
Categorising Example Sentences in Dictionaries for Research Purposes*

Lorna Hiles, Cape Town, South Africa (lorna@hiles.co.za)

Abstract: Examples in dictionaries come in many different forms. They may be sentences or phrases. They may be corpus-based or made up by the lexicographer. They may contain the lemma in its uninflected form or they may contain an inflection of the lemma. In some dictionaries the function of examples is to provide contextual support to the meaning of the headword, and in others the grammatical support that they provide is more important.

While there is literature on the usefulness of examples, and on whether examples should be corpus-based or not, there is very little on what makes one example more useful than another. I have set out to find out what sort of examples South African school users identify as most helpful. In this article, I look at whether examples in five South African school dictionaries do provide suitable contextual or grammatical support.

I have constructed a table to classify example sentences according to different criteria. I filled in this table with randomly selected words and their examples which have been taken from five different South African school dictionaries.

The goal of this research is to present characteristics of examples in a way that makes them easier to analyse and compare. This should help lexicographers in future dictionaries check whether they have written or selected the best possible examples for their users' needs.

Keywords: EXAMPLE SENTENCES, ILLUSTRATIVE EXAMPLES, SCHOOL DICTIONARIES, TABLE OF CATEGORIES

Opsomming: Kategorisering van voorbeeldsinne in woordeboeke vir navorsingsdoeleindes. Voorbeelde in woordeboeke verskyn in baie verskillende vorme. Hulle kan sinne of frases wees. Hulle kan korpusgebaseer of deur die leksikograaf geskep wees. Hulle kan die lemma in sy onverboë vorm bevat of hulle kan 'n verbuiging van die lemma bevat. In sommige woordeboeke is die funksie van voorbeelde om kontekstuele steun vir die betekenis van die trefwoord te verskaf, in ander is die grammatiese steun wat hulle verskaf, belangriker.

Terwyl daar literatuur oor die nuttigheid van voorbeelde bestaan, en of voorbeelde korpus-

* This article is a combination of two papers respectively presented at the Fourteenth International Conference of the African Association for Lexicography, organised by the Xhosa Department, University of the Western Cape, Bellville, South Africa, 6-8 July 2009 and the Fifteenth International Conference of the African Association for Lexicography, held at the University of Botswana, Gaborone, Botswana, 19-21 July 2010, both based on research for an M.Phil. Thesis *Examples in South African School Dictionaries: From Theory to Practice*, which was accepted by Stellenbosch University, Stellenbosch, South Africa, November 2009.

gebaseer moet wees of nie, bestaan daar baie min oor wat een voorbeeld nuttiger maak as 'n ander. Ek het begin om uit te vind watter soort voorbeelde Suid-Afrikaanse skoolgebruikers aanwys as die nuttigste. In hierdie artikel kyk ek of voorbeelde in vyf Suid-Afrikaanse skoolwoordeboeke wel geskikte kontekstuele of grammatiese steun verskaf.

Ek het 'n tabel opgestel om voorbeeldsinne volgens verskillende maatstawwe te klassifiseer. Ek het hierdie tabel voltooi met lukraak gekose woorde en hul voorbeelde wat uit vyf verskillende Suid-Afrikaanse skoolwoordeboeke geneem is.

Die doel van hierdie navorsing is om eienskappe van voorbeelde op 'n manier aan te bied wat dit makliker maak om hulle te ontleed en te vergelyk. Dit behoort leksikograwe in toekomstige woordeboeke te help om te kontroleer of hulle die bes moontlike voorbeelde geskryf of gekies het vir hul gebruikers se behoeftes.

Slutelwoorde: VOORBEELDSINNE, TOELIGTENDE VOORBEELDE, SKOOL-WOORDEBOEKE, TABEL VAN KATEGORIEË

While doing research into example sentences in school dictionaries I found a lot of theory about what examples are meant to achieve (proof of a word's use and existence in historical dictionaries, a complement to the definition, and an illustration of grammatical features of the word), but very little information on evaluating examples and what makes a good example.

