

## **Towards a New Type of Bilingual Dictionary**

**Paul Bogaards<sup>1</sup>, Mike Hannay<sup>2</sup>**

Leiden University

Choorstraat 65

NL - 5211 KZ 's-Hertogenbosch

\*\*\*\*\*

Free University, Amsterdam

Dept of English

De Boelelaan 1105

NL - 1081 HV Amsterdam

### **Abstract**

This paper contains a proposal for a new type of bilingual dictionary. This dictionary is conceived as an electronic device that will contain one central database including all relevant information for receptive and productive use of the L2. For productive purposes the lexical units (not words) of the L1 will be linked to the corresponding lexical units of a L2 learners' dictionary so as to make available all the information about pronunciation, meaning and use that can be found in these dictionaries. In a comparable way the lexical units of the L2 will be linked with the equivalent lexical units of the L1 of the users. A third kind of use that will be catered for is L2 – (L1) – L2, i.e. a situation in which the user already knows the word of the L2 he needs but can use some information in his mother tongue.

### **1. Introduction**

Compared to the achievements over the past twenty years in monolingual lexicography for language learners, the domain of bilingual dictionaries is strikingly immobile. This is rather surprising when one takes into account that learners still use bilingual dictionaries on a massive scale. They may be urged by their teachers to use monolingual dictionaries, but 'such urgings often fall on deaf ears' (Wakeley 1998: 428).

One development that should, however, be mentioned is that of the 'bilingualized' learners' dictionary. This type of dictionary combines the advantages of the monolingual learners' dictionary with the ease of access and use that is characteristic of the bilingual dictionary. In bilingualized dictionaries the definitions, examples, and grammatical information of the monolingual dictionary are partly translated into the mother tongue of the users. Marello (1998) gives a clear overview of the various types of bilingualized or 'semi-bilingual' dictionaries that have been published since the early seventies (see also Kernerman 1996).

In her conclusion Marello (1998: 310) also mentions some of the limitations that are inherent in bilingualized dictionaries. For us, the most important of these limitations is the idea that they are essentially *readers'* dictionaries and that *encoding* is not a priority. In fact, all bilingualized or 'semi-bilingual' dictionaries take the macrostructure of some kind of L2 monolingual dictionary as their starting point and go from there to the L1 of the user. Only a small number of them include some sort of L1-L2 dictionary, but this never constitutes a full bilingual production dictionary. The *Passport English Learner's Dictionary* (2002)

published by K Dictionaries (Israel), for instance, contains what is called a 'L1- English index', and Collin's *Mono-bilingual learner's dictionary* (1989) for speakers of Italian includes a section 'italiano-inglese' that looks like a very poor bilingual list, sometimes giving up to ten possible translations for a given Italian word, but without any further specification. Another weak point of the bilingualized dictionaries, inherited from the monolingual dictionaries they are based on, is that they do not take into account the specific contrastive difficulties that users of a given mother tongue may experience.

In most bilingualized dictionaries the elements of the L2 are presented with clear grammatical descriptions. When such dictionaries are used for comprehension, much of this information is not really needed. When it comes to production, however, this rich information is essential, but in most bilingualized dictionaries the user has difficulty in finding the words he needs. In a traditional bilingual dictionary, on the other hand, he can easily find the words he needs, but these are not given with full grammatical details.

The solution to this problem would seem to consist in adding the riches of the modern learners' dictionary to a well organized set of data concerning the L1 of the learners. In other words, the 'left side' of a bilingual production dictionary could be translated using the very rich data that are available in the L2 learners' dictionary. This is what we are planning to do at Van Dale Lexicography in the Netherlands, and the aim of this paper is to provide a sketch of the approach we are adopting.

In section 2 we set out the steps designed for three look-up procedures, involving a reception mode and two different kinds of production mode, one which proceeds from an L1 element and one from an L2 element. In section 3 we then take a brief look at specific editorial issues that emerge when linking two kinds of data.

## **2. Linking Data from Different Types of Dictionary**

In a standard Dutch-English bilingual production dictionary one can find things like this:

- verontschuldigen** <wk. ww.> 1. [excuses aanbieden] *apologize*  
2. [...] ....

