

Willem Botha,
Bureau of the Woordeboek van die Afrikaanse Taal

The Lemmatization of Expressions in Descriptive Dictionaries

ABSTRACT: Of the three types of lexical items, namely the sublexical, the lexical and the multilexical, it seems that the multilexical is neglected most by descriptive dictionaries. Descriptive dictionaries, however, do not consistently include multiword lexical items as lemmas, but more often treat them as microstructural elements in the articles of lexical lemmas. This paper discusses the lemmatization of a very special group of multiword lexical items, namely expressions. The theoretical component of lexicography has long been in favour of the awarding of lemma status to expressions, yet has until recently not presented feasible suggestions on how this should be undertaken. Two possible ways of lemmatization are discussed. Guidelines are given for the microstructural treatment of expressions by the development of a computer field structure for onscreen entry. This discussion could help the practising lexicographer to present expressions in a more satisfactory and lexicographically correct way.

1. Multiword lexical items

Of the three types of lexical items, namely the sublexical, the lexical and the multilexical, it seems that the multilexical is neglected most by descriptive dictionaries. Multiword lexical items are seldom included as lemmas but are mostly included as microstructural elements in the articles of lexical lemmas.

A multiword lexical item functions as a unit and its meaning does not reflect the meanings of its individual components. Therefore they should be awarded lexical item status, i.e. be included in dictionaries as lemmas in their own right. Well-known subcategories of multiword items are prepositional phrases ON BEHALF OF, foreign phrases SINE QUA NON, ENFANT TERRIBLE and expressions CARRY COALS TO NEWCASTLE, STRIKE THE IRON WHILE IT IS HOT (Gouws 1989a, 59-67; 1989b, 97; 1990, 265-283; Ponelis 1989, 48-54; Zgusta 1971, 144-151).

What follows is a discussion of the lemmatization of expressions. The aim of the discussion is to find a practical solution for the treatment of expressions in descriptive dictionaries.

2. Expressions

In this article "expression" is used in a broad sense to include what is called "idioms", "fixed expressions" and "phrases". An expression is a special construction which is either not formed or cannot be interpreted according to the ordinary grammatical rules of a language. It is special with regard to meaning and/or morphology and/or syntax (Combrink 1989, 55-77; Moon 1985, 107-115).

As Gates (1985, 99-106) and Benson (1985, 61-68) have so clearly pointed out English and American descriptive dictionaries generally display a lack in consistency on the inclusion of expressions as main entries or subentries. It is not unusual to find that a dictionary includes an expression like *AT FIRST HAND* as a microstructural element in the article of the lexical lemma *HAND*, but treats *CASTLE IN THE AIR* as a main entry. There is a reluctance to make main entries of expressions not having a noun or an adjective as first component.

Main entries starting with *AT*, *IN*, *ON*, etc. are hard to find. An expression like *IN COLD BLOOD* is entered in the article of *COLD* but *COLD SHOULDER* is included as a main entry. Expressions consisting of more than two components are rarely included as main entries and so you could find *WET BEHIND THE EARS* entered as a separate sense in the article of *WET*, but *WET BLANKET* as a main entry.

The inconsistency in the inclusion of expressions as main entries or subentries is user unfriendly as the user can never be sure where he should look up an expression. It can also be a very tedious task finding an expression under the many different senses of a long article.

The Afrikaans comprehensive dictionary *WOORDEBOEK VAN DIE AFRIKAANSE TAAL* sub-enters terminological phrases like *ARCTIC CIRCLE* and *COLD FRONT* under an appropriate sense of the first main component. All other expressions are included at the end of an article of a lexical lemma in a separate section under the heading *UIT-DRUKKINGS (EXPRESSIONS)*. They are arranged in an alphabetical order which is explained in the guide to the use of the dictionary.

None of the above mentioned treatments of expressions really satisfies lexicographically, because of 1. inconsistency; 2. poor findability and 3. neglect of the lexical item status of expressions. Two possible alternative treatments of expressions are suggested.

3. Alternatives for the lemmatization of expressions

3.1 Alphabetical lemmatization according to the first component

Considering that an expression functions as a whole and that its meaning cannot be inferred from its components, the logical way of presenting all expressions would be as main entries in the alphabetical order of a dictionary. This alternative works well in the case of noun phrases not having a preposition as first component, e.g.:

cold cathode
cold feet
cold shoulder

cold sweat
cold war

The user is able to get a complete picture of noun phrase expressions starting with COLD at a glance. Most dictionaries limit the lemmatization of expressions to the noun phrase type. The reason for the reluctance to lemmatize all expressions is that other types of expressions pose some serious lexicographical problems for the lexicographer:

(i) There would be a very user unfriendly accumulation of expressions starting with the definite and indefinite article and certain pronouns and prepositions. There are literally hundreds of expressions starting with "It is a/an ..." and "It is ...". If the user was for example looking for the expression IN THE NICK OF TIME and TIME was the only component he was sure of, it would be impossible to find the expression.

