

Bibliografie

- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. 1999. Longman grammar of spoken and written English. London: Longman.
- Bible study tools. 2014. New Testament Greek lexicon.
<http://www.biblestudytools.com/lexicons/greek/nas/metoche.html> Toegang verkry: 6 Desember 2013.
- Booij, G. 2002. The morphology of Dutch. Oxford: Oxford University Press.
- Booij, G. 2006. Inflection and derivation. (*In Encyclopedia of Language & Linguistics*, 2nd ed., 5:654-661).
- Booij, G. 2010a. Construction morphology. Oxford: Oxford University Press.
- Booij, G. 2010b. Construction morphology. *Language and linguistics compass*, 4(7):543-555.
- Botha, J.P. & Van Aardt, J.M.H. 1987. Afrikaans vir die praktyk. Johannesburg: Lexicon Uitgewers.
- Botha, T.J.R., Ponelis, F.A., Combrink, J.G.H. & Odendal, F.F., *reds.* 1989. Inleiding tot die Afrikaanse taalkunde. 2de uitg. Pretoria: Academica.
- Breed, C.A. 2012. Die grammatikalisering van aspek in Afrikaans: 'n semantiese studie van perifrastiese progressiewe konstruksies. Potchefstroom: Noordwes-Universiteit. (Proefskrif – PhD).
- Broccias, C. & Hollmann, W.B. 2007. Do we need summary and sequential scanning in (Cognitive) grammar? *Cognitive linguistics*, 18(4): 487-522.
- Bybee, J. 1985. Morphology: a study of the relation between meaning and form. Amsterdam: Benjamins.
- Canakis, C. 1997. Pragmatics vs. Cognitive Semantics. *Papers from The Main Session of the Chicago Linguistic Society, April 17-19:15-29.*

Carstens, W.A.M. & Bosman, N., *reds.* 2014. Kontemporêre Afrikaanse taalkunde. Pretoria: Van Schaik Uitgewers.

Carstens, W.A.M. 2011. Norme vir Afrikaans – enkele riglyne by die gebruik van Afrikaans. 5de uitg. Pretoria: Van Schaik Uitgewers.

Cognitive linguistics. 2013. <http://www.degruyter.com/view/j/cogl> Toegang verkry: 25 September 2013.

Combrink, J.G.H. 1989. Afrikaanse morfologie: 'n oorsig. (*In Botha, T.J.R., Ponelis, F.A., Combrink, J.G.H. & Odendal, F.F. reds.* Inleiding tot die Afrikaanse taalkunde. Pretoria: Academica. pp. 220-254).

Combrink, J.G.H. 1990. Afrikaanse morfologie. Pretoria: Academica.

Comrie, B., Haspelmath, M. & Bickel, B. 2008. The Leipzig glossing rules: conventions for interlinear morpheme-by-morpheme glosses. University of Leipzig: Department of Linguistics of the Max Planck Institute for Evolutionary Anthropology & the Department of Linguistics. <http://www.eva.mpg.de/lingua/pdf/LGR08.02.05.pdf> Toegang verkry: 18 April 2014.

Croft, W. 2012. Verbs: aspect and causal structure. Oxford: Oxford University Press.

CTexT. 2013a. National Centre for Human Language Technology corpus: Afrikaans (voorbekendstellingsweergawe). Potchefstroom: Noordwes-Universiteit.

CTexT. 2013b. Protocol: lemmatisation (Afrikaans). NCHLT project: text resources (voorbekendstellingsweergawe). Potchefstroom: Noordwes-Universiteit.

CTexT. 2013c. Protocol: part-of-speech tagging (Afrikaans). NCHLT project: text resources (voorbekendstellingsweergawe). Potchefstroom: Noordwes-Universiteit.

De Villiers, M. 1948. Woordsoorte, werkwoorde en tye. Kaapstad: Nasionale Pers Beperk.

De Villiers, M. 1983. Afrikaanse grammatika vir volwassenes. Kaapstad: Nasou Beperk.

De Wit, A. & Brisard, F. 2013. A cognitive grammar account of the semantics of the English present progressive. *Journal of linguistics*, 10:1-42.

http://journals.cambridge.org/articla_S0022226713000169 Toegang verkry: 12 Oktober 2013.

Donaldson, B.C. 1993. A grammar of Afrikaans. Berlin: Mouton de Gruyter.

Du Toit, P.J. 1986. Taalleer vir onderwyser en student. Pretoria: Academica.

Elektroniese woordeboek van die Afrikaanse taal (eWAT). 2009.

Expert Advisory Group on Language Engineering Standards [EAGLES]. 1996. Recommendations for the Morphosyntactic Annotation of Corpora. EAGLES Document EAG – TCWG – MAC/R. Version of March 1996.

Evans, V. & Green, M. 2006. Cognitive linguistics: an introduction. Edinburgh: Edinburgh University Press.

Geerts, G., Haeseryn, W., De Rooij, J. & Van der Toorn, M. 1984. Algemene Nederlandse spraakkunst. Groningen: Wolters-Noordhoff.

Greenberg, J. 1963. Some universals of grammar with particular reference to the order of meaningful elements. (*In Greenberg, J., ed. Universals of Language. Cambridge: MIT Press. p.73-113*).

Groenewald, H.J. & Van Huyssteen, G.B. 2008. Outomatiese lemma-identifisering vir Afrikaans. *Literator*, 29(1):65-91.

Huddleston, R. 1988. English grammar – an outline. Cambridge: Cambridge University Press.

Huddleston, R. & Pullum, G.K. 2005. A student's introduction to English grammar. Cambridge: Cambridge University Press.

Jenkinson, A.G. 1993. Die probleem van fleksie en afleiding in Afrikaans. *South African journal of linguistics*, 11(18):100-122.

Kempen, W. 1982. Samestelling, afleiding en woordsoortelike meerfunksionaliteit in Afrikaans. Kaapstad: Nasou Beperk.

Kempen, W. s.j. Woordvorming en funksiewisseling in Afrikaans. Kaapstad: Nasionale Boekhandel Bpk.

