*B.T.S. Atkins, Oxford University Press*

# Bilingual Dictionaries Past, Present and Future

**Abstract**

The past is print dictionaries; the present is print dictionaries with some electronic versions of the same text; the future must be print dictionaries and truly electronic dictionaries, compiled afresh for the new medium, enriched with new types of information the better to meet the needs of the multifarious users. The paper sets out the various aspects of the bilingual dictionary which must be taken into account if the new dictionaries are to be different from (and better than) the old. A design for a new electronic bilingual dictionary is sketched out, applying a frame semantics approach to corpus analysis. A demonstration of the prototype multilingual hypertext Dictionary of the Future will be given.

## 1. Looking at today's dictionaries

Change is not something that people tend to associate with dictionaries. Changing these highly labour-intensive products is not to be undertaken lightly. (Here I am talking about large-scale, radical change, not simply updatings and corrections.) The heavy cost of dictionary production, and the penalty to be paid for errors of judgement, have made it almost impossible for any radically new dictionary to come into being. Of course, our dictionaries of the present do look a little different from their predecessors, and do behave a little better (it is becoming rarer now to find dictionaries with hermetically sealed nuggets of information coded up to defy interpretation by all but the dogged few); they may even come to you on a CDROM rather than in book form, but underneath these superficial modernizations lurks the same old dictionary. Some of the more innovative may introduce a few new types of information (corpus frequencies are the flavour of the month), but when it comes to setting out the meanings of words, giving them definitions or equivalents in another language, including examples, idioms, pronunciations, usage notes, cross-references and the score or so of other kinds of information, tradition rules supreme. Most dictionaries are sublimely unaffected by the highly relevant work currently being done by linguists, especially in lexical semantics. The dictionary of the present is at heart little different from the dictionary of the past. Will the dictionary of the future simply

blip its little electronic way off into the sunset dazzling its readers with the speed which it dishes up the same old facts on a technicolor screen? It is up to us to take up the real challenge of the computer age, by asking not how the computer can help us to produce old-style dictionaries better, but how it can help us to create something new: to look at the needs of dictionary users of every language, and every walk of life, users as diverse as people themselves, and give them the kind of information they need for whatever they are using the dictionary for, and not simply the popular selection of facts that will pack semi-legibly inside book covers. I respect and admire the achievements of our great predecessors. But if they were here today, I put it to you that they would not be simply reproducing the achievements of their elders, or revising the great works of the past: they would be rooting for a new kind of dictionary, one in which the computer plays its rightful, creative role.

Our particular present is a good time for taking stock: we have behind us a long tradition of dictionary-making, a rich heritage of reference works to study and analyse; we the lexicographers are ourselves dictionary users and know the frustrations; in electronic corpora, now fairly freely available, we have a wealth of lexical evidence undreamt of in the past; we have friends and colleagues in academia whose work we can learn from, and use in our own; new research (much of it by EURALEX members) is telling us about the way in which people use dictionaries, and what they use them for; and now at last we are liberated from the straitjacket of the printed page and alphabetical order. If we are to exploit these propitious circumstances, if we are to create a new kind of dictionary, there are a few questions to be answered: first, questions about what our current dictionaries are, and why they are like that, and if they can be improved; then, questions about the new dictionary, who it is for, what they will want from it, and how we can provide that. In this paper I will be looking particularly at dictionaries for bilingual use, but (for reasons which I hope will become clear) I do not want to limit the discussion to "bilingual dictionaries" as such.

### 1.1 The organization of current bilingual dictionaries

A systematic approach to the study of what a bilingual dictionary does and how it does it must take account of the following aspects of the entry[1]:

- type of data (e.g. "headword", "indicator", "example");
- type of information it carries (what it is telling the user);

- function of that information (what the user can use it for);
- mode of expression (how it is expressed);
- type of user: Source Language (SL) speaker or Target Language (TL) speaker;
- purpose of use (encoding into a foreign language, or decoding into one's own language).

It is easy to confuse **type of data** with **type of information**, and care must be taken to distinguish these two concepts. Moreover, they must both be differentiated from the **function** of the piece of information, that is, what the user can do with it. The material in Tables 1 and 2, together with the following example, will clarify the distinctions.

A francophone wishes to know how to translate *une couche d'argile* into English and looks up *couche* in a French-English dictionary. Figure 1 shows an extract from the *Oxford-Hachette French-English Dictionary* (1994) entry from which the example phrases and other items have been omitted:

**I** *nf* **1** (de vernis, peinture, d'apprêt) coat; (d'aliments, de poussière, neige) layer;
**2** (strate) stratum, layer.
**3** Sociol sector.
**4** (pour bébés) nappy, diaper.

Figure 1: Abridged entry for couche

The underlined segments (our underlining) of the entry do not all constitute the same **data types**, nor do they carry the same **types of information**, but they all have the same **function**, that of helping the francophone (SL) user to select the correct equivalent of the headword for the context it is to be used in (*couche d'argile* is *layer of clay*). The **mode of expression** of the first two items and the fourth (*(de vernis ...)*, *(strate)* and *(pour bébés)*) is the SL, while that of the third item (*Sociol*) is a code, in this case common to both SL (*sociologie*) and TL (*sociology*). The situation is summarized[2] in Table 1:

| Item | Data Type | Information Type | Function |
|---|---|---|---|
| (*de vernis peinture d'apprêt*) | complementing sense indicator | SL collocates of *couche* | pinpoints relevant sense of *couche* |
| (*strate*) | substituting sense indicator | synonym of *couche* | pinpoints relevant sense of *couche* |
| Sociol | diatechnical label | semantic domain | pinpoints relevant sense of *couche* |
| (*pour bébés*) | complementing sense indicator | real-world fact | pinpoints relevant sense of *couche* |

Table 1. Types of data, information and functions

Table 2 gives an overview of the organization of the traditional bilingual dictionary entry[3] (see the next pages). The planning and design of future bilingual dictionaries must take account of all of these factors.

### 1.2 Evaluation

The information in Table 2 allows us to evaluate an imaginary best example of our current bilingual dictionaries[4]. If we are to design the dictionary of tomorrow, we need to be able to build on the good and improve the less good aspects of today's dictionaries. Looking at the various aspects of bilingual dictionaries set out in Table 2, we must consider what is good and must be retained, and what is less good, and must be improved.

#### 1.2.1 Strengths

In the best of today's bilingual dictionaries, as Table 2 shows, there are many things to praise. I shall list these briefly:

**(a) Wealth of information**

- Semantics: lexical items[5] are carefully analysed and explained, and their various TL equivalents are set out clearly and helpfully.
- Grammar: there is a commitment to include enough information (albeit often couched in opaque codes) to allow the foreign language expressions to be used correctly.

Table 2: The organization of a bilingual dictionary entry

| | Data Type | Mode | Information Content | Function | User |
|---|---|---|---|---|---|
| 1 | lemma forms | SL | lexical form(s) of the HW[1]/subheadword | helps user find the information being sought | enc[2] SL<br>dec[3] TL |
| 2 | phonetic transcription | code IPA | how the HW is pronounced | helps the non-native speaker pronounce the word correctly | enc TL |
| 3 | grammar form | code | part of speech, gender, etc. of HW | helps user find the information being sought | enc SL<br>dec TL |
| 4 | sense or /subsense + counter | alph /num code | this is a distinct sense or subsense of the HW | helps user find the information being sought | enc SL<br>dec TL |
| 5 | grammar usage item | SL + TL | grammatical complementation of HW in this sense & its translation | helps TL user use SL item correctly<br>helps SL user identify the sense of the HW | enc TL[4]<br>enc SL |
| 6 | TL equivalent | TL | this is TL equivalent of HW in this sense | helps TL user understand<br>helps both users translate | dec TL<br>enc SL |
| 7 | gloss | TL | an explanation of HW in this sense | helps TL user understand<br>helps both users translate | dec TL<br>enc SL |
| 8 | typical example[5] + translation | SL + TL | this is how the HW in this sense is typically used & translated | helps SL user identify the sense of the HW<br>reassures SL user trying to translate SL item<br>helps TL user use SL item correctly | enc SL<br>enc TL |
| 9 | problematic example[6] + translation | SL + TL | the HW in this context has a specific TL equivalent | helps SL user identify the sense of the HW<br>helps SL user avoid translating error | enc SL |
| 10 | idiomatic example[7] + translation | SL + TL | the HW and context have this specific TL equivalent | helps TL user understand<br>helps both users translate | dec TL<br>enc SL |
| 11 | diatechnical label | code | HW in this sense belongs to this semantic domain of (Music, Science etc.) | helps both users select correct TL equivalent<br>helps SL user identify the sense of the HW | dec TL<br>enc SL |
| 12 | stylistic label | code | using the SL or TL item in this sense is in (literary etc.) style | helps both users translate<br>helps TL user understand<br>helps SL user identify the sense of the HW | enc SL<br>dec TL |