The Dutch word is described on the level of grammar ('wk. ww.' meaning reflexive verb) as well as on the level of meaning (the words between square brackets specifying the meaning of the entry word). However, nothing is said about the verb that is given as a translation. Thus the implicit message to the user seems to be that 'if this reflexive verb in this particular sense is the element you want to translate, just use the translation given and you will end up with a correct and appropriate sentence in English'. This may invite users to think that, taking into account the general rules of syntax and morphology that pertain in the other language, whatever is possible with *zich verontschuldigen* in Dutch will be equally possible with the English verb that is given as an equivalent. As we all know, this may in many cases lead to erroneous or else infelicitous utterances.

Now let us see what a learners' dictionary could add in the case of *apologize* as a translation of *zich verontschuldigen*. The idea is that the Dutch verb would then be linked with the English verb, but not with the complete entry, as this contains a lot of information that has nothing to do with what the user is looking for. In order to help and guide the user better, the

lexicographer has to link the Dutch lexical unit to a well defined lexical unit in the L2. This might result in a structure such as Figure 1, which has been adapted from the *Macmillan English Dictionary for Advanced Learners* (MEDAL):

<p><b>zich verontschuldigen</b> [excuses aanbieden] – <i>apologize</i> /pronunciation/</p> <p>to tell someone that you are sorry for doing something wrong or causing a problem</p> <p>&lt;intransitive verb&gt;</p> <p><b>There's no need to apologize</b></p> <p>+ for <i>apologize for taking so long to reply</i></p> <p>+ to <i>You should be apologizing to your customers for wasting their money</i></p> <p><b>apologize profusely/sincerely/unreservedly</b> <i>The newspaper has apologized unreservedly for printing false allegations about her</i></p>
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Figure 1: A basic coupling of L1 and L2 data

In the case of MEDAL there is a cross reference to a table giving ‘ways of **apologizing**’, something that might also be added to the entry given above.

The advantages of this format are as follows:

1. grammatical information about the L2 word is given in an explicit way;
2. there are concrete examples of use;
3. models of frequent constructions are given and exemplified;
4. a clear definition allows the user to verify whether the equivalent really has the same meaning as the word to be translated;
5. for those who want to use the word in an oral context, the pronunciation is given.

## 2.1 The Production Mode

In the Van Dale project, we are working on an electronic dictionary where this type of

de <b>film</b>		NV-1
a. mijn film is vol [filmrolletje]	- my <b>film</b> is full	
b. hij houdt van film	- Br. he likes <b>films</b>	
	Am. he likes <b>movies</b>	
c. film is een belangrijke kunstvorm	- <b>Cinema</b> is an important form of art	

coupling between the lexical units of two languages can be realized quite efficiently on the basis of a single integrated database. The productive user chooses a Dutch word, for instance by clicking on a word in a text, and will then be given a first screen. The first screen for the Dutch word *film* would be as in Figure 2 below:

Figure 2: First production screen for *film*

This screen presents three senses of the Dutch word *film* illustrated by three short sentences which are translated into English. Clicking on one of the senses will take the user to the corresponding second screen. The basic lay-out of a second screen is given in Figure 3:

NV-2		
[sentence]	[translation]	
[canonic form]	[other forms]	[pronunciation]
[definition including basic lexico-grammatical frames (in bold)]		
[L2 examples without translation]		
[L1-elements] + [translations]		
ATTENTION !	OTHER WORDS	L2-LEMMA

Figure 3: Basic structure for NV-2 screens

For the first sense of **film**, this second screen might look like this:

a. mijn film is vol [filmrolletje]	my <b>film</b> is full	NV-2a
<b>film</b> [pronunciation]	<b>films</b> [pronunciation]	
a <b>film</b> is the narrow roll of plastic that is used in a camera to take photographs		
The photographers had already shot a dozen rolls of film. ...35 mm film ...colour film		
...She put a new film in her camera. ...to have a film developed		