(ii) It would often not be possible for the user to get a complete picture of all expressions that have a certain word as main component. The different expressions with time as main component, for instance, would be pages apart in a dictionary, e.g.:

ahead of one's time
behind the times
for the time being
in good time
pass the time of day
time out of mind

Although alphabetical lemmatization according to the first component is theoretically a sound policy it seems to be user unfriendly and therefore not a feasible solution.

3.2 Marked lemmatization in a separate section

The practice of entering expressions in a separate section under a lexical lemma has distinct advantages of which greater prominence and findability are the most important. There can be no doubt that it serves the convenience of users best.

The main objection raised against this practice is that the lexical item status of expressions is not recognized (Gouws 1990, 265-283). What follows, is an explanation of a presentation which gives main entry status to each expression in a separate section.

3.2.1 *Placing and division of the section EXPRESSIONS (EXPR.)*

The section *EXPR.* is placed at the end of an article, or if a lemma has more than one part of speech, after each different part of speech. *WORK WHILE IT IS DAY* would be included in the section following the verb *WORK* and *ALL WORK AND NO PLAY MAKES JACK A DULL BOY* in the section following the noun *WORK*.

The section *EXPR.* is divided into two parts, namely *SPECIALISED* and *IDIOMATIC*. Specialised expressions include terminological phrases like *ARCTIC CIRCLE* and *COLD FRONT* as well as phrases of a less technical nature like *THE KISS OF LIFE* and *THE FACTS OF LIFE*.

Specialised expressions have literal meanings and the components usually still have literal ties with the lexical lemmas. They are placed first in the section under the heading

A. *SPECIALISED*, followed by the idiomatic expressions under the heading B. *IDIOMATIC*.

3.2.2 *Lemma at which an expression is entered*

Expressions are entered under the lexical lemma of the first main component (also called “major invariable word”) of an expression. A main component can either be a noun, adjective or any form of a finite verb. *IN THE LONG RUN* and *THE LONG ARM OF COINCIDENCE* would be entered under the lexical lemma *LONG*.

3.2.3 *Determining the marker and arranging the expressions in the section EXPR.*

The second main component (noun, adjective or form of a finite verb) in an expression acts as a marker. It remains in the form in which it appears in the expression (e.g. in the plural or past participle form) and is placed first followed by a colon and the expression as a whole. This constitutes a marked lemma. The whole is in bold print and stands in the normal main entry position. The different expressions in the section are arranged alphabetically according to the markers.

The following idiomatic expressions would be entered as marked lemmas in the section *EXPR.* of the lexical lemma *MAKE*:

B. *IDIOMATIC*

bones: make no bones
 head: make head or tail
 make: make it
 mincemeat: make mincemeat (of)
 rings: make rings round a person

If there is no second main component in an expression the first component (which determines the lexical lemma where the expression is entered) is also used as marker, e.g. *make: make it*.

The inclusion of expressions as marked lemmas in a separate section was developed with Afrikaans descriptive dictionaries in mind. If the suggested treatment were applied to expressions with *LEWE* (life) as first main component the section *EXPR.* of the lexical lemma *LEWE* (life) would appear as follows:

LEWE II

1

LEWE II, Uitdr.

le'we

II s.n.w., lewes of lewens; lewetjie of soms lewentjie.

UITDR.

A. GESPESIALISEERD

lewe: na die lewe Met iem. of iets as model: *Na die lewe beeldhou, skilder of teken.* In 1901 het Anton (van Wouw) die Boergeneraals, Botha, De la Rey en Beyers ... direk na die lewe geteken (A.E. Duffey: Van Wouw, 1981, 9). *Sy eerste groot opdrag was die standbeeld vir pres. Kruger, wat hy ... na die lewe gemaak het* (Afr. Kernens., 1965, 849). Vgl. LEWENSGETROU.

voel: lewe voel Die eerste fetale bewegings voel: *Om en by die 16 de tot 18 de week van swangerskap begin die vrou lewe voel* (M. v. Rensburg: Lewe, s.j., 14).