*continued ...*

Table 2 Part 2

| | Data Type | Mode | Information Content | Function | User |
|---|---|---|---|---|---|
| 13 | register label | code | using the SL or TL item in this sense is in (literary etc.) style | helps both users translate<br>helps TL user understand<br>helps SL user identify the sense of the HW | enc SL<br>dec TL |
| 14 | diatopic label | code | the SL or TL item in this sense belongs to X regional variety of the language | helps both users translate<br>helps TL user understand<br>helps SL user identify the sense of the HW | enc SL<br>dec TL |
| 15 | diachronic label | code | the SL or TL item in this sense is (obsolete / old-fashioned etc.) | helps both users translate<br>helps TL user understand<br>helps SL user identify the sense of the HW | enc SL<br>dec TL |
| 16 | evaluative label | code | using the SL or TL item in this sense is (pejorative etc.) | helps both users translate<br>helps TL user understand<br>helps SL user identify the sense of the HW | enc SL<br>dec TL |
| 17 | sense indicator | SL | synonym or paraphrase of HW in this sense / other brief sense clue | helps SL user identify the sense of the HW | enc SL |
| 18 | collocators | SL | typical subjects /objects of HW verbs, nouns modified by HW adjectives etc. | helps both users translate<br>helps SL user identify the sense of the HW | enc SL<br>dec TL |
| 19 | collocators | TL | typical subjects /objects of TL equivalent verbs, nouns modified by TL equivalent adjectives etc. | helps both users translate | enc SL<br>dec TL |
| 20 | cross-reference | SL | this other definiendum is relevant to the HW in this sense | helps users find the information being sought | enc SL<br>dec TL |

Notes on contents of Table 2

1 headword
2 encoding (translating into or writing in the foreign language)
3 decoding (understanding or translating from the foreign language)
4 a TL speaker who stores the information for later use in encoding.
5 an SL example sentence in which the headword and context are amenable to virtually a word-to-word translation into the TL
6 an SL example sentence which is easily understandable for the TL speaker but presents translation problems for the SL speaker
7 a multiword expression (MWE) in which the headword figures, or an example containing such an MWE; the meaning of the MWE is idiomatic, and thus the SL item is semantically opaque to the TL user and not amenable to straightforward translation by the SL user.

- Collocation: this type of information is often drawn from corpora, and the tendency now is towards including this wherever possible.
- Peripheral linguistic information, regarding style, register, region, currency, semantic domain and so on: dictionaries are very rich in this.
- Pragmatics: this type of information often appears in the form of usage notes, or of extra-textual information in the front or back matter.
- Up-to-date language: this is a priority for most publishers, and the tendency is more and more for corpora to supplement editors' card-index files.

### (b) Excellent scholarly work

- Lexicographical: the planning, design and implementation of today's top bilingual dictionaries are often excellent, and the editors of new dictionaries on the market are hard put to it to devise anything better in the same size and price range as their competitors.
- Linguistic: the summary list (in Section (a) above) of the types of information painstakingly gathered, ordered, compressed and presented intelligibly gives enough evidence of this.

### (c) User's needs are paramount

- The lexicographers had a clear idea of the competence, objectives and needs of the users they were writing for, and this is evident from the content and presentation of the dictionary.
- The explanatory material is rich and well thought out, and the metalanguage is tailored to the user who needs the information.
- The front and back matter, also, is well planned and informative, often including verb tables, other tabular information, and annotated sample pages to help the user to get the most out of the work.
- Today's bilingual dictionaries are a pleasure to use : the books are clearly printed and attractively bound, and the text carefully designed to best serve the purpose of the publication.
- Finally, today's dictionaries are excellent value for money. Few other books contain so much information per square centimeter, or entertain the discerning reader so well.

### 1.2.2 Weaknesses

We take a constructive approach to the task of identifying weaknesses in the bilingual dictionaries of today: it is from these flaws (often imposed by the limited technology of our immediate past) that we may draw our inspiration for the dictionary of tomorrow.

#### (a) Redundancies

As Table 2 shows, every entry is too rich for any one reader. It is layered with pieces of information which the reader does not need (what is actually redundant for any individual reader depends of course on the particular circumstances); this makes the dictionary harder to use. Research[6] has shown that many dictionary users, particularly the less motivated, give up before finding the information they need, even when that information is reasonably prominent in the entry. The ideal dictionary should be tailored, or at least tailorable, to one particular type of user.

#### (b) Gaps in coverage

Ironically, in view of these redundancies, no current dictionary, however large, can hope for anything like comprehensive coverage, even if its scope is limited by date or regional variety. Space considerations are not the only reason for this inadequacy: certain linguistic phenomena make it impossible for a static dictionary (such as a print dictionary, or the same on CD-ROM) to predict semantic or lexical variants which may occur as single words or multi-word expressions (MWEs). A list of such phenomena would be long, but would certainly include the following (shown here with brief examples of each, taken from the Oxford lexicographical corpora[7]):

- **neologisms** e.g. (from the dozen or so examples in the OUP US reading programme corpus[8])*By introducing a certain gene a spare may be grown if a part of the anatomy is bobbitted. The Washington Times praises ... Bush for 'the bobbitting of both Saddam Hussein and Manuel Noriega'.*
- **systematic polysemy** (a much discussed topic[9]: the following corpus sentences exemplify the lexical implication rule "animal -> its meat") e.g. *It's not a porcupine, it's a hedgehog. That woman nearly had hedgehog stew.*

- **variation in MWEs** e.g. (chosen from nine attested variants) *whether he has taken a sledgehammer to crack a nut : accused of trying to crack a nut with a sledgehammer; the use of a sledgehammer for the cracking of a smallish nut*;
- **creative exploitation of MWEs** e.g (chosen from many more variants[10]) *the Dean shook in his shoes : unlikely to make any of the teams shake in their boots : Corman has every reason to quake in his boots : I'm quivering in my boots at these problems : made the Redskins quake in their Doc Marten's.*

### (c) Limited user involvement in equivalence selection

It is very hard for a bilingual dictionary user to tell if a word in Language A means the same as an unknown word in Language B, far less whether they diverge in style, register, collocational potential etc. It could be argued that, like true synonyms in a single language, such cross-linguistic synonyms do not exist[11]. Approximation in many, probably most, of the equivalences is inevitable. The lexicographer has to make decisions which rightly should be made by the dictionary user, who is the only person to know exactly what is needed in the other language[12]. The ideal dictionary should offer the skilled user the chance to make his or her own judgement on equivalences, by scanning examples of the TL items (grouped according to meaning) in various types of context, as well as – for contrastive checking purposes – examples of the relevant meaning of the SL item in a wide variety of contexts.

### (d) Distortion of SL analysis by needs of TL

The "left-hand side" of a bilingual dictionary (the SL items) is never simply the same material as is to be found in a monolingual dictionary of the same size. The SL material is subtly distorted by the TL, in order to make the bilingual dictionary better, allowing, for instance, a very brief entry in cases where all or most of the senses of the SL item have the same TL equivalent[13]. Such devices clearly make the dictionary much easier to use, and compaction of information allows more detail elsewhere. It does, however, prevent the dedicated user from getting a clear view of the potential of the SL item, which must be sought in a monolingual work. The ideal bilingual dictionary would be able to cater for all needs: impossible, of course, in a printed work.

**(e) Restricted information**

We often find when we are using a dictionary that we need more information either about a word in our own language or more often about an expression in the foreign language: research described in Atkins and Varantola (in press) shows that people often turn to a monolingual dictionary during a bilingual search. The ideal dictionary should offer monolingual functions (definitions, etymologies, usage notes) to the bilingual dictionary user. It should cater for the dictionary browser, as well as the user intent upon one task.

