Creating a Multilingual Data Collection for Bilingual Lexicography from Parallel Monolingual Lexicons

Abstract

In the DELIS project\(^{1}\), a set of parallel monolingual lexicon fragments have been designed. They are parallel in two ways: (1) they cover the same fragment (the most general verbs of sensory perception and of speech act), and (2) they are based on the same theoretical approaches and on comparable classifications and descriptive devices.

It is claimed in this paper that such parallel fragments, formalized and represented in a modular and access-neutral way, constitute a lexical data collection which can serve as a pre-dictionary fact base from where bilingual dictionaries can be derived. We discuss examples of the procedures by which raw material for bilingual dictionaries can be derived from the fact base (semi-)automatically.

1. Metalexicographic introduction: monolingual and bilingual dictionaries – the role of a pre-dictionary fact base

In metalexicography, there has been some discussion about directional as opposed to non-directional bilingual dictionaries. Directional dictionaries – as advocated and illustrated by [Kromann 1989] and [Kromann/Riiber/Rosbach 1989] – aim at efficiency of presentation, taking the users’ perspective and the users’ mother tongue as a starting point. For an “active” translation dictionary of the directional type (from mother tongue to the users’ “foreign” language), the main objective is to make those cases clear where the target language differs considerably from the source language. If the target language lexical items display the same properties as the source language items, the lexicographer can leave their description (partly) underspecified, to save space in the article; it is not felt necessary, for example, to describe reading distinctions which are parallel in both languages.

The non-directional dictionary\(^{2}\) aims at explicitness more than at economy of space: ideally, the relevant distinctions of the source language are made explicit in any case: even if they happen to exist in a parallel fashion in the target language\(^{3}\). The non-directional approach seems to more readily support the “reuse” of lexical descriptions: the
experiments carried out on the Van Dale dictionaries by [Al 1988], [Heid 1990] and [Martin/van der Vliet 1992] demonstrated that non-directional lexical descriptions can quite easily be reused.

The difference between directional and non-directional dictionaries can thus be paraphrased (if we allow ourselves some simplification) as a trade-off between efficiency and explicitness. It thus has to do, mostly, with the presentational side of lexicographic work, not so much with its descriptive side, although both are to some extent influenced by the choice between directionality and non-directionality.

It is possible to create a bilingual data collection from where both a directional and a non-directional dictionary can be derived. A similar observation has been made by [Martin/Al 1988]: they claim that there should be a distinction between a collection of lexicographic data (called a “data base” by them) and the actual dictionaries derived from there (called “front-end dictionaries”). According to [Martin/Al 1988], the front-end dictionaries have to be user-oriented, whereas the data base must not be user-oriented.

In the DELIS project, such a collection of parallel lexical descriptions has been produced. The language-specific descriptive work is exclusively done on a monolingual basis, by native speaker lexicographers. When the monolingual fragments were available, a contrastive comparison has been performed, with the goal of semi-automatically identifying and listing semantically motivated correspondences between items form the different languages.

In the remainder of this paper, we describe the monolingual dictionaries (section 2) and their joint use as a basis for bilingual and multilingual dictionaries (section 3).

2. The monolingual dictionary fragments underlying the multilingual data collection

In DELIS, the description of lexical items is based on Frame Semantics and on a syntactic description inspired by (but not formalized in terms of) Head Driven Phrase Structure Grammar (HPSG, cf. [Pollard/Sag 1994]).
2.1 The fragments

2.1.1. Coverage

The work in DELIS deals with two "lexical semantic classes", perception and speech act verbs. The fragments have been covered evenly in several European languages. Table 2 in the appendix summarizes the perception fragment, from where our examples are taken.

2.1.2. Parallel fragments as a prerequisite for a multilingual data-collection

The construction of both monolingual and bilingual lexical specifications heavily relies on the notion of parallel lexicon fragments, which are a precondition for the contrastive work.