Figure 4: Second screen for *film*

Where there is both a British and an American translation, two different screens can be offered:

b. hij houdt van film	-Br. He likes <b>films</b>	NV-2b (Br.)
<b>film</b> [pronunciation]	<b>films</b> [pronunciation]	
<p>A <b>film</b> consists of moving pictures that have been recorded so that they can be shown at a cinema or on television. A film tells a story, or shows a real situation.</p> <p><i>Everything about the film was good. Good acting, good story, good fun. ...a government health film about the dangers of smoking ...Let's go to the cinema – there's a good film on this week. ...Let's stay in and watch a film. ... a horror/documentary/feature film ... a silent film (= one recorded without sound) ... an international film festival ...a film crew/critic/director/producer ...the film version of the novel</i></p>		
Een korte film	a short, a short film	
Een film draaien	to show a film	
We gaan naar de film	we are going to see a film	

Figure 5: second screen for *film* (British English version)

b. hij houdt van film	Am. He likes <b>movies</b>	NV-2b (Am.)
<b>movie</b> [pronunciation]	<b>movies</b> [pronunciation]	
<p>A <b>movie</b> is a film [AM; also BRIT, INFORMAL]</p> <p><i>In the first movie Tony Curtis ever made, he played a grocery clerk ... a horror movie ... Have you seen the latest Tarantino movie? A famous movie director/star</i></p>		
Een korte film	a short, a short movie	
Een film draaien	to show a movie	
We gaan naar de film	we are going to see a movie	

Figure 6: second screen for *film* (American English version)

Some remarks are in order. First, each screen contains all and only the relevant information for a given lexical unit. The English definitions given in the screens above have been taken from Cobuild (2001); the examples come from various learners' dictionaries. The L1-elements have been chosen because of the translation problems they may cause to Dutch learners of English. The pronunciation of the words (and where relevant of irregular plurals or verb forms) will be given as audio files. The screen will only show a symbol. The ATTENTION!, OTHER WORDS and L2-LEMMA buttons in Figure 3 are only present when some kind of special information is to be given. The first button will be active when

there is some contrastive difficulty, the second one when a choice may have to be made between synonyms or when extra information is available, for instance on formulas, as in the case of *apologize*. The last button provides a link to the whole entry of the English word; this will only be implemented for very advanced users because in most cases, especially for intermediate learners, seeing other senses of the L2 word will only confuse them. As an illustration of the extra buttons, consider the following sequence of three screens for the Dutch word *lekker*.

<b>lekker</b>	<b>NV-1</b>
b. dat ijs is lekker	- that ice cream is <b>good</b>
c. dat is een lekkere stoel	- that is a <b>nice</b> chair
d. ik voel me niet lekker-	I don't feel too <b>well</b>

Figure 7: first production screen for *lekker*

a. dat ijs is lekker	- that ice cream is <b>good</b>	
<b>NV-2</b>		
<b>good</b> [pronunciation]	<b>better</b> [pron.] <b>best</b> [pron.]	
of a high quality or level		
<i>The food at this restaurant is very good</i>		
Is het lekker?	- Do you like it?	
Ja, het heeft me lekker gesmaakt	- Yes, I enjoyed it	
Lekker (kunnen) koken	- Be a good cook	
<table border="1"><tr><td>OTHER WORDS</td></tr></table>		OTHER WORDS
OTHER WORDS		

Figure 8: Second production screen for *lekker*

<p><b>OTHER WORDS</b></p> <p>In plaats van <b>good</b> kun je ook gebruiken:</p> <p><b>nice:</b> pleasant  <i>We could go to the coast tomorrow, if the weather is nice</i></p> <p><b>tasty:</b> food which is <b>tasty</b> has a good flavour and is nice to eat  <i>They do a really tasty chicken and mushroom soup</i></p> <p><b>yummy:</b> if food or drink is <b>yummy</b>, it tastes very good  dit woord is meer spreektaal!</p> <p><b>delicious:</b> if food or drink is <b>delicious</b>, it smells or tastes extremely good  <i>this soup is absolutely delicious</i></p>	<p><b>NV-3</b></p>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------

Figure 9: 'Other words' screen for *lekker*

The English material for these three screens has been drawn from the *Cambridge Learner's Dictionary* (2001).

## 2.2 The Reception Mode

We now turn briefly to the reception part of the dictionary. Basically, this part will present the items listed in the monolingual learners' dictionary with their translations into Dutch. The first screen will contain the forms, their pronunciation, basic grammatical information and a list of possible translations. An example is given below in Figure 10 for the verb *learn*.