B. IDIOMATIES

afhang: asof jou lewe daarvan afhang Na die beste van jou vermoë: *Die jong moeder het haar kind in die hof verdedig asof haar lewe daarvan afhang.* Niël is 'n gedugte hokkiespeler, want hy speel altyd asof sy lewe daarvan afhang.

afkom: met jou lewe daarvan afkom Bly lewe: *Hy was gelukkig genoeg om met sy lewe daarvan af te kom* (W.G.H. en S. Vivier: Hooyvlakte, 1969, 25).

afle: jou lewe afle (gee, neerle) vir Sterf vir: *Die patriotiese jong manne is bereid om hul lewe vir hul land af te le.*

boet: met jou lewe boet Sterf: *Weer het 'n jongeling in die bloei van sy jare vir sy wangstuk met sy lewe geboet* (S.J. en C.F. du Toit: Daeraad³, 1942, 73).

bring: in die lewe bring Geboorte gee aan: *As 'n vrou kinderloos bly ... sal iemand anders ... gestuur word om kinders vir die onvrugbare eggenote in die lewe te bring* (N.J.J. Olivier: Privaatreg, 1969, 101).

bring: om die lewe bring Doodmaak: *Kan dit moontlik wees dat Saul gewaag het om die diensknegte van die Allerhoogste om die lewe te laat bring?* (F.J. Steenkamp: Dawid, 1948, 168). *Dit was dieselfde Dingaan wat agt jaar tevore sy halfbroer Tsjaka om die lewe gebring het* (Kinders v.d. Wêreld VIII, s.j., 392).

dood: op (om, vir) lewe en dood Vir die behoud v.d. lewe: *Op ... jeugdige leeftyd stry hy reeds 'n stryd op lewe en dood* (Vad., 5 Febr. 1973, 1). *Ons (het) aanskou hoe die Prins vir dood en lewe geveg het* (L. Maré, vert.: Admiraal, 1925, 188).

■ Ook fig.: 'n Seksie wou aan Engels wasklou op lewe en dood (I.L. de Villiers in Koninkryk, 1981, 35).

Rugbyspelers (worstel) om lewe en dood om 'n drie agter die teenstanders se doellyn te gaan druk (F.P. v.d. Merwe: Einders, 1952, 54).

dreig: iemand met sy lewe dreig Iem. m.d. dood dreig: *Hy is al verskeie kere met sy lewe gedreig en daar is mense wat vrees dat daar ... geprobeer sal word om hom om die lewe te bring* (Burg., 21 Sept. 1966, 13).

gee: (die) lewe gee aan Geboorte gee aan: *Die nag toe sy aan haar kind die lewe gegee het, het haar eie gees gesterf* (J.H. Naudé: Oog, 1944, 51).

gee: jou lewe gee aan Jou totaal toewy aan: *Die seun het reeds sy lewe aan Christus gegee.* *Hy gee sy lewe aan die Afrikaanse toneelbedryf.*

gee: jou lewe gee (afle, neerle) vir Sterf vir: *Daar is pragtige verhale van ouers wat hul lewe vir hul kinders gegee het, omdat hul kinders se voortbestaan vir hulle belangriker as hul eie was.*

hand: jou lewe in jou hand neem (dra) Jou lewe waag: *Hy neem letterlik sy lewe in sy hand toe hy dit digby genoeg waag om oor die rand te kan afkyk* (S. Roux: Vlam, 1961, 29). *Hy (het) as Kaapse rebel dag na dag sy lewe in sy hand gedra* (C.M. v.d. Heever: Nooit, 1949, 49).

help: in die lewe help Hulp verleen met 'n bevalling: *Los vliegtuig en help baba in die lewe* (Oost., 5 Jul. 1967, 9 — opskr.). *Sy hoor weer Buks se eerste skreeue toe tant Poppie hom in hierdie selfde slaapvertrek in die lewe gehelp het* (F.A. Venter: Geknelde Land, 1960, 315).

■ Ook fig.: *Sara was haar lewe lank vroeërvrou op Die Dal.* *Sy het Die Dal in die lewe gehelp.* *Want wat is Die Dal sonder sy mense?* (A.H. de Vries: Dubbeldoor, 1963, 17).

hoop: solank daar lewe is, is daar hoop Al is omstandighede hoe swak, hoop 'n mens steeds op beterskap: *Ons sal ons nie deur die droogte laat onderkry nie, want solank daar lewe is, is daar hoop.*

hou: die lewe hou Bly lewe: *Wat ek sal sê as ek die lewe hou tot daar 'n derde druk verskyn, moet ek nog oor dink* (C.J. Langenhoven: Werke, 1956, 8).

inskiet: die lewe inskiet (laat) Sterf: *Dit kon maklik gebeur het dat hy saam met die perd die lewe ingeskiet het* (C.R. Swart: Agterryer, 1939, 57).