**(f) Lack of collocational options**

Space constraints make it impossible for users to see a wide range of collocational partners of the foreign language word they want to use. The ideal dictionary should allow the user to browse through genuine attested examples of the foreign expression in use in various types of texts.

**(g) Restricted metalanguage: abbreviations, codes and symbols**

Owing again to space constraints, much metalinguistic information is expressed in the form of abbreviations ("Naut", "Archit" etc.), codes ("*vt*" "*npl*", "*+to-infin*") or symbols (asterisks, daggers, bullet points etc.). For the less motivated dictionary user, these can be hard to understand.

**(h) No formal thesaural functions**

Lack of space and commercial pressures during the editing[14] prevent a systematic semantics-based approach to compiling, and hence exclude the possibility of a full thesaurus as an integral part of a dictionary. "Dictionary and Thesaurus" works usually consist of a small dictionary packaged with a selection of word-based synonymic material.

**(i) No multilingual dimension**

Multilingual dictionaries tend to be simple listings of equivalences across three or more languages. The most useful of these focus on specific semantic domains and technical terms. Again, lack of space and commercial pressures make a true multilingual dictionary impossible, but, even if these obstacles were removed, the bilingual dictionaries of

today could not be transformed into multilingual dictionaries, because of the distortion of the SL analysis by the needs of the TL (discussed above). If a multilingual dictionary is to be compiled, we have to devise an analysis technique common to all the languages involved, and capable of recording without distortion the linguistic phenomena occurring in each language.

## 2. Devising tomorrow's dictionary

As the evaluation in 1.2 shows, even the best of current bilingual dictionaries suffer from serious deficiencies, but I would argue that lexicographers are now in a position to address almost all of them. Many of the obstacles to the creation of tomorrow's improved bilingual dictionary have been removed in the past few decades by the advent of the computer (computer-assisted lexicography, rich electronic text corpora as sources of lexicographical evidence, computational searches of dictionaries, and so on) and advances in linguistic theory, in particular – in my view at least – the development of frame semantics[15] as a theoretical tool for multilingual contrastive descriptions. However, the greatest obstacle to the production of the ideal bilingual dictionary is undoubtedly cost[16]. While we are now, I believe, in a position to produce a truly multidimensional, multilingual dictionary[17], the problem of financing such an enterprise is as yet unresolved.

### 2.1 Users and their needs

Every good dictionary starts from a clear idea of who its users are and what they are going to do with it. User profiles for bilingual dictionaries must of course include the user's native language. The new-style bilingual dictionary must cater equally well for speakers of Language A, and speakers of Language B. All metalanguage should be in the user's mother tongue (L1). This will obviously involve reduplication of effort at the compiling stage, but in an online dictionary should not result in redundant information at the point of use.

In a discussion of multilingual electronic dictionaries, it is important to distinguish between the content language and the presentation language. The **content language** constitutes the object of the lexicographical analysis and description: a monolingual database contains facts about one content language; a bilingual English-French dictionary involves two content languages, and so on. The **presentation language**

is the language in which all metalinguistic information is couched, and also other types of information: in a monolingual French dictionary (one in which the content language is exclusively French), if English is selected as the presentation language the definitions as well as instructions for using the dictionary and the metalinguistic information might well be expressed in English. An electronic bilingual dictionary is able to offer the user a choice of presentation language, as well as of content language; it is indeed possible to envisage a situation where a Japanese speaker wishing to compare English and French consults the bilingual English and French dictionary in contrast mode and elects Japanese as the presentation language[18].

Furthermore, definitions, explanations and other metalinguistic information must be transparent: abbreviations, codes and symbols should be avoided. The familiar "telegraphese" style of definitions and explanations may be abandoned. The new dictionary should be a pleasure to read.

It must serve the following types of user activity[19]:

- understanding L2 (written and spoken)
- translating L2->L1
- translating L1->L2
- expressing oneself in L2 (written and spoken)
  (all four well known), and in addition:
- learning more about L2
- learning more about L1-L2 equivalences and contrasts

For some of the above tasks, some types of data will not be appropriate. For instance, a user trying to read in a foreign language will want the minimum of information, in order not to interrupt the concentration of the reading process[20]. On the other hand, someone studying the language will want more detail, and someone with time to spare may simply wish to browse.

The new dictionary must give its users the opportunity to make their own decisions about equivalences: they should be able to consult as many examples as they need of words used in their various senses, each in a variety of contexts with a variety of collocate partners. They should be able to call up monolingual definitions for these words, to learn about their semantic relationships (of hyponymy, synonymy, antonymy etc.) with other items in the language and with items in the other language. The new bilingual dictionary will provide for its users an accurate reflection of the various meanings of a word, independent of the needs of TL equivalences.

Finally, the new bilingual dictionary must not overwhelm its user. This means that the user must have a say in what information the dictionary offers, and how it presents it. When, as will now be proposed, the dictionary is held in hypertext, it also means that serious thought must be given to making sure users can orient themselves effectively: it is easy to get lost in hypertext.

## 2.2 Exploiting new computational resources

The new-age bilingual dictionary must exploit the advantages of the electronic medium, of which the following are the principal (the letters in brackets below indicate the weakness or weaknesses, set out in 1.2.2, which the particular item addresses):

- hypertext functionality eliminating linear text restrictions and opening the way to new types of information by offering new ways of presenting it (a, b, c, d, f, h, i);
- no space constraints other than the need to avoid swamping the user (e, f, g, h, i);
- no distortion of the source language description by the needs of the target language (d);
- flexible compiling liberated from alphabetical order (h);
- alternative ways of presenting information, as for example graphics (e);
- rapid access to large amounts of lexicographical evidence in corpora (b, c, f);
- large-scale user customization (a, c).

Various consequences for the new-style dictionary design are discussed below. Today's CD-ROM dictionaries, being little more than reincarnation of print dictionaries, do not exploit any of these opportunities. Computerized functions and processes currently realized or realizable, such as accessing virtually unlimited corpus material, or the computerization of compiling, typesetting and so on, are omitted from this discussion.

### 2.2.2 Real databases, real links and virtual dictionaries

One of the priorities of the new bilingual dictionary is to avoid the distortion to the source language analysis (noted in 1.2.2 d) by the pull of

the target language equivalences to be offered in the entry: the more sense overlap there is in the SL and TL lexical equivalents, then the greater the distortion. This should not happen in our proposed new dictionary, which consists of two types of material: (a) the databases (compiled independently for each language) and (b) the dictionaries (the hypertext links, metalinguistic explanations, and instructions for use which are created separately for each dictionary). Figure 2 sets out the relationships: the shaded oblongs are **dictionaries** for human users; these are created by the four **processes** (simple extraction, partial translation, comparison and alignment) carried out on the two **databases** (which in our prototype hold analyses of English and of French).

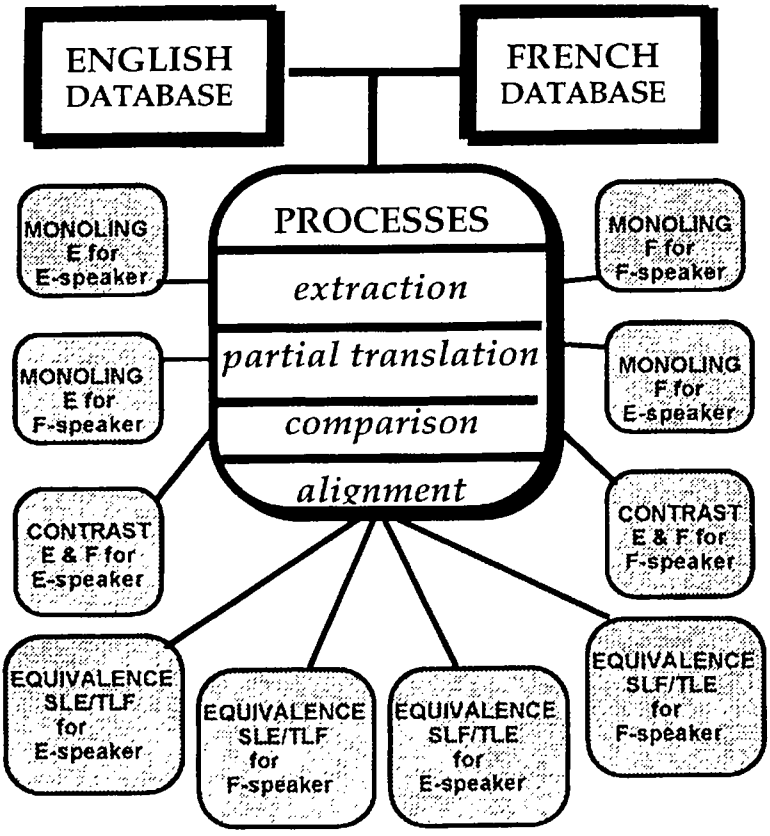



Figure 2. Real databases, processes, and virtual dictionaries

### (a) The databases

A monolingual database is created for each individual language, completely independently of any other, except that all the monolingual databases are compiled within the same theoretical framework (see 3.1) and most of the linguistic facts they hold are inter-compatible, allowing matching of equivalents according to a variety of criteria. This feature enables the production of various types of dictionaries (see (b) below) by adding hypertext links and explanatory material to the monolingual databases.

