Monolingual, bilingual, «interlingual» description. Some remarks on a new method for the production of bilingual dictionaries*

Ulrich Heid

1. Introduction

In a paper published in 1988, Bernard Al[1] proposed a method for the semi-automatic preparation of contrastive lexicographic descriptions which could serve as an input to the construction of bilingual dictionaries. The basic idea of his approach, developed within the framework of the Van Dale series of monolingual and bilingual dictionaries, is to combine two bilingual dictionaries, X → A and X → B, in order to get material for A → B and B → A dictionaries. The A ↔ B material can be retrieved automatically by a «conversion» procedure.[2]

Within the framework of a preliminary study, we had the opportunity to analyse the output of the conversion system on the basis of two extracts from German/French, French/German raw material.[3] The goal of the analysis was to estimate the feasibility of a large-scale application of the conversion procedures for the production of input to lexicographic work. But the results of the feasibility study are of some interest not only for the application of the conversion procedures in lexicographic practice, but also for the discussion of more general questions pertaining to the re-use of dictionary material in natural language processing. Some of the results are of particular relevance for the construction and use of multilingual translation dictionaries.

Within the present paper, we proceed from more specific problems of the actual conversion experiments to more general questions concerning the re-use of dictionaries in natural language processing.

We first discuss the input to the dictionary conversion and the methodology of Bernard Al (1988).

* The research described in this paper has been carried out under a contract from the publishing house Dictionnaires Le Robert, Paris, and in collaboration with Van Dale and Le Robert. The work described here has been done together with Miriam Scheytt. Without her contribution, this presentation would not have been possible. The present text has benefited a lot from discussions of previous versions with Bernard Al, Alain Duval and Willy Martin, whom we would like to thank for their constructive remarks. All errors or imprecisions are of course our own.

1. This is Al (1988).

2. The term «conversion» will henceforth be used for the (set of) procedure(s) described by Al (1988) which mainly reorganize lexical descriptions contained in the Van Dale dictionary databases. (Al 1988) does not use this term himself.

3. We would like to thank Van Dale publishers and especially Bernard Al for giving us the opportunity of analysing the material.

Then the results of the preliminary study are presented, from both a quantitative and a qualitative point of view (section 3).

The fourth section deals with the main problems encountered in the assessment of the feasibility of the approach; there are some general limitations inherent to the philosophy underlying the Van Dale dictionary series and the conversion routines; on the other hand, we also encountered a certain number of problems originating from inconsistencies in the use of coding conventions, or simply from the use of lexicographic text-condensing devices.

The last section of this paper is devoted to a generalization of the results from different points of view: the Van Dale approach to bilingual dictionaries, the ongoing discussion about the structure of bilingual dictionaries, and finally, the discussion about re-use of dictionary material in multilingual natural language processing.

2. The input to dictionary conversion

Before discussing the actual conversion procedure proposed by Al (1988), we should briefly describe the input to this procedure. We start with a rough description of the microstructure of the relevant Van Dale dictionaries, then discuss the philosophy of the Van Dale series of monolingual and bilingual dictionaries, and finally outline the conversion procedure.

2.1. The microstructure of the monolingual Van Dale dictionary

We cannot describe here the complete microstructure of Van Dale dictionaries, and thus limit ourselves to the four main types of «information packages»:

1. The headword;
2. A syntactic description[4] which can be generalized over all «senses» of the headword;
3. A semantic description of different «senses» of the headword (keeping track of the syntactic properties);
4. A description of the combinatory properties of the headword, organized according to the «senses», with collocations and example sentences.

As an example, we reproduce the entry s.v. *stalling*, from the Dutch monolingual dictionary (Van Sterkenburg/Pijnenburg, 1984). We have separated the information packages according to the above categories.

4. Optionally, other types of information can complete this information package, e.g. variation or subject field marks, indications of the language from which a lexeme is borrowed, etc.

'stalling

(de ~ (v.); -en)

0.1 loods, garage enz. waarin rijwielen of auto's worden gestald **0.2** het op stal brengen of zetten van dieren, met name paarden **0.3** het in een loods of garage onderbrengen van rijwielen, auto's enz. **0.4** geld dat men betaalt voor het stallen ⇒*stalgeld* ◆

1.3 *gelegenheid tot ~ van rijwielen* **2.1** *een overdekte ~ bij het station* **3.4** *ik kom u de ~ betalen.*

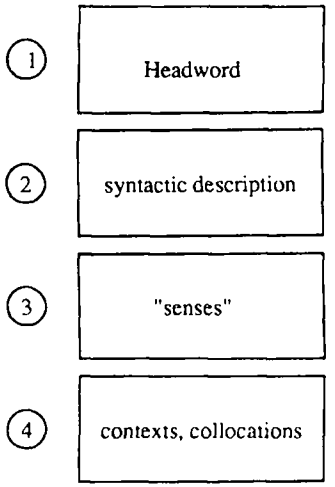

Figure I: Components of an entry of the monolingual dictionary

We now have to take a closer look at the semantic description (i.e. at information package 3), because it is central to the argument of the present paper. As do most defining dictionaries, Van Sterkenburg/Pijnenburg (1984) also uses different devices of meaning description. They range from definition-like texts (specifying the *genus proximum* and the *differentia specifica*) over indications of synonyms to usage notes in the sense of short descriptions of the application context of a given lexical item (often interpretable as selection restrictions or domain-specific indications). No particular definition style (as in *COBUILD*) is used, and it seems from a rough glance at some examples that the devices described above (the list of which is by no means exhaustive) may co-occur in any combination. As a general rule, every «sense» has to have one (and only one) meaning description.[5]

5. The situation may be different for other Van Dale dictionaries; the use of more rigidly definable «definition models» within the English → Dutch dictionary has been argued for by W. Martin, personal communication.