- Langacker, R.W. 1987a. Foundations of cognitive grammar – volume I: theoretical prerequisites. Stanford: Stanford University Press.
- Langacker, R.W. 1987b. Nouns and verbs. *Language*, 63:53-94.
- Langacker, R.W. 1991a. Foundations of cognitive grammar – volume II: descriptive application. Stanford: Stanford University Press.
- Langacker, R.W. 1991b. Concept, image, and symbol: the cognitive basis of grammar. New York: Mouton de Gruyter.
- Langacker, R.W. 2000. A dynamic usage-based model. (In Barlow, M. & Kemmer, S., eds. Usage-based models of language. Stanford: CSLI Publications. p. 1-63).
- Langacker, R.W. 2008a. Cognitive grammar: a basic introduction. Oxford: Oxford University Press.
- Langacker, R.W. 2008b. Sequential and summary scanning: a reply. *Cognitive linguistics*, 19(4):571-584.
- Langacker, R.W. s.a. The functional organization of English clauses. San Diego: University of San Diego.
- Le Roux, T.H. 1952. Die voltooide deelwoord sonder *ge-* by werkwoorde op *-eer*. Pretoria: Van Schaik Beperk.
- Matthews, P.H. 1997. The concise Oxford dictionary of linguistics. Oxford: Oxford University Press.
- Odendal, F.F. & Gouws, R.H. 2005. Verklarende handwoordeboek van die Afrikaanse taal. 5de uitg. Kaapstad: Pearson Education South Africa.
- Oxford advanced learner's dictionary. 2000. 6th ed. Oxford: Oxford University Press.
- Pheiffer, R.H. 1989. Afrikaans en standaard-Nederlands. (In Botha, T.J.R., Ponelis, F.A., Combrink, J.G.H. & Odendal, F.F. eds. Inleiding tot die Afrikaanse taalkunde. Pretoria: Academica. pp. 72-95).
- Pilon, S. 2005. Outomatiese Afrikaanse woordsoortetikettering. Potchefstroom: Noordwes-Universiteit. (Verhandeling – MA).

- Ponelis, F. 1979. Afrikaanse sintaksis. Pretoria: Van Schaik.
- Ponelis, F. 1993. The development of Afrikaans. Frankfurt: Peter Lang.
- Ponelis, F.A. 1989. Sintaksis. (In Botha, T.J.R., Ponelis, F.A., Combrink, J.G.H. & Odendal, F.F. *reds.* Inleiding tot die Afrikaanse taalkunde. Pretoria: Academica. pp. 255-357).
- Raidt, E. 1989. Ontwikkeling van vroeë Afrikaans. (In Botha, T.J.R., Ponelis, F.A., Combrink, J.G.H. & Odendal, F.F. *reds.* Inleiding tot die Afrikaanse taalkunde. Pretoria: Academica. pp. 255-357).
- Reichenbach, H. 1947. Elements of symbolic logic. New York: Macmillan.
- Taalkommissie (samest.). 2009. Afrikaanse woordelys en spelreëls. 10de uitg. Kaapstad: Pharos.
- Teknia. 2013. Greek dictionary. <http://www.teknia.com/greek-dictionary/metoche>
Toegang verkry: 6 Desember 2012.
- Tuggy, D. 1985. The inflectional/derivational distinction. *Workpapers of the Summer Institute of Linguistics at the University of North Dakota XXIX. Grand Forks: Summer Institute of Linguistics*, p. 209-222.
- Van Eynde, F. 2004. Part of speech tagging en lemmatisering van het corpus gesproken Nederlands. Leuven: Centrum voor Computerlinguïstiek.
- Van Huyssteen, G.B. 2000a. Die reduplikasiekonstruksie in Afrikaans: enkele aspekte van 'n kognitiewe gebruiksebaseerde beskrywingsmodel vir Afrikaans. Potchefstroom: Noorwes-Universiteit. (Proefskrif – PhD).
- Van Huyssteen, G.B. 2000b. Teoretiese vooronderstellings van 'n kognitiewe gebruiksebaseerde beskrywingsmodel vir die Afrikaanse grammatika. *Southern African linguistics and applied language studies*, 20(4):303-323.
- Van Huyssteen, G.B. 2005. 'n Kognitiewe gebruiksebaseerde beskrywingsmodel vir die Afrikaanse grammatika. *South African linguistics and applied language studies*, 23(2): 125-137.

Van Huyssteen, G.B. 2010. (Re)defining component structures in morphological constructions: a cognitive grammar perspective. (*In Michel, S. & Onysko, A., eds. Cognitive approaches to word-formation. Berlin: Mouton de Gruyter. p. 97-126*).

Van Huyssteen, G.B. 2014. Morfologie. (*In Carstens, W.A.M. & Bosman, N., eds. Kontemporêre Afrikaanse taalkunde. Pretoria: Van Schaik Uitgewers. pp. 171-208*).

Van Schoor, J.L. 1983. Die grammatika van standaard-Afrikaans. Kaapstad: Lex Patria.

NCHLT Project: Text Resources

2010-2012

Protocol: Lemmatisation (Afrikaans)

Introduction

Aim of this protocol

The aim of this protocol is to guide annotators in the process of annotating the Afrikaans corpus on the lemmatisation layer for the NCHLT project: Text Resources. This document explicates some general principles to be followed in annotating the corpus, explains the process of annotation, and provides a detailed explanation of the categories to be used.

This is a dynamic, “living” document, and should be adapted following the procedure as explicated in the section “Changing the protocol”.

Concept and definitions

The aim of annotation on the lemmatisation layer is to identify the lemma (i.e. headword-form or base-form) of inflected words. The following definitions apply:

- **Lemmatisation:** The process of normalising all inflected forms of a lexical word to its common headword-form (lemma).

- **Examples:**

honde hond

katjie kat

gespeel speel

pragtige pragtig

was is

but:

hondagtig hondagtig

skrywer skrywer

vreeslik vreeslik

Note that these last three examples should not be normalised to “hond”, “skryf” and “vrees” respectively, since these are derivational forms.

- **Lemma:** The simplest form of a word as it would appear as headword in a dictionary; also in morphological terms referred to as the base-form, base or canonical form. Note that what is considered to be a lemma might differ from language to language.