<p><b>learn</b></p> <p>werkwoord</p> <ol style="list-style-type: none"> <li>1. leren</li> <li>2. studeren</li> <li>3. horen, vernemen</li> <li>4. iets leren van</li> <li>5. <b>learn your lesson</b> - leergeld betalen</li> </ol>	<p>/pronunciation/</p>	<p>verleden tijd en voltooid deelwoord <b>learnt</b>          /pronunciation/ of <b>learned</b> /pronunciation /</p>	<p><b>VN-1</b></p>
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------	--------------------------------------------------------------------------------------------------------------------------	--------------------

Figure 10: First screen for *learn*

Again, by clicking on one of the possible translations, a second screen will follow, the basic lay-out of which is as in Figure 11:

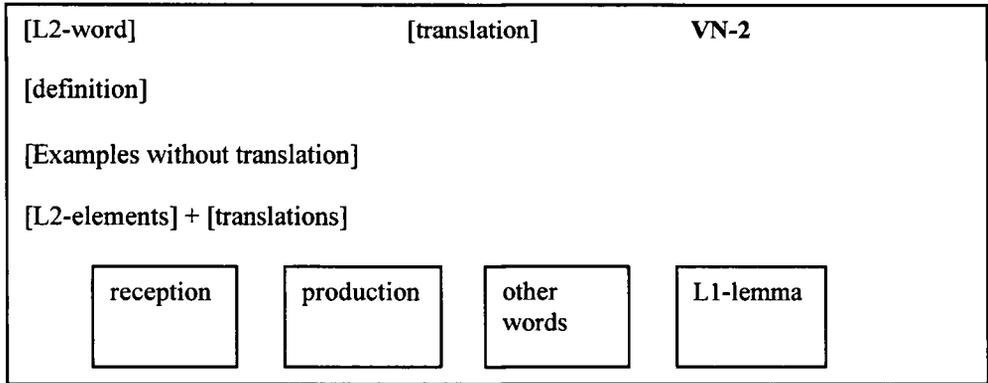


Figure 11: Basic structure for VN-2 screens

For the first sense of *learn* this will give something like Figure 12:

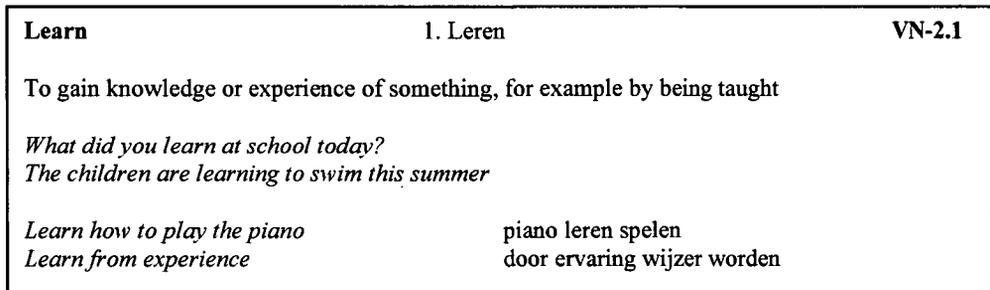


Figure 12: Second screen for *learn*

The L2-elements that are translated at the bottom of the screen have been chosen because of the specific contrastive difficulties they present for Dutch learners of English.

The four buttons at the bottom of the screen in Figure 11 can each be clicked to reveal a new screen. The reception screen gives information about the L2 item. This may be cultural information, or alternatively a warning about the Dutch translation. The production screen is in fact only relevant in the production mode, and is discussed in section 2.3 below. The ‘Other words’ screen gives information on synonyms on a par with the ‘Other words’ screen in the production mode, and finally, clicking on ‘L1 lemma’ takes the user back to the first NV production screen for the main translation of the L2 headword.