DELIS dictionaries are parallel in two ways:

- the lexical items described are potential equivalents (i.e. could be used in sentence pairs which are translations of each other);
- the lexical specifications of all languages analyzed follow a common descriptive approach with common inventories of descriptive categories, wherever possible; language-specific variation is kept track of, e.g. at the levels of morphosyntax and of subcategorization classes.

2.2. Descriptive Devices

2.2.1 Frame Semantics as a framework for lexical semantic description

The main descriptive devices of Frame Semantics are frames and frame elements; Fillmore has given the following definition of the main objectives of Frame semantics ([Heid (Coord.) 1995]:22):

One of the basic tasks of frame semantics is the schematic description of the situation types associated with the use of particular predicating words (here we concentrate on verbs) by discovering and labeling elements of such situations (the frame elements) in so far as these can be reflected in the linguistic structures that are built around the word being analyzed.
The semantic frame associated with a particular verb has something in common with the notion subcategorization frame except that it includes the subject, and it has something in common with such notions as theta grid and argument structure, with the important exception that the list of frame elements includes not only what gets represented as the arguments (narrowly conceived) of a verb, but also the frame-specific adjuncts. The working methodology underlying the DELIS dictionaries is inspired by onomasiology: frames are characteristic of lexical classes; our implementation of frames as types allows for an easy construction of frame hierarchies.

The lexical encoding of verbs in DELIS makes use of Frame Element Groups (FEGs): the Frame Element Group combines the description of the participants (in terms of “roles” and possibly sorts) with a syntactic description (in terms of grammatical functions and syntactic categories). Figure 1 schematically depicts a lexical entry template for a verb with two frame elements.

To describe perception verbs, the following roles have been used; we give an intuitive definition here, although the roles are defined more in detail through tests, and through their interrelationships with syntactic phenomena (which can be observed text corpora):

- The **experiencer** (exp): the individual who perceives something (the perceiver).
- The **percept** (pct, p-): the entity, event or phenomenon perceived.

Roles can have subtypes, depending on features or on sortal restrictions. Subtypes of **experiencer** according to the feature [INTENTION] have been introduced, whereas **percept** has subtypes defined by sortal restrictions. The subtyping of the **experiencer** role corresponds to a subdivision of the verb class of sensory perception into the subclasses of (1) ‘pure’ **perception** and (2) **attention**, depending on the intentional or non-intentional participation of the experiencer in the perception process: verbs of attention (e.g. EN look at, watch, listen to, etc.) imply that the experiencer acts such that the perception can take place. Verbs of ‘pure’ perception, however (e.g. EN see, hear), exclude this implication: the event happens without any action on the side of the experiencer.
2.2.2. Role constellations – semantic classes

Situation types (frames) are characterized by typical constellations of frame elements. The field of perception verbs falls into a number of subfields characterized each by a given frame element constellation, expressed in terms of roles. The following are a few examples:

- "attention verbs": <experiencer-intentional percept>: FR écouter, regarder; EN listen [to], watch; IT ascoltare, guardare;
- "(non-intentional, "pure") perception verbs": <experiencer-nonintentional percept>: FR entendre, voir; EN hear, see; IT sentire, vedere;
- "emission+perception verbs": <percept (judgement-evaluation)>: FR embaumer; EN smell; DE duften.

For the perception domain, the major role constellations are summarized and exemplified with EN data in table 29.
2.2.3. Syntactic description

The syntactic description makes use of a small inventory of grammatical functions: as in Head-Driven Phrase Structure Grammar ([Pollard/Sag 1994]), the distinction between subject, complement and adjunct is made\(^{10}\). In addition, a traditional classification of syntactic categories (noun phrase, adjective phrase, subclause, infinitival) has been adopted; for the different languages, a common basic inventory of grammatical functions and of syntactic categories has been devised which is extended to cater for language-specific phenomena.