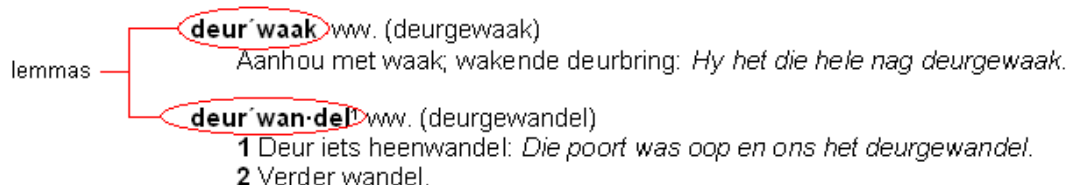


Figure 1: Example from HAT

- **Inflection:** “The modification of a word to express different grammatical categories such as tense, mood, voice, aspect, person, number, gender and case. Conjugation is the inflection of verbs; declension is the inflection of nouns, adjectives and pronouns.” (<http://en.wikipedia.org/wiki/Inflection>).

General principles

The following general principles apply:

- Annotation on the lemmatisation layer should be in excess of 98%
- The categories set below is based on:
 - Groenewald, HJ & Van Huyssteen, GB. 2008. Outomatiese Lemma-identifisering vir Afrikaans [Automatic Lemmatisation for Afrikaans]. *Literator*. 29(1): 65-91.
- The general principles of the Text Encoding Initiative P5 Guidelines (TEI P5; <http://www.tei-c.org/Guidelines/>) should be followed when in doubt.
-

Technical specifications

- All documents and annotation layers should be saved in UTF-8 format.
- Depending on the annotation software, XML encoding should be used. This will be determined by the software developers; annotators should in general not be concerned about appropriate syntax for the XML encoding.
-

Authoritative references

The following language-specific works of reference should be considered authoritative when in doubt:

- Odendal, F.F. & Gouws, R.H. (reds.). 2005. *Verklarende handwoordeboek van die Afrikaanse taal*. Vyfde uitgawe. Pinelands: Maskew Miller Longman.
- Taalkommissie. (samest.). 2009. *Afrikaanse Woordelys en Spelreëls*. Tiende uitgawe. Kaapstad: Pharos Woordeboeke.

Also consider the following references:

- Brits, JH. 2006. *Outomatiese Setswana Lemma-identifisering [Automatic Setswana Lemmatisation]*. Unpublished dissertation. Potchefstroom: North-West University.
- Brits, JC, Pretorius, RS & Van Huyssteen, GB. 2006. Automatic Lemmatisation in Setswana: Towards a Prototype. *South African Journal of African Languages*. 25: 37-47.
- Erjavec, T & Džeroski, S. 2004. Machine Learning of Morphosyntactic Structure: Lemmatizing Unknown Slovene Words. *Applied Artificial Intelligence*. 18(1): 17-40.
- Groenewald, HJ. 2006. *Automatic lemmatisation for Afrikaans*. Unpublished dissertation. Potchefstroom: North-West University.
- Groenewald, HJ. 2007. *Handleiding vir die annotasie van afrigtingsdata vir die ontwikkeling van LIA (Lemma-identifiseerder vir Afrikaans)* [Manual for the annotation of training data for the development of LIA (Lemmatiser for Afrikaans)]. Unpublished technical report. Version 2.0.3. Potchefstroom: North-West University.
- Groenewald, HJ & Van Huyssteen, GB. 2008. Outomatiese Lemma-identifisering vir Afrikaans [Automatic Lemmatisation for Afrikaans]. *Literator*. 29(1): 65-91.
-

Process

The following process should be followed:

- 1.

Categories

For purposes of lemmatisation in Afrikaans, we distinguish the following categories (and only these categories):

Nouns

1. Plural (e.g. *tafels*)
2. Diminutive (e.g. *tafeltjie*)

Adjectives

3. Comparative degree (e.g. *mooier*)
4. Superlative degree (e.g. *mooiste*)
5. Attributive *-e* (e.g. *interessante person*)
6. Partitive genitive (e.g. *iets moois*)

Prefixes that indicate negation could also be identified, but are not used for lemmatisation in this corpus (in other words, words in these categories remain as they are). They usually consist of the prefixes *on-*, *non-* and *nie-*, as well as *wan-* in nouns.

7. Negative (e.g. *onmoontlik*; also in some nouns, e.g. *wanhoop*)

Verbs

8. Infinitive *-e* (e.g. *iets te drinke*)
9. Past tense (e.g. *geskop*)
10. Present participle (e.g. *skreeuend*)
11. Past participle *ge-...-t/-d* (e.g. *gemeganiseerd*; also strong past participles like *gebroke*)

Changing the protocol

Since this protocol is considered to be a dynamic document, it is subject to change from time to time. For purposes of good governance, the following principles and procedures should be followed:

- The language expert (in consultation with the project leader) has the final say in changing this procedure document, and is the only authoritative person who might change the protocol.
- If an annotator (or other project member) wishes change a specification, s/he should log a request on Bugzilla, indicating the nature of the change request. The language expert then considers the request, and adds the change to the protocol document. The protocol document is saved with a new version number. All other annotators are then informed of the change, and requested to change previous annotations, where applicable.
- If an annotator wishes to add specific examples to the protocol, s/he sends an email to the language expert with the suggested examples. The language expert then adds the examples to this document, and when a significant number of examples have been added, saves the document with a new version number. All other annotators are then informed of the new version of the protocol document.

Roles and contact details

The following roles in this project can be identified:

- Project leader: Responsible for successful technical delivery of project.
 - Martin Puttkammer
martin.puttkammer@nwu.ac.za
082...
- Project manager: Responsible for timeous delivery of project, within budget. If you suspect that you will not be able to deliver on a set milestone, contact the project manager.
 - Ané Bekker
ane.bekker@nwu.ac.za
082...
- Language expert: Responsible for language specific requirements and specifications. This person also controls the protocol document.

