

A new dictionary model for closely related languages: The Dutch-Afrikaans dictionary project as a case-in-point

Willy MARTIN and Rufus GOUWS,

Amsterdam, The Netherlands and Stellenbosch, South Africa

Abstract

In this paper a new model and approach is presented for the lexicographic linking and presentation of two closely related languages. As a case-in-point Dutch and Afrikaans are taken.

The proposed model differs from the traditional bilingual dictionary in that the type of word pair determines the presentation and treatment, e.g. *tafel / tafel (table)* will be treated differently from *aanschuiven / aanskuif (shuffle along, draw up)* because of, among others, formal characteristics. Semantic and pragmatic differences will also come into play.

Although in the database itself the frame for the datacategories remains the same all over, the actual representation of the data in the dictionary, will be determined by the type of word, the type of user as well as the type of word pair. With this new model, applicable beyond the Dutch-Afrikaans case itself, the authors hope to stimulate research in this neglected domain of bilingual lexicography.

1 The problem

Bilingual dictionaries can be regarded as one of the most commonly used lexicographic types in monolingual and multilingual societies. In spite thereof the history of metalexigraphy and dictionary research indicate a limited amount of research directed at this dictionary type. More in particular, in the research that has been undertaken up till now little attention was given to the problems arising when two closely related languages are treated in a bilingual dictionary. The aim of this paper is to identify and discuss a model formulated for the compilation of a bilingual dictionary with Dutch and Afrikaans, two closely related languages, as language pair. Although the focus will be on aspects of the lexicographic treatment of these two specific languages, the model and the suggestions made in this paper could also be used when planning a dictionary in which two other closely related languages, e.g. South African languages from the Nguni or Sotho family groups respectively, are treated.

2 The need

Afrikaans developed from seventeenth century Dutch but in its development other European languages with which the speakers of early Afrikaans came into contact, e.g. English, French, German and Portuguese, had a huge influence on this young language. This influence can still be detected in modern-day Afrikaans, cf. [Ponelis 1993]. Due to the constant language contact in the multilingual and multicultural South African society the indigenous languages of South Africa made a substantial contribution to the development of Afrikaans. This language contact situation still exists and in the new South Africa it has been formalised by means of

an official policy of multilingualism. Where the pre-1994 South Africa had two official languages, i.e. Afrikaans and English, South Africa now hosts eleven official languages. This in itself confronts the lexicographers with tremendous challenges and opportunities, cf. [Gouws 2000a]. As a developing language in this multilingual society Afrikaans is constantly the target of an ongoing process of language change and these changes occur primarily under the influence of English and, to a lesser degree, the indigenous African languages. As a result of both the constant language contact and the natural but isolated growth of Dutch and Afrikaans, the divergence between Dutch and Afrikaans has become much more salient over the last few decades.

The new political situation in South Africa has revitalised relations with Holland and Belgium resulting in a growing demand for Afrikaans literature in Holland and Belgium. The last few years saw an increasing need for the translation of Afrikaans texts to Dutch because of the problems Dutch and Flemish speakers encounter when reading Afrikaans texts. Both the translators and the Dutch readers choosing the original Afrikaans texts have expressed their need for a bilingual dictionary with Dutch and Afrikaans as language pair. In South Africa the interest in Dutch art, culture and literature is expanding on both an academic and a popular level, but the process is impeded by the lack of a bilingual dictionary coordinating these two closely related languages.

