Fashion Words in Afrikaans Dictionaries: a Long Walk to Lexicographic Freedom or just a Lexical Fly-by-Night?

Danie J. Prinsloo
Department of African Languages, University of Pretoria, Pretoria 0002. South Africa
danie.prinsloo@up.ac.za

Rufus H. Gouws
Department of Afrikaans and Dutch, University of Stellenbosch

Abstract
To date compilers of dictionaries of Afrikaans based the allocation of labels such as neologism/fashion words on intuition on the assumption that such words would either disappear from the language in a relative short time or that they may indeed get established and become unmarked entries in future dictionaries. Without a corpus from which the history of frequency of occurrence could be studied such labeling was indeed subjective and compilers were well aware of the arbitrariness of such labeling. In this paper it will be illustrated that studying the frequency of occurrence of such words in a corpus that merely reflects two and a half decades of Afrikaans, enables the lexicographer to draw conclusions on the rise and fall of fashion words.

1 Brief theoretical conspectus

Lexicographic labels are frequently employed in the comment on semantics to give explicit contextual guidance. As pragmatic markers labels are used to relate an item in a dictionary to the world outside the dictionary and they can be used to mark either a macro- or a microstructural item in a dictionary article. When a label is used to mark a lemma sign it implies that all the senses of the word represented by the lemma sign fall within the scope of the label. A label could also be used to mark a specific microstructural item, e.g. a specific item giving the pronunciation or one sense in the treatment of a polysemous item. This implies that only one item falls within the scope of the label.

Although dictionaries use their own domestic labels the majority of these labels can be categorised in a few well-established classes of lexicographic labels. Four major classes of labels that are commonly used in dictionaries are *regional labels, subject field labels, stylistic labels* and *chronolectic labels* (cf. Gouws 1989, Gouws & Prinsloo 2005).

Labels are used to mark deviations form the neutral or default value prevailing in a dictionary. In a special field dictionary dealing with terms from the field of chemistry the lemmata will not be labeled as belonging to the field of chemistry because this is the default value of the items selected for inclusion in the macrostructure of this specific dictionary. How-

ever, when a term from the field of chemistry is selected for inclusion in the macrostructure of a general dictionary it will typically be labeled to indicate that it deviates from the default value of the dictionary, i.e. the general lexicon. The default value of any dictionary represents the zero point of reference. In a general dictionary, either monolingual or bilingual, this zero point of reference is the here, the now and the neutral, i.e. the unmarked centre (cf. Hausmann 1989). A dictionary entry, whether it is a macro- or a microstructural entry, can be regarded as representative of the zero point of reference (a) if it represents the general language as it is used at the time of compilation, (b) if it is used throughout the speech community for whom the dictionary is compiled, and (c) if it is used by the typical member of a well-identified target user group of the envisaged dictionary. These entries would necessarily constitute the majority of the dictionary entries. They form the deictic centre of that specific dictionary. As pragmatic markers labels participate in a process of lexicographic deixis by indicating a deviation from the deictic centre with reference to the spatial, temporal or stylistic nature of the relevant entry, cf. Gouws (1988).

2 Deviating from the deictic centre

Identifying the boundaries of the deictic centre of a specific dictionary and the point where lexicographic labels have to be allocated to an entry is not always an objective assignment. In a general dictionary the lexicographers have little difficulty in determining a deviation in the direction of languages for special purposes. Knowing when a subject field label should be used usually does not confront the lexicographer with too many problems. Likewise objective criteria can be employed to mark a deviation in the direction of regional variety. Where a word is only used in a specific region that deviation can be determined unambiguously and can be indicated by means of a regional label. When it comes to stylistic values distinctions between neutral on the one hand and values like formal, informal, colloquial or different positions on the taboo hierarchy like vulgar, obscene, coarse, etc. demand something more from the lexicographer. Criteria need to be formulated to distinguish between say neutral and informal, between informal and slang or to distinguish between different forms of taboo. In this regard the lexicographer will do well to devise a taboo hierarchy and to allocate different labels, e.g. *vulgar, obscene* and *derogatory* to different positions on this taboo scale. The application of these labels then needs to be applied consistently and the front matter text containing the user's guidelines should include an explicit discussion of this important procedure.