- Afrikaans: Gerhard van Huyssteen
gerhard@trifonius.co.za
082 562 4814
- Annotator: Responsible for annotating the corpus, according to the above procedures and principles.
 - Afrikaans: XXX
XXX
XXX

Appendixes

Some general principles

Punctuation and diacritics

- Ignore all **punctuation**, such as commas, full-stops, exclamation marks, etc.
- Indicate **compulsory full-stops** as part of abbreviations.
 - *bv* *bv.*
 - *ens* *ens.*
 - Note that full-stops are not compulsory in abbreviations of titles, such as *mnr*, *prof*, *dr*, *me*, etc.
- Indicate **compulsory commas** between main verbs in complex sentences.
 - *Dat hy moeg is is 'n feit.* *Dat hy moeg is, is 'n feit.*
 - *Aangesien sy kan sing sal sy wen.* *Aangesien sy kan sing, sal sy wen.*
- Indicate **compulsory parentheses**.
 - *(soos koffie, tee en melk.* *(soos koffie, tee en melk).*
 - *"Hy is moeg.* *"Hy is moeg."*
- Ignore **optional hyphens**, but correct erroneous use of hyphens.
 - *radio-stasie* *radio-stasie*
 - **ra-dio-stasie* *radiostasie*
- In words where **sentence stress** is indicated with diacritics, the word is changed to the unstressed form.
 - *Jy moet dit só doen.* *Jy moet dit so doen.*
 - *Die pérd was moeg.* *Die perd was moeg.*
 - Note the following cases where the diacritics are compulsory:
 - The definite pronoun *dié* *dié*
 - *ná* 'volgende op' *ná*
 - *vóór* 'alvorens, voordat' *vóór*
 - *óf... óf* *óf... óf*
 - *én...én* *én...én*
 - *nóg...nóg* *nóg...nóg*

Capitalisation

- Lemmas are always written in **small letters** (e.g. even when a word is used at the beginning of a sentence, the lemma is still written in small letters; *Seuns* -> *seuns*; *Mnr.* -> *mnr.*). However, when the lemma is a proper name (*Bothas* -> *Botha*), or contains an acronym (*ATKV-lede* -> *ATKV-lid*), the capital letters are retained.
- In cases where an ordinary word functions as (part of) a proper name (*Wet op Organisasies sonder Winsmotief* – which is a multi-word named entity), that word is treated as an ordinary word.
 - *Wet* *wet*
 - *op* *op*
 - *Organisasies* *organisasie*
 - *sonder* *sonder*
 - *Winsmotief* *winsmotief*

 - *Registrateur* *registrateur*
 - *van* *van*
 - *Maatskappye* *maatskappy*

 - *Departement* *departement*
 - *Korrektiewe* *korrektief*
 - *Dienste* *diens*

- *Port* *port*
- *Elizabeth* *Elizabeth*

- *Republiek* *republiek*
- *van* *van*
- *Suid-Afrika* *Suid-Afrika*

- *Standard* *Standard* <Other Language>
- *Bank* *bank*

Note cases where **actual proper names** should be retained:

- *Piet* *Piet*
- *Burger* *Burger*

Abbreviations

- **Abbreviations** (*ens.*, *bv.*, *d.i.*) and **acronyms** (*ATKV*, *Sanlam*) are not normalised to their full forms, but are kept as they are.
- Plural forms of abbreviations (*mnre.*; *mee.*) or acronyms (*GOW's*) are lemmatised to represent the singular form (*mnr.*; *me.*; *GOW*)

Other languages

- Words in **other languages** (e.g. English) are marked as “Other Language”, and are left unchanged (including any capital letters).
 - *Client Types*
 - *Client* *Client*
 - *Types* *Types*
 - *Acceptance of trust as executor*
 - *Acceptance* *Acceptance*
 - *of* *of*
 - *trust* *trust*
 - *as* *as*
 - *executor* *executor*
- If the first part (modifier) of a compound is in another language, the whole word is treated as an Afrikaans word.
 - *sustainability-kongresse sustainability-kongres*

Variants

- Variant orthographic forms are kept in their original form.
 - *afgrawe* *afgrawe*
 - *belowe* *belowe*
 - *beloof* *beloof*
 - *skrywe* *skrywe*
 - *skryf* *skryf*
 - *lewe* *lewe*
 - *leef* *leef*
 - *skuiwe* *skuiwe*
 - *skuif* *skuif*
 - *strewe* *strewe*
 - *streef* *streef*
 - *sterwe* *sterwe*
 - *sterf* *sterf*
 - *chronies* *chronies*
 - *kronies* *kronies*
 - *foutopspoorstelsel* *foutopspoorstelsel*

- *foutopsporingstelsel* *foutopsporingstelsel*
- *more* *more*
- *môre* *môre*

Named entities

- **Telephone numbers** (or parts of telephone numbers) and **dates** should be marked as named entities.
 - *012-840-3456*
 - *012*
 - *1997*
 - *14 Maart 2010*
- **URLs and email addresses** should be marked as named entities.
 - *www*
 - *www.ctext.co.za*
 - *info@ctext.co.za*
- **Derivations of named entities** should not be lemmatised, unless an inflection can be removed.
 - *Suid-Afrikaans* *Suid-Afrikaans*
 - *Suid-Afrikaanse* *Suid-Afrikaans*
 - *Suid-Afrikaner* *Suid-Afrikaner*
 - *Suid-Afrikaners* *Suid-Afrikaner*
- If the **first part (modifier) of a compound** is a named entity, the whole word is treated as an Afrikaans word.
 - *Gateway-sentra* *Gateway-sentrum*

Spelling errors

- Only correct spelling errors that would be spelling errors without taking context into consideration.
 - *...wat grond **onwettige** besit...* *...wat grond **onwettige** besit...*
 - *onmidelik* *onmiddellik*
- Correct **repeated words**.
 - *...die die...* *die*
- Correct **wrong word-order**.
 - *...as jy nie die appel **nie** self eet.* *...as jy nie die appel self eet **nie**.*
- The **valence morpheme** is most often not obligatory; only correct cases where it is clearly wrong.
 - **regsbank* *regbank*
 - **besigheidkamer* *besigheidskamer*
 - *kortingitem* *kortingitem*
 - *kortingsitem* *kortingsitem*

Fixed expressions

- Afrikaans has a large number of fixed expressions where relicts of inflection processes have been fossilised. In such cases, **all the words in the archaic fixed expression are left unchanged** (i.e. treated as lemmas).
 - *ter* *ter*
 - *harte* *harte*
 - *neem* *neem*
- The following examples are taken from, amongst others, Botha, JP & Van Aardt, JMH. 1978. *Afrikaans vir die praktyk*. Johannesburg: McGraw-Hill.