In the past Afrikaans and Dutch lexicographers have failed to make any serious endeavours to compile bilingual dictionaries with Dutch and Afrikaans as language pair. Their arguments have always been that these two languages are so closely related that no dictionary of this kind is needed. In a way they were right, in the sense that no traditional bilingual dictionary would have been adequate for the situation at hand. Neither were the word lists containing false friends and language specific items. During 1998 ZASM (Zuid Afrikaanse Spoorwegmaatschappij), a Dutch Foundation working for better relations between the Netherlands and South Africa, had identified the need for a more adequate instrument and consequently commissioned the compilation of a pilot and feasibility study to establish the extent and nature of the need. The pilot study, done during 1998 and 1999 cf. [Martin et al 1999a], re-iterated the need and made some suggestions regarding the overall lexicographic process. It soon became clear that this dictionary would have to be the result of an innovative lexicographic approach which makes provision for ways to treat both the similarities and the differences between Dutch and Afrikaans in a scientifically founded way. The pilot study therefore included the formulation of a theoretically based model to guide the lexicographers in this compilation process. This paper will give an account of some of the major features suggested for this dictionary and of ways and means to achieve them.

3 The underlying database

In line with the recommendations formulated in a recent feasibility study carried out by order of the European Commission (DG XIII) on the construction of bilingual dictionaries, cf. [Martin/Heid 1998], the proposed dictionary will be based on underlying monolingual databases. The starting-point is a Dutch database, the so-called RBN (Referentie Bestand Nederlands = Reference Database Dutch), cf. [Martin et al. 1999], to which Afrikaans lexemes are linked. This 'linkage' implies two things. Firstly, both the Dutch and Afrikaans items are not mere

Form Units (FUs), but lexemes, Lexical Units (LUs) [forms + specification of their meaning, e.g. in the form a resume, cf. [Martin/Heid 1998]. Secondly, the relationship between the LUs in both languages is specified by means of, among other things, semantic and pragmatic parameters, cf. [Martin/Tamm 1996] for a full treatment, so that a reversal calculus can take place. The advantages of this approach are the following:

- Starting from one monolingual lexical database (the RBN), a second one (the RBA = Reference Database Afrikaans) is created during the linking process, although the latter one needs, of course, some post-editing.
- The explicit way in which the linking is carried out, does not only guarantee an A-B part (Dutch-Afrikaans in our case), but a B-A part as well. This reverse can be generated automatically, in a very accurate way, be it with the same proviso as for the preceding point, viz. that some kind of post-editing is unavoidable.
- Given the fact that linkable reference lexicons are constructed makes it feasible to use the resources in a hub-and-spoke configuration. This model has been presented for the first time in [Martin 1996]: it stipulates the conditions under which two or more languages (spokes) can be linked to a common one (hub), so that, automatically, links between the spokes can be derived. In a multilingual context such as the EU and South Africa each with eleven official languages, this Hub-and-Spoke Model becomes an attractive means to solve the (bilingual) dictionary needs.
- Above all, however, the approach where a differentiation is made between an underlying database and a front-end one (a dictionary) opens the possibility to derive from the fundamental non-user oriented underlying database as many user-oriented front-end databases (dictionaries) as there are types of dictionaries, cf. [Martin/Al 1988]. In our case this means that, from the two underlying lexical databases (RBN and RBA) linked to each other, we cannot just derive a 'classical' bilingual dictionary, but more in particular, one which takes into account the specific relationship between the two languages and is adequate to carry out the functions aimed at by the target groups.

In what follows first a brief characterisation will be given of the planned dictionary, followed by its most prominent and innovative features on both macro- and microlevel.

4 The dictionary profile

The planned dictionary should take into account the close relationship as well as the differences between Dutch and Afrikaans. These differences do not only lie on the level of words and denotation, but can also reside in the fact that 'cognate' languages show differences on the level of co-text (combinations) and context (pragmatics, connotation). The Dutch and Afrikaans items *tafel* (*table*) are identical in form and meaning but differ on the level of co-text, e.g. the Dutch form combines with the verb *afruimen* and the Afrikaans form with the verb *afdek* (*to clear a table*). The word *kar* (*car*) has the same reference in Dutch and Afrikaans. In Afrikaans it is

used in a neutral register whereas in Dutch it belongs to a very informal register. The combinatory possibilities also differ. In Afrikaans it is used with the particle verbs *inklim/uitklim* (*to get in/out*) whereas in Dutch it is used with the particle verbs *instappen/uitstappen*. Differences of this nature have not received ample attention in existing sources. The dictionary should furthermore take into account that the target users are Dutch, respectively Afrikaans speaking groups, that as a rule, *understand* each other's language relatively well in normal general conversation; however, they cannot properly *produce* each other's language without making many errors.