<table>
<thead>
<tr>
<th>Subclass</th>
<th>Role Constellation</th>
<th>Examples (EN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>perception</td>
<td>([_ROLE: \text{exp-nonint}]\ [ROLE: \text{pct-act}])</td>
<td>John saw the light. [ ROLE: \text{exp-nonint}] [ ROLE: \text{pct-act}] &gt; John saw the car.</td>
</tr>
<tr>
<td>attention</td>
<td>([ROLE: \text{exp-int}]\ [ROLE: \text{pct-act}])</td>
<td>John watched the car. [ ROLE: \text{exp-int}] [ ROLE: \text{pct-act}] &gt; John looked at the book.</td>
</tr>
<tr>
<td>attention-target</td>
<td>([ROLE: \text{exp-int}]\ [ROLE: \text{pct-tgt}])</td>
<td>John looked for a pen. [ ROLE: \text{exp-int}] [ ROLE: \text{pct-tgt}] &gt;</td>
</tr>
<tr>
<td>judging-ver.</td>
<td>([ROLE: \text{pct-act}]\ [ROLE: \text{jud-ver}])</td>
<td>The juice tastes sweet. [ ROLE: \text{pct-act}] [ ROLE: \text{jud-ver}] &gt;</td>
</tr>
<tr>
<td>judging-eval.</td>
<td>([ROLE: \text{pct-act}]\ [ROLE: \text{jud-eval}])</td>
<td>The juice tastes awful. [ ROLE: \text{pct-act}] [ ROLE: \text{jud-eval}] &gt;</td>
</tr>
<tr>
<td>judging-inf.</td>
<td>([ROLE: \text{pct-act}]\ [ROLE: \text{jud-inf}])</td>
<td>She looks tired. [ ROLE: \text{pct-act}] [ ROLE: \text{jud-inf}] &gt;</td>
</tr>
</tbody>
</table>

Figure 2: Role constellations defining subclasses of the class of **perception** verbs

2.3. Dictionary Architecture

The lexical specifications have been formally represented by means of a constraint-based computational formalism (Typed Feature Structures, TFS; cf. [Emele 94]); this formal modeling enforces consistency: the well-formedness of individual descriptions can be checked automatically.

The following principles have been applied in the construction of the lexicon fragments:

1. Modularity:

   • the individual monolingual lexicons are modules which can be combined to form a multilingual lexicon;
   • each monolingual specification is broken up into modules which implement one level of linguistic description (here: lexical semantics, functional syntax and syntactic categories (phrase level), see
the interaction between the levels is expressed by means of relational statements which implement "linking rules"

• for each level-specific module, an inventory of descriptive devices is defined (here: a role inventory, an inventory of grammatical functions and an inventory of phrase types);

2. Classificatory approach:

Each level of the specification is organised hierarchically to encode classifications of the relevant elements

<table>
<thead>
<tr>
<th>Construct →</th>
<th>Descriptive Devices</th>
<th>Constellations (Classes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>lexical semantics</td>
<td>ROLES</td>
<td>ROLE CONSTELLATIONS</td>
</tr>
<tr>
<td>functional syntax</td>
<td>GRAMM. FUNCTIONS</td>
<td>TOPMOST SYNTACTIC CLASSES</td>
</tr>
<tr>
<td>categorial syntax</td>
<td>SYNTACTIC CATEGORIES, PHRASE TYPES</td>
<td>SPECIFIC SYNTACTIC CLASSES</td>
</tr>
</tbody>
</table>

Table 1: Summary of components and classes

3. Access-neutrality:

• None of the levels of linguistic description is dependent on or prioritary with respect to any other;
• although the individual levels are hierarchically structured, we do not see any advantage in a hierarchy of individual readings of single verbs, and thus avoid to such a structure.

Figure 3, below, summarizes the architecture: Each descriptive level is a separate, usually hierarchical component of the lexical specifications; single "readings" (symbolized by a black circle, in figure 3) inherit from the relevant classes of each component.

Figure 3: Dictionary architecture: separate hierarchies contributing to the definition of a reading
2.4 Examples

The lexical specifications define the types of information which together determine the readings contained in the lexical fragments. In the following, we give a few examples of the internal format of the entries. The examples 1 and 2 serve to encode the readings of the verb EN *to notice* which are present in the sentences 3 and 4 respectively.