The contents of these databases should as far as possible be formalized, in order to facilitate access by computers, both those serving information to the dictionaries, and hence to the dictionary users, and those populating automatically lexicons being built for other systems.

### (b) The dictionaries

These will be of at least three **types**: monolingual, bilingual and multilingual, and indeed when enough dictionaries have been compiled the user will be able to switch dictionary types at will.

Each type of dictionary will offer the user various **levels** of information, from brief and simple to long and complex. We may think of these as Level 1, Level 2 and so on.

Monolingual dictionaries may be used in two distinct ways: **look-up mode**, where the user is in search of a specific piece of information, and **browsing mode**, where a more relaxed reading takes place. Dictionary browsing is an activity to be specifically catered for in the dictionary of tomorrow, and the electronic medium offers new ways of making this type of dictionary use even more informative and agreeable.

Bilingual and multilingual dictionaries may function in at least two different **modes**: the traditional mode of bilingual dictionaries, which we term "equivalence mode", and a new mode designed to satisfy the scholar or browser, "contrast mode".

**Equivalence mode** is intended to help users who have to perform specific tasks (such as translation, comprehension or self-expression; see 3.1); by a process of **alignment** (see Figure 2) expressions in Language A, which in our prototype dictionary is English, and Language B (French) are aligned on the basis of one or more specified condition(s), both the traditional ones (synonymy or near-synonymy, antonymy, style, register etc.), and the criterion pioneered in DELIS and described in Heid (1994) and Heid and Krüger (1996), namely the matching of frame

elements expressed in the context of the words in question. Section 3.1 (5) explains the term *frame element.*

**Contrast mode** (or **bilingual browsing mode**) is intended for the person who wishes to find out more about how selected items compare in two or more languages. This produces the dictionary of the browser. It is compiled by the process termed **comparison**, and offers ways of contrasting the meaning and syntactic behaviour of chosen words across languages, with both textual and graphical explanations of similarities and differences.

### (c) The processes and links

As Figure 2 shows, in the creation of the various types of dictionaries for human users, four different **processes** are envisaged. The diagram oversimplifies these processes but is useful for the purposes of explanation. All four processes involve the introduction of hypertext links, and, for each link, the concomitant metalinguistic information and operating instructions and guidelines.

**Extraction** is the name given to the process of selecting and linking items within one content language, and so results in a monolingual dictionary. Here extraction subsumes a certain amount of comparison and alignment of items in the same language, since the dictionary which is created by this process includes functions such as the matching and differentiation of near-synonyms.

**Partial Translation** is the name given to the process of creating a monolingual dictionary by the extraction process and bilingualizing various sections of it so that the language of presentation is different from the language being analysed and described, thus making it more accessible to speakers of other languages.

**Comparison** is the name given to the process of creating a "contrast dictionary" where items in two or more languages are compared along various axes, such as meaning, syntax, style, register, collocational patterns etc., and particularly the way in which the elements of the semantic frame get expressed in the context of the words in question. The resultant dictionary allows the browing user to discover and evaluate real similarities and differences between the items.

**Alignment** is the name given to the process of establishing equivalence links between items in two or more languages. This process involves designating one language as the "departure" or "source" language (the SL) and one as the "arrival" or "target" language (the TL) and the resulting "equivalence" dictionaries are very close in function to the bilingual dictionaries we know today.

The term **links** is intended to cover the **hypertext links** themselves, together with any linguistic **metalanguage** compiled by the lexicographers in order to structure the information for the user, and any navigation **instructions** written by the lexicographers in order to help the user get the best out of the dictionary.

Thus, in brief, the proposal is for a multilingual hypertext lexical resource in which

- the monolingual databases are *real*;
- links (including metalanguage and instructions) between database items are *real*;
- the dictionaries themselves are *virtual*.

### 2.2.3 Customizing

Another function to come into its own in the dictionary of the future is the ability to customize the dictionary to suit one's own circumstances; at present, dictionaries on CD-ROM allow a minimal amount of customization of inessentials, mainly in computational environment and selection of data types to be included in a search.

Users will be able to customize the new dictionary according to their individual needs; in the case of the bilingual dictionary, the customization will bear largely on their knowledge of their own and the foreign language, and the task they are performing with the help of the dictionary. (Too much information in a bilingual dictionary is as bad as too little.) The following aspects affect the type, amount and complexity of data to be returned in response to a query, and also the way in which it is presented, and must be amenable to user preferences:

- content language
- presentation language
- type
- mode
- level.

## 2.3 Exploiting new linguistic resources

Computer-assisted compiling and online dictionaries offer the lexicographer the opportunity of creating a much fuller, more accurate

and easier to use dictionary, whether it is monolingual or bilingual. As already noted, CD-ROM versions of print dictionaries do not (and cannot) take full advantage of the electronic medium. However, there are already in existence a number of techniques and functions which must be exploited in our dictionary of the future; these will not be discussed in detail here. They include the use of corpus analysis during the editing process, and the accessing of corpus citations from the appropriate dictionary sense by the user of the dictionary.

Similarly, not all the types of information which the new dictionary will offer will be the subject of our discussion. An online dictionary will naturally include all the types of information already available, albeit selectively, in current dictionaries (see Table 2 for the bilingual dictionary list, but there are others currently implemented, such as corpus frequency information in some learners' dictionaries of English), and these will not be discussed further in this paper. Here, I shall consider only one of the major lexicographical resources which tomorrow's electronic dictionaries must exploit: the growing body of relevant theoretical linguistic work.

The type of lexicographical analysis that has been implemented in the prototype *Dictionary of the Future* devised by Atkins et al. (1994) and in its bilingual successor (Atkins et al. (1996)) was based on the principles discussed in Atkins, Fillmore and Heid (1995). The technique was pioneered in Delis, and is described in Heid and Krüger (1994)[21]. Space considerations prevent a detailed account here of the analysis of the motion frame which gave rise to the prototype entry for the English verb *crawl* in the first hypertext dictionary and the French verb *ramper* in the second. However, in order to introduce the demo of the prototype of tomorrow's bilingual dictionary, I include now, as exemplification, some brief extracts from the work on the motion frame, and from the hypertext entry for *crawl.*

## 3. Creating monolingual databases

The lexicographical analysis resulting in the monolingual database is a threefold operation: (1) the description of the frame; (2) the compilation of the lexical entries; and (3) the compilation of the thesaurus, involving a feature analysis of the lemmas in each frame. These are briefly outlined below.

### 3.1 Description of frame

The principal steps in this stage of the lexicography are:

1. Selection of semantic domain to work on, and identification of frames to be described within the domain.
   **Example** In the semantic domain of space, one might expect to describe amongst other frames the frame of motion, perhaps in terms of the subframes of locomotion, positional change etc.; within locomotion itself one might wish to distinguish the subframes of manner of motion (*crawl, limp*), speed of motion (*dash, amble*), sound of motion (*clatter, roar*) and so on.

2. Preliminary description of frame and compilation of working list of *frame elements* with which the verbs' behaviour may be comprehensively described.