To comply with its genuine purpose this dictionary should therefore be:

- contrastive (on the level of words, their meaning as well as their combinations and use);
- multifunctional, serving the needs of both decoding and encoding Dutch and Afrikaans users;
- well-structured and informative, without being redundant.

Given this profile, the most characteristic and innovative features of this proposed dictionary will now be specified.

5 The macrostructure

5.1 A single amalgamated central list

One of the most striking features of the new Dutch-Afrikaans dictionary, in comparison with other bilingual dictionaries, is the fact that there is but one, single, amalgamated, alphabetical macrostructure, motivated by the close relationship between the two languages. This embedding of two macrostructures is the result of an innovative process of outer textual condensation, for an explanation of the term, cf. [Wiegand 1996]. The macrostructural similarity forms as it were the starting-point for the description of the two languages. Such a point-of-view facilitates the contrastive description greatly: cognates are kept together in one article, 'false cognates' are entered as separate lemmata albeit in close vicinity, meaning extensions (in either or both languages) become immediately clear. Moreover, redundancy is kept to a minimum: one does e.g. not have to repeat in an Afrikaans-Dutch volume in the article of the lemma *tafel* (*table*) what one has already mentioned under the same lemma in a Dutch-Afrikaans volume.

This comprehensively condensed macrostructure will necessarily influence the data distribution structure, cf. [Bergenholtz et al 1999], of the articles. The macrostructural embedding due to the process of textual condensation results in the occurrence of combined lemmata, consisting of both a Dutch and an Afrikaans lexical item as treatment unit. Although a combined lemma as such functions as head of the article, both its components have a dual function. The Dutch form, typically the first component of the majority of combined lemmata, functions as source language entry and as guiding element determining the alphabetical ordering of the combined lemma within the relevant article stretch. The Afrikaans form, usually the second component of a combined lemma, functions as target language entry but also as translation equivalent of the preceding source language entry.

5.2 Different lemma types

This new approach will lead to the inclusion of articles headed by different lemma types, e.g.

- Absolute cognates (identical form and meaning), e.g. Dutch *tafel* / Afrikaans *tafel* (*table*) will result in a condensed combined lemma, consisting of the Dutch form *tafel*, entered as guiding element of the article, and a structural marker, the symbol "#", substituting the Afrikaans lemma and indicating that the Afrikaans item is identical in form and meaning to the Dutch. The condensed combined lemma will be entered as *tafel/#*.
- Absolute cognates with systematic morphological difference, e.g. Dutch *ontsnappen*/Afrikaans *ontsnap* (*escape*), will appear as combined lemmata with the Dutch form functioning as guiding element: *ontsnappen/ontsnap*. Besides its occurrence as second component of the combined lemma, the Afrikaans form will also be entered in its relevant alphabetical slot as the guiding element of a reference article which includes the Dutch form as the reference address entry: *ontsnap* → *ontsnappen*.
- Equivalents, which display no formal identity, will be entered as two separate combined lemmata with both the Dutch and the Afrikaans form functioning as the respective guiding element, e.g. *verkeersdrempel/spoedwallejtjie* and *spoedwallejtjie/verkeersdrempel* (*speedbump*). The Afrikaans form will not only be entered as guiding element of a reference article but, in compliance with the norms of user-friendliness, will introduce an article in which a full treatment of the combined lemma is presented. This diminishes the extent of the mediostructural procedures and prevents a totally Dutch-biased macrostructure.
- Partial cognates differing in polysemy: polysemy is a language specific feature and although a relation of formal identity may prevail between a source language item and a translation equivalent applicable for a given co-text, it does not imply that this target language lexical item will display the same polysemous paradigm as the source language item. The lemmatisation of polysemous Dutch lexical items with different target language forms for their different polysemous senses will lead to an additional lemmatisation procedure.