(1) notice-perception-thing
    [LEMMA:“notice”],
    FEG: <fe
        [FE: exper-n,
        GF: subj,
        PT: np]  
    fe
        [FE: p-actual-ent,
        GF: comp,
        PT: np]>,
    EXPL: “They noticed the marine sergeant.”,
    EVENT: vis-mod[MODALITY:vis]].

(2) notice-that
    [LEMMA:“notice”],
    FEG: <fe
        [FE: exper-n,
        GF: subj,
        PT: np]  
    fe
        [FE: p-actual-prp,
        GF: comp,
        PT: that-cl[COMPLT:that-compl]]>,
    EXPL: “I noticed that two of them were pairing off.”,
    EVENT: vis-mod[MODALITY:vis]].

(3) They noticed the marine sergeant.

(4) I noticed that two of them were pairing off.

To ease the lexicographer’s work with the contents of the lexicons, the internal format can automatically be translated into a tabular format which is easier to handle; this includes the metalanguage: role names can be paraphrased in the user language; the output has a fixed order of roles and indicates the syntactic category of each role (unless the role is
realized as an *np*). Optionally, it contains examples; the English DELIS examples are mostly taken from the BNC. Figure 4 contains the main readings of EN *to notice* in this format.

\[
\begin{align*}
\text{[PERCEIVER non-intentionally]notice} & : \text{[actual entity PERCEIVED]} \\
\text{[PERCEIVER non-intentionally]notice} & : \text{that-clause [actual situation PERCEIVED]} \\
\text{[PERCEIVER non-intentionally]notice} & : \text{sent [actual situation PERCEIVED]} \\
\text{[PERCEIVER non-intentionally]notice} & : \text{wh-clause [actual situation PERCEIVED]} \\
\text{[PERCEIVER non-intentionally]notice} & : \text{v-ing: [actual event PERCEIVED]} \\
\end{align*}
\]

Figure 4: The readings of EN *to notice* in a lexicographer-oriented table-like format

3. A multilingual data collection: combining monolingual lexical specifications

3.1. Principles

For the purpose of constructing the DELIS multilingual data collection, we assume that the *frames* described by *Frame Semantics* – and with them their formal representations as DELIS Frame Element Groups (FEGs) – can be generalized across the languages compared: this is in line with the claim of *Frame Semantics* that the frames cover situation types (partly independent from a given language).

Readings of verbs in DELIS are defined by means of the interrelationship between role constellations and lexical and syntactic means of a given language to realize these role constellations; consequently, when searching a dictionary, or when comparing items from two languages, one can query the data collection with either one, taking a semasiological or an onomasiological perspective, respectively. When contrastively using the dictionaries, we take an onomasiological view:

- For a monolingual dictionary, we assume that one frame element constellation defines one "reading" of an item in question; a listing of all possible frame element constellations of a given lemma gives a semasiological dictionary entry.

- For contrastive work, we onomasiologically compare readings from different languages which share a frame element group. These are equivalence candidates; similarly, intralingual synonym candidates have the same frame element constellations.
We illustrate this (both contrastively and monolingually) with a schematic entry for DE *duften* and the FR verbs *embaumer* and *fleurer* taken from [Schwenger 95] (figure 5): in the middle, we indicate the frame elements horizontally (<percept-actual (judgement-evaluative) >); on the left hand side, the German item *duften* is indicated (as in *Das Parfüm duftet (angenehm)*), on the right hand side, the (quasi-synonymous) French items *fleurer* and *embaumer* are listed.