2.2. The relationship between the monolingual and the bilingual dictionary

In the previous section, we have described the microstructure of entries of Van Sterkenburg/Pijnenburg (1984). Abstracting away from the collocational part (information package 4), we can individuate pairs composed of:

- A Dutch headword and its syntactic (...) properties (= information packages 1 and 2), and
- A Dutch meaning description (= information package 3).

The central idea of the series of bilingual dictionaries with Dutch as a source language is now to use the set of those pairs (headword & meaning description, henceforth called «concept») «as a starting-point for the macro-structure» of the bilingual dictionaries.[6] This means, roughly speaking, that all Dutch → X dictionaries contain (more or less) the same set of «concepts» to be lexicalized in the respective foreign languages. The idea behind this procedure is to view the Dutch → X translation dictionaries as inventories of lexicalization candidates for situations or states of affairs which Dutch-speaking people would like to speak about.[7] Therefore, all bilingual dictionaries of the series have the same microstructural organization as the monolingual one, with the only difference being that the meaning description (information pack-age 3) contains a shorthand for the meaning description text used in the monolingual dictionary, including, of course, a list of (one or more) equivalents in the target language.

As to the collocational and contextual description, part of the collocations and examples of the monolingual dictionary are carried over to the bilingual dictionaries and translated into the respective target languages.

Schematically, an entry of the monolingual dictionary can be represented as follows:

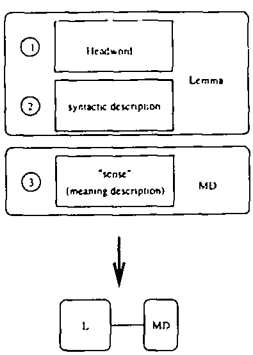

Figure 2: Schematic representation of a monolingual entry

6. This is the formulation used by Van Sterkenburg/Martin/Al (1982): 228.

7. For a discussion of this approach, see Al (1988: 21), Van Sterkenburg/Martin/Al (1982): passim. I have recently compared this approach with the «directional» approach to bilingual dictionary construction, as advocated, for example, by Kromann/Riiber/Rosbach (1984), Kromann/Riiber/Rosbach (1989). Cf. Heid (1990).

A bilingual entry, Dutch → French for example, may then be schematized in the following way:

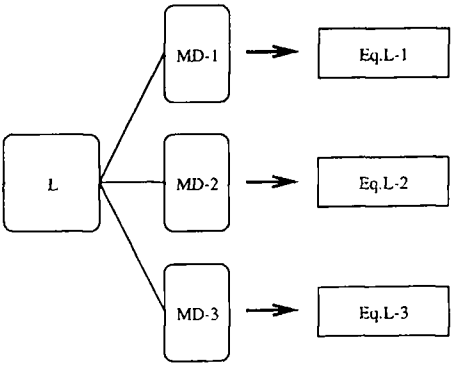

Figure 3: Schematic representation of a bilingual entry

2.3. The proposal of Al (1988): relating different bilingual dictionaries

The procedure proposed by Al (1988) consists of combining two or more of the bilingual descriptions in the format discussed above. Ideally, the Dutch → French and Dutch → German dictionary entries for the same Dutch headword will have the same sense division according to the respective Dutch meaning descriptions:

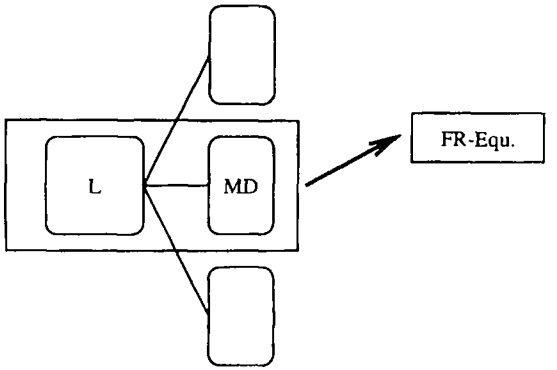

Figure 4: Schematic representation of a Dutch → French entry

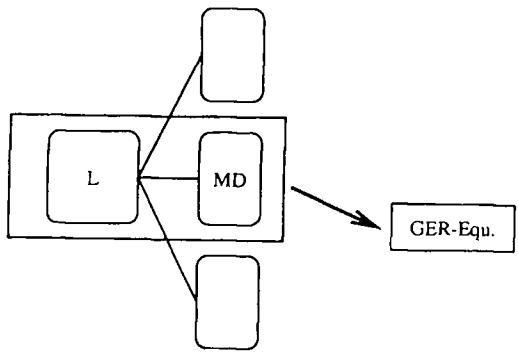

Figure 5: Schematic representation of a Dutch → German entry

Thus a merging is assumed to be possible which will yield a synopsis of French and German lexical means for expressing the concepts described by the pairs composed from a Dutch lemma (L) and a meaning description (MD), as illustrated in figure 6:

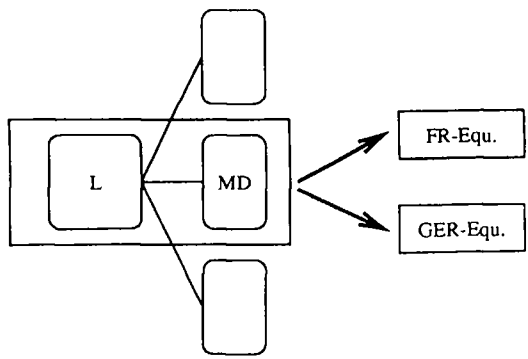

Figure 6: Combining different bilingual dictionaries

This «synopsis», as Al (1988: 22) calls it, uses Dutch as a «métalangue» in that it states equivalences of French and German with respect to a «conceptual» description of states of affairs given by the pairs of Dutch lemmas and meaning descriptions.

The guiding principle of the construction of this synopsis is near to the «interlingua-based» approach to machine translation. The combination of the monolingual description and the bilingual dictionaries yields a multilingual lexical description where the Dutch part serves as an interlingual representation of conceptual units. Consequently, Al (1988) discusses the possibility of combining more than two bilingual dictionaries in order to arrive at a true multilingual lexical database.

When it comes to realizing this conception technically, the input to the dictionary conversion process consists of different bilingual Van Dale dictionaries with Dutch as a source language. The output of the conversion routine is a synopsis in the sense described above.

Al (1988) foresees several semi-automatic ordering devices which allow to easily extract subsets from the synopsis. A more sophisticated one consists in running the synopsis against the macrostructure of the French → Dutch Van Dale dictionary. And since the Van Dale dictionaries with Dutch as a target language obey the same organization principles as the other bilingual dictionaries, in that they have four information packages, it is even possible to make partial use of this matching for the ordering of senses in the bilingual extract of the synopsis to be created: the sense division in the French → Dutch or in the German → Dutch dictionaries is based on criteria applying to French or German and is thus independent of the Dutch language.[8]

Our evaluation has been based on two subsets taken from a German ↔ French (↔ Dutch) synopsis, one ordered alphabetically by German items, the other by French items.

3. An overview of the results of the dictionary conversion

In this section, we will describe the results of the analysis of the sample output of the dictionary conversion. This comprises a quantitative estimation (coverage of the macrostructure of a dictionary which could be derived by means of the conversion procedures) as well as a qualitative one (what information would such a dictionary contain? how reliable would it be?). Phenomena which cause problems in the conversion process are discussed in section 4.

3.1. Quantitative results

The Van Dale dictionaries used as an input to the conversion are comprehensive one-volume dictionaries; the conversion should thus allow the construction of dictionaries of roughly the same format.

We have evaluated two samples from the conversion results:

- French → German: *g - gare*;
- German → French: *abwerten - Alleinhandel.*

The French → German alphabetical sample covers about 0.5% of the number of pages of the usual bilingual French → German and French → English dictionaries.[9] The German sample is equivalent to 0.9% of the number of pages of usual dictionaries.[10]

For the French → German part, the conversion results add up to around 80% of the macrostructure of Al *et al.* (1985); the German → French part is even richer. This indicates (as expected) that the richness of the input dictionaries will be inherited by those which could be produced out of the multilingual database.

8. There has been some more detailed discussion, within Van Dale and elsewhere, about optimizing the conversion procedure; an alternative to the method described by Al (1988) has been proposed by Martin (1989). We will not go into details about this topic here.

9. We have compared the results with exhaustive lists from Sachs/Villatte, Weiss/Mattutat, Robert/Collins, Robert/Collins-junior, as well as with Al *et al.* (1985).

10. Comparisons with Sachs/Villatte, Weiss/Mattutat, as well as Bertaux/Lepointe which has a vast amount of outdated items absent from the conversion results and Cox *et al.* (1986).

These quantitative results are important from the point of view of a potential reuse of already existing dictionary descriptions for the construction of new dictionaries: it is possible to transfer the macrostructural richness from existing dictionaries to new ones.

3.2. Qualitative results: the reliability of the results

It is not sufficient just to have a large amount of data, these data should also be reliable. If the conversion results are to be used for the construction of a new bilingual dictionary, it has to be made sure that the results allow one to construct a sufficiently detailed microstructure. Furthermore, we have to make sure that the conversion does not introduce deficiencies at the level of descriptive adequacy, for example by wrongly relating items which should not, descriptively speaking, be related. In order to assess the microstructural richness of a dictionary which would be produced on the basis of the conversion results, we produced ourselves entry-like constructions out of the information available in the conversion results and compared them to entries of existing dictionaries.