Old genitive

<i>allersins</i>	<i>enigsins</i>	<i>smiddags</i>
<i>brood ter smarte</i>	<i>geensins</i>	<i>smorens</i>
<i>deesdaags</i>	<i>goedsmoeds</i>	<i>snags</i>
<i>desgelyks</i>	<i>grotendeels</i>	<i>soggens</i>
<i>desnoods</i>	<i>kind des doods</i>	<i>steen des aanstoots</i>

<i>desondanks</i>	<i>kwelling des geestes</i>	<i>stem des roepende</i>
<i>destyds</i>	<i>myns insiens</i>	<i>sweet des aanskyns</i>
<i>eersdaags</i>	<i>onses insiens</i>	<i>syns insiens</i>
<i>engel des ligs</i>	<i>saans</i>	<i>syns weegs</i>
<i>Gods water</i>		

- Note that the partitive genitive *-s* should be removed in cases like:
 - iets goeds* *iets goed*
 - niks nuuts* *niks nuut*

Old dative

<i>agterweë</i>		
<i>byderhand</i>	<i>by magte</i>	<i>by monde van</i>
<i>by wyse van</i>	<i>by ontvangs van</i>	
<i>dienaangaande</i>	<i>deser dae</i>	
<i>in der ewigheid</i>	<i>in der mine</i>	<i>in dier voege</i>
<i>in deser voege</i>		
<i>met dien verstande</i>	<i>met referte tot</i>	<i>mettertyd</i>
<i>metterwoon</i>		
<i>onder andere</i>	<i>op heter daad</i>	<i>onder meer</i>
<i>sedertdien</i>	<i>sindsdien</i>	
<i>te berde bring</i>	<i>te dien effekte</i>	<i>te dien einde</i>
<i>te gelegener tyd</i>	<i>te goeder trou</i>	<i>te lande</i>
<i>te velde trek</i>	<i>ten aanhore van</i>	<i>ten bate van</i>
<i>ten behoefte van</i>	<i>ten beste</i>	<i>ten dele</i>
<i>ten eerste, tweede, ens.</i>	<i>ten einde</i>	<i>ten enemale</i>
<i>ten gevolge van</i>	<i>ten/te gronde gaan</i>	<i>ten gunste van</i>
<i>ten huise van</i>	<i>ten koste van</i>	<i>ten laste</i>
<i>ten minste</i>	<i>ten nadele van</i>	<i>ten onregte</i>
<i>ten opsigte van</i>	<i>ten regte</i>	<i>ten seerste</i>
<i>ten slotte</i>	<i>ten spyte van</i>	<i>ten strengste</i>
<i>ten volle</i>	<i>ten voordele van</i>	<i>ter aarde bestel</i>
<i>ter bekwamer tyd</i>	<i>ter dood (veroordeel)</i>	<i>ter elfder ure</i>
<i>ter ere van</i>	<i>ter harte neem</i>	<i>ter insae</i>
<i>ter insage</i>	<i>ternouernood</i>	<i>ter ore kom</i>
<i>ter perse gaan</i>	<i>ter plaatse</i>	<i>(wel) ter ruste</i>
<i>ter sake</i>	<i>ter sprake</i>	<i>ter tafel</i>
<i>wel ter tale</i>	<i>ter verduideliking</i>	<i>ter wille van</i>
<i>uit hoofde van</i>		
<i>van aller harte</i>	<i>van harte (geluk)</i>	<i>van nature</i>

Distant compounds (“afstandsamestellings”)

- If the first part of a distant compound is written disjunctively (e.g. *honde- en kattedos*), that part should not be lemmatised.
 - honde-* *honde-*
 - en* *en*
 - kattedos* *kattedos*
- If the second part of a distant compound is written disjunctively (e.g. *hondebeddens en -hokke*), that part should be lemmatised.
 - hondebeddens* *hondebed*
 - en* *en*
 - hokke* *-hok*

Notes per category

Nouns

- Note that deverbal nouns like *gelaggery* and *gespelery* are nouns, and should not be lemmatised as verbs.
- Also be careful to **not lemmatise** other nominalisations like *voorkoms*, *binnekoms*, *uitkoms*, *opvangs*.
- **Feminine forms** should also not be lemmatised – e.g. *kelnerin*, *aktrise*, *sekretaresse*.

Plural

-e	<i>feeste</i>	<i>fees</i>
-ci	<i>politici</i>	<i>politikus</i>
-e	<i>gasse</i>	<i>gas</i>
-ers	<i>kinders</i>	<i>kind</i>
-der	<i>beendere</i>	<i>been</i>
-s	<i>mans</i>	<i>man</i>
-a	<i>stadia</i>	<i>stadium</i>
-‘s	<i>pa’s</i>	<i>pa</i>
-‘e	<i>m’e</i>	<i>m</i>
-ns	<i>vermoëns</i>	<i>vermoë</i>
-ens	<i>kwajongens</i>	<i>kwajong</i>
-te	<i>dienste</i>	<i>diens</i>
-ere	<i>goedere</i>	<i>goed</i>

- *Bloes* has a variant form *bloese*, as well as a plural form *bloese*. In cases like this, the words should be lemmatised to the simplest form – in this case *bloes*.
 - *uitkomstes/uitkomste* *uitkoms*
 - *klagtes/klagte* *klag*
 - *bylaes/bylae* *bylaag*
 - *toelaes/toelae* *toelaag*
 - *tesame/tesaam* *tesaam*
- Some words, like *juweliersware*, do not have singular forms.
 - *juweliersware* *juweliersware*
 - *skoolure* *skoolure*
 - *lewensomstandighede* *lewensomstandighede*
 - *kontakbesonderhede* *kontakbesonderhede*
 - *regeringsgeledere* *regeringsgeledere*
 - *media ‘pers’* *media*
 - *verrigtinge ‘amptelike gebeurtenis’* *verrigtinge*
 - *betrekkinge* *betrekkinge*
 - *klere* *klere*
 - *gebeure* *gebeure*
- In some words, like *ambagslui* and *seeliede*, the last part of the word changes completely when lemmatised.
 - *ambagslui* *ambagsman*
 - *seeliede* *seeman*
 - *paaie* *pad*
 - *uitsprake* *uitspraak*
 - *stede* *stad*
 - *oë* *oog*
 - *stowe* *stoof*
- Note cases of deadjectival nouns, derived using the **-e** morpheme – these should not be confused with the attributive form of the adjective. For example, *domme* in *die domme swaap* should be lemmatised as *dom*, because the **-e** is being used attributively. *Dommies* (*die lede van die span is ’n spul dummies*)

should not be lemmatised as *dom*, but as *domme*. Remember that the part of speech category should not be changed.

- The lemma of *bure* is *buur*.

