

Prefix	Non-conforming words
DE-	DEEG, DEEL, DEEM, DEER, DEES, DEFTIG, DEIK, DEKBL, DEKG, DEKH, DEKK, DEKLAAG, DEKM, DEKOES, DEKP, DEKRIET, DEKS, DEKT, DEKV, DELF, DELG, DELS, DELT, DELW, DEMP, DENI, DENK, DENN, DENS, DENT, DEPP, DERM, DEUG, DEUN, DEUS, DEUT
DIA-	DIAR
DIS-	DISEN, DISPENS, (except for DIS-PENSASIE)
EEN-	EENDS, EENDV, EENSD, EENSG, EENSK, EENSL
EKS-	EKSA, EKSE, EKSİ, EKSO, EKSPAN, EKSPEK, EKSPERT, EKSPI, EKSTA, EKSTER (except for EKS-TERI, EKS-TERR), EKSTR (except for EKS-TRAD, EKS-TRAH, EKS-TRO, EKS-TRU)
EKWI-	none
EN-DO-	ENDOSS
ER-	ERD, ERE, ERF (except for ER-FENIS), ERI, ERNS, ERO, ERTJIE, ERTS, ERU
ET-NO-	none
GEO-	none
GE-	GEE (except for GE-Ë), GEI (except for GE-Ï), GEKH, GEKK, GEKLIK (except for GE-KLIKKLAK), GEKS, GELD, GELL, GEMM, GEMS, GENRE, GENS, GENT, GEREËL, GERF, GERM, GESPE (except for GE-SPESI), GESTE, GESTIK, GEU (except for GE-ÛR)
HEK-SA-	HEKSAAN
HE-MI-	none
HER-	HERDS, HERE (except for HER-EK, HER-EN), HERF, HERO", HEROUT
HE-TE-RO-	HETEROPS
HIER-	HIERTS
HI-PER-	none
HI-PO-	HIPOST
HO-MO-	none
IDIO-	IDIOO
....	
etc.	

5.2 Prefix Combinations

Multiple prefixes can occur. Kempen (1977) does discuss the topic in § 142 but does not give an exhaustive list. He appears to believe that only pairs occur in practice, and that duplications are rare e.g. HER-HER-SIEN. However, we have seen ON-AAN-GE-KONDIG which comprises three prefixes, but presumably is also a rare feature. Those combinations listed below are some of the possibilities that occur in our source word list and also in Marais (1970).

AAN-GE-bede, AAN-BE-vole

AARTS-VER-borgenheid

AF-GE-handel

BE-ANT-woord, BE-GE-lei

GE-DE-moraliseer, GE-HER-kou, GE-~~PRE~~-mediteer, GE-RE-formeer,

GE-TRANS-formeer

HER-AAN-pas, HER-BE-draad, HER-OOR-weeg, HER-ONT-dek, HER-OP-bou,

HER-OWER-baar, HER-VER-hoor, HER-UIT-send, HER-VER-deel, HER-WAAR-deer

HIPER-GE-voelig

MIS-GE-was, MIS-VER-stand

OER-BE-woner, OER-GE-steente

ON-AAN-GE-kondig, ON-BE-antwoord, ON-ER-kentlik, ON-GE-nooi,

ON-HER-stelbaar, ON-MIS-kenbaar, ON-ONT-kombaar, ON-VER-hoor

VER-GE-wis, VER-ON-reg, VER-ONT-skuldig

WAN-BE-stuur, WAN-GE-spoor, WAN-VER-hoor

Since there are many more that could be compounded from our prefix list, we do not show them all. It is sufficient to observe that multiple prefixes do occur and therefore prefixes must be searched for several times during the algorithm.

Chapter 6 SUFFIX RESEARCH

6.1 Suffixes

What do we mean by a 'suffix'? Again, we are really concerned with a morpheme which forms the end of a word, that can stand alone as a pronounced item, but is not a root. Thus, although S is a morpheme, it is not a pronounceable syllable and is therefore excluded. Likewise, although E can be a syllable as in VRA-E, our typographic rules preclude its acceptability.