Temporal or chronolectic labels are used when a given entry deviates from the deictic centre's focus on the language as it is used during the time of compilation (in a general synchronic dictionary) or the specific period at which the lexicographic treatment is directed. In a dictionary dealing with a period belonging to the past the deviation could be in two directions – forms representing an older use than the period covered by the treatment, or forms introduced in the final part of that period and actually representing the language of a subsequent period. The present paper is not directed at dictionaries of this type but rather at general synchronic dictionaries dealing with language representing the current time, i.e. the time at which the dictionary is being compiled. In these general synchronic dictionaries temporal de-

viation can only be in one direction, i.e. in the case of words or senses of which the typical use belong to the past but where they still qualify for inclusion in the dictionary on other grounds, e.g. to fulfill the lexicographic function of text reception, cf. Schmidt (1989:659). In this regard labels like *obsolete* or *archaic* mark the relevant deviation. Labeling entries in this way may never be done in a haphazard way and the dictionary conceptualisation plan should determine points on the time scale where a word or sense crosses the boundary to exit the deictic centre. The category of temporal labels in Afrikaans monolingual dictionaries include the label *verouderd* (obsolete) and *verouderend* (antiquating, i.e. moving towards being obsolete). These labels have too often in too many dictionaries been allocated on an arbitrary basis with at most only a vague indication in the user's guidelines text that the first label marks an entry as *verouderend* if it is according to the lexicographer's intuition believed to be on its way to become *verouderd*. A much clearer indication and method of validation are needed to ensure an unambiguous identification by the lexicographer of entries belonging to these categories. This issue will not be discussed in this paper – for a comprehensive discussion see Prinsloo (forthcoming).

In many dictionaries lexicographers also use chronolectic labels to mark entries located on the other end of the temporal scale, i.e. words or senses belonging to the category of neologisms and fashion words. This implies that the deviation from the deictic centre could be in two directions – towards the old (e.g. *verouderd* and *verouderend*), and towards the new. From the perspective of lexicographic deixis this is an invalid implication and lexicographic deixis has no need for this kind of labeling. The inclusion of entries from the pole of the temporal scale opposite to antiquating and obsolete entries does not represent a deviation from the deictic or neutral centre. These entries are included in the dictionary on account of their frequency of use and this constitutes their position as occurring within the deictic centre. Their inclusion implies that they have already been accepted by the speech community (cf. Schmidt 1989:660). Temporal labels marking these entries merely indicate their relative newness but do not imply a deviation from the default value of the dictionary and its deictic centre. From the perspective of lexicographic deixis this labeling demonstrates a different approach compared to the use of the other types of labels.

3 Explicit and condensed labeling

Because dictionaries have to reflect the real language and the development of the lexicon lexicographers often feel compelled to employ labels marking neologisms in an attempt to offer their users some additional guidance. In some cases neologisms are labeled to emphasise an aspect of language development (cf. Schmidt 1989:661). Quite often lexicographers do not explicitly label a word as a neologism or a fashion word but by means of a highly condensed procedure of labeling the paraphrase of meaning contains a formulation allocating that status to the given word. Whether explicit or condensed labeling is used, the lexicographer is yet again faced with the challenge of objectivity. When is a word still a neologism, and therefore a candidate for labeling, and when should it be regarded as an established item of the lexicon with no label-bearing potential? Also in this endeavour it is extremely important that a set of criteria should be developed which can be applied consistently to ensure the proper treatment of these entries.

Utilizing a system of lexicographic labels places a very definite responsibility on the shoulders of the lexicographer. This responsibility regards the compilation of a new dictionary but also the revision of an existing dictionary. Language is dynamic and changes constantly – this does not only apply to the lexicon where new words and senses appear and old ones disappear but it also applies to the network of relations holding within the lexicon and to the pragmatic relation between a word or sense and the extra-linguistic world.