### 2.3 The Production Mode with L2 Input

In a foreign language learning environment the input for a foreign language production task will often come from the first language. However, it will be clear that this by no means needs to be the case. The learner may already have decided which lexical unit in the foreign language he wants to use, but may wish to have more information. In the system we are developing, this specific production function can be neatly dovetailed with the reception function and the standard production function. The opening screen will be the same as that for reception (see Figure 10), with the Dutch translations now serving to help the user check which meaning of the English lexical unit in question he is interested in; in other words they are now functioning as meaning descriptions rather than actual translations. Clicking on the relevant translation leads to the second reception screen, as given in Figure 11. But from this point the production mode becomes more important. Clicking the production button produces a new screen, the structure of which is given in Figure 13 below.

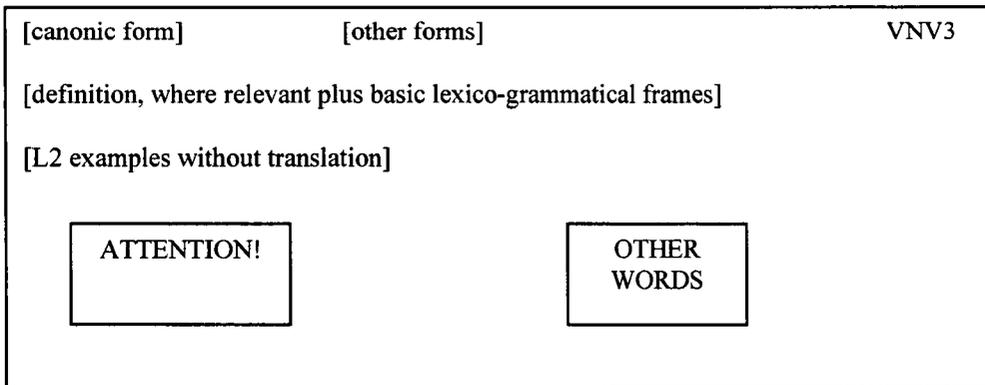


Figure 13: Third production screen with L2 input

This contains the definition of the lexical unit in question, plus the basic lexico-grammatical frames which it operates in, as well as a set of English examples. These elements are also present in the NV-2 screen (Figure 3), as are the ATTENTION! and OTHER WORDS buttons, both of which provide information on aspects of use.

### 3. Editorial Considerations

Obviously the linking of an existing set of monolingual data to an existing set of bilingual data constitutes a major editorial task. On the one hand it is important to strive for an efficient editing procedure, but on the other hand it is equally important to present the data in a fashion which promotes maximum accessibility for the user, who in the case of the project we are working on is a learner in the 12-15 age range. In this section we mention a few

points which need to be taken into consideration when presenting the translation and usage profiles which constitute the initial screens.

The Dutch usage profile in the production mode is of particular value for language learners at the elementary and intermediate levels since it is designed to include not only short sentences, as is the case with *film* above, but also fixed combinations, idiomatic expressions and other multiword lexical units. This leads to a richer and more accessible usage profile than is customary in bilingual dictionaries, where all kinds of multiword expression are given a more subordinate role. As an illustration of our approach consider the Dutch noun *loop*. One of the traditionally identified senses relates to the expressions *op de loop zijn* and *op de loop gaan*, which can be translated as ‘to be on the run’ and ‘to flee’ respectively. A standard approach in a bilingual dictionary would be to define a separate sense for this use of *loop* which requires its own meaning description and its own translation. However, this gives the user little help in actually identifying the meaning intended and even less help in finding a really useful translation equivalent, since the sense involved is so closely related to the two fixed expressions. The first screen for *loop* might then look like this:

<b>De loop</b>	<b>NV-1</b>
a. <i>In de loop van</i> de wedstrijd werd Ajax sterker	In the course of the game Ajax got stronger
b. Dit geweer heeft een korte loop	This gun has a short barrel
c. De dieven zijn <i>op de loop</i>	The robbers are on the run
d. De daders waren <i>op de loop gegaan</i>	The culprits had fled

Figure 14: First production screen for *loop*

A major editing task thus consists of identifying what elements of varying kinds actually make up the usage profile.

