or paraphrase of the headword (pool, OALD7, senses
1, 4 and 5) but may also offer a superordinate of the headword (pool, OALD7, sense 6) or an
indication of the domain or subject matter (pool, OALD7, sense 7). For example:

```
pool ... noun ...

FOR SWIMMING 1 [C] = SWIMMING POOL: Does the
hotel have a pool? ◊ relaxing by the pool

OF WATER 2 [C] a small area of still water, especially one
that has formed naturally: freshwater pools ◊ a rock pool (=
between rocks by the sea)

OF LIQUID/LIGHT 3 [C] ◊ (of sth) a small amount of
liquid or light lying on a surface: The body was lying in a
pool of blood. ◊ a pool of light

GROUP OF THINGS/PEOPLE 4 [C] ◊ (of sth) a supply
of things or money that is shared by a group of people and
can be used when needed: a pool of cars used by the firm’s sales
force ◊ a pool car 5 [C] ◊ (of sth) a group of people available
for work when needed: a pool of cheap labour

GAME 6 [U] a game for two people played with 16
coloured balls on a table, often in pubs and bars. Players use
cues (= long sticks) to try to hit the balls into pockets at the
edge of the table: a pool table ◊ to shoot (= play) pool

FOOTBALL 7 the pools [pl.] = FOOTBALL POOLS: He
does the pools every week. ◊ a pools winner ...
```

<table>
<thead>
<tr>
<th>pool ... noun ...</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 area of liquid</td>
</tr>
<tr>
<td>2 name of game</td>
</tr>
<tr>
<td>3 light shining on area</td>
</tr>
<tr>
<td>4 competition</td>
</tr>
<tr>
<td>5 group someone/something chosen from</td>
</tr>
</tbody>
</table>

OALD7: 1169

MED2: 1150

Figure 7. Shaded examples pinpoint signposts (OALD7) and a menu (MED2).

It should be pointed out that in many cases the context in which the user has met an unknown
word will prompt the choice of signpost. Consequently, users should usually be able to select
the right sense paragraph to read fully without having to read all the details in several other
paragraphs first.

\(^4\) A signpost is a term used generically to refer to signposts, guidewords, short cuts or items in a menu as used in different dictionaries.
4. Equivalent Differentiation in a Bilingual Decoding Dictionary

So far, various methods of sense disambiguation in mono- as well as bilingual dictionaries in general have been presented, while monolingual learners’ dictionaries have been dealt with more specifically in the previous section. At this point, I would like to focus on the purposes of the inclusion of equivalent differentiation in a bilingual dictionary for decoding. Careful indication of the meanings of the word being handled is necessary for various reasons and it is important especially in complicated and complex entries. It may make a look-up process quicker and easier in long entries but very often it is necessary because of partial equivalence, lexical gaps or culture-specific items.

Another dilemma faced by compilers of bilingual dictionaries only is which language to use for sense disambiguation. Is it appropriate to use the source language or the target language of the dictionary? Or should the native language of the user be chosen as the metalanguage of the dictionary, thus also for equivalent differentiation? These questions should be answered before starting work on a bilingual dictionary. Without a shadow of a doubt, it is the target users and their needs that are of the utmost importance. Consequently, it can be claimed that the metalanguage should always be the target users’ mother tongue.

I would now like to present some sample entries from a bilingual English-Slovene dictionary intended for decoding purposes. All these entries employ at least one method of equivalent differentiation but very often several of the types of equivalent differentiation will have to cooperate in order for the desired result to be achieved.

Equivalent differentiation is used in highly polysemous entries. The primary purpose of equivalent differentiation is to help the user quickly identify the translation equivalent that fits his/her context. Equivalent differentiation can be included in the form of indications of context:

<table>
<thead>
<tr>
<th>natural prid.</th>
<th>1. naraven</th>
<th>2. (človek) rojen</th>
<th>3. (lastnost) prороjen</th>
<th>4. (sin, hči) roden, pravi</th>
<th>5. GLAS. (ton) z razvezajem, z vračajem</th>
<th>6. ZASTAR. nezakonski</th>
</tr>
</thead>
</table>

Figure 8. Equivalent differentiation in the form of indications of context and domain labels.

In sense 2, the indication of context implies that this sense is restricted to a person, sense 3 to a characteristic, sense 4 to a son or daughter and sense 5 to a tone. Indications of context may indicate typical referents or reference domains. Apart from the indication of context, the domain labels (GLAS. = MUSIC, ZASTAR. = ARCHAIC) are used and sense 5 illustrates equivalent differentiation by cooperation of a domain label and indication of context.