4 Labeling and the revision of a dictionary

When revising a dictionary lexicographers have often been accused of committing plagiarism in alphabetical order. Unfortunately an uncritical approach to existing dictionary entries prevails too often when the revision process is under way. This is especially true with regard to the lexicographer's approach towards existing lexicographic labels. Labels are regarded as permanent fixtures qualifying a given entry. In stead, lexicographers need to be aware of the fact that labels only reflect the pragmatic anchoring of a given entry at the time of compilation and this allocation needs to be re-evaluated whenever a dictionary is revised.

Because of language dynamics words move closer to or further away from the deictic centre of a given dictionary and this ever-changing process should be reflected by means of the labeling system of that dictionary with not only the labeling but also the non-labeling as well as the unlabeling of entries giving evidence of the linguistic reality. Even when it comes to the use of so-called objective labels, e.g. subject field or regional labels, the lexicographer may not assume that they should have an unchanged occurrence in the dictionary. A monosemous word belonging to a language for special purposes may acquire a new and general sense which needs to be excluded from the scope of a prevailing subject field label, or the use of a word with a former regional restriction may expand beyond the regional borders and should consequently lose its regional label. These changes in the labeling cannot be done by merely responding to the gut feeling of the lexicographer. A much more objective instrument is needed. A well-balanced and representative corpus offers these possibilities to the lexicographer.

5 Using a corpus in the revision of neologisms and fashion words.
5.1 Distinguishing between neologisms and fashion words

When a dictionary uses chronolectic labels like *neologism* or *fashion word* this allocation needs to be re-evaluated whenever the dictionary is revised. In this regard a corpus has to be regarded as the most suitable instrument to guide the lexicographer. A corpus can assist the lexicographer to identify changes in the relative proximity of any given word or sense to the deictic centre, as determined for a specific dictionary. Entries that had been unlabeled in a previous edition of a dictionary may have become obsolete or the corpus may indicate a far lesser frequency of usage that moves the entry to the domain of becoming obsolete. Using a corpus and fixed points on the time scale, e.g. not having been used more than X times in Y years, will help the lexicographer to employ a label or to change the current label of a given entry.

If the lexicographer feels the use of labels to mark neologisms and fashion words as newcomers to the domain of the deictic centre could enhance the value of a dictionary as an in-

strument reflecting the real and actual language use, a corpus should play an important role in these decisions. If there is a time span of say ten years between two editions of a given dictionary the classification as neologism in the first edition should not be valid any longer in the second edition and if the specific entry still qualifies to maintain its inclusion in the dictionary it should lose the label *neologism*. On the other hand the inclusion of new entries, whether new lemmata or new senses of an existing lemma, should also be considered as carriers of the label of neologism.

An important function of a corpus regards the re-evaluation of the labeling allocated to existing entries in a dictionary. In this regard it could be beneficial to make a distinction between neologisms and fashion words. In Afrikaans dictionaries like the *Woordeboek van die Afrikaanse Taal* (henceforth abbreviated as WAT) and the *Verklarende Handwoordeboek van die Afrikaanse Taal* (henceforth abbreviated as HAT) a clear distinction is made between neologisms and fashion words. When the definitions for *neologisme* 'neologism' and *modewoord* 'fashion word' in WAT and HAT are compared it is clear that *neologismes* are regarded as new words or phrases but that *modewoorde*, although also defined as 'new words', exhibit certain (even negative) characteristics such as 'not expected to remain for a long period as part of the vocabulary', 'somewhat uncommon', 'replacing other words because it sounds "learned/fancy"', 'private creations' frequently used for a limited time. Neologisms being "a new word or expression in a language, or a new meaning for an existing word or expression", cf. COBUILD 1995, typically become established members of the language where fashion words often have a limited occurrence as active participants in a given language. They often have a rapid rise but also a rapid fall in their usage frequency. While rising or still being used they need to be included in a dictionary.