   Table 3 shows a list of the motion frame elements currently used in the analysis of *crawl*, verb and noun, together with corpus examples in which the expression instantiating the frame element is capitalized.

| Frame Element | Corpus Sentence |
|---|---|
| MOVER | THE SURVEYOR will ... crawl into the loft. |
| AREA | Some bees were already crawling OVER THE EARLY CLOVER. |
| PATH | It can only escape by crawling ALONG A NARROW CHANNEL. |
| SOURCE | Exhausted fugitives crawl FROM THE LAKE. |
| GOAL | She was crawling INTO THE TENT when she heard the sound. |
| DISTANCE | It took him fifty minutes to crawl FIFTY YARDS. |
| MANNER | He crawled ON TOES AND ELBOWS round the Land-Rover. |
| SOUND | I pictured them crawling SILENTLY through the mud. |
| SPEED | I crawled SMARTLY after him. |
| VEHICLE | We crawled through the city IN HIS CAR. |
| MEDIUM | You crawl ON THE GROUND looking for worms. |
| EVENT | It was A LONG CRAWL back to where he had left the tent. |

Table 3. Expression of motion frame element in context of *crawl*

The motion frame elements were identified by means of an analysis of a number of sentence subcorpora, sets of sentences containing a representative lexical unit[22] evoking the frame (for instance, a high-frequency verb or nominalization). First, the frame elements were identified in the sentences; next each was associated with its instantiating sentence constituent and the grammatical phrase type and sentence function of each was noted.

**Example** Figure 3 shows the links between frame elements and their lexical and grammatical realizations in a corpus sentence; each complex description constitutes one *valence formula.*

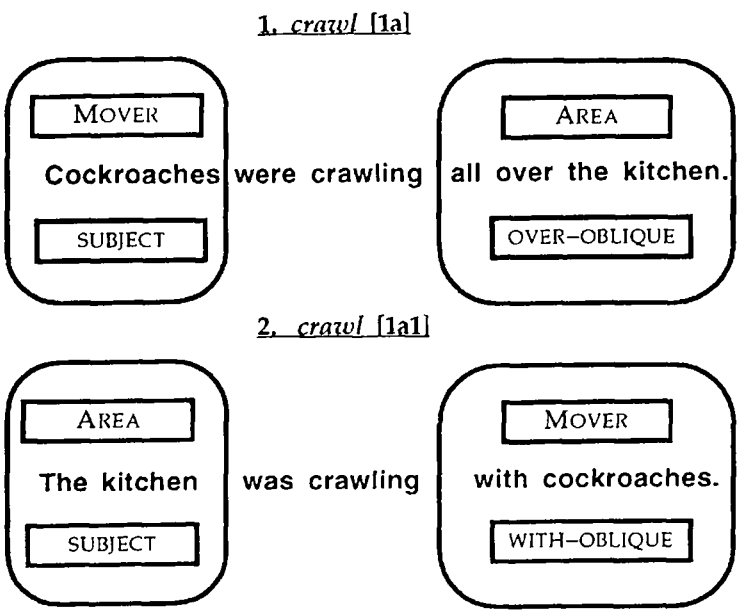



Figure 3. Motion frame elements: two realisations

3. Listing of lemmas which (in one or more of their meanings) evoke this frame, and hence for which lexical entries are to be written in terms of the elements of the frame.
   **Example** The list of verbs evoking the motion frame would run to many hundreds, of which some examples are *walk, run, swim, fly, pass, go, come, leave* and so on.

4. For each of the lemmas, an analysis of the corpus data and recording of the way in which the frame elements are expressed in the context of each lemma. See Table 3, and Figure 3. Such an

analysis normally results in a refinement of the preliminary frame description, as new phenomena are discovered.

5. Listing of its *valence formulas* (see Figure 3) for each of the lexical units analysed. A valence formula comprises

   (a) a *Frame Element Group* (FEG), that is, a configuration of *frame elements* that co-occur in a given structure (e.g. phrase, clause, sentence) headed by that lemma (see the example sentence in Figure 3, which realizes the FEG "Mover, Goal") ; and, for each frame element in the group,
   (b) a specification of *sortal features* (indicating the "selectional" properties of the constituents that can instantiate it); and
   (c) its possible *grammatical realization.*

   The group of valence formulas associated with one sense of a lemma constitute its *valence description.*
   **Example** Table 4 shows a valence formula for *crawl.*

| *The surveyor will crawl into the loft.* | | |
|---|---|---|
| Valence formula:<br>[MOVER / subject / NP / person] *crawl* [Goal / Adjunct / PP-in / direction] | | |
| Frame element | MOVER | GOAL |
| Grammatical function | subject | adjunct |
| Phrase type | NP | PP-in |
| Sortal features | person | direction |

Table 4: A valence formula for crawl [2c]

6. Refinement of the frame description, and definition of the formal metalanguage (frame element names, grammatical codes etc.) to be used for the description of phenomena within the frame.

### 3.2 Compilation of lexical entries

This stage of the lexicography involves the following tasks, in respect of each of the lexical units listed in 3.1(3):

1. Scrutiny of corpus sentences, working with (i) the description of the frame and a checklist of frame elements finalized under 3.1(6) (see Table 3), and (ii) the list of FEGs identified in the representative lexical units analysed under 3.1(4) (described at point 5a above, and see Figure 3).

2. For each sentence: identification of frame elements realized by its constituents, and markup of valence formulas (see point 5 above, and Table 4), associating each element with its appropriate sortal feature(s) and and its grammatical realization in the sentence.

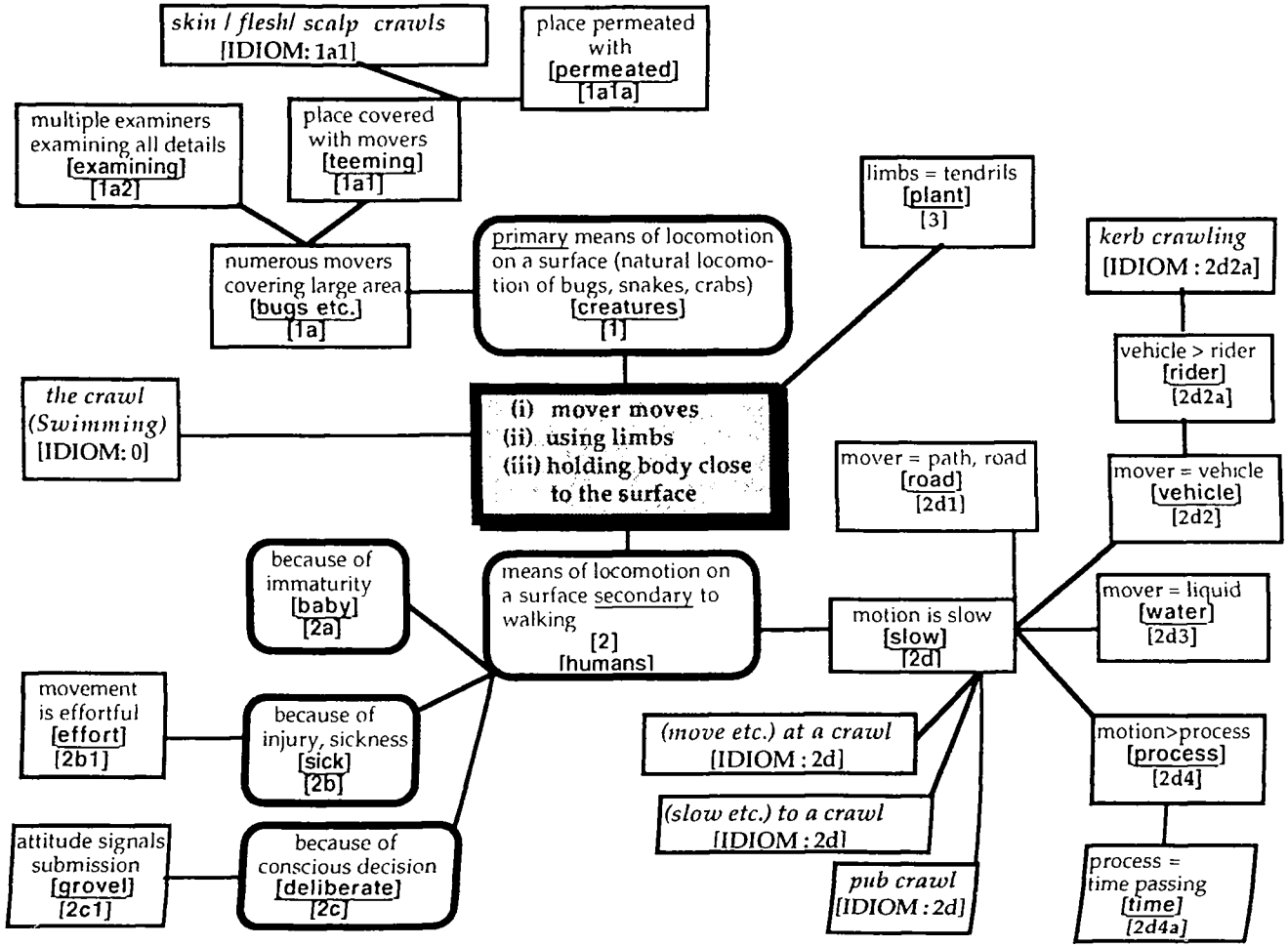
3. Post-editing of computationally extracted valence description (set of valence formulas, see Table 4) for the lexical unit (i.e. the lemma, or headword, in that particular sense), each formula linked to the annotated corpus sentences from which it was derived, and other sentences assigned to that lexical unit from the corpus sentences including that lemma. The valence description of that lexical unit forms part of the database, and dictionary, entry for the word.