The results are encouraging: we reproduce here the entry which could be constructed out of the information available s.v. *Ader* in the German → French conversion output:

Ader in the German ⟶ French conversion output:

Ader
(bloedvat) *veine*
(doorgang in de aardkorst) *fontaine, point d'eau,* (stroompje) *ruisselet*
(laag in de aardkorst) *veine, filon*
(kronkelige streep) *veine*
(draad in een kabel) *conducteur*
(biol.) *nervure*
◊ **zur Ader lassen**
(bloed aftappen) *saigner,* (wet.) *phlébotomiser*
((afzetten) *écorcher*
◊ **poetische Ader, dichterische Ader** *inspiration poétique, veine poétique, verve (poétique)*

What is missing in this entry, besides the reading «**Verkehrsader**», present in all dictionaries[11] used for the comparison, are collocations: only *poetische Ader, dichterische Ader* are present. We will come back to this problem later; it has to be seen in connection with the fact that only the information packages 1, 2 and 3, but not the

11. The same ones as indicated above: Sachs/Villatte, Weiss/Mattutat, Bertaux/Lepointe.

collocational and contextual information, have been converted. Improvements could easily be reached by also converting information from package 4 (collocations and examples). Part of the items of this information package which appear in the Van Dale bilingual dictionaries are inherited from the monolingual dictionary. In those cases where they are inherited «in parallel» by two languages, the conversion could apply without problems, since:

- every collocation or context in the Dutch → X dictionaries appears in both languages, the Dutch part being identical for all language pairs;
- every collocation or context is related with exactly one sense and can thus be unequivocally related with a sense under conversion.

If the reliability of the results arrived at with the methods currently implemented is already decent, we can hope that a more sophisticated software would allow one to increase further the usefulness of the conversion results.

4. Limitations and problems of the dictionary conversion

There are practical as well as theoretical limitations to the use of the conversion system. The practical problems (which will be discussed first) have something to do with the consistency of dictionary coding, the use of abbreviatory conventions in dictionary texts and the reflexion of user-specific standard assumptions in the raw material used as an input to the conversion.

The theoretical problems are related with the discussion about interlingual systems and «directionality» of bilingual dictionaries.

4.1. Practical limitations

4.1.1. Multiword lists

The results of our analysis show that the current implementation of the conversion routines is word-form oriented. It can handle lists of items in a straightforward way. For example, the Dutch → French entry s.v. *misgreep* («blunder, mistake») contains the equivalent *bévue* and pointers to *erreur, maladresse* and *gaffe*. The Dutch German dictionary gives the equivalents *Missgriff, Fehlgriff* for *misgreep*. The French → German conversion result will propose for each French item of the «meaning» **[misgreep]** all of the German equivalents:

bévue	**[misgreep]**	→	*Missgriff, Fehlgriff*
erreur	**[misgreep]**	→	*Missgriff, Fehlgriff*
gaffe	**[misgreep]**	→	*Missgriff, Fehlgriff*
maladresse	**[misgreep]**	→	*Missgriff, Fehlgriff*

This procedure supports the «inheritance» of information from related items. Another example of this kind of «inheritance» of information is illustrated by the

conversion of the Dutch word *garage* and its synonyms. The Dutch monolingual as well as the Dutch → X bilingual entries s.v. *garage* indicate two readings which give rise to the following situation:

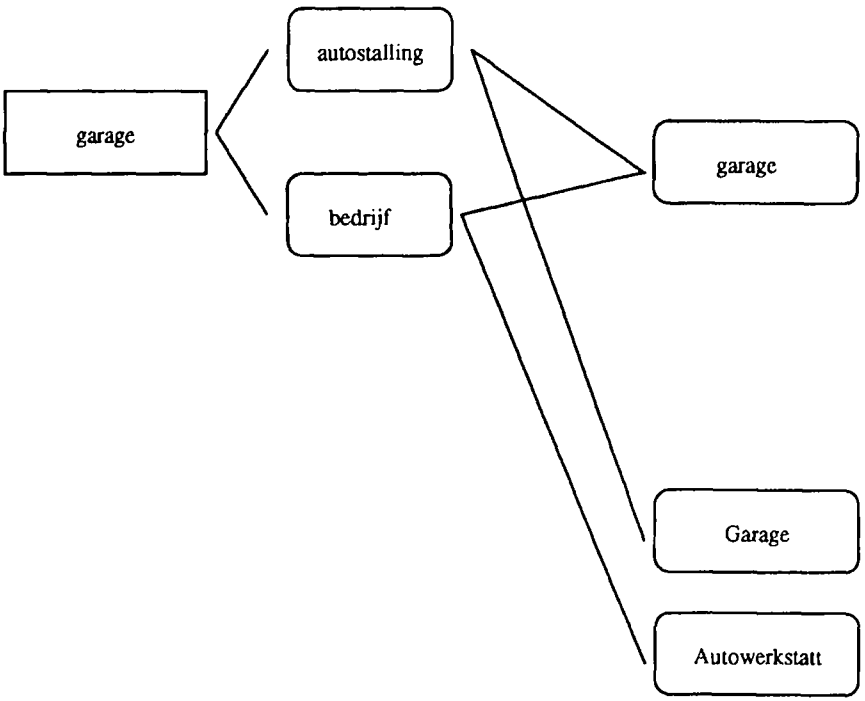

Figure 7: Equivalents of the Dutch word *garage*

The entry of the Dutch word *stalling* distinguishes four meanings (see above, in Figure 1), two of which are translated into French as *garage*. The French entry *garage* is therefore related with four concepts, for each of which German equivalents exist:

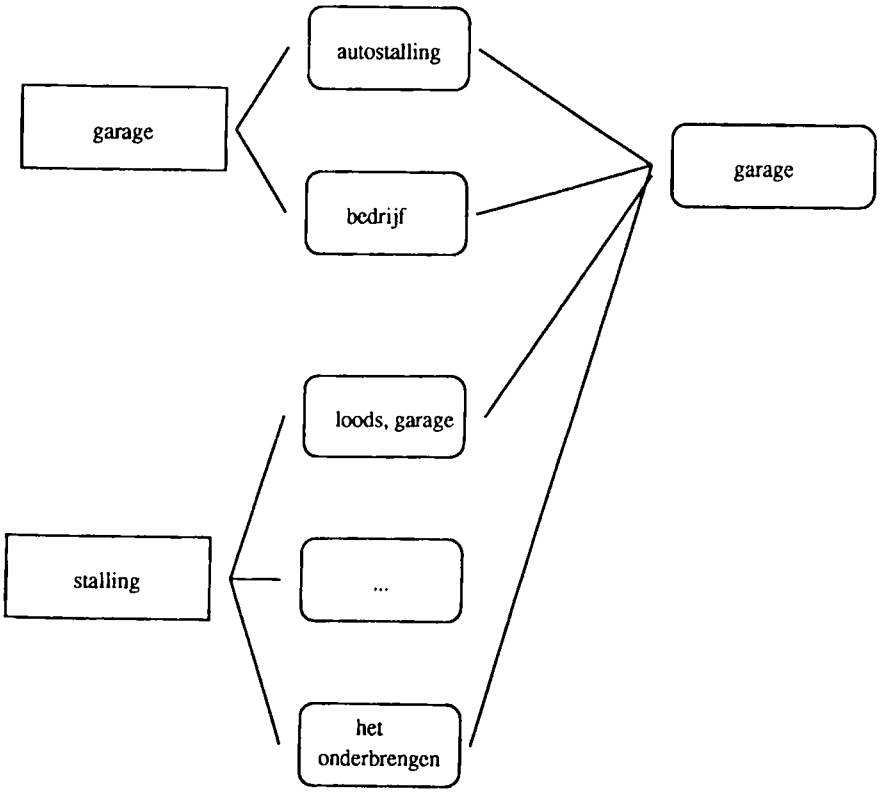

Figure 8: Equivalents of the Dutch word *stalling*

And inheritance from synonyms is still possible, since French *garage* has synonyms, such as *hangar, remise, dépôt*, which themselves have German equivalents.

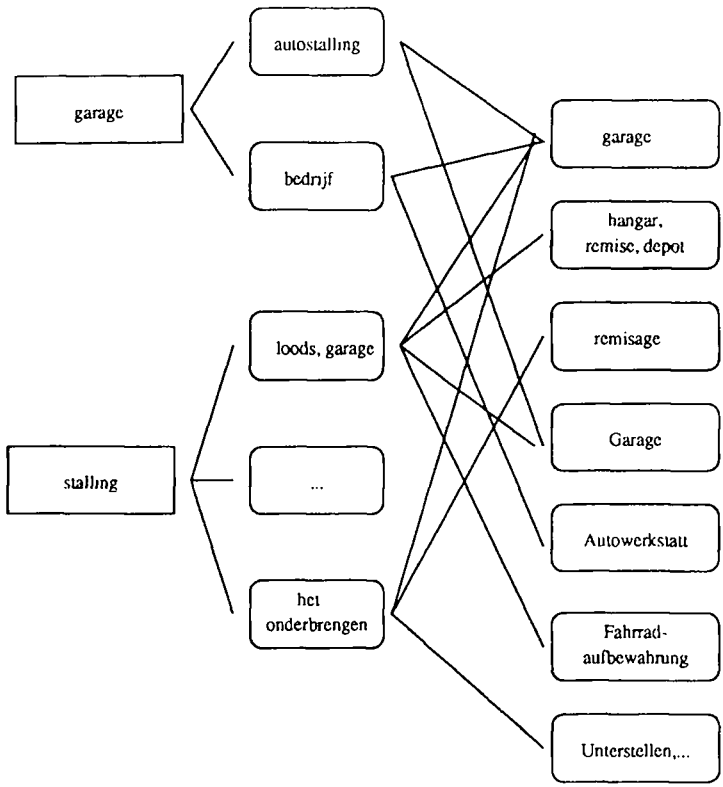

Figure 9: The French word *garage* inherits from Dutch synonyms

In these cases, the underconstrained conversion mechanism, which takes each word form as an entry point for a sorted list of conversion results, is quite useful. By navigating within the networks of Dutch lexemes, meaning descriptions and foreign language equivalents, information can be accumulated, and especially near-synonymous and equivalence relations can be tentatively established.

This underconstrained mechanism is much less useful, however, when it comes to the conversion of multiword lexemes of any type. They are not recognized for the moment as building *one* lexical unit. Consequently, all information produced by the conversion appears under each of the elements of a multiword unit. This is not problematic as long as true noun compounds or even collocations are concerned: the lexicographer gets for free a reminder allowing him to include or to point to a multiword unit in each of the articles of the components of such a unit. An example is the inclusion of *mettre sur une voie de garage* in the material sorted s.v. *garage*.

This device is problematic, however, when it comes to cases where only paraphrases exist. Each word of the paraphrase will receive, during the conversion, the information pertaining to the element for which the paraphrase stands. This drawback could be avoided by introducing a specific marking convention for paraphrases and for multiword units.