As a source presenting the evidence from actual language usage in an unbiased way a standard general dictionary may discriminate between neologisms and fashion words but it may not discriminate against fashion words. If the occurrence of any given word in the corpus adheres to the minimum criteria for inclusion the lexicographer is compelled to include that word as a lemma in the dictionary. By including neologisms and fashion words that qualify as lemma candidates in a dictionary the lexicographer has obeyed one half of his/her responsibility. If the system of labeling followed in the specific dictionary employs labels like *neologism* and *fashion word* the lexicographer has to determine which one of these labels applies to a specific entry – criteria for the distinction between fashion words and neologisms will not be discussed here. In the treatment of neologisms and fashion words lexicographers are well aware of the fact that they normally lack sufficient evidence to allocate labels and that delicate differentiation is required in the allocation of labels with surety and that they are dependent on their intuition.

When revising a dictionary the lexicographer has to re-evaluate all the labels allocated to entries in the dictionary. In the case of neologisms the typical procedure could be to unlabel these words and keep them included as unmarked entries – implying that they are established entries functioning within the boundaries of the deictic centre. Re-evaluating the status of fashion words may sometimes see these items having become established members of the lexicon of the specific language but quite often the corpus could indicate that these items

have moved through the deictic centre and are on their way out, contrary to the situation with many neologisms that enter the language and display an often slow but increasing and sustained frequency of use. If labeled as neologisms the typical procedure in a subsequent revision of the dictionary would be the unlabeling and consequent implication that the word has become part of the established vocabulary of the language whereas many fashion words burst onto the scene and enjoy a brief but energetic spell before fading into oblivion. The re-evaluation of a fashion word may see that word qualifying to maintain its position in the macrostructure and be elevated to the status of an unlabeled entry. However, it may also witness the word failing to make the cut for continued inclusion in the dictionary or having to change labels to indicate that the fashion word is now regarded as an obsolete form. Drastic changes like these are only possible with evidence from a balanced and representative corpus. Where a fashion word becomes an unlabeled entry it has had a long walk to lexicographic freedom; where it is omitted or labeled as obsolete the corpus gives evidence that the word has been nothing more than a lexical fly-by-night.

The remainder of this paper focuses on evidence from a corpus to show the trajectory of a number of neologisms and fashion words in Afrikaans. It is shown how the trajectory of a number of function words as well as a few core vocabulary items bear witness to their sustained occurrence in Afrikaans. The names of a few politicians show their respective ups and downs while some neologisms establish their position in the lexical team while a number of fashion words witness an early exit.

5.2 Utilizing an Afrikaans corpus to detect trends and changes in the use of neologisms/ fashion words

The Afrikaans corpus utilised for this study is the Media 24 archive, from 1985 to 2003. It contains the newspapers *Beeld, Rapport* and *Die Burger* as well as the magazine *Landbou Weekblad.* This archive/corpus contains 749,000,000 tokens and 2.8 million types of which sligthly more than 50% are hapaxes. In order to detect trends and changes in the use of neologisms / fashion words, the corpus was divided into four sections roughly covering five-year periods.

	Block 1	Block 2	Block3	Block 4
Period:	Up to 1989	1990-1994	1995-1999	2000-2003
Tokens	59.0 million	171.9 million	253.9 million	264.4 million

Table 1. An arbitrary division of the *Media24* Afrikaans Archive

From Table 1 it is clear that although the blocks are relatively similar in respect of the number of years covered there are substantial differences in the number of tokens per block. So, for example Block 4 reflects 264 million tokens which is four times the number of Block 1 and can therefore not be directly compared since a frequency count of say 300 in Block 4 will still be relatively less than a count of 100 in Block 1. Blocks 1-4 have therefore been recalculated to reflect the frequencies of the words per 50 million words. Compare Table 2 illustrating the frequency and consistency of some top function words in Afrikaans:

| Word	89/50M	94/50M	99/50M	2003/50M
DIE 'the'	3461460	3541101	3469709	3269432
VAN 'of'	1476852	1466624	1381944	1273209
HET 'has, have'	1229819	1163842	1126305	1076879
EN 'and'	1137866	1145804	1113085	1096586
IN	1048957	1086543	1084421	1059205
'N 'a, an'	1017268	1066178	1044576	999591