<table>
<thead>
<tr>
<th>Lemma</th>
<th>FEG</th>
<th>Lemma</th>
<th>Lemma</th>
</tr>
</thead>
<tbody>
<tr>
<td>duften</td>
<td>P-act</td>
<td>fleurer</td>
<td>Subj</td>
</tr>
<tr>
<td>GF:</td>
<td>Subj</td>
<td></td>
<td>NP</td>
</tr>
<tr>
<td>PT:</td>
<td>NP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(gut)</td>
<td>(J-eval)</td>
<td>(bon)</td>
<td>(bon)</td>
</tr>
</tbody>
</table>

Figure 5: Comparing French and German verbs: the case of DE *duften*, FR *fleurer*, *embaumer* (following [Schwenger 95])

### 3.2 Equivalence conditions

The contrastive comparison can be automatically performed by the same means as monolingual queries to the lexicon: the TFS system allows to retrieve lexical objects that fulfil a set of criteria specified in the query. Queries involve conditions from one or from two languages.

The following are the minimal conditions for equivalence between items of two languages:

(a) The lexical items from the two languages have identical role constellations.

(b) If an event classification is used, the lexical items belong to the same event subtype.

(c) If sortal restrictions on frame elements are encoded in the monolingual lexicons, the sortal properties of the lexical items compared must be identical or in a subsumption relation (to identify partial equivalence).

In addition, in a large-scale application, lists of equivalent candidate lemmas from the two (or more) languages compared should be available, as can be found in any bilingual dictionary. This set of equivalence...
conditions abstracts away from language-specific syntactic and morphosyntactic properties; it follows the onomasiological orientation of the dictionaries and thus provides a means to also relate items which have different syntactic patterns.

3.3 Using the multilingual data collection for lexicographic work

The internal encoding of the lexical specifications is not very handy for lexicographic work; thus, a reformatter is used to transform the internal representation into a notation which is easier to manipulate (see above, section 2.4). In the examples below, we use Italian metalanguage with Italian data, and Dutch metalanguage with Dutch data.

Assume a lexicographer is interested in comparing the Italian and Dutch verbs of visual perception. He searches for Dutch equivalents of those Italian verbs which can have a subcategorized *wh*-clause, i.e. IT *guardare* ("[to] watch") and *vedere* ("[to] see"). Both verbs are of the attention type, in this reading, and thus quasi-synonyms; they both are translated by NL *kijken* (examples 5 and 6):

(5) [**ESPERIENTE intenzionale**] guardare se [**SCOPO della percezione**]
    [example: "non restera'che guardare se il prezzo del passaggio sara'
    giusto."]

=> [**actieve WAARNEMER**] kijken indir. vraag: [intentioneel
    WAARNEMINGSOBJEKT]

    [example: "Hij kijkt of er post is."]

(6) [**ESPERIENTE intenzionale**] vedere se [**SCOPO della percezione**]
    [example: "per vedere se laggiu'ci fosse ancora qualche segno."]

=> [**actieve WAARNEMER**] kijken indir. vraag: [intentioneel
    WAARNEMINGSOBJEKT]

    [example: "Hij kijkt of er post is."]

When we look for Italian equivalents of some of the readings of NL *luisteren* ("[to] listen"), the differences in the second frame element (and consequently in its syntactic behaviour) have an impact on the equivalent choice: 7 and 8 lead to different Italian verbs, *sentire* and *ascoltare*.

(7) [**actieve WAARNEMER**] luisteren indir. vraag: [intentioneel
    WAARNEMINGSOBJEKT]

    [example: "Hij luistert wie er zo'n lawaai maakt."]

=> [**ESPERIENTE intenzionale**] sentire se [**SCOPO della percezione**]
    [example: "va tu dal babbo e senti se ha bisogno di qualcosa."]
To get a general picture of a given lexical semantic field across languages, the lexicographer can automatically compare all entries of a given subset of the DELIS lexicon, for two languages. The system will provide four lists; to illustrate these, we continue to use the NL/IT example:

- List of source language items and their verbal equivalents in the target language (L1: source language, L2: target language; e.g. all IT verbs with verbal equivalents in NL);
- List of source language items and their verbal equivalents, with source and target language switched around (e.g. all NL verbs with verbal equivalents in IT);
- List of items from L1 without equivalents in L2;
- List of items from L2 without equivalents in L1.