4. When all senses of a lemma have been analysed[23], write definitions, complete the various sections of the entry, and draw up the semantic network of that lemma.

   **Example**
   Figure 4 shows the semantic network for *crawl*, verb and noun (see the next page).

In the diagram in Figure 4[24], each sense is (i) numbered in such a way as to reflect semantic relationships, and (ii) assigned a mnemonic ("[humans]", "[plant]", "[time]" and so on) to help users to navigate more easily round the hypertext entry. The meanings discerned for *crawl*, after the study of over 700 sentences from the OUP current English corpus, are illustrated by the following corpus citations, each linked to one of the sense identifiers:

[1] *A ladybird crawled up a dry stalk.*
[1a] *The feeling of insects crawling on the skin ...*
[1a1] *By the time he got back, our room was crawling with cops.*
[1a1a] *He is simply crawling with money nowadays.*
[1a2] *Its members had been crawling inside details of federal grant programmes.*
[2] *I spent ages crawling around the hotel's foundations.*
[2a] *Fit stair gates before your baby starts crawling.*


skin / flesh/ scalp crawls
[IDIOM: 1a1]
place permeated with
[permeated]
[1a1a]
multiple examiners examining all details
[examining]
[1a2]
place covered with movers
[teeming]
[1a1]
limbs = tendrils
[plant]
[3]
kerb crawling
[IDIOM : 2d2a]
numerous movers covering large area
[bugs etc.]
[1a]
primary means of locomotion on a surface (natural locomotion of bugs, snakes, crabs)
[creatures]
[1]
vehicle > rider
[rider]
[2d2a]
the crawl (Swimming)
[IDIOM: 0]
(i) mover moves
(ii) using limbs
(iii) holding body close to the surface
mover = path, road
[road]
[2d1]
mover = vehicle
[vehicle]
[2d2]
because of immaturity
[baby]
[2a]
means of locomotion on a surface secondary to walking
[2]
[humans]
motion is slow
[slow]
[2d]
mover = liquid
[water]
[2d3]
movement is effortful
[effort]
[2b1]
because of injury, sickness
[sick]
[2b]
(move etc.) at a crawl
[IDIOM : 2d]
motion>process
[process]
[2d4]
(slow etc.) to a crawl
[IDIOM : 2d]
attitude signals submission
[grovel]
[2c1]
because of conscious decision
[deliberate]
[2c]
pub crawl
[IDIOM : 2d]
process = time passing
[time]
[2d4a]

[2b] *He is so weak he has to crawl upstairs.*
[2b1] *Too tired to do anything more, he crawled into bed.*
[2c] *The surveyor will pull up carpets and crawl into the loft.*
[2c1] *Let's stop trying to get women to support us by crawling to them.*
[2d] *Dark heavy clouds were crawling across the sky.*
[2d1] *hugging the road that crawls around the mountains ...*
[2d2] *In the blackout the train crawled exasperatingly.*
[2d2a *He dipped his headlights and began to crawl round the bends.*
[2d3] *We watched the wide waves crawling in from the Atlantic.*
[2d4] *She was having friendly chats as she crawled down the list.*
[2d4a] *The days before Christmas seemed to crawl past.*
[3] *He looked at the dark green ivy crawling up the walls.*

In Figure 4, the shading of the central rectangle indicates that the meaning distinction described there is so general as not to be directly lexicalized, but gives rise to all the senses developing from it. The "core" meaning of *crawl*, verb and noun, may be thought of as the tripartite sense contained in the centre shaded rectangle.

Rectangles with rounded corners and bold lines refer to "literal uses" of this word: the first division is "primary means of locomotion" (for snakes, bugs and other creatures) versus "secondary means" (for human beings, whose natural means is walking), the latter being further sub-divided according to the reason behind the adoption of a secondary means of moving.

Regular rectangles with roman type refer to "extended senses" of the word *crawl*, and these extended senses develop from different literal uses: the network diagram was devised in order to show these relationships. In addition, although there is not room on the diagram to include these, the various lines linking meanings can each be labelled, according to the type of semantic change involved. The label for the line linking [2c : deliberate] to [2c1 : grovel] is "Metaphor"; that linking [2d2 : vehicle] to [2d2a: rider] is "Metonymy: riders as their vehicles" (see below for examples sentences illustrating the various uses).

Regular rectangles with italic typeface refer to idioms. Our design assumes an Idioms Database, with hypertext links from each item there to the appropriate sense of the various component words.

### 3.3 The compilation of the thesaurus

Compiling the thesaural sections of the new dictionary involves (i) the selection of a frame to work on; (ii) a feature analysis of each of the words which evoke that frame[25]. This is not to claim any theoretical value for a decompositional approach to word meaning; however, in the prototype dictionary it has proved a useful method of differentiating among semantic neighbours, in this case the co-hyponyms of the verb *move*. In the prototype dictionary, comparing verbs and feature sets[26] is an interactive process: the lexicographers' task is to compile the feature set for each verb (and noun, and adjective etc.) in the frame, ensuring that the contrastive descriptions which result from this reflect not only the native speaker's intuition about the core sense of the word but also insights gained from a study of the word's behaviour in the corpus evidence.

The components of meaning of each verb are recorded in the form of semantic features attached to the elements of the frame. Thus, in the case of *crawl*, for instance, the Manner frame element (see Table 3 for the elements so far discerned for the motion frame) is noted as being "body-angle: horizontal", Medium as being "ground", rather than "air" or "liquid", Speed as being typically "slow", and so on. In this approach, a verb may be marked or unmarked in respect of any frame element; if marked, then the options depend on the element in question. For instance, *crawl* is marked for Manner (you cannot crawl upright, or erect, or on tiptoes); *enter*, on the other hand, is not marked for Manner (you can enter somewhere erect or on all fours, gracefully or awkwardly, and so on). Verbs like *sidle* or *crabcrawl* are marked for Path (which is lateral, rather than forward, backward, up, down, etc.); verbs like *crawl*, *enter* and *swim* are not.

| Semantic feature | | Verbs | | | | | |
|---|---|---|---|---|---|---|---|
| | | *crawl* | *creep* | *amble* | *wriggle* | *swim* | *fly* |
| EVENT | continuous | x | x | x | x | x | x |
| EVENT | non-continuous | | | | | | |
| EVENT | salient | | | | | | |
| EVENT | non-salient | | x | | | | |
| MANNER | body-angle: horizontal | x | | | x | | |
| MANNER | body-angle: vertical | | | | | x | x |
| MANNER | surface contact: constant | | | | x | | |
| MANNER | surface contact: low | | | | | x | x |
| MANNER | motion: autonomous | x | x | x | x | x | x |
| MANNER | motion: non-auton. | | | | | | |
| MEDIUM | air | | | | | | x |
| MEDIUM | ground | x | x | x | x | | |
| MEDIUM | liquid | | | | | x | |
| SOUND | level: loud | | | | | | |
| SOUND | level: soft | | x | | | x | |
| SPEED | level: fast | | | | | | |
| SPEED | level: slow | | | x | x | | |

Table 5: Feature-based contrasts

Table 5 shows a partial contrastive analysis based on some of the features which are used to differentiate the meaning of verbs evoking the frame of motion. In the hypertext dictionary, the process is a dynamic one. Words may be contrasted according to their semantic features (as in Table 5), but it is also possible to submit an arbitrary group of semantic features and receive a listing of the verbs whose meaning incorporates them.