Table 2. Consistency in the frequency of use of the top function words in Afrikaans.
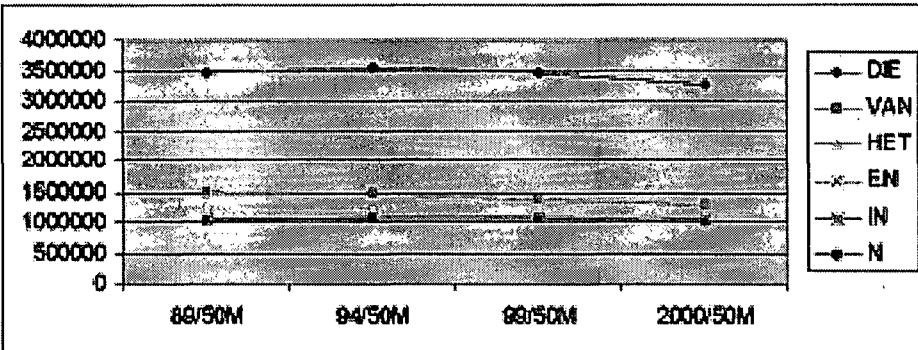
Consider also Table 3 for similar results for randomly selected core vocabulary items.

Word	89/50M	94/50M	99/50M	2003/50M
EET	2352	2595	2660	3123
WERK	27180	32593	35614	35470
MENS	26405	27830	26936	28346
BAIE	68789	70691	69849	70826

Table 3. Consistency in the frequency of use of *eet* (eat), *werk* (work), *mens* (person) and *baie* (many)

307
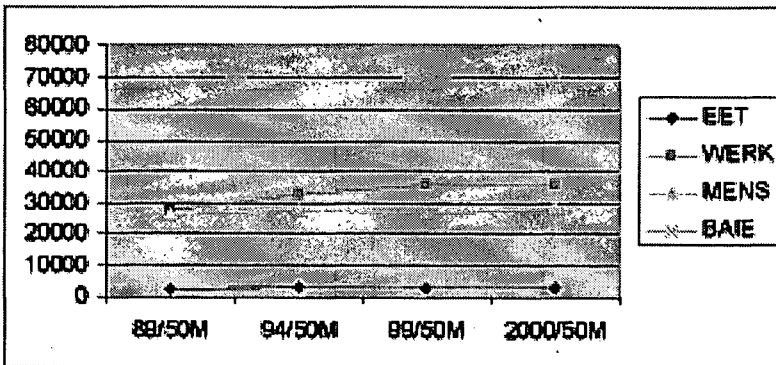
In Afrikaans the word *missie* (mission) is no neologism. As a polysemous lexical item it has displayed senses like "important assignment", "representation of a state at another government" and "mission work under the heathen". During the past decade a new sense was added, i.e. "the goals and objectives of a person or institute". The Afrikaans dictionary HAT labels this sense as *modewoord* (fashion word). The Afrikaans woord *bosberaad* (bush deliberation – a meeting held in a secluded place) has burst onto the lexical scene during the last decade. *Regstellende* (affirmative) in the Post-Apartheid political sense of 'regstellende aksie' (affirmative action), dominated the restructuring of the South African society after the 1994 elections. The corpus gives the following evidence of the occurrence of *missie* (no indication is given of the specific sense that applies) *bosberaad* and *regstellende*:

			Total	Block 1	Block 2	Block3	Block 4
Period:				Up to 1989	1990-1994	1995-1999	2000-2003
Frequency:	missie		5694	127	1099	2678	1786
Frequency per 50m. tokens:				108	320	527	338
Frequency:	bosberaad		2385	0	738	1019	628
Frequency per 50m. tokens:				0	205	201	119
Frequency:	regstel-lende		22575	53	4764	11155	6599
Frequency per 50m. tokens:				45	1386	2196	1248

Table 4. Spreading of the frequencies of *missie* and *bosberaad* in the *Media24* archive

It is clear that the frequency of use of *missie* in the late nineties was 5 times higher than in the late eighties but that it's frequency has declined in the first few years of the new millennium to the levels of the early nineties. Exactly the same pattern is observed for *bosberaad* and *regstellende*.