The example that follows illustrates the entry for nanny goat, whose translation equivalent in Slovene is koza. The noun koza is a polysemous word in Slovene (in the Dictionary of Standard Slovene it has 5 senses) and the gloss in brackets that precedes the translational equivalent tells the user that in this particular case it refers to a female goat – a piece of information that is helpful
even though it is perhaps not absolutely necessary if we take into account that the user comes across the word in context.

| **nanny goat** sam. (samica) koza |

*Figure 9. Gloss narrowing down the specific sense of the TL polysemous word.*

There are several common types of gloss, i.e., a (near) synonym of the lemma, a short paraphrase of the particular (sub-)sense of the lemma, a hyperonym of the lemma or a typical hyponym. In the above example, the gloss used is a hyperonym of the lemma. The same holds true of sense 1 (woman, mother) of the sample entry below (old woman), whereas the gloss provided in sense 2 (cowardly man) is an example of a short paraphrase.

| **old woman** sam. POG. 1. (žena, mati) stara 2. SLABŠ. (strahopeten moški) baba |

*Figure 10. Glosses in the form of a hyperonym and short paraphrase.*

Very often, indications of context suggest possible collocations (cf. *lay*, senses 2–6, 8–10). Another factor that has to be highlighted is that one sense in a monolingual dictionary does not necessarily correspond to one sense in a bilingual dictionary. For example, *lay* may be defined as ‘to put sth down, especially on the floor, ready to be used’ (OALD7, sense 2). The examples of use that illustrate this particular sense are: to lay a carpet/cable/pipe • *The foundations of the house are being laid today.* • (figurative) They had laid the groundwork for future development. The collocates (carpet/cable/pipe, foundations, groundwork) of the English lemma produce quite different Slovene translations because of collocational requirements in Slovene. We can see that the first example is translated by means of the verb *položiti* (sense 1), whereas the second and the third ones are translated in the same way, i.e. as *postaviti* (sense 3).

| **lay gl.** položiti 2. (miza) pripraviti 3. (temelj) postaviti 4. KNJIŽ. (breme, odgovornost) naložiti 5. (zakon) predložiti 6. KNJIŽ. (strah) premagati 7. staviti 8. SL. položiti (žensko) 9. (jajce) leči, znesti, nesti 10. (ogenj) podkuriti, narediti |

*Figure 11. Indications of context suggesting possible collocations.*

It is helpful for users to include additional information in some entries (e.g., nouns denoting animals) although this is already a piece of information that can be regarded as an encyclopaedic one. But in this way the user quickly learns more about the lemma (e.g. *anchovy* is a fish (= riba) rather than a mammal, bird, snake, etc.).

| **anchovy** sam. ZOOL. (riba) inčun |

*Figure 12. Encyclopaedic information in nouns denoting animals.*

Encyclopaedic information is also necessary in culture-specific items where the concept of the SL lemma is unknown to the TL speaker. For example:
The encyclopaedic information in brackets following the translation equivalent tells the users that, firstly, there is a difference between this unit for measuring liquid in Great Britain and in the USA, and, secondly, the users learn the unit of measurement it equals.

Similarly, encyclopaedic information is provided in some other culture-specific items. The translation equivalents offered in the entries National League and Sally Army can certainly be used in context, but the encyclopaedic information helps the users who are not familiar with what the National League (i.e. one of the two leagues in the United States-based professional Major League Baseball organization) and the Sally Army (i.e. the international Christian organization that helps the poor) are to understand the meaning of these two entries.

A kind of encyclopaedic information may also be represented by different symbols (e.g. a symbol used in email addresses) shown in brackets after the translation equivalent.

As has been seen from the above examples, different ways of equivalent differentiation are a powerful force in matching equivalents across languages, and therefore including different types of equivalent differentiation are an absolute must in bilingual dictionaries.

5. Conclusion

The function of sense indicators and equivalent differentiation is to guide the user to the appropriate sense in a monolingual dictionary and to the appropriate translation equivalent in a bilingual dictionary, respectively. For this reason, they are an indispensable part of a dictionary entry, especially a dictionary entry consisting of several senses or providing several translation equivalents, since they enable the user to find the right sense or the right translation equivalent more quickly. The lexicographers should strive to provide short but informative sense indicators. If sense indicators are used (e.g. in a monolingual learner’s dictionary), they should be those words with which users are probably most familiar. If they are used in a bilingual dictionary to disambiguate different translation equivalents, they should be written in the native language of
the target audience, since the users will immediately know which type of equivalent differentiation guides them to the correct translation equivalent of the headword.

Bibliography

A. Dictionaries


B. Other literature