Diminutive

-etjie	<i>balletjie</i>	<i>bal</i>
-kie	<i>koninkie</i>	<i>koning</i>
-pie	<i>boompie</i>	<i>boom</i>
-tjie	<i>vroutjie</i>	<i>vrou</i>
-jie	<i>ritjie</i>	<i>rit</i>
-'tjie	<i>foto'tjie</i>	<i>foto</i>
-ie	<i>huisie</i>	<i>huis</i>
-ies	<i>koekies</i>	<i>koek</i>
-jies	<i>gebedjies</i>	<i>gebed</i>
-kies	<i>garinkies</i>	<i>garing</i>
-pies	<i>besempies</i>	<i>besem</i>
-'tjies	<i>tablo'tjies</i>	<i>tablo</i>
-tjies	<i>kleurtjies</i>	<i>kleur</i>
-etjies	<i>mannetjies</i>	<i>man</i>

- The following classes of diminutives should not be lemmatised:
 - Lexicalised forms: These words might have been originally diminutives (or other constructions) in Dutch, but the diminutive form became lexicalised in Afrikaans. They therefore only occur as diminutives in Afrikaans.
 - *polfyntjie, kolwyntjie, dankie, donkie, mandjie, bietjie, meisie*
 - *graatjie* ‘maer persoon’; *hokkie* ‘soort sport’; *strokie* ‘betalingsbewys’
 - Institutionalised forms: These words are usually compounds with a diminutive suffix, and are often names for plants, animals or body parts. They usually only occur as diminutives.
 - *mooinoientjie, bobbejaantoontjie, middelloorbeentjie, depositostrokie*
 - Diminutive as nominaliser.
 - *bruintjie, kleintjie*
 - Diminutive as adverbialiser.
 - *effentjies, fyntjies, saggies, skoontjies*
- Some institutionalised words could occur with or without the diminutive. In such cases, one should use frequency information to determine what the most common form is.
 - *hondepisbossie, karoobossie, rosyntjie, boontjie, klontjie, glippertjie*

Adjectives

- **Ordinal numbers** (*eerste, tweede, derde*) are not lemmatised.
- Be careful for **adjectival forms** that should not be lemmatised – e.g. *onderduims, ondermaans, buitemuurs, ondergronds, boaards*.
- Note that morphemes that express negation **should not be removed**.
 - *onmoontlike* *onmoontlik*
 - *nie-vervangbare* *nie-vervangbaar*
- Note the following specific cases:
 - *lang, langer, langste* *lank*
 - *jong, jonger, jongste* *jonk*
 - *ouer<ADJ>, oudste, oud, ou, oue, ouder (gewoonte)* *oud*
 - *ouer<N>* *ouer*
 - *wyer, wydste, wye* *wyd*
 - *politieke* *politiek*
 - *politiese* *polities*
 - *fisieke* *fisiek*
 - *fisiese* *fisies*

Comparative degree

-er	<i>mooier</i>	<i>mooi</i>
-der	<i>verder</i>	<i>ver</i>
-ër	<i>hoër</i>	<i>hoog</i>

- Where the comparative degree is expressed with the construction *meer X* (e.g. *meer betroubaar*), the **meer** should be lemmatised as **baie**.
 - Be careful with **meer**, since it has numerous senses, as well as numerous POS categories. Only when it is used as adverb, should it be lemmatised as **baie**.
- Note that **eerder** should not be lemmatised.

Superlative degree

-ste	<i>mooiste</i>	<i>mooi</i>
------	----------------	-------------

- Where the superlative degree is expressed with the construction *mees X-e* (e.g. *mees betroubare*), the **mees** should be lemmatised as **baie**.
- Note that **enigste** is lemmatised as **enig**.
- With regard to the superlative degree, note that some forms could differ significantly from the base-form.
 - *meeste* *baie*
 - Be careful with **meeste**, since it has numerous senses, as well as numerous POS categories. It is mostly lemmatised as **baie**.
 - *naaste<ADJ>* *naby*
 - *naaste<N>* *naaste*

Attributive -e

-ë	<i>leë</i>	<i>leeg</i>
-ë	<i>vroeë</i>	<i>vroeg</i>
-e	<i>moeilike, verre</i>	<i>moeilik, ver</i>

Partitive genitive

-s	<i>moois, leliks</i>	<i>mooi, lelik</i>
----	----------------------	--------------------

Verbs

Past tense

ge-	<i>gehelp</i>	<i>help</i>
aange-	<i>aangehelp</i>	<i>aanhhelp</i>
aanmekearge-	<i>aanmekeargeskryf</i>	<i>aanmekearskryf</i>
agterge-	<i>agtergebly</i>	<i>agterbly</i>
agternage-	<i>agternagesit</i>	<i>agternasit</i>
bangge-	<i>banggepraat</i>	<i>bangpraat</i>
binnege-	<i>binnegekom</i>	<i>binnekom</i>
byge-	<i>bygebly</i>	<i>bybly</i>
deurge-	<i>deurgeloop</i>	<i>deurloop</i>
deurmekaarge-	<i>deurmekaargeklits</i>	<i>deurmekaarklits</i>
droogge-	<i>drooggemaak</i>	<i>droogmaak</i>
inge-	<i>ingelê</i>	<i>inlê</i>
klaarge-	<i>klaargespeel</i>	<i>klaarspeel</i>
koudge-	<i>koudgelei</i>	<i>koudlei</i>
kwaadge-	<i>kwaadgestook</i>	<i>kwaadstook</i>
losge-	<i>losgelaat</i>	<i>loslaat</i>
misge-	<i>misgegooi</i>	<i>misgooi</i>
nage-	<i>nagelaat</i>	<i>nalaat</i>
natge-	<i>natgelei</i>	<i>natlei</i>
omge-	<i>omgegooi</i>	<i>omgooi</i>
onderge-	<i>ondergedompel</i>	<i>onderdompel</i>
oopge-	<i>oopgemaak</i>	<i>oopmaak</i>
oorge-	<i>oorgelaai</i>	<i>oorlaai</i>
openbaarge-	<i>openbaargemaak</i>	<i>openbaarmaak</i>
opge-	<i>opgetooi</i>	<i>optooi</i>
platge-	<i>platgeslaan</i>	<i>platslaan</i>
rondge-	<i>rondgestaan</i>	<i>rondstaan</i>
saamge-	<i>saamgeflans</i>	<i>saamflans</i>
skerpge-	<i>skerpgemaak</i>	<i>skerpmaak</i>
skrikge-	<i>skrikgemaak</i>	<i>skrikmaak</i>
stilge-	<i>stilgemaak</i>	<i>stilmaak</i>
teenge-	<i>teengegaan</i>	<i>teengaan</i>
teëge-	<i>teëgegaan</i>	<i>teëgaan</i>
tegemoetge-	<i>tegemoetgegaan</i>	<i>tegemoetgaan</i>
tentoonge-	<i>tentoongestel</i>	<i>tentoonstel</i>
terugge-	<i>teruggeskryf</i>	<i>terugskryf</i>
toege-	<i>toegestaan</i>	<i>toestaan</i>
uitge-	<i>uitgevang</i>	<i>uitvang</i>
vasge-	<i>vasgesluit</i>	<i>vassluit</i>
verbyge-	<i>verbygegaan</i>	<i>verbygaan</i>
voorge-	<i>voorgeloop</i>	<i>voorloop</i>
voortge-	<i>voortgegaan</i>	<i>voortgaan</i>
vryge-	<i>vrygelaat</i>	<i>vrylaat</i>
wegge-	<i>wegge loop</i>	<i>wegloop</i>