There are also many common endings for words, e.g. -LAND, that should also be considered as suffixes but, as we discussed in Chapter 5, most of these features comprise meaningful words on their own and are necessarily excluded at this stage.

Taking the major list in Kernpen (1977), together with the items listed in Kotze (1979) and Eksteen (1980) and amalgamating these, we arrive at a workable set of suffixes.

The suffix elements are listed below in alphabetical order of their final letter together with all the words that do not conform to the appropriate rule for breaking each suffix. These exceptions were determined by inspection of our word list reproduced in Gee (1987). The suffixes cannot necessarily be used in the order shown because, for example, a test for JIE will produce -JIE, whereas the word may be PYLTJIE, which breaks PYL-TJIE. Thus TJIE must be tested before JIE.

The single letters -E and -S need to be treated separately. We could list here every possible combination of suffix with these two elements, but an easier approach is to use the following empirical rule. "If the unit ends in an S, then observe the morpheme before it and if that is a valid suffix, remove the suffix and the S together." This will also cover the fact that although a word may end in HEID, HEDE and HEDES, no word may end in HEIDS. However, HEIDS does occur *within* compounds and multiple suffixes.

A similar rule can be used for E endings. We also note that the ending ES can occur, and that both these rules and the suffix rules must be applied in that case.

In the notation used below, v represents a vowel, and c a consonant.

Suffix	Non-conforming words
-AARD	BAARD, DAARD, GAARD (except for GIERIG-AARD), HAARD, JAARD, BLAARD, KLAARD, EWENAARD, PAAFD, VAARD, WAARD
-cEERD	SJEERD
v-EERD	none
-cERD	none
vv-ERD	none
-HEID	none
-HE-DE	none
-KUN-DE	none
-DE	none
-RI-GE	none
-LO-GIE	none
c-TJIE	BULTJIE, PRENTJIE, RANTJIE, PLANTJIE, KAARTJIE, HARTJIE, KWARTJIE, ERTJIE
AT-JIE	BATJIE, MAMATJIE, OUMATJIE, PAPATJIE, OUPATJIE, STRATJIE
E-TJIE	KADETJIE, PARKIETJIE, MIETJIE, GRIETJIE, SPRIETJIE, KIEWIETJIE, ETIKETJIE, KROKETJIE, TAMELETJIE, SNOETJIE, VOETJIE
IT-JIE	AAITJIE, OOITJIE, EITJIE
OT-JIE	FOTOTJIE
U-TJIE	NUUTJIE
Y-TJIE	none
-JIE	none
v-ASIE	none
vE-SIE	none
vi-SIE	none
c-SIE	none
-cvSIE	BLASIE, PLASIE, SPASIE, GRASIE, KRASIE, TRASIE, STASIE (except for vS-TASIE), SPEDISIE, KRESIE, KLUSIE
-cIE	VLIE, KNIE, GAPIE, UAPIE, BRIE, DRIE, TRIE
-cIEKE	BLIEKE, PLIEKE, BRIEKE, GRIEKE, KRIEKE, TRIEKE
v-IEKE	none
-LI-KE	none

Suffix	Non-conforming words
-RY-KE	none
-cISME	SCISME, NGISME, CHISME, DHISME, KHISME, PHISME, TRISME, RXISME
-cIN-NE	none
-SKAP-PE	none
-cA-RE	SKARE, BLARE, PLARE, TSARE
-TO-RE	none
-cANSE	FRAAN'SE
-TRI-SE	none
-LO-SE	none
v-AAN-SE	none
-cAAN-SE	none
-cES-SE	none
-SEUSE	none
-GE-WY-SE	none
c-STE	BEKRANSTE, OMKRANSTE, DIENSTE, GUNSTE, VERFLENSTE, GEFRONSTE
vS-TE	none
-TE	none
IE-RIG	none
-ERIG	none
-RIG	VAANDRIG, LANGORIG, RUMOERIG, WRIG
-AG-TIG	MAGTIG, KRAGTIG, PRAGTIG, WRAGTIG
-TIG	GESTIG, KAMSTIG, BYKOMSTIG, HERKOMSTIG, EENKOMSTIG, OOREENKOMSTIG, GOEDGUNSTIG, BENAARSTIG, WEERBARSTIG, ONTSTIG
-cIG	cPLIG (except for LAMP .IG)
...	
etc	