We display in figure 6 those NL readings from the data collection, for which no IT equivalent has been found. It contains many judging-readings: Italian does not have ways to express these verbally (an exception being *Questo strumento suona strano*).

### 4. Evaluation and Conclusions

The usefulness of this device, for a lexicographer, is in getting an immediate overview of the correspondences between verb readings of two (or more) languages, and, in parallel, lists of verb readings for which no verbal equivalents are found in the language compared. These lists are produced, irrespective of syntactic or other divergences (term used by [Dorr 1993], except “categorial divergences”); the syntactic description of source an target language is made available in the synopses (cf. figure 6, above), although not taken as a criterion in the contrastive comparison.
The system is still restricted: it only deals with verbal equivalents from the frame of perception; there are however no principled problems for the encoding of nouns or adjectives. It is clear, however, that a frame-based lexicon would only deal with these word classes, and that much effort would have to go into the preparation of a sufficiently large frame lexicon (see the discussion in the paper by Atkins, this volume).

Even though only a limited vocabulary could be dealt with by means of a frame-based analysis, we see a considerable advantage in the fact that the parallel fragments can easily be related and compared. Thereby a part of the contrastive work underlying the creation of bilingual dictionaries is carried out automatically, and the lexicographers can free more time for dealing with mismatches and other problematic cases.
Appendix 1: Perception Verbs analysed in DELIS

<table>
<thead>
<tr>
<th>Language → Type ↓</th>
<th>EN</th>
<th>NL</th>
<th>FR</th>
<th>IT</th>
<th>DK</th>
</tr>
</thead>
<tbody>
<tr>
<td>general perception</td>
<td>notice</td>
<td></td>
<td>percevoir</td>
<td>percepire</td>
<td>bemærke</td>
</tr>
<tr>
<td>auditory-perc</td>
<td>hear</td>
<td>horen</td>
<td>entendre</td>
<td>udire</td>
<td>høre</td>
</tr>
<tr>
<td>auditory-att</td>
<td>listen</td>
<td>luisteren</td>
<td>écouter</td>
<td>sentire</td>
<td>lytte</td>
</tr>
<tr>
<td>auditory-jud</td>
<td>sound</td>
<td>klinken</td>
<td>-</td>
<td>ascoltare</td>
<td>lyde</td>
</tr>
<tr>
<td>visual-perc</td>
<td>see</td>
<td>zien</td>
<td>voir</td>
<td>vedere</td>
<td>se</td>
</tr>
<tr>
<td>visual-att</td>
<td>look</td>
<td>kijken</td>
<td>observer</td>
<td>guardarre</td>
<td>se på</td>
</tr>
<tr>
<td>visual-jud</td>
<td>watch</td>
<td>kijken</td>
<td>-</td>
<td>sembrare</td>
<td>kigge</td>
</tr>
<tr>
<td>olfactory-perc</td>
<td>smell</td>
<td>ruiken</td>
<td>respirer</td>
<td>sentire</td>
<td>se</td>
</tr>
<tr>
<td>olfactory-att</td>
<td>smell</td>
<td>ruiken</td>
<td>sentir</td>
<td>annusare</td>
<td>se på</td>
</tr>
<tr>
<td>olfactory-jud</td>
<td>smell</td>
<td>ruiken</td>
<td>respirer</td>
<td>avere</td>
<td>kigge</td>
</tr>
<tr>
<td>gustative-perc</td>
<td>taste</td>
<td>proeven</td>
<td>gouter</td>
<td>sentire</td>
<td>luge</td>
</tr>
<tr>
<td>gustative-att</td>
<td>taste</td>
<td>proeven</td>
<td>-</td>
<td>il sapore</td>
<td>luge</td>
</tr>
<tr>
<td>gustative-jud</td>
<td>taste</td>
<td>smaken</td>
<td>-</td>
<td>gustare,</td>
<td>-</td>
</tr>
<tr>
<td>tactile-perc</td>
<td>feel</td>
<td>foelen</td>
<td>sentir</td>
<td>assaporare</td>
<td>-</td>
</tr>
<tr>
<td>tactile-att</td>
<td>feel</td>
<td></td>
<td>toucher</td>
<td>assaggire</td>
<td>-</td>
</tr>
<tr>
<td>tactile-jud</td>
<td>feel</td>
<td></td>
<td>tâte</td>
<td>sapere di</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2: Main perception verbs of EN, FR, NL and IT analysed in DELIS