In a similar way to the first screen in the productive mode, the left column in the first screen of the reception mode provides a translation profile of the lexical unit in question. This may include other lexical units which the headword is part of. An important case in point for English is phrasal verbs. These need to be included in the opening screen for the base verb since the user needs to be shown that the phrasal verbs constitute lexical units in their own right. One way to do this is to include references in the translation profile, for instance in the case of polysemous phrasal verbs. A possible first reception screen for the verb *shut* is given in Figure 15:

<b>shut</b>	/pronunciation/	verleden tijd en voltooid deelwoord <b>shut</b>	VN-1
<b>Werkwoord</b>			
1.	dicht doen		
2.	sluiten		
3.	>> <b>shut down</b>		
4.	>> <b>shut out</b>		
5.	>> <b>shut up</b>		

Figure 15: First reception screen for the verb *shut*

This again differs from the translation profile in standard bilingual reception dictionaries, where links to other headwords tend to appear later in the lemma.

A rather different situation obtains with a verb like *border*. In one of its senses it can be used either transitively or intransitively, in which case it invariably occurs with the fixed preposition *on*. The *Cambridge Learner's Dictionary* captures these two uses in the same way, and this means of representation can be retained when the Dutch data are incorporated because the Dutch translation is the same for both uses. The first screen for the verb *border* might look like this:

<b>border</b>	/pronunciation/	verleden tijd en voltooid deelwoord	VN-1
		<b>bordered</b> /pronunciation/	
werkwoord			
1. begrenzen, omranden			
2. [ook border on] grenzen aan			
3. <b>this borders on the ridiculous</b> - dit grenst aan het absurde			

Figure 16: First reception screen for the verb *border*

For both these reception cases the implication again is that the translation profile may differ significantly from that provided in standard bilingual dictionaries, but in the reception case the structure of the English learner's dictionary provides the editor with a lot more help, since phrasal verbs, fixed combinations and the like are already given a relatively prominent status within the lemma.

One final editorial task worth mentioning in the context of opening screens relates to the choice of Dutch translations in the first reception screen. Particularly with a target group of elementary and intermediate learners it is important to limit the number of translations per sense as much as possible, preferably to one and at the most to two. In a well organized bilingual database the main translational possibilities will have already been noted, but an additional constraint arises when the illustrative examples from the learner dictionary are taken into account. It may thus be valuable to carry out an editorial check to ensure that these examples can indeed be translated using the translations offered for the headword. If this is not the case, there are two options available: either the choice of main translation can be reviewed, or those examples which would not receive a 'model' translation can be instead be offered lower down on the second reception screen (see Figures 11 and 12) together with their specific translation. To facilitate this process it would make sense to work with a

relatively rich English-Dutch database rather than one which has already been reduced to fit in with the requirements of the less experienced learner.

#### **4. Summary**

We hope to have shown that a bilingual dictionary which is conceived as an electronic dictionary can provide the user, in a single sequence of connected steps, with information which can help with three different tasks: a reception task and two different production tasks, taking either L1 or L2 as a starting point. Such a dictionary avoids the shortcomings of bilingualized learners' dictionaries, which have a relatively weak encoding function. A good and efficient way of building a bilingual learners' dictionary with a powerful production component is to link the kind of information given in existing monolingual learners' dictionaries to the relevant elements from a standard bilingual lexicographical database.

#### **References**

##### **A. Dictionaries**

- Cambridge learner's dictionary*. 2001. Cambridge: Cambridge University Press.  
Collin, P.H. 1989. *Mono-bilingual learners' dictionary. Dizionario inglese monolingue con traduttori italiani e sezione italiano-inglese*, Bologna: Zanichelli.  
Rundell, M. et al. 2002. *Macmillan English Dictionary for Advanced Learners*. Oxford: MacMillan (MEDAL).  
Sinclair, J. et al. 2001. *Collins COBUILD English Language Dictionary*. London & Glasgow: Collins (Cobuild).

##### **B. Other literature**

- Kernerman, L. 1996. 'English learners' dictionaries: how much do we know about their use?' In M. Gellerstam et al. (eds.), *Euralex '96 Proceedings*. Vol. II, 405 – 414.  
Marello, C. 1998. 'Hornby's bilingualized dictionaries.' *International Journal of Lexicography* 11/4, 292 – 314.  
Wakeley, R. 1998. 'The treatment of French reflexive verbs in bilingual dictionaries.' In T. Fontenelle et al. (eds.), *Euralex '98 Proceedings*, Université de Liège, Vol. II, 421 – 430.