kom: in die lewe kom Wakker word of in pas kom m.d. werklikheid: *Kom in die lewe, jy weet al weer nie wat aangaan nie!* *Daardie dosent moet in die lewe kom; sy lesings boei nie die studente nie.*

kort: die lewe (tyd) is kort, die kuns is lank Teenoor die

kortstondigheid v.d. lewe staan die blywende waarde v.d. kuns: *Die beroemde Hippocrates sê: die lewe is kort; die kuns is lank* (Star, 19 Sept. 1966, 22).

kry: lewe kry Wakker, aktief of bedrywig word: *Moenie net daar staan nie, kry 'n bietjie lewe!* *Die stad kry eers laat in die aand lewe.* *By Strandfontein begin die galjoene ook nou lewe kry* (Burg., 13 Apr. 1962, 8).

laat: die lewe laat (inskiet) Sterf: *Die oormag was vir kol. Ournford te groot sodat ... 700 (manskappe) die lewe laat* (Afr. Kernens., 1965, 926).

lei: iemand 'n lewe lei Vir iem. die lewe moeilik maak: *Die man lei sy vrou 'n verskriklike lewe en dring daarop aan dat sy voltydys huisvrou moet wees.*

lewe: dit is [die] lewe Die ideale wyse om te lewe of die hoogste geluk: *Hy staar oopmond na die bruingebrande meisies op die strand en prewel saggies: "Dis die lewe!"* *Die lieflike geur van spekvet gebraaide vis hang in die lug ... "Dit is mos nou vir jou lewe manne!"* (J.A. Lückhoff: Visterm. Duik, 1968, 54).

lewe: in lewe Tydens iem. se lewe: *In lewe was Boela se pa 'n visserman* (C. Barnard: Boela, 1962, 3). *Vanoggend is Sy Eksellensie, dr. E.G. Jansen, in lewe Goewerneur-genl. ... ter ruste gelê* (A.N. Pelzer: Verwoerd, 1963, 298).

lewe: in [die] lewe Lewend: *Gaan verneem ... by die ander laers ... wie van hul vriende nog in die lewe is* (C.R. Swart: Agterryer, 1939, 40). *"Oom Louis was toe nog in lewe, nè?"* (A.P. Brink: Gebondenens, 1958, 87).

lewe: in die lewe! Uitroep van verbasing of verontwaardiging: *Het sy tien kinders binne tien jaar gehad? In die lewe!*

lewe: jou lewe wees Dit wees waaraan jy jou lewe wy of waarin jy jou hoogste geluk vind: *Boerdery is sy lewe en sy hele akademiese loopbaan was ook daarop gemunt om eenmaal self te gaan boer* (A.J. du Plessis: Man, 1937, 5). *Die heling en versorging van hierdie diere is sy lewe* (Landbw., 13 Jul. 1965, 28). *Die plaas is sy lewe sê hy.* *Dit is die plek waar hy hoort* (Landbw., 13 Jul. 1965, 23).

lewe: van jou lewe nie (noot) Nooit: (Hulle) hebbe my gesoebat om van myn lewe nie weer na die Kaap te rye (C.S. Nienaber: Afr. Vr. Jare, 1971, 12). *Dit kan elke skoolseun verstaan, maar hy sou van sy lewe nooit daartoe gekom het ... om dit self uit te dink nie* (A.H. Jonker: Roman, 1942, 64).

lewe: vir jou lewe Ten einde die lewe te behou: *Baklei, hardloop, stry, veg vir jou lewe.* (Die tier)

3.2.4 Advantages of marked lemmatization in a separate section

Some of the advantages of marked lemmatization in a separate section are:

(i) The lexical item status of expressions is recognized as they are entered as marked lemmas in the section *EXPR*. It solves the problem of how longer expressions should be entered as lemmas.

(ii) The suggested treatment does not require a drastic adjustment by the user as he presently looks up expressions under the lexical lemmas of the main components.

(iii) Expressions are conveniently arranged so that they are easy to find and expressions containing the same main component (e.g. LIFE) are grouped together.

(iv) Since the expressions are entered as marked lemmas it is possible to develop a computer field structure for on-screen entry into the dictionary. Such a field structure facilitates the microstructural treatment of expressions. The following field structure has been developed for Afrikaans idiomatic expressions:

<Idiomatic marker/> beer: <Idiomatic marker\>

<Idiomatic marker no./> <Idiomatic marker no.\>

This field is only filled in when there are more than one expression with the same marker. The number remains in hidden text and is never printed, but is useful in data search operations.

<Idiomatic expr./> life is not all beer and skittles <Idiomatic expr.\>

<Idiom. register/> <Idiomatic register\>

Sociolinguistic labels are entered in this field, e.g. (informal), (racist), (sexist), (vulgar), etc.