I needed to analyse and evaluate examples in school dictionaries and to compare the examples in dictionaries with what the publisher set out to achieve with them, and with what learners wanted from example sentences in their dictionaries.

To this end, I created a table which I used to compare the features of examples from different dictionaries and extract related statistical data.

This made it easier to compare the dictionaries to each other and find patterns within the dictionaries. I was also able to compare what the dictionaries actually contain with what they were intended to contain, according to the introduction in the dictionary, the How to Use page, or discussions with the editor.

Before creating this table, I chose five South African school dictionaries: three monolingual and two bilingual.

The monolingual dictionaries are: South African Oxford School Dictionary (SAOSD) second edition, 2004, South African Oxford Secondary School Dictionary (SAOSSD) 2006, and Longman South African School Dictionary (Longman) 2007.

The bilingual dictionaries are *Oxford Afrikaans–Engels/English–Afrikaans Skoolwoordeboek/School Dictionary* (Tweetalige) 2006, and *Pharos Aanleerderswoordeboek vir Skole* (Pharos) third edition, 2006.

There are two Oxford school dictionaries because they are aimed at different users and they are quite different dictionaries, as the tables will show. The *South African Oxford School Dictionary* is for learners in Grades 4 to 10, while the *South African Oxford Secondary School Dictionary* is for learners in Grades 8 to 12.

So while there is an overlap of grades, the dictionaries are sufficiently different to warrant the inclusion of both.

The table contains fourteen columns of data, arranged in six sections. The first section is the example number, actual example sentence or phrase with the headword highlighted, and the part of speech. The second section indicates whether the example is a phrase or a sentence. The third section is for whether the example is a definition, whether it provides a context, or neither. The fourth section provides information about grammatical support: whether the example provides grammatical support, and if so, how. The fifth section shows whether the headword is simple or inflected in the example. The final section shows whether the example provides any other support, such as cultural information or usage restrictions not mentioned in the other columns.

		POS	Sentence	Phrase	Definition	Contextual support	No context	Grammatical support	Comments	Simple	Inflected	Additional support	Comments
	SAOSD												
1	a cake of soap	n		1			1	1	cake of	1			

Each column is added up, with totals provided for each dictionary, and for all the dictionaries together. These totals are used to generate statistics of the examples.

I selected 24 words randomly, keeping to the same words in each dictionary, where the dictionary contained the word as a lemma. I only chose nouns, verbs and adjectives, so that I could compare these part of speech categories. If a word was polysemous, I used all senses in that entry. For example, the verb **keep** has five senses in the SAOSD, four of which are exemplified. In the Longman dictionary there are seven senses of **keep**, six of which are exemplified. I used all of the senses in this study, rather than choosing one example from an entry, as in some cases the examples are meaningless on their own, but are useful when compared to other examples in that entry. I did not use examples from idioms or expressions. For example, in the Pharos dictionary, **keep** has eight senses followed by 17 fixed expressions. The expressions may contain valuable examples, but these are not used in this study. Where a lemma does not feature in the category table, it means either that the lemma is not treated in that dictionary or that it is not exemplified.

These 24 words generated a total of 146 examples across the five dictionaries. The results for all dictionaries are shown in the table below.

	POS	Sentence	Phrase	Definition	Contextual support	No context	Grammatical support	Comments	Simple	Inflected	Additional support	Comments
Total	146	114	32	3	98	45	62		103	43	60	
		78%	22%	2%	67%	31%	43%		71%	29%	41%	

The columns I was most interested in are the columns about the contents of the examples, and whether they assist the learner in any way (with contextual, grammatical or additional support).

The **Contextual support** section was the most difficult to keep consistent, since the allocation of values is determined by my intuition or rules. The basis for deciding whether an example contained contextual support was that if the headword could only be replaced by a synonym or near synonym in the sentence, or a word in the same lexical set, it was marked as providing contextual support. If the headword could be replaced by any other word of the same part of speech, then I marked it in the **No context** column. Other researchers could use their own policies for this. Consistency and being able to defend each decision are the keys to ensuring accurate data.