Compare the following table of the words *gonswoord* (buzz word), *scenario* (scenario), both labeled as neologisms in HAT and the words *uitset* (output) and *bemagtig* (empower), both words having a relatively new but unlabeled sense:

Word	89/50M	94/50M	99/50M	2003/50M
SCENARIO	58	236	265	325
UITSET 'output'	19	57	67	86
GONSWOORD 'buzzword'	0	1	12	24
BEMAGTIG 'empower'	1	44	459	632

**Table 5.** Fashion words and neologisms *scenario, uitset, bemagtig* and *gonswoord*
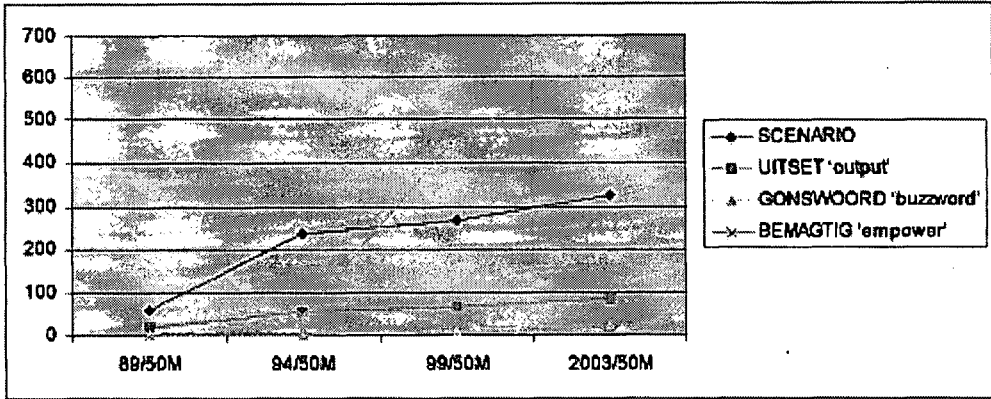
In contrast to *missie and bosberaad*, the graphs of *scenario, uitset, bemagtig* and *gonswoord* all indicate a slow but steady growth and the difference in frequency of occurrence between these words. Based on these statistics the compilers of the specific dictionary could consider labeling the specific sense of *missie*, and *bosberaad* in a revised edition e.g. as *verouderend* (antiquating) whereas *scenario, uitset, bemagtig* and *gonswoord* can be unlabeled.

Validation for rising and falling trends in frequency as seen in Tables 4 and 5 can be found in the following three test cases involving the names of political leaders in South Africa.

Word	89/50M	94/50M	99/50M	2003/50M
MBEKI	445	1172	6383	11437
KLERK	9328	17860	5515	3652
BUTHELEZI	1331	3394	1775	771
MANDELA	4507	15178	15025	4943

Table 6. The rise and fall of political leaders in terms of frequency in the corpus

309
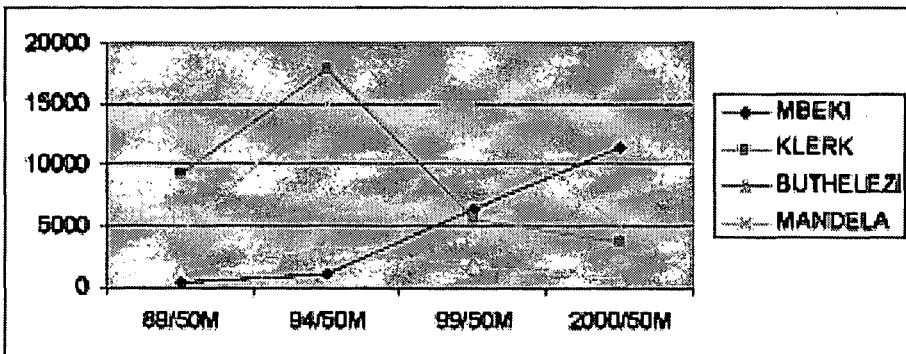
From these graphs it is clear how the media interest in political parties has declined since the high of the 1994 elections and the diminishing attention to former President F.W. de Klerk of the now dissolved National Party and Mangusuthu Buthelezi of the Inkatha Freedom Party in contrast to reigning President Thabo Mbeki. Former presidents Nelson Mandela and F.W. de Klerk still enjoy much attention though much lower than the consistent high enjoyed during the time of their presidency.