<Idiom. temporal labels/> <Idiom. temporal labels\>

Temporal labels like (archaic) and (obsolete) are entered in this field.

<Idiom. other labels/> <Idiom. other labels\>

These may include labels referring to frequency, regional speech, etc.

<Idiom. etymology/> <Idiom. etymology\>

<Idiom. definition/> <Idiom. definition\>

<Idiom. synonym/> <Idiom. synonym\>

<Idiom. opposition/> <Idiom. opposition\>

<Idiom. geographical occurrence/> <Idiom. geographical occurrence\>

<Idiom. cross ref./> <Idiom. cross ref.\>

3.2.5 Marked lemmatization outside a separate section *EXPR*.

Expressions which cannot be entered in the section *EXPR*. of a lexical lemma on account of morphological or semantic considerations, are included as marked lemmas in the

alphabetical order of lemmas. Also in this group are expressions containing words which do not occur outside the expressions. These words are either historic relics or morphological structures without independent word status (Gouws 1989b, 100).

The first main component acts as marker:

QT: on the QT
rack: rack and ruin

Lexicographically it would be wrong to enter a component like RACK as a lexical lemma as it only exists in the multiword item RACK AND RUIN.

Marked lemmatization also solves the problem of entering expressions containing Proper Names as main components, e.g.:

Lancashire: Lancashire thinks today what all England will think tomorrow

A lexical item classification (instead of a part of speech classification) can be given in the microstructure of marked lemmas, e.g.: rack: rack and ruin idiom. expr. Expressions entered under the heading *EXPR.* do not require further classification.

4. In conclusion

The treatment of expressions in descriptive dictionaries is unsatisfactory, mainly because of the inconsistency in awarding lemma status and the very limited microstructural treatment. The suggestions on marked lemmatization put forward in this discussion could help the practising lexicographer to present expressions in a more satisfactory and lexicographically correct way.

Bibliography

Dictionaries

- COLLINS, V.H. (1964): *A Book of English Idioms*. Longmans, London.
- COLLINS, V.H. (1965): *A Book of English Proverbs*. Longmans, London.
- GOVE, P.B. (1961): *Webster's Third New International Dictionary of the English Language*. Merriam-Webster, Springfield, Massachusetts:
- McLEOD, W.T. (1986²): *The New Collins Concise Dictionary of the English Language*. Collins, Glasgow.
- MURRAY, J. et al/BURCHFIELD, R. (Ed.) (1884/1933; 1972/1986): *The Oxford English Dictionary*. Clarendon Press, Oxford.
- SCHOONEES, P.C./SNIJMAN, F.J./HAUPTFLEISCH, D.C. (Eds.) (1951-1983): *Woordeboek van die Afrikaanse Taal*. State Press, Pretoria.

Other sources

- BENSON, M. (1985): "Collocations and Idioms". In *Dictionaries, Lexicography and Language Learning*. Ed. by R. Ilson. Pergamon Press, Oxford.
- COMBRINK, J. G. H. (1989): "Die Aangawe van Idiome in Tweetalige Woordeboeke. Wesenlike elemente, Ekstras en Vulsels".
- GATES, E. (1988): "The Treatment of Multiword Lexemes in Some Current Dictionaries of English". In: *ZüriLEX'86 Proceedings*. Ed. by M. Snell-Hornby. Francke Verlag, Tübingen.
- GOUWS, R. H. (1989a): "Setsels as Makrostruktuurelemente van Afrikaanse Woordeboeke". In: *S. A. Journal of Linguistics* 7(2). Bureau for Scientific Publications, Pretoria.
- GOUWS, R. H. (1989b): *Leksikografie*. Academica, Kaapstad.
- GOUWS, R. H. (1990): "Vaste Uitdrukings as Multeleksikale Lemmas in Verklarende Afrikaanse Woordeboeke". In: *Tydskrif vir Geesteswetenskappe* 30(4). Suid-Afrikaanse Akademie vir Wetenskap en Kuns, Pretoria.
- MOON, R. (1988): "'Time' and Idioms" In: *ZüriLEX'86 Proceedings*. Ed. by M. Snell-Hornby. Francke Verlag, Tübingen.
- PONELIS, F.A. (1989). "Leksikale Elemente". In: *Leksikografie en Leksikologie*. Ed. by T. J. R. Botha. Serva, Pretoria.
- ZGUSTA, L. (1967): "Multiword Lexical Units". In: *Word* 23.
- ZGUSTA, L. (1971): *Manual of Lexicography*. Mouton, Den Haag.