For example, in the hypothetical example *I love peaches* where **peaches** is the headword, I found there to be no clue, as the lemma could easily be replaced by any other noun: *I love soccer, I love animals*. In the sentence *I enjoy eating peaches* the options for replacing the headword are limited to the lexical set of food. In the sentence *Peaches are my favourite fruit* the options have been further narrowed down to the lexical set of fruit.

My main aim for looking for contextual support in the examples was if a learner does not understand the definition, will the example help? An example of how this has worked in this study is an example from SAOSD: *Defeats are offset by successes*. The definition (counterbalance or make up for something) contains difficult vocabulary and could be misinterpreted, but the example clearly shows what **offset** means and how it is used.

This support is especially valid in South African schools, where learners may not have the language skills and fluency for a monolingual dictionary aimed at first language English speakers. Their dictionaries need to give as much support as necessary.

As can be seen by the results table above, 67% of the examples (across the monolingual and bilingual dictionaries) examined do contain contextual support.

Grammatical support is another purpose of example sentences, specifically for the text production function of a dictionary. Most of the information in

the **Grammatical support** columns concerns collocations, and whether the example sentence provides suggestions for common collocations. For example, the sentence *The success of the product depends on marketing* shows that **depends** can collocate with **on**.

I have also looked at whether the example shows that verbs can be used transitively or intransitively. This is difficult to quantify because an example cannot show how a word cannot be used. I have indicated transitive/intransitive use where there are two or more examples for the same sense of a verb, and where the examples show different uses.

Out of 146 sentences examined 55 (38%) provide this type of grammatical support. It is difficult to draw conclusions and say whether this is significant, because many of the headwords do not need grammatical support of this kind.

The **Additional support** columns contain information that may be helpful to the learner, but cannot be described by the previous columns. These sentences may contain cultural data, which would include racial and gender information. This may be more useful in a bilingual dictionary where the cultures of the different languages are different. For example, in an English–Northern Sotho school dictionary, the lemma **stick** is exemplified by sentences about hitting children or animals, all from a Northern Sotho corpus. For example, *The teacher beats the children with a stick*. A reader is unlikely to find these sentences in a monolingual English school dictionary. This is considered cultural information.

Gender information would include whether gender roles are stereotyped or not, such as *His wife nagged him to paint the wall*. While it is difficult to avoid gender altogether, some dictionaries use strategies to avoid gender in example sentences, for example, by using "I", "we" or "you" as the subject of a sentence, instead of "he" or "she" or names. Another strategy is to use names that are not gender specific, such as Jo, Lee, or Terry. Gender information can also be neutral, such as *He painted a picture*.

Other information that fits in the **Additional support** section is whether the word is used figuratively or idiomatically. This is valuable support, especially for a second language user of a dictionary.

I have also included whether the sentence is a command as some teachers prefer school dictionaries to not be too instructive. If learners see their dictionary as a set of instructions, they may be discouraged from using it. The examples are supposed to get learners' attention and interest, without being didactic.

The information found in the **Additional support** section is also what makes an example, and therefore a dictionary, engaging to a learner. If a learner consulting a dictionary finds the examples interesting and varied, they are more likely to respond to the dictionary positively. In this study 41% of the examples contained additional support in some form. Again, these are difficult statistics to work with, since it is unlikely that every example can contain additional support. However, one must be aware when interpreting these results that support in the **Additional support** column may not be valuable as actual support in the dictionary.

I also compared the dictionaries according to various criteria, such as how many of the headwords are treated in each dictionary, how many senses are treated, and how many have examples. I also looked at the differences in nouns, verbs and adjectives, to see if there were patterns in their treatment.