- *doeltreffende* (wat **die doel tref**)
 - *ondergenoemde* (wat **onder genoem is**)
 - *bogenoemde* (wat **bo genoem is**)
 - *nouspannende* (wat **nou span**)
 - *voormelde* (wat **voor vermeld is**)
 - *voornoemde* (wat **voor genoem is**)
 - *hoogaangeskrewe* (wat **hoog aangeskryf is**)
 - *regsgebaseerde* (wat op die **reg gebaseer is**)
 - *doelgerigte* (wat op 'n **doel gerig is**)
- doeltreffend*
ondergenoemde
bogenoemde
nouspannend
voormelde
voornoemde
hoogaangeskrewe
regsgebaseerd
doelgerig
- **Prepositions**
 - *gedurende*
 - *ingevolge*
 - *rakende*
 - Note that *rakende* should be lemmatised as *raak* in the context of *rakende hoeke*.
 - **Participles with noun stems**, and where the noun could not be interpreted as a verb (conversion)
 - *getiteld*
 - *gevueld*
 - *gemiddeld*
 - *geletterd*
 - *gesyferd*
 - *getand*
 - *bebaard*
 - *bebloed*
 - *beblaard*
 - Note cases such as *ontmande* and *beboste* where *ontman* and *bebos* are the respective lemmas.

Present participle

-end	<i>lopend</i>	<i>loop</i>	<i>water wat loop</i>
	<i>oorblywend</i>	<i>oorbly</i>	<i>geld wat oorbly</i>
	<i>beduidend</i>	<i>beduie</i>	<i>resultate wat beduie</i>
-ende	<i>lopende</i>	<i>loop</i>	<i>water wat loop</i>
	<i>volgende</i>	<i>volg</i>	<i>voorbeeld wat volg</i>
	<i>voldoende</i>	<i>voldoen</i>	<i>bewyse wat voldoen</i>
	<i>bestaande</i>	<i>bestaan</i>	<i>gevalle wat bestaan</i>
	<i>gepaardgaande</i>	<i>gepaardgaan</i>	<i>voorbeeld wat gepaardgaan</i>

Weak past participle ge-...-t/-d

ge-...-d	<i>geregistreerd</i>	<i>registreer</i>	<i>wat geregistreer is</i>
Ø-...-d	<i>vermoeid</i>	<i>vermoei</i>	<i>wat vermoei is</i>
ge-...-de	<i>geregistreerde</i>	<i>Registreer</i>	<i>wat geregistreer is</i>
Ø-...-de	<i>erkende</i>	<i>erken</i>	<i>wat erken is</i>
	<i>gematigde</i>	<i>gematig</i>	<i>wat gematig is</i>
	<i>bemande</i>	<i>beman</i>	<i>wat beman is</i>
	<i>beplande</i>	<i>beplan</i>	<i>wat beplan is</i>
	<i>ontbeende</i>	<i>ontbeen</i>	<i>wat ontbeen is</i>
	<i>ontblaaide</i>	<i>ontblaar</i>	<i>wat ontblaar is</i>
ge-...-t	<i>gelapt</i>	<i>lap</i>	<i>wat gelap is</i>
Ø-...-t			
ge-...-te	<i>gelapte</i>	<i>lap</i>	<i>wat gelap is</i>

Ø-...-te	<i>beperkte</i>	<i>beperk</i>	<i>wat beperk is</i>
	<i>gepaste</i>	<i>gepas</i>	<i>wat gepas is</i>
	<i>ontboste</i>	<i>ontbos</i>	<i>wat ontbos is</i>
inge-...-d(e)	<i>ingedraaide</i>	<i>indraai</i>	<i>wat ingedraai is</i>
deurge-...-d(e)	<i>deurgesnyde</i>	<i>deursny</i>	<i>wat deurgesny is</i>

- Take note of words that might look like past participles, but that are not. They should not be lemmatised.
 - *gebeente, gedoente, gedermtte*, etc.

Strong past participle

<i>aangewese</i>	<i>aanwys</i>
<i>beslote</i>	<i>besluit</i>
<i>betrokke</i>	<i>betrek</i>
<i>oortrokke</i>	<i>oortrek</i>
<i>versweë</i>	<i>verswyg</i>
<i>geskrewe</i>	<i>skryf</i>
<i>voorgeskrewe</i>	<i>voorskryf</i>
<i>gegewe</i>	<i>gee</i>
<i>gesogte</i>	<i>soek</i>

Note that some strong past participles cannot be lemmatised:

- *bestorwe*
-

NCHLT Project: Text Resources

2010-2012

Protocol: Part-of-Speech Tagging (Afrikaans)

Introduction

Aim of this protocol

The aim of this protocol is to guide annotators in the process of annotating the Afrikaans corpus on the part-of-speech layer for the NCHLT project: Text Resources. This document explicates some general principles to be followed in annotating the corpus, explains the process of annotation, and provides a detailed explanation of the tag set to be used. This is a dynamic, “living” document, and should be adapted following the procedure as explicated in the section “Changing the protocol”.