4.1.2. Consequences of the use of text-condensing devices

Most lexicographers make use of a variety of text-condensing devices. One of them, used in the dictionaries which serve as an input to the conversion, is the truncation of noun compounds, if two or more of them are listed which have the same head. The entry s.v. *afbetalingssysteem* may serve as an example: it indicates three equivalents, two of which are truncated:

(2) afbetalingssysteem (het) 0.1 *Abzahlungs-, Ratenzahlungs-, Teilzahlungssystem* (o.11).

Since the conversion routine does not reconstruct the full forms of the respective lexemes, the truncated forms are used in the synopsis and in the lists of conversion results. This could be avoided by an expansion of truncated forms before the conversion.

Another type of condensing device is the use of generalizations in the dictionary entries: the information given in package 2 is supposed to be relevant for all «senses» distinguished at the level of the information package 3. Ideally, the conversion routine should incorporate a device allowing it to inherit the syntactic information present in package 2 downwards to each «sense».

4.1.3. Consequences of user-orientation of the dictionary

The conversion operates on dictionaries which have been conceived for a Dutch public.[12] Consequently, the linguistic properties of Dutch lexemes are supposed to be well known to the user of the Dutch → X dictionaries. The dictionary authors have taken this situation into account in the description of the syntactic properties of lexical items, specifying a «default rule», according to which only those foreign language items differing from their Dutch equivalents are explicitly marked for their syntactic properties. To take an example from verbal construction[13] the Dutch → French entry (in Al *et al.* (1985)) s.v. *bereiken* **[aankomen in]** (3) appears as follows:

bereiken ⟨ov.ww.⟩ **0.1** [aankomen in] ***atteindre*** ⟨→t58⟩ ⇒*arriver à, gagner, parvenir* ⟨→t32⟩ **0.2** [komen tot] ***atteindre*** ⇒*parvenir à, arriver à*, ⟨schr.⟩ *atteindre à* **0.3** [contact krijgen met] ***atteindre*** ⇒*joindre* ⟨→t58⟩, *toucher* **0.4** [mbt. een leeftijd] ***atteindre*** ⇒*parvenir à* ◆

12. This is at least the intention of the dictionary series. Hausmann (1988) argues about this in his discussion of the treatment of collocations. His point is that the collocational description is equally useful for different groups of users, at the price of some redundancy.

13. The problem becomes most evident when it comes to the description of syntactic construction properties of verbs (see Heid (1990) for details).

It specifies the preposition to be used with *arriver* but does not specify anything for *atteindre* and *gagner*.[14] These two verbs are taken to behave in a structurally isomorphic way to *bereiken*:

- *een bestemming bereiken*
- *atteindre une destination*
- *gagner une destination*

Problems may occur, however, when the Dutch description disappears, as in the conversion. Then no fixed point is available with respect to which the default can be evaluated. This situation occurs in the Dutch → French → German system each time one language behaves in parallel with Dutch but the other doesn't. In this case the «exception» will be marked. The marking will be carried through the conversion, but so will the unmarked description will be carried through also. In our example, in the worst case, *arriver à* will be marked but *erreichen* will not; does the lack of an indication then mean that *erreichen* takes a preposition, as *ankommen* does?

When constructing a bilingual dictionary which will translate the marked case into the unmarked one, it will not be clear *with respect to which default* the item is unmarked. For sure, *not* with respect to the marked case. Syntactic construction properties are not the only case of the use of defaults: the same device is used throughout the dictionary for usage marks (e.g. «old use», «informal», etc.).

Interestingly, here the use of defaults and dependencies between items of different languages cause problems of interpretation. If all items of all languages were explicitly described without depending on other items, no problem would arise.

This treatment in the Van Dale dictionaries is in some sense contrary to Martin/Al (1988)'s proposal of keeping the underlying lexicographic data description free from requirements of individual applications. They argue that user-specific requirements should only be kept track of in «front end dictionaries» derived from an underlying database. This database should be application-independent. We agree completely with this claim and add a further requirement: since «leaving out what everybody knows» is dependent on the user of the intended application, the underlying database should not use defaults or any implicit descriptions, but should explicitly give access to all relevant information, even if it is then slightly more redundant.

4.2. Theoretical limitations

We now discuss general limitations which are inherent to the way how the conversion procedure is designed, and, more far-reaching, inherent to the concept of translation underlying the whole Van Dale series of bilingual dictionaries.

We will first discuss the problem of equivalence gaps and then come to a more general discussion of the «interlingua» problem.

14. The lack of a preposition after *parvenir* is an error.

4.2.1. A concrete problem: equivalence gaps

The starting point of the conversion procedure, as of all Dutch → X bilingual dictionaries, is the inventory of Dutch lexemes contained in (Van Sterkenburg/Pijnenburg 1984). Some of these lexemes have lexicalized equivalents in other languages; some may have lexicalized equivalents only in some of the languages under consideration, or even in none of them. The Dutch → X dictionaries then give paraphrases in the target language. This is rather frequent with germanic derivatives which are lexicalized in Dutch and German, but not in French. Here are some examples.

NL: spelbreker;
GER: Spielverderber;
FR: personne qui gâche le plaisir des autres.

NL: houdbaar;
GER: haltbar; lagerfähig;
FR: qui se conserve; qui se garde (eetwaren).

There are, however, also cases where neither French nor German has a lexicalized equivalent:

NL: toonvast;
GER: den Ton haltend;
FR: qui garde, tient le ton.

NL: bokpaal;
GER: A-förmiges Gestänge;
FR: poteau télégraphique en forme de A.