5.2.1 Compairing word pairs

The frequency counts can be very useful in the monitoring of word pairs reflecting loans versus coined words. Table 7 indicates that *CD* 'compact disk' is winning the race against *laserskyf* 'compact disk' in spectacular fashion. Where *laserskyf* has been used five times more than *CD* at the end of the eighties it is now used 200 times less than *CD*.

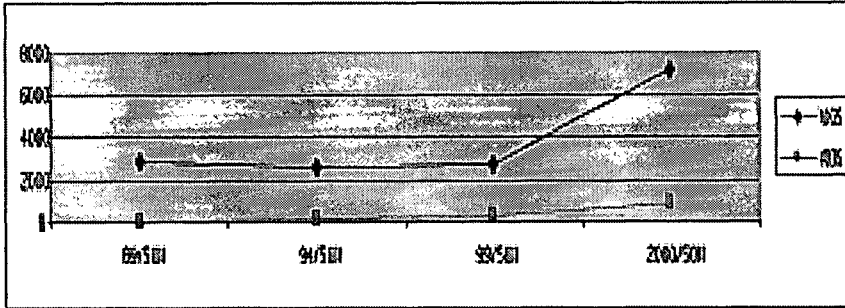| Word	89/50M	94/50M	99/50M	2003/50M
LASERSKYF	152	196	39	21
CD	32	647	3872	4679

Table 7. *Laserskyf* versus *CD*
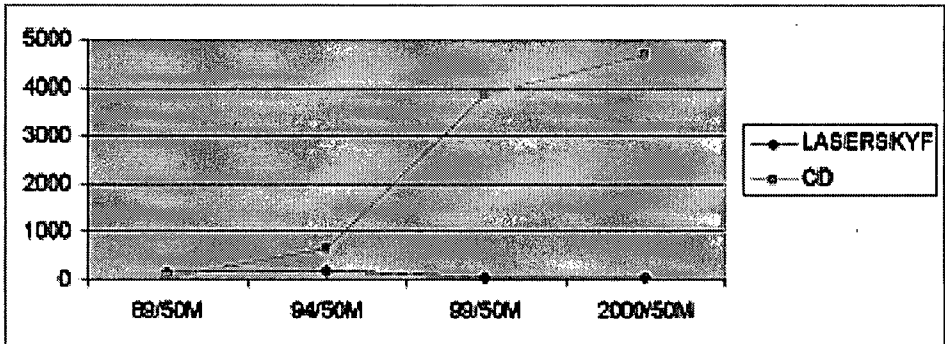
From Table 8 it is clear that both *Vigs* and *Aids* (Aids) have increased their frequencies by 400% though the overall use of *Aids* is still ten times lower than that of Vigs.

| Word	89/50M	94/50M	99/50M	2003/50M
VIGS	2836	2586	2700	7176
AIDS	53	82	223	790

Table 8. *Vigs* versus A*ids*

As is the case with other languages Afrikaans has also been bombarded with acronyms and abbreviations. During the last decade or two numerous acronyms entered the lexicon of Afrikaans. Many of them had a short spell whilst others became established items of the lexicon. A corpus also assists the lexicographer in his/her decision regarding the continued inclusion of a specific acronym or abbreviation in a dictionary. From Table 9 it is clear that acronyms referring to political parties/initiatives, peace forces, tax systems, etc. that were discontinued rapidly fall in frequency.