The Dutch form will be entered as guiding element and the treatment will include the presentation of a subcomment on semantics for each one of the polysemous senses. The search zone of each subcomment on semantics will be indicated by a numerical polysemy marker, followed by a brief paraphrase of meaning to identify the relevant sense (given in square brackets).

The Afrikaans form to be used as a translation equivalent for the given sense of the Dutch form, is then presented, cf. the entry *toilettring* in the second subcomment on semantics of the article of the Dutch lemma sign *bril* (given below). As part of the treatment unit this entry forms part of a remote condensed combined lemma. Where formal identity prevails for a given sense between the Dutch and Afrikaans forms the Afrikaans lemma component is substituted by a hash sign (#), cf. the first subcomment on semantics of the article of the Dutch lemma sign *bril* (given below). For each subcomment on semantics

the first component of the combined lemma remains the same, but the second component may change from sense to sense.

The Dutch and Afrikaans forms constitute a combined lemma because together they function as the primary treatment unit of the article.

The Afrikaans forms included as part of the presentation of remote condensed combined lemmata will also be entered in their relevant alphabetical slots as the guiding elements of either reference articles, with the Dutch form as the reference address entry, or articles displaying a fully-fledged treatment. Reference articles will guide the user from an Afrikaans form to its partial cognate, e.g. Dutch *bril* (*spectacles*) has a wider paradigm of polysemous senses than its Afrikaans cognate *bril*, albeit that the two words display absolute formal identity. The Afrikaans word will be entered as the guiding element of a reference article leading the user to the relevant entry in the subcomment on semantics of the article of the Dutch lemma sign : *bril* → *bril 1*. The Afrikaans form *toiletring* will be entered as a guiding element of an article headed by a combined lemma, which functions as the address of an extensive lexicographic treatment.

<p>bril 1 [om te kijken] # ... (to look) 2 [van w.c.] <i>toiletring</i> ... (of a wc) bril → bril 1 toiletring / bril (bril 2)</p>
--

To assist the target user on the search path of the external access structure typographical structural markers will identify the lemmata as belonging to a specific member of the treated language pair. In this multifunctional dictionary the amalgamated macrostructure creates a dictionary consultation environment where both differences and similarities become apparent in an efficient and contrastive way.

5.3 False friends

False friends (complete and partial) confront the users of closely related languages with communication problems. The amalgamated macrostructure proposed in this dictionary model restricts the extent of this problem due to the lemmatisation (in close vicinity) of both members of a pair of complete false friends, complemented by the inclusion of combined lemmata with the respective equivalents of the false friends as guiding elements. A relation of false friends exists between the Dutch and Afrikaans lexical items *mus*. The Dutch form (= English *sparrow*) is the guiding element of a combined lemma with the Afrikaans form *mossie* as second component. The Afrikaans item *mus* (*cap*) is also entered as the guiding element of an article with a combined lemma *mus/muts*. Likewise the respective Afrikaans and Dutch forms *mossie* and *muts* will be entered in the relevant article stretches as the guiding elements of combined lemmata: *mossie/mus* and *muts/mus*. This presentation emphasises the lack of semantic identity in spite of the formal identity holding between the members of a pair of complete false friends. To assist users with a rapid identification of false friends the relevant lemmata will be accompanied

by a non-typographical structural marker, i.e. "@". Partial false friends are accounted for on microstructural level by means of lexicographic commentary.

6 The microstructure

6.1 Focusing on similarities and differences

The innovative features of this dictionary model include various ways and means to illustrate similarities and differences between the members of the language pair. This will be done in such a way that the target users can rely on the dictionary as a functional instrument in the enhancement of communicative equivalence between Dutch and Afrikaans. The text book structure as well as the text list structure, the article structure, the different search zones and the article internal data distribution structure will work towards this goal.