These cases are irrelevant for the bilingual French ↔ German dictionaries, since nobody ever would search them.

Another case, for which we cannot give examples here, occurs where one of the other languages has a lexeme, and *Dutch* has not. These cases will never be part of the multilingual Van Dale database, since it has been produced on the basis of Dutch lexemes. This shortcoming might be a serious problem; we cannot quantify it for the moment. This problem leads to the more general question of which concepts are lacking in the multilingual description and how they can be added. A practical solution which of course leaves untouched the theoretical question of what the status of the Dutch interlingua is, would be to run an automatic comparison whith the macro-structure of Van Dale dictionaries with other languages than Dutch as a source language.

4.2.2. Problems of an «interlingual» representation

Here we do not take up the discussion about the advantages and problems of inter-

15. See e.g. Tucker (1987), Nirenburg (Ed.) (1987), etc.

lingual approaches to (machine) translation.[15] As a matter of fact, the dictionary architecture which is the basis of the conversion experiments, is very close to that of an interlingua-based system. It shares with true interlingua systems one of their major problems: the distinctions available at the conceptual level may turn out to be insufficient for covering all of the distinctions needed for a bilingual dictionary. This implies a certain amount of imprecision in critical cases: it may happen that the bilingual dictionaries indicate two or more equivalents which are not exchangeable, and that they do not specify usage conditions. Making use of the collocational and contextual information may be helpful to bridge some of the gaps which may occur.

The approach to bilingual dictionaries which underlies the Van Dale dictionaries, essentially starts from facts and situations Dutch-speaking people want to express in foreign languages. It does not lay much emphasis on the constraints which in translation may originate from properties of the target language. The Danish group of lexicographers has recenty advocated a strong orientation of bilingual dictionaries towards constraints from the target language.[16] A combination of both approaches is possible as far as syntactic properties of target language items are concerned, through a combination of detailed monolingual descriptions with a «concept-based» approach such as that of Van Dale. This leaves the question open, however, of how detailed the interlingual representation has to be in order to cover semantic distinctions of the target language.

The bilingual dictionaries which are used as input to the conversion, sometimes, give particular usage indications: for example in the Dutch → French entry for *stalling* 0.1., *dépôt* is restricted to buses, trains, tramways, etc. It would of course be technically possible to add such usage restrictions manually each time when the conversion leads to lists of equivalents. The possibilities of inheriting information from synonyms, within the conversion process, produce in many cases a relatively large number of equivalents. The sheer number of equivalence candidates, however, need not be a major advantage of a bilingual dictionary; this richness is only really valuable when it is possible to distinguish between the candidates, on the basis of clear-cut criteria.

5. Conclusion - relation to other approaches

In this paper, we have tried to comment on an analysis of results of the dictionary conversion procedure proposed by Al (1988).

We have described its methodological basis which borders architectures for interlingua-based machine translation dictionaries. But no machine translation system so far, almost none of the monolingual or multilingual natural language generation systems which are based on conceptual descriptions and relations between concepts and lexical items, have ever treated statistically relevant quantities of items. The present preliminary study or a more detailed analysis in its sense would probably allow to assess somewhat more realistically the chances of a concept-based dictionary architecture than would the experience gathered up to now in small-scale domain-dependent NLP applications. Our statistical results as well as the quality of the output are rather encouraging.

16. Cf. Kroman (1989) and Kroman/Riiber/Rosbach (1989).

Some problems are due to the fact that dictionary encoding sometimes does not obey strict rules governing the use of the different types of indications in an entry; similarly, all presentation conventions used in lexicographic practice in order to save space, require an intelligent cooperative user and do not allow for straightforward electronic processing. The dictionaries of the Van Dale series are remarkably highly standardized, and still the use in the conversion shows some (minor) deficiencies at this level. This especially concerns the separation between the basic lexicographic description (what Martin/Al (1988) call «underlying database») and the user-oriented application-specific presentation of the descriptive results in an actual dictionary.

In recent discussions about descriptive linguistic resources for natural language processing, emphasis has been laid on the aspect of re-using existent dictionaries; it has become clear that sizeable dictionaries are needed for any application of natural language processing, and that dictionary compilation is among the most costly and time-consuming activities in this field.

Al (1988) shows how it is possible, starting from explicitly «concept-based» bilingual dictionaries, to obtain material which could serve as an input to new bilingual dictionaries. There is no doubt that these new bilingual dictionaries will be «implicitly concept-based»; technically, it is no problem to carry over most of the features of the Van Dale dictionaries, and the conversion will thus lead to Van Dale-type bilingual dictionary material. It is even possible to carry over part of the collocational information contained in Van Dale's dictionaries.

But still, from the point of view of a re-use within natural language processing, the fact that the input to the conversion routine is not completely explicit (use of defaults, implicit inheritance of features from information package 2 to all «senses» or information package 3, etc.), may constitute a minor technical obstacle. The orientation towards the «concept-based» approach means that the conversion results would be best suited for an «interlingua-based» application. From this point of view, an experiment on a larger basis would be most promising; the structure of Van Dale entries could easily be modelled, and the (semi-automatic?) addition of some more syntactic information should allow for more detailed experimentation.

References

AL, Bernard P. F. (1983): «Dictionnaire de thème et dictionnaire de version», in: *Revue de phonétique appliquée*, 66-68 (Mons, Belgien) 1988: 201-211.