Footnotes

1 DELIS stands for "Descriptive Lexical Specifications and tools for corpus-based lexicon building". DELIS (February 1993 through December 1995) is a shared-cost project partly funded by the DG XIII E of the Commission of the European Community, Luxembourg, under its LRE programme (Linguistic Research and Engineering, project no. 61.034). The project brings together expertise from system builders (Sonovision ITEP Technologies (Paris, first phase of the project), Lingsoft (Helsinki)), from (computational) linguists (Universities of Amsterdam (VUA), Clermont-Ferrand, Copenhagen, Pisa, and Stuttgart), Linguacubun Ltd. (London) and from lexicographers (B.T.S. Atkins, Lewes, UK) and dictionary publishing (Den Danske Ordbog (Copenhagen), Oxford University Press and Van Dale Lexicografie (Utrecht)).
2 As realized to some extent, in the Van Dale series of bilingual dictionaries with Dutch as a source language, cf. [Van Sterkenburg/Martin/A11982]. See also the discussion in [Hausmann 1989].

3 This can easily be verified in articles on internationalisms, such as s.v organisation: a directional dictionary will have one equivalent, a non-directional one will distinguish several “readings” (e.g. “the action of organizing”, “the fact/way of being organized; the structure”, “a group”), albeit with identical translation equivalents.

4 In this case to combine two bilingual dictionaries in order to produce raw material for a third one (e.g. NL/EN and NL/FR, in order to produce an EN - FR data collection).

5 These concepts are close to Quémada’s distinction between “travail lexicographique” and user-oriented presentational work (“dictionnairique”); the “data base” suggested by [Martin/Al 1988] is close to Quémada’s “base de données pré-dictionnairique” (we will use the term “pre-dictionary fact base”, in the following). It is also in line with the kind of data collection suggested by Atkins (this volume) as a basis for new types of bilingual and multilingual dictionaries.

6 The author would like to thank all DELIS colleagues who have contributed to the linguistic work in the project: the responsibles are Beryl T. Sue Atkins, Anna Braasch, Gabriel Bès, Karine Baschung, Laura Bloksma, Nicoletta Calzolari, Willy Martin, Nicholas Ostler and Maurice Vliegen; most precious advice was given by Charles Fillmore, in several meetings and many mailings; Sonja Schwenger has worked out a detailed French/German case study, from where some of the examples discussed below have been taken; Katja Krüger has contributed many details of this paper, in discussions of earlier versions; all errors and misconceptions in this paper are of the author’s responsibility.

7 See e.g. [Fillmore 1993a] and [Fillmore 1993b], [Atkins/Fillmore94]. We do not want to motivate the choice of this approach, here, in detail; one of the reasons underlying this choice is the fact that it supports explicit descriptions of the interaction between syntax and semantics.

8 We do not want to justify here the use of Frame Semantics. It has, a. o., been chosen because of it contains explicit statements on the interaction between syntax and semantics, and because it can easily be used in onomasiologically oriented work.

9 Abbreviations used in this table:
   - exp - experiencer; int - intentional; nonint - non-intentional;
   - pct - percept; pct-act - percept-actual; pct-tgt - percept-target;
   - jud - judgement; jud-ver - veridical judgement; jud-eval -evaluative judgement; jud-inf - inference judgement.

10 In addition, special complement subtypes for predicatives (predicative: This smells strange and for controlled constructions (xcomp: Ha detto a Maria di venire a mezzogiorno) have been introduced.

11 This has been simulated in our experiments by restricting the analysis to two distinct lexical semantic fields.
References