6.2 Semantic data

Although this dictionary primarily falls within the typological category of translation dictionaries a brief paraphrase of meaning is entered in each subcomment on semantics in the article of a lemma sign representing a polysemous lexical item. The article of the lemma sign *dageraad* (daybreak) which represents a polysemous Dutch lexical item, includes the following entries:

"1. [aanbreken van de dag]... " (the breaking of the day)... "2. [begin]... " (start) ...

From a typological perspective the brief paraphrase of meaning may be seen as giving the dictionary a hybrid character. In terms of the genuine purpose of this dictionary it is a functional text segment in the subcomment on semantics because it assists the user in distinguishing the different senses of the source language item to ensure the correct choice of a translation equivalent.

In traditional translation dictionaries the translation equivalent is usually regarded as the most salient entry in the comment or subcomment on semantics, cf. [Gouws 1996]. In this dictionary, however, the entry functioning as translation equivalent usually forms part of the combined lemma and therefore falls within the domain of the comment on form. A procedure of lemmatic addressing in the comment on semantics will consequently also have the translation equivalent as an address and will necessarily enhance the quality of equivalent discrimination, cf. [Gouws 2000]. Lemmatic addressing, albeit in a slightly untraditional way, will prevail in many articles. However, this dictionary will often deviate from a traditional lemmatic bias in order to establish a topic switch by promoting a variety of entries in the comment on semantics, e.g. collocations, other illustrative examples as well as idiomatic expressions, to function as secondary treatment units. In doing so the occurrence of source and target forms in co-text receives much more emphasis. To enable the user to distinguish between these entries different search zones will be allocated to them and each search zone will be indicated by means of a non-typographical structural marker.

6.3 Combinatory Data

Focusing on the occurrence of source and target forms in co-text, the articles of this dictionary includes a variety of combinatory data displaying, in a fixed order, different search zones

accommodating non-contrastive combinations, contrastive combinations and idiomatic expressions respectively. The combinatory data represents the core of the lexicographic presentation.

For non-contrastive combinations only the Dutch example is entered, implying that Afrikaans has a similar collocation. This category of illustrative examples is preceded by the non-typographical structural marker "-". Where the Afrikaans form shows a difference which is regarded as contrastively relevant the translation is entered. This category is marked by a "•". Idiomatic expressions (i.e. idioms, proverbs, pragmatic formulae) are always provided with translations. This category is indicated by the structural marker "◇". In all these categories Dutch is the source language. If, however, the treatment of the Afrikaans component of the combined lemma demands the inclusion of an Afrikaans collocation or idiomatic expression not provided for by the Dutch entries, this Afrikaans form may be entered as a source language form and provided with a Dutch equivalent, translation or meaning paraphrase. Such a switch from Dutch to Afrikaans as source language is indicated by the structural marker "▷". The article of the lemma sign *bril* illustrates the use of combinatory entries:

<p>bril 1 [om te kijken] # – een <u>bril</u> hebben/dragen; hij heeft zijn <u>bril</u> niet op; ... • <inf.> een <u>bril</u> moeten 'n <u>bril</u> moet kry ... ◇ elk ziet door zijn eigen <u>bril</u> <i>elkeen kyk deur sy eie bril</i> deur 'n roze <u>bril</u> <i>kijken deur 'n rooskleurige bril kyk</i> ▷ <i>iemand 'n bril op die neus sit</i> iemand te grazen nemen 2 [van w.c.] <i>toiletring</i>...</p>
--

6.4 Lexicographic Commentary

Provision is made for the inclusion of lexicographic commentary in its own separate search zone which can occupy different positions within an article. The lexicographic commentary is used where contrastive differences lead to interpretation or usage problems, e.g. due to the occurrence of partial false friends, culture bound items or language-specific discourse markers. The nature of those types of lexicographic commentary which appear regularly will be indicated by means of a specific structural marker. The comment on partial false friends will be preceded by the marker "!" and the commentary addressed at culture bound items will be identified by the marker "%". To illustrate this: the article of the condensed combined lemma *toe/# (closed)* will include the following entry: "! In Afrikaans word *toe* ook gebruik waar in Nederlands *dicht* gebruik word". (In Afrikaans *toe* is also used where Dutch uses *dicht*.)