## 4. Creating dictionaries

A further task for the monolingual lexicographer is to decide on the functionality required in the monolingual dictionaries to be extracted from the database (see Figure 2), and to set up the hypertext links, design the screen displays, and compile the metalinguistic explanations and user guidelines for each function and each display. Because of the user customization imperative (see 2.2.3), it is planned to offer for most types of information (definition, syntax, examples, etc.) levels of complexity and amounts of data which depend on the user's declared objective in using the dictionary, standard of competence in language and degree of interest in the dictionary contents. Consequently, the editorial design and the lexicography needed to implement that design are extremely detailed and complex.

The bilingual lexicography is (as it always is) vastly more complicated than the monolingual. Bilingual and monolingual lexicographers must work together if it is decided to produce a version of a monolingual dictionary for users whose mother tongue is not the content language. For these users, metalinguistic explanations, user guidelines and even conceivably the definitions themselves must be translated.

When the dictionaries being compiled from the databases are to be contrast dictionaries or equivalence dictionaries, then the functionality becomes even more complex. The designers' tasks are to ask themselves what would users of various degrees of competence, with different objectives and needs, want from this resource? How is the best way to display the information without swamping the reader? How best can the user customize each aspect of the dictionary? The planning stage will involve lexicographers, linguists, computational linguists and computer scientists. Creating the dictionaries by establishing the hypertext links, and writing the explanations and guidelines, is the task of the bilingual lexicographers. They have to study the contents of the two databases, to decide whether Item X in Database A and Item Y in Database B should be linked as equivalents; to select and manipulate all the other types of information to be extracted from the databases and edited into the dictionaries. All the types of information listed in Table 2 will obviously be included, but there will of course be an added dimension of thesaural cross-linguistic contrasts and equivalences.

## 5. Envoi

We have at our disposal the knowledge to plan, and the computational and linguistic capabilities to implement, a radically new type of bilingual dictionary. It will demand more of the lexicographers, more energy for sifting lexicographical evidence and more intellectual effort to understand and systematize what is found there. It will require the collaboration of linguists and linguistically aware computer scientists, and can be produced only if there is a continuous and efficient dialogue between them and the lexicographical team. It will undoubtedly cost more initially than any standard print dictionary. But in this forum, if not yet in publishers' planning meetings, let us look beyond the currently possible and set our sights on the distant ideal. A demonstration will be given of a prototype dictionary of the future (Atkins et al (1996)), conceived as a multilingual hypertext dictionary, which will subsequently be available for consultation on the World Wide Web[27]

## Notes

1 My analysis of the bilingual dictionary entry was carried out within the EC Compass project (LRE 62-080) and is taken from Deliverable 24 of that project: *Adapting Bilingual Dictionaries for On-line Comprehension Assistance*, Atkins et al (1996).

2 The names of the data types are taken from Compass Deliverable 1: *Terminology for Bilingual Dictionaries in Computational Lexicography*, Elisabeth Breidt.

3 This encompasses the maximal entry in a bidirectional bifunctional dictionary, i.e. one designed to be used by speakers of either of its two languages, for encoding or decoding (see Al (1983)), and consequently highly redundant for any particular user. Individual dictionaries vary of course from this model (for instance, in the subset of the data types which constitute an entry, or the choice of SL or TL for metalanguage), but where the book is to be sold in two markets most standard bilingual dictionaries offer most of these data types, and overall hold much the same types of information.

4 This assumes that every design decision made for the hypothetical dictionary is the best possible, and that the editorial policy was carried out during the editing process in the best possible way.

5 The terms *lexical item* and *item* in this sense are intended to cover both single- and multi-word expressions.

6 See Atkins and Varantola (in press, and in preparation).

7 These are the corpus resources used by the lexicographers of Oxford University Press, and include the British National Corpus, the various corpora created by the OUP reading programme, and historical corpora.

8 My thanks to John Simpson for these examples. For the reader who needs elucidation, two further citations from the same corpus might be helpful: *To bobbitt – short for to depenistrate*, and *Last month, a Taiwanese wife bobbitted her husband with a pair of scissors after learning of his affairs with other women.*

9 One of the seminal works is Apresjan (1973); more recently, see Nunberg and Zaenen (1992); for a study of this phenomenon in the context of computational lexicography, see Copestake and Briscoe (1995). The expression "lexical implication rule" was coined in Ostler and Atkins (1992).

10 My thanks to Rosamund Moon, who drew my attention to these examples.

11 See Marello (1989) Part I, Chapter 2, for an excellent discussion of this, also Zgusta (1984), Snell-Hornby (1984, 1986 and 1990), and Duval (1991).

12 Krista Varantola (personal communication) points out however that lexicographers are often better linguists than the person using the dictionary, and care must be taken to avoid abdication of responsibility towards the less skilled dictionary user. The advanced dictionary users of course are those who will benefit from selective access to corpus data (see Varantola (1994)).

13 The OHFD entry for *column* contains two senses ("gen colonne *f*' and "Journ rubrique *f*"); the entry in the Concise Oxford Dictionary (1995), with a similar-sized headword list, is set out in six senses, three of which are further subdivided. See Kromann (1989), and Kromann et al. (1984, 1989) for further discussion of this.

14 Compiling entries for words in semantic sets entails an additional pass through the wordlist, greatly increasing the time and expense of dictionary production. For instance, the English adjective *civil* would require to be compiled in the "Military", and the "Social Behaviour" sets, as well as figuring in compounds like *civil servant* and *civil engineering*, which themselves belong to different semantic sets. When all such uses had been compiled individually, the final version of the entry would have to be assembled. Reducing this to the correct length might then have a knock-on effect on the various sets involved. Editors have nightmares of an infinite loop.

15 See Fillmore (1985, 1993a and b, 1995), Fillmore and Atkins (1992, 1994) and Atkins (1995) for a discussion of frame semantics and its application to lexicographical analysis.

16 Preliminary budgeting suggests that a monolingual hypertext dictionary of the type discussed here would be equivalent in editorial costs to a similar very large multivolume monolingual scholarly dictionary. Such works are never undertaken for commercial reasons. Bilingual and multilingual versions would be proportionately more expensive.

17 The results of a frame-semantics-based multilingual analysis may be seen in the prototype five-languages lexicon of Perception and Speech Act verbs produced during the DELIS project and described in Heid and Krüger (1996), while the *Dictionary of the Future* presentation (see Atkins et al (1995)) demonstrated a prototype entry in a multidimensional hypertext dictionary.

18 Since adding a presentation language to the multilingual database will involve a lot of work it has to be assumed that this situation is quite a long way into the future.

19 This topic is well discussed in the literature: see Hausmann (1977), Al (1983), Kromann (1987) and Bogaards (1990), among others.

20 The objective of the Compass Project (LRE 62-080), now successfully completed, was to develop the prototype of just such a dictionary; see Breidt and Feldweg (ms).

21 See Cowie (forthcoming) for a discussion of the application of a frame-based approach to the analysis of idioms for lexicographical purposes.

22 This term denotes a lemma in one of its meanings.

23 Each lexical unit may evoke a different frame and consequently a polysemous word is likely to participate in the analysis of many frames.

24 The network is the idea of Charles Fillmore and this description of the meanings of *crawl* is to a considerable extent his work.

25 As the work in DELIS indicated (although this aspect was not fully developed during the project), wordclasses other than verbs also evoke frames; see Fillmore <<REF: Hard Road >> for a description of applying frame semantics to the analysis of nouns.

26 So far, this operation has been performed only for verbs.

27 My thanks go to Marie-Hélène Corréard, Ulrich Heid, Carla Marello and Krista Varantola for their valuable comments on an earlier version of this paper, and I acknowledge with gratitude the unique contribution to the design of the hypertext dictionary by J. B. Lowe, whose computational expertise called it into being, and Charles Fillmore, whose ideas it attempts to embody. The WWW version may be found at: http://www.linguistics.berkeley.edu/hyperdico/.