6.2 Suffix Combinations

Multiple suffixes can occur. Kempen (1977) discusses the topic in § 142 and gives much detail but not an exhaustive list. Those combinations listed below are some of the possibilities that occur.

In general a suffix sequence can comprise up to 5 components, $S^1 S^2 S^3 S^4 S^5$.

S^1 can be one of the possibilities listed above.

S^2 comprises: -HEID, -IE, -JIE, -TJIE, -ASIE, -AGTIG, -ERIG, -ING, -DOM, -SKAP, -BAAR, -LOOS, -S, ...

S^3 comprises: -HEID, -IE, -JIE, -TJIE, -ING, -DOM, -SKAP, -BAAR, -S, ...

S^4 comprises: -IE, -JIE, -TJIE, -S

S^5 can be only -S.

The following words are examples of using combinations from our suffix list, after applying correct hyphenation rules to separate the resulting syllables.

leuen-AGTIG-HEID
onverdun-BAAR-HEID, debat-TEER-BAAR-HEID
eien-DOM-LIK, eien-DOM-LI-KE
moeg-ERIG-HEID
bewe-GING-LOOS
heer-LIK-HEID
huis-LOOS-HEID, betekenis-LOOS, betekenis-LOOS-HEID
klae-RIG-HEID, ongehoor-SAAM-HEID
vriend-SKAP-LIK, vriend-SKAP-LI-KE, vriend-SKAP-LIK-HEID
krag-TE-LOOS, beslui-TE-LOOS-HEID
onderwyse-RES-SE

Care is needed when suffixes such as -DE and -TE are to be deleted. Consider the word ONGEDEERDE. If the -DE is first deleted, the DEERD suffix will not be found. Instead, the E can be temporarily be removed and suffixes ending in D searched for. In this case one is found resulting in ONGE-DEER-DE, which is acceptable.

Chapter 7 COMPOUND BOUNDARIES

Having identified separable affixes, two more problems have to be investigated: the syllable structure of root words, and the various root compounds that can occur.

7.1 Root Syllables

Syllable structure is always a consonant cluster followed by a vowel cluster, the simplest of which is CV. A tabulation follows of clusters detected. Here we are only concerned about syllables within stems, not prefixes, suffixes, or compounds.

The root word may be single- or multi-syllable and the structure of these may vary according to its position in the word. If we start with the premise that a typical syllable consists of Consonant and Vowel clusters e.g. ...CCVV..., then it should be possible to formulate rules concerning these breaks, and the words or clusters that do not conform. In order to test these cases we referred to Eksteen (1980), used De Villiers (1976) as the source material, and also an ordered list of vowel consonant substitutions in our source word list, an extract of which is included in the Appendix. This last was found to be of limited use since it was not verified for spelling mistakes before conversion, and it is not possible to inspect the source word against it.

A reference list is given below. See also de Villiers (1976, p 130).

VC	algemeenste
VCC	enggeestig
CV	eksamen
CVC	eksamen
CVCC	lomp vang
CVCCC	aanbevelings
CCV	waarsku
CCVC	skyn slim
CCVCC	stand
CCVCCC	kwarts
CCCV	skrywer
CCVCV	skryf
CCCVCC	streng

CVV	verbouding
CVVC	vervoer waar
CVVCC	standaard
CVVCCC	afwaarts
CCVV	steuring
CCVVC	kwaad bruid slaan
CCVVCC	staats
CCCVV	skree
CCCVVC	spleet spraak
CVVV	leeu
CCVVV	sneeu
CCCVVV	strooi skroei
CCCVVVC	stroois

7.1.1 Vowel Clusters

Certain vowel clusters can never be broken, and some always. The borderline cases are rare and have not been elaborated, but tend to be foreign words. In the table below we give typical examples of both types. Whenever a diaeresis occurs, a break is allowed before it, whatever the CV configuration.