AL, Bernard P. F. *et al.* (1985): *Van Dale groot woordenboek Nederlands-Frans* (Utrecht/Antwerpen: Van Dale) 1985.

AL, Bernard P.F. (1988): «Langue source, langue cible et métalangue», in: Landheer (Ed.) (1988): 15-29.

ATKINS, Beryl T., DUVAL, Alain, MILNE, Rosemary C (1987): *Robert & Collins Dictionnaire Français-Anglais, Anglais-Français* (Paris: Le Robert/Glasgow: Collins) 1987.

BERTAUX & LEPOINTE: *Dictionnaire Allemand-Français, Français-Allemand* (Paris: Garnier)s.d.

COX, H. L. *et al.*: *Van Dale groot woordenboek Nederlands-Duits* (Utrecht/Antwerpen: Van Dale) 1986.

GOETSCHALCKX, J., ROLLING, L. (Eds. 1982): *Lexicography in the Electronic Age*, Proceedings of a Symposium held in Luxemburg, 7-9 July, 1981 (Amsterdam: North-Holland) 1982.

GOLAN, Igal, LAPPIN, Shalom, RIMON, Mori: «An Active Bilingual Lexicon for Machine Translation», in: *Proceedings of COLING-88, Bonn* (Bonn: IKP), 1988: 205-211.

HAUSMANN, Franz Josef (1988): «Grundprobleme des zweisprachigen Wörterbuchs», in: Hyldgaard-Jensen/Zettersten (1989).

HEID, Ulrich (1990): «Syntactic information in (machine) translation dictionaries — towards a modular architecture for bilingual dictionaries», to appear in: Hyldgaard-Jensen/Zettersten (1991).

HYLDGAARD-JENSEN, Karl, ZETTERSTEN, Arne (Eds.) (1989): *Symposium on Lexicography III* (Tübingen: Niemeyer), 1989.

HYLDGAARD-JENSEN, Karl, ZETTERSTEN, Arne (Eds.) (1991): *Proceedings of the fifth international Symposium on Lexicography,* København, May 3-5, 1990 (Tübingen: Niemeyer), to appear 1991.

KROMANN, Hans Peder (1989): Neue Orientierung der zweisprachigen Wörterbücher, in: Mary Snell-Hornby/Esther Pöhl (Eds.): *Translation and Lexicography, Paintbrush*, 1989.

KROMANN, Hans-Peder, RIIBER, Theis, ROSBACH, Poul (1984): «Active and Passive Bilingual Dictionaries: The Scerba Concept Reconsidered», in: *Proceedings of Lexeter-83, Bd. II* (Tübingen), 1984: 207-215.

KROMANN, Hans-Peder, RIIBER, Theis, ROSBACH, Poul (1989): Principles of bilingual lexicography, in: Franz-Josef Hausmann, Oskar Reichmann, Herbert-Ernst Wiegand, Ladislav Zgusta (Eds.): *Dictionaries, Dictionnaires, Wörterbücher, Ein internationales Handbuch* (Berlin: de Gruyter) 1989.

LANDHEER, Ronald (Ed.) (1988): *Aspects de linguistique française, Hommage à Q.I.M. Mok* (Amsterdam: Rodopi) 1988.

VAN DER LINDEN, Erik-Jan (1988): «The Representation of Lexical Objects», in: *Proceedings of BudaLEX'88.*

MARTIN, Willy, AL., Bernard P.F. (1988): «User-orientation in dictionaries: 9 propositions», in: *Proceedings of BudaLEX'88.*

MARTIN Willy (1989): «Over computationele lexicologie», in: *Handelingen van het 40ste Nederlands Filologencongres*, 1989.

NIRENBURG, Sergei (Ed.) (1987): *Machine Translation — Theoretical and methodological issues* (Cambridge: Cambridge University Press) 1987.

NIRENBURG, Sergei (1989): «KBMT-89. Project Report.», ms. (Pittsburg, Pa: Center for Machine Translation, Carnegie Mellon University) 1989.

SACHS & VILLATTE: *Wörterbuch Deutsch-Französisch* (Stuttgart: Klett).

SACHS & VILLATTE: *Wörterbuch Französisch-Deutsch* (Stuttgart: Klett).

SCERBA, L. V. (1940): «Versuch einer allgemeinen Theorie der Lexikographie», in: Wolski (1982): 17-62.

VAN STERKENBURG, Piet, MARTIN, Willy, AL., Bernard P.F (1982): «A new Van Dale project: bilingual dictionaries on one and the same monolingual basis», in: Goetschalckx/Rolling (Eds.) (1982): 221-237.

VAN STERKENBURG, Piet, PIJNENBURG, Willem J.J.: *Van Dale groot woordenboek van hedendaags Nederlands* (Utrecht/Antwerpen: Van Dale) 1984.

TUCKER, Allan B. (1987): «*Current Strategies in Machine* Translation Research and Development», in: Nirenburg (Ed. 1987): 22-41.

WEISS & MATUTTAT: *Wörterbuch Französisch-Deutsch, Deutsch-Französisch* (Stuttgart: Klett).

WOLSKI, Werner: *Aspekte der sowjetrussischen Lexikographie* (Tübingen: Niemeyer), 1982 (= Germanistische Linguistik, 43).