Word	89/50M	94/50M	99/50M	2003/50M
TBVC	152	573	175	21
AVB	773	374	63	33
KONTRALESA	0	19	51	11

Table 9. *TBVC, AVB* and *KONTRALESA*
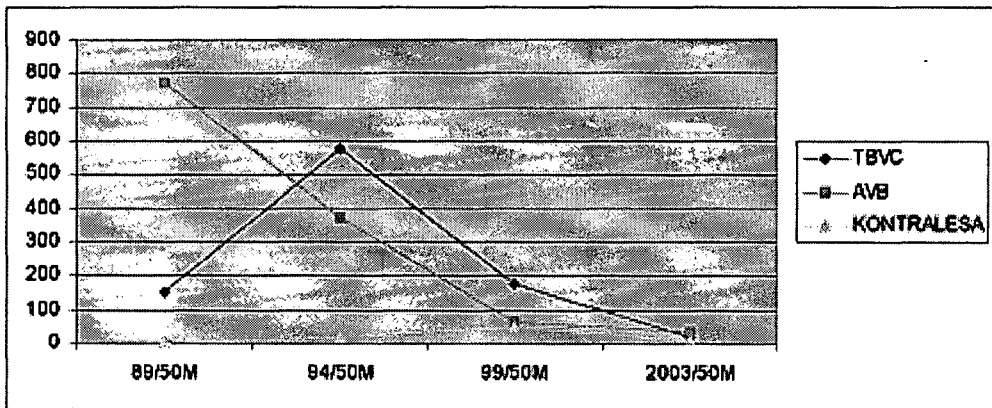
6 In conclusion

Given the availability of corpora reflecting language use over a number of decades, the allocation of temporal labels such as neologism/fasion should no longer be subjectively or

solely determined on the intuition of the lexicographer. Statistics drawn from such corpora can give a clear indication to the lexicographer of rising and falling trends in the use of such words and labeling should be adjusted accordingly with each revision of the dictionary. In this way a clear distinction can be made between those words that completed the long walk to unmarked lexicographic freedom, the lexical fly-by-nights that deserve only a limited and labeled stay in a dictionary and those words that remain in dictionaries, displaying a well-established deviation from the deictic centre, sentenced to life long labeling.

References
A. Dictionaries
(COBUILD) Sinclair, J. (ed.) (1995), *Collins COBUILD English Dictionary.* London: HarperCollins.
(HAT) Odendal, F.F & R.H. Gouws (2005), *Verklarende Handwoordeboek van die Afrikaanse Taal.* Cape Town: Pearson Education.
(WAT) Van Schalkwyk, D.J. et al (eds.) (1996; 2000) – *Woordeboek van die Afrikaanse Taal.* Bureau of the WAT: Stellenbosch.

B. Other Literature
Gouws, R.H. (1988), 'Die gebruik van etikette as leksikografiese hulpmiddel', *S.A. Journal of Linguistics*, Occasional Papers, nr. 6.
Gouws, R.H. (1989), *Leksikografie.* Cape Town: Academica.
Gouws, R.H., Prinsloo D.J. (2005), *Principles and Practice of South African Lexicography.* Stellenbosch: African Sunmedia.
Hausmann, F.J. (1989), 'Die Markierung im allgemeinen einsprachigen Wörterbuch: eine Übersicht', in Hausmann, F.J. et al. (eds.) (1989-1991), pp. 649-657.
Hausmann, F.J. et al. (eds.) (1989-1991), *Wörterbücher. Dictionaries. Dictionnaires. An International Encyclopedia of Lexicography.* Berlin: Walter de Gruyter.
Media24. *http://www.media24.co.za*
Prinsloo, D.J. (Forthcoming). *Using corpus data in the application and validation of the temporal labels verouderd and verouderend in Afrikaans monolingual dictionaries.*
Schmidt, G.D. (1989), 'Diachronische Markierungen im allgemeinen einsprachigen Wörterbuch', in Hausmann, F.J. et al. (eds.) (1989-1991), pp. 657-661.