Table: Single Vowel Clusters - always break after vowel

	v cv
A	BA-NE
E	BE-KER
I	TI-PE
O	SLO-TE
U	MU-RE
Y	BY-BELS

Table: Two Vowel Clusters - break where indicated

A	A	E	I	O	U	Y
A	BAAT	VRA-E	r	r	r	r
E	BE-AMPTE	LEES	PEIL	TEORIEË	NEUS	r
I	r	VIER	r	r	r	r
O	r	VOER	TOI-LET	VOOR	HOUT	r
U	SITU-ASIE	BRÛE	RUIMTE	r	HUUR	r
Y	r	r	r	r	r	r

where r means rare or never

Table: Three or more Vowels

	cvvv
AAI	BLAAI-
AAIE	PAAI-E
AIE	BAI-E
EEU	LEEU-
EIE	KEI-E
EUE	LEU-EN
IAA	KI-AAT
IEE	FINANSI-EEL
IOE	PENSI-OEN
OEI	BLOEI-SEL
OEIE	BLOEI-ER
OIE	NOI-ENS
OII	TOI-INGS
OOI	MOOI-
OOII	MOOI-IGHEID
OUE	HOU-E
UII	PLUI-INGS

Inspection of these three tables reveals reliable rules for hyphenating vowel clusters.

7.1.2 Consonant Clusters

Consonant cluster that begin words or roots are listed as **Initial**. Those that occur within words as **Medial**; those clusters that occur at the ends of roots are called **Final**.

Initially: (Anlaut)

Singles B D F G H J K L M N P R S T V W

Pairs BL BR

CH

DR DW

FL FR

GL GR

KH (rare) KL KN KR KW

MN (rare)

PL PR PN (rare) PS

SF (rare) SG (rare) SJ SK SL SM SN SP ST SW

TJ TR TS (rare) TW

VL VR

WR

Triples SKL (rare) SKR SPL SPR STR

Medially: (Inlaut) There seem to be a large number of combinations that occur within roots, and we have been unable to trace all of them. Some of the readily accessible ones are listed below with examples of their use. For the purpose of identifying compound boundaries, it may be clearer to identify actual syllables rather than medial consonants.

BB BL BR BS

DD DJ DR

FF FL FR FT

GG GN GR GT

KK KL KR KS KW

LG LJ LK LL LS

ML MM MP

ND NG NK NN NS NT

PL PP PR

RD RG RP RS RT

SF SJ SK SM SP SS ST

TR TS TT

dubbel, publick, fabriek, abses

padda, adjektief, adres

snuffel, refleks, refrein, deftig

oggend, magneet, program, agter

sakkeroller, baklei, akrobaat, aksent, lukwart

algebra, baljaar, elke, hulle, kalsium

gomlastiek, jammer, trompet

ander, honger, tinkinkie, nonna,

bensien, drentel

diploma, knuppel, depressie

harder, argief, torpedo, kursief, artikel

fosfor, pasja, biskop, jasmyn, respek,

passief, kaste

katrol, bitsig, letter

Triples	FFR KST KTR	affront/saffraan, tekstuur, aktrise
	LGR MPL NCH	pelgrim, amplitude, sinchronies
	NGL NGR NKR	anglisme, kongres, bankrot
	NKT NTR PPL	puntuasie, sentraal, supplement
	RST RTR SKR STR	worstel, portret, diskrimineer, astronoom

Finally: (Auslaut)

Singles	B D F G K L P R S T
Pairs	DS
	GD GT
	KS KT
	LD LF LG LK LM LP LS LT
	MD MP
	ND NG NS NT
	RD RF RG RK RM RP RS RT
	TS
	FTS KTS LKS LPS NDS NGK* NGS NKS RFS RTS
Triples	

7.2 Root Compounds

Afrikaans is fond of glueing words together to form new words or new meanings or both. The words used in these compounds can be simple or complex; that is to say, they can vary from a pure root such as BOU, through intermediate forms such as AFVAL, to roots with multiple prefixes and multiple suffixes such as WAARSKYNLIKHEID. Nor is this the end of the complexity. The 'word' may itself be already compounded.