## Bibliography

Al, B. P. F. (1983b). Dictionnaire de thème et dictionnaire de version. *Revue de Phonétique Appliquée*, Vols. 66–68: pp. 203–211.

Apresjan, J. D. (1973). *Regular polysemy.* Linguistics, 142. Mouton, The Hague.

Atkins, B. T. S. (1995). Analyzing the verbs of seeing: a frame semantics approach to corpus lexicography, in: S. Gahl, C. Johnson, A. Dolbey (eds.) *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistic Society, 1994.* BLS, University of California, Berkeley, CA.

Atkins, B. T. S., Charles J. Fillmore, John B. Lowe and Nancy Urban (1994) The Dictionary of the Future: A Hypertext Database. Pre-entation and on-line demonstration at the Xerox-Acquilex Symposium on the Dictionary of the Future, Uriage, France. (ms).

Atkins, B. T. S., Charles J. Fillmore and Ulrich Heid (1995) Lexicographical Relevance in Corpus Evidence, Deliverable D-IX-2 of DELIS Project (LRE 61.034).

Atkins, B. T. S., Charles J. Fillmore, John B. Lowe, Marie-Hélène Corréard and Nancy Urban (1996) The Dictionary of the Future: A Multilingual Hypertext Dictionary. Sample entry available on WWW.

Atkins, B. T. S. & K. Varantola (in press): Monitoring Dictionary Use, *International Journal of Lexicography.*

Atkins, B. T. S. & K. Varantola (in preparation*):* Language Learners Using Dictionaries: The Final Report of the EURALEX and AILA Research Project into Dictionary Use.

Bogaards, P. (1990a). *Deux langues, quatre dictionnaires.* Lexicographica, 6. eds. F. M. Dolezal et al, Niemeyer, Tübingen.

Breidt, Elisabeth and Helmut Feldweg (unpublished ms.) Accessing Foreign Languages with Compass, submitted to forthcoming special issue of *Machine Translation.*

Copestake, A. A. and E. J. Briscoe (1995) Regular polysemy and semi-productive sense extension, in *Journal of Semantics*, Vol. 12, pp 15–67.

Cowie, A. P. (forthcoming) Semantic frame theory and the analysis of phraseology, in *Proceedings of the Second International Symposium on Phraseology, Moscow, 2–4 April 1996*, L. Minaeva (ed.). Moscow: Moscow State University.

Duval, A. (1991). L'équivalence dans le dictionnaire bilingue. in Hausmann, F. J., Reichmann, O., Wiegand, H. E., and Zgusta, L., eds., *Wörterbücher. Ein internationales Handbuch zur Lexikographie Band*

*3, Handbücher zur Sprach- und Kommunikationswissenschaft*, 2817–2823. Berlin and New York..

Fillmore, Charles J. (1985) Frames and the Semantics of Understanding, in *Quaderni di Semantica* VI:2.

Fillmore, Charles J. (1993a) "Corpus Linguistics" or "Computer-aided armchair linguistics", in *Nobel Symposium Proceedings*. Stockholm.

Fillmore, Charles J. (1993b) Frame semantics and perception verbs, in: Hans Kamp and James Pustejovsky (eds.), Universals in the Lexicon: At the Intersection of Lexical Semantic Theories, ms., Dagstuhl.

Fillmore, Charles J. (1995) The Hard Road from Verbs to Nouns, in *In Honor of William S-Y. Wang* (eds. Matthew Chen and Ovid Tzeng): Pyramid Press

Fillmore, Charles J. & B. T. S. Atkins (1992) Towards a Frame-based Lexicon: the Semantics of RISK and its Neighbors, in A. Lehrer & E. F. Kittay (eds.), *Frames, Fields and Contrasts: New Essays in Semantic and Lexical Organization*, pp. 75–102. Lawrence Erlbaum Associates: Hillsdale, New Jersey.

Fillmore, Charles J. & B.T.S. Atkins (1994) Starting where the Dictionaries Stop: the Challenge of Corpus Lexicography, in B. T. S. Atkins & A. Zampolli (eds.), *Computational Approaches to the Lexicon*, pp. 349–393. Oxford University Press, Oxford UK.

Hausmann, F. J. (1977). Einführung in die Benutzung der neufranzösischen Wörterbücher, in *Romanistische Arbeitshefte*, 19. Tübingen: Max Niemeyer Verlag.

Hausmann, F. J. (1988). Grundprobleme des zweisprachigen Wörterbuchs, in: Hyldgaard-Jensen, K. and Zettersten, A., (eds.), *Symposium on Lexicography III. Proceedings of the Third International Symposium on Lexicography May 14–16, 1986 at the University of Copenhagen*, Lexicographica Series Maior 19, 137–154, Tübingen.

Heid, Ulrich and Katja Krüger (1994) On the DELIS Corpus Evidence Encoding Scheme (CEES), Deliverable D-III-0 of DELIS (LRE 61.034), January 1994.

Heid, Ulrich and Katja Krüger (1996) A multilingual lexicon based on Frame Semantics, in *Proceedings of AISB96 Workshop on Multilinguality in the Lexicon*, (eds.) Lynne Cahill and Roger Evans. University of Sussex, UK.

Heid, Ulrich (1994) Contrastive Classes – Relating Monolingual Dictionaries to Build an MT Dictionary in: Ferenc Kiefer, Gabor Kiss, Julia Pajzs (eds.), *Papers in Computational Lexicography – Complex '94*, (Budapest: Linguistics Institute, Hungarian Academy of Sciences), 115–126.

Kromann, H.-P. (1987) Zur Typologie und Darbietung der Phraseologismen in Übersetzungswörterbüchern, In. Korhonen, J. (ed.) *Beiträge zur allgemeinen und germanistischen Phraseologie-forschung: Internationales Symposium in Oulu, Juni 1986.* Oulu: Universität Oulu.

Kromann, H.-P. (1989). Neue Orientierung der zweisprachigen Wörterbücher. Zur funktionellen zweisprachigen Lexikographie. In: Snell-Hornby, M. and Pöhl, E., (eds.), *Translation and Lexicography,* Paintbrush.

Kromann, H.-P., Riiber, T., and Rosbach, P. (1984). 'Active' and 'passive' bilingual dictionaries: the Scerba concept reconsidered. In: Hartmann, R. R. K., (ed.), *LEXeter '83 Proceedings. Papers from the Inter-national Conference on Lexicography at Exeter, 9–12 September, 1983*, Lexicographica Series Maior 1, 207–215, Tübingen. Niemeyer.

Kromann, H.-P., Riiber, T., and Rosbach, P. (1989). Principles of bilingual lexicography. In: Hausmann, F. J., O. Reichmann, H. E. Wiegand and L Zgusta (eds.) *Dictionaries, Dictionnaires, Wörterbücher, Ein internationales Handbuch*, 2711–2728. Berlin: de Gruyter.

Marello, C. (1989). *Dizionari bilingui.* Zanichelli, Bologna.

Nunberg, G. and Zaenen, A. (1992). Systematic polysemy in lexicology and lexicography. In: Varantola, K., Tommola, H., Salmi-Tolonen, T., and Schopp, J., eds., *The Proceedings of EURALEX '92, Tampere, Finland.* Department of Translation Studies, University of Tampere.

Snell-Hornby, M. (1984). The bilingual dictionary help or hindrance? In Hartmann, R. R. K., (ed.), *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9–12 September, 1983*, Lexicographica Series Maior 1, Tübingen. Niemeyer.

Snell-Hornby, M. (1986). The bilingual dictionary victim of its own tradition. In: Hartmann, R., ed., *The History of Lexicography.* John Benjamins, Amsterdam and Philadelphia.

Snell-Hornby, M. (1990b). *Dynamics in meaning as a problem for bilingual lexicography.* In: Tomaszczyk, J. and Lewandowska-Tomaszczyk, B., eds., Meaning

Varantola, Krista (1994) The Dictionary User as Decision Maker, in Martin, W., W. Meijs, M. Moerland, E. ten Pas, P. van Sterkenburg and P. Vossen (eds.), *Euralex 1994 Proceedings, Amsterdam*: Vrije Universiteit Amsterdam. pp. 606–611.

Zgusta, L. (1984). Translation equivalence in the bilingual dictionary. In: Hartmann, R. R. K., ed., *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9–12 September, 1983*, Lexicographica Series Maior 1, Tübingen. Niemeyer.