Lemmata representing culture specific words will generally not have a corresponding form in the other language to participate in a combined article. Such lemmata will be marked and their comment on semantics will include a brief paraphrase of meaning of the translation equivalent. Where an Afrikaans item is the guiding element of such an article the treatment will not be restricted to a reference address entry guiding the user to a Dutch form, e.g.

biltong % Gedroogde vlees dat . . . (dried meat that ..)

7 In conclusion

From the above it will have become clear that the model focuses on the fact that closely related languages like Dutch and Afrikaans do not display so many denotative differences but that the major differences are rather on the connotative and combinatory level. This necessitates innovative lexicographic strategies. An amalgamated macrostructure brings lemmata from both languages within the scope of a single access structure and puts the user in a position to differentiate between single and combined lemmata; between absolute cognates, partial cognates and translation equivalents; between reference articles and articles displaying a fully-fledged treatment and between non-contrastive combinations and those which are contrastively relevant. The use of typographical and non-typographical structural markers adds to an unambiguous inner access structure and enhances the quality of the dictionary consultation procedures. The underlying principles to which this model adheres, creates the opportunity of a more general application to other pairs of related languages and could assist to enhance the quality of bilingual lexicography.

References

- [Bergenholtz et al. 1999] Henning Bergenholtz/Sven Tarp/Herbert E. Wiegand: "Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern", in Lothar Hoffmann et al. (Eds.) *Fachsprachen. Languages for Special Purposes*, (Berlyn: De Gruyter) 1999: 1762-1832.
- [Gouws 1996] Rufus H. Gouws: "Bilingual Dictionaries and Communicative Equivalence for a Multilingual Society", in *Lexikos*, Vol. 6 1996: 14-31.
- [Gouws 2000] Rufus H Gouws: "Doeltaalgerigtheid teenoor lemmagerigtheid in vertalende woordeboeke", in *Tydskrif vir Geesteswetenskappe*, Vol.40 (1) 2000: 39-47.
- [Gouws 2000a] Rufus H Gouws: "Toward the formulation of a metalexically founded model for national lexicography units in South Africa", in Herbert E Wiegand (Ed.). *Wörterbücher in der Diskussion IV*, (Tübingen: Max Niemeyer) 2000: 109-133.
- [Martin 1996] Willy Martin: "Lexicographical Resources in a Multilingual Environment : an Orientation", in *Lexikos*, Vol. 6 1996: 199-214.
- [Martin/Al 1988] Willy Martin/Bernard Al: "User-orientation in Dictionaries : 9 Propositions", in T. Magay et al. (Eds.): *Budalex '88 Proceedings*, (Budapest) 1988:393-399.
- [Martin/Heid 1998] Willy Martin/Ulrich Heid: *On the construction of bilingual dictionaries*. (CLVV, The Hague) 1998.
- [Martin/Tamm 1996] Willy Martin/A. Tamm: "OMBI: An editor for constructing reversible lexical databases", in M. Gellerstam et al. (Eds.): *Euralex '96 Proceedings I-II*, (Göteborg University) 1996: 675 – 685.
- [Martin et al. 1999] Willy Martin et al. *Referentie Bestand Nederlands. Documentatie*. (Amsterdam: Vrije Universiteit, Lexicologie) 1999.
- [Martin et al. 1999a] Willy Martin/Rufus Gouws/Luc Renders: *Haalbaarheids- en Definitiestudie voor een woordenboek Afrikaans-Nederlands/Nederlands-Afrikaans*,(Amsterdam: Vrije Universiteit) 1999.
- [Ponelis 1993] Friedrich A. Ponelis: *The Development of Afrikaans*, (Frankfurt: Peter Lang) 1993.

[Wiegand 1996] Herbert E. Wiegand: "Textual Condensation in Printed Dictionaries: A Theoretical Draft", in *Lexikos*, Vol. 6 1996: 133-158.