In formal BNF notation, the structure of an Afrikaans word is defined as follows.

```

<Afrikaans word> ::= <compound word> | <simple word>
<compound word> ::= <compound word><simple word> | <simple word>
<simple word> ::= <root> | <prefix cluster><root><suffix cluster> |
                <prefix cluster><root> | <root><suffix cluster>
<prefix cluster> ::= <prefix>3
<suffix cluster> ::= <suffix>5

```

* listed by de Villiers (1976) on p92, but we can find no example and have assumed it is a mistake.

<prefix> ::= listed in Chapter 5

<suffix> ::= listed in Chapter 6

where the superscripts indicate the maximum number of elements of this type.

Examples are

LANGTERMYNPROJEK

root + root + root

VOORTREKKERBEWEGING

prefix + root + suffix + prefix + root + suffix

For the purposes of this work, we need to be able to identify the boundaries between *simple words*.

There are four cases that concern the boundary between <simple word> and <simple word>. The first element may end with a suffix and the second be an unaffixed root. Then the first element may be a root, and the second begin with a prefix. Thirdly, the boundary may be between a suffix and a prefix. Lastly, there may be no affixes, leaving only a root followed by a root.

These are addressed below for each instance.

7.2.1 Suffix/Root

At any point in our search through the word for syllables, it should be possible to undertake a suffix check without difficulty. Thus when we encounter the feature *suffix* followed by *root* as in

BURGERLIKE-KANTOOR

the suffix LIKE will be found and we can hyphenate *after* it as well as before and within it.

7.2.2 Root/Prefix

It should be possible to undertake a prefix check within a compound word. Thus when we encounter the feature *root* followed by *prefix* as in

LAA'S-GENOEMDE

the prefix GE will be found and we can hyphenate *before* it as well as after it.

7.2.3 Suffix/Prefix

When we encounter the feature *suffix* followed by *prefix* as in

BURGERLIKE-BESKERMING

we can adopt the same methodology as above for the Suffix/Root feature. Thus the suffix LIKE will be found and we can hyphenate *after* it.

7.2.4 Root/Root

This is a really difficult area because almost any letter can terminate a word, and any letter begin the following word. Thus the compound break can only be detected if the consonant clusters provide a unique determination. This clearly requires further investigation. We see 4 cases

consonant-consonant	STREEKS-KANTOOR, TAAL-SEGMENT
consonant-vowel	HOOG-OOND, KNIP-OOG
vowel-consonant	SEE-VAREND, HOU-VAS
vowel-vowel	EEUE-OUE

This last is the only one that we can currently see a solution for, by using the tables given earlier in this chapter.

Author Gee Quentin H

Name of thesis Automatic Hyphenation Of Afrikaans. 1987

PUBLISHER:

University of the Witwatersrand, Johannesburg

©2013

LEGAL NOTICES:

Copyright Notice: All materials on the University of the Witwatersrand, Johannesburg Library website are protected by South African copyright law and may not be distributed, transmitted, displayed, or otherwise published in any format, without the prior written permission of the copyright owner.

Disclaimer and Terms of Use: Provided that you maintain all copyright and other notices contained therein, you may download material (one machine readable copy and one print copy per page) for your personal and/or educational non-commercial use only.

The University of the Witwatersrand, Johannesburg, is not responsible for any errors or omissions and excludes any and all liability for any errors in or omissions from the information on the Library website.