An interesting result of these statistics is that in both of the bilingual dictionaries that I looked at, each entry contains example sentences, sometimes more than one, while the monolingual dictionaries did contain entries without examples. One can also see that while the monolingual dictionaries made use of example phrases in some instances, the bilingual dictionaries only used full sentences. These statistics also show that the monolingual dictionaries treated more words, with the SAOSSD and Longman containing the highest number of headwords treated for the 24 words.

The monolingual dictionaries and bilingual dictionaries yield different statistics, as shown in the table below. The bilingual dictionaries can be seen to provide more support across **Contextual** and **Additional support**.

	Sentence	Phrase	Definition	Contextual support	No context	Grammatical support	Comments	Simple	Inflected	Additional support	Comments
Monolingual	63%	37%	1%	54%	45%	44%		66%	34%	39%	
Bilingual	100%	0%	3%	86%	10%	41%		78%	22%	44%	

I analysed the category table further to find out how many examples provide support in more than one area (contextual, grammatical and additional). Pharos has the most with 21%, followed by Tweetalige (16%), SAOSD (14%), Longman (10%) and SAOSSD (4%). Again, one needs to be wary of using these statistics as some headwords may not require additional or grammatical support. However, it is interesting to note these numbers, as many examples could be changed to contain as much support as possible.

The Pharos dictionary performs well in all the categories showing that it provides the most support out of all the dictionaries. It is clear that the examples were chosen or written with this in mind. However, it must be noted that this is offset by the space that these supportive sentences take up. Without doing a word count comparison of the dictionaries, one can see that the Pharos examples are longer than those of the other dictionaries. Many of the Pharos examples contain two clauses or phrases, or even sentences, making them more complicated, and less accessible to learners of English (who do not speak Afrikaans).

The data supports the conclusion that phrases are less useful at providing support than sentences, and my recommendation is that unless there is a good reason for using a phrase in a particular instance, that a sentence is used instead. A phrase can be used with a sentence, but it is recommended that in school dictionaries, using phrases as the only way of illustrating a sense is avoided.

To conclude, I feel that this table and its corresponding statistics offer great value in analysing and comparing different dictionaries according to very specific criteria. For my research, in particular, I used it to determine whether learners' needs in example sentences were being met. I also used it to provide recommendations for future school dictionaries.

The table of categories can be used in different ways, depending on who is using it. One can populate the table with examples from one dictionary, and find areas where the dictionary is lacking sufficient support for the headwords. One can also use the table to make sure that the examples are balanced — for example, in terms of gender references or location, and that the examples are not weighted to a particular subject matter such as crime.

Another way this table can be used is to compare dictionaries according to different criteria — as I have used it, but with different information in the columns. One can determine whether there is a pattern in the structure of examples: how do they use tenses, are they written actively or passively, are the sentences simple or complex?

One can also use a similar table of categories for whole entries in a dictionary in order to find out what support each entry provides. Other forms of support include artwork, synonyms and antonyms, usage notes, etymology, cross references, and labels. These can be used in conjunction with word frequency data and a defining vocabulary to find out which words do need more support. Words that are frequent and in the defining vocabulary may not need as much support as more difficult, less common words.

Bibliography

- Bullon, S. et al. (Eds.).** 2007. *Longman South African School Dictionary*. First Edition. Harlow, Essex: Pearson Education Ltd.
- Du Plessis, M. (Ed.).** 2006. *Aanleederswoordeboek vir Skole/Learner's Dictionary for Schools*. Third Edition. Cape Town: Pharos.
- Justus, J.A. (Ed.).** 2004. *The South African Oxford School Dictionary*. Second Edition. Cape Town: Oxford University Press.
- Louw, P.A. et al. (Eds.).** 2006. *Oxford Afrikaans–Engels/English–Afrikaans Skoolwoordeboek/School Dictionary*. First Edition. Cape Town: Oxford University Press.
- Reynolds, M. et al. (Eds.).** 2006. *South African Oxford Secondary School Dictionary*. First Edition. Cape Town: Oxford University Press.