Concept and definitions

The aim of annotation on the part-of-speech layer is to identify the part-of-speech category of each token, as identified on the tokenisation layer (i.e. usually words, multi-word expressions, and punctuation). The following definitions apply:

- **Part-of-speech:** The function that a word fulfils in a sentence (i.e. in a grammatical context). It is also sometime referred to as the lexical category, word class or lexical class of a word. The part-of speech of a word is determined by the morpho-syntactic behaviour of the word in the specific context – e.g. what happens to the words morphologically within certain syntactic contexts. In linguistics, we usually distinguish between open classes (i.e. classes to which new words can be added, and in that sense productive classes; usually nouns, verbs, adjectives and adverbs), and closed classes (i.e. classes that are generally not expanded through productive, everyday processes; e.g. pronouns, determiners, etc.).
 - **Examples:**
 - Die* <article; definite>
 - grootste* <adjective; superlative>
 - skrywer* <noun; singular>
 - van* <particle>
 - die* <article; definite>
 - Renaissance* <noun; proper>
 - was* <verb; copula; past tense>
 - William* <noun; proper>
 - Shakespeare* <noun; proper>
 - .* <punctuation>
 - Note that the word *was* can also be a noun (*Die was van die kers*), but in this syntactic context, it is a verb.
- **Word:** A string of characters with whitespaces on both sides, which could consist of alphanumeric characters, hyphens, apostrophes, full-stops, parentheses and/or symbols, but no other punctuation marks. Abbreviations are considered to be words.
 - **Examples:**
 - Hy*
 - het*
 - bv.*

tragedies
 , [Punctuation; not a word]
komedies
en
historektomieë
geskryf
 . [Punctuation; not a word]
Romeo
and
Juliet
is
een
voorbeeld
 . [Punctuation; not a word]

- **Multi-word expression:** A string of words (two or more words) that is considered to be one lexical unit. Various multi-word expressions can be considered, including idioms, collocations, separable verbs, etc. For purposes of this project, only the following categories are distinguished:
 - **Named entities:**
 - Romeo and Juliet*
 - San Salvador*
 - Sint Kitts en Nevis*
 - Nelson Mandelarylaan*
 - Pretoriase Kunsmuseum*
 - Departement Kuns en Kultuur*
 - Department of Arts and Culture* (also marked as “Other Language”)
 - **Distance compounds:**
 - adjunk- eerste minister*
 - hoof- uitvoerende beampte*
 - langtermyn- maatskaplike beplanning*
 - stads- elektriese ingenieur*
 - Parts like *adjunk-* and *hoof-* are annotated as **RWD**.
- **Punctuation:** “Symbols which indicate the structure and organization of written language, as well as intonation and pauses to be observed when reading aloud.” (<http://en.wikipedia.org/wiki/Punctuation>). Typical punctuation marks include the full-stop, comma, bracket/parenthesis, colon, semi-colon, dash, ellipsis, exclamation mark, question mark, quotation mark (single or double), and back-slash
 - Note that the forward slash and various symbols (e.g. currency symbols) should be **tagged as symbols and not as punctuation**. The guideline here is: if you can pronounce a symbol/punctuation mark while reading a sentence, then it is should be tagged as symbol.
- **Primary/secondary tags:** Due to the complexity of language and the multi-layeredness inherent to language data, a number of categories are annotated on multiple layers. With regard to POS tagging, the assigned POS is considered the **primary tag**, while **secondary tags** could be used to annotate on other, orthogonal dimensions. The following secondary tags could be assigned, by checking the relevant check boxed in the software environment (see tokenization protocol for definitions):
 - **Run-on**
 - Words that are written as one word, but that should actually be two words (e.g. *dietafel*)
 - **Split**
 - Words that are written as two words, but that should actually be one word (e.g. *sake plan*)
 - **Confusable**

- Words that could often be confused, e.g. homophones (*wei* and *wy*)
 - **Spelling mistake**
 - Words that do not conform to the official orthography of Afrikaans (e.g. *onmideliks*)
 - **Named entity**
 - Words that are used in context to refer to a single, unique instantiation through convention, an institutionalised process or process, and of which the orthography is determined by law, or by some authority (Puttkammer, 2005)
 - Romeo and Juliet*
 - San Salvador*
 - Sint Kitts en Nevis*
 - Nelson Mandelarylaan*
 - Pretoriase Kunsmuseum*
 - Departement Kuns en Kultuur*
 - Department of Arts and Culture* (also marked as “Other Language”)
 - **Fixed expression**
 - A commonly used phrase, often featuring a preposition (e.g. *in der ewigheid, ten opsigte van*)
 - The lemmas of ALL the words in the fixed expression remain unchanged (i.e. stay the same as the token words), e.g.
 - token = lemma = tag (flag)
 - in = in = SVS (and flagged fixed expression)
 - der = der = LB (and flagged fixed expression)
 - ewigheid = ewigheid = NA (and flagged fixed expression)
 - **Participle**
 - All verbs that are NOT base-forms should also be indicated as participles.
 - *Die kinders **speel*** base-form
 - *Die kinders **het gespeel*** participle
 - *Hy **ophou sing*** base-form
 - *Hy **het ophou sing*** participle
 - **Ambiguous**
 - Words that might oftentimes have ambiguous POS categories should be marked as “ambiguous”.
 - *gebied* **NSE or VVHOG**
 - *bank* **NSE or VTHOG or VTHOO**
 - **Other language**
 - Words that are not Afrikaans words in the specific context (e.g. *Bank of England, African National Congress*)
 - Note: All words that are tagged as RV should also be marked on the secondary level as “other language”.
 - Note: Words and acronyms in another language that are part of named entities should be tagged as such (**NEE/NEM/NEED/NEMD**), and not as RV, but flagged as “other language”.
 - *DAC* **RK; “Other Language”; “Named Entity”**
 - *Long* **NEE; “Other Language”**
 - Walk* **NEE; “Other Language”**
 - To* **NEE; “Other Language”**
 - Freedom* **NEE; “Other Language”**
 - *Sun* **NEE; “Other Language”**
 - City* **NEE; “Other Language”**

- *African* *NEE; “Other Language”*
 - National* *NEE; “Other Language”*
 - Congress* *NEE; “Other Language”*
- **Other**
 - The tag “other” could be used for cases that are not covered by the protocol, and should be used sparingly.
- **Unsure**
 - Use “Unsure” to indicate cases that should be discussed.

General principles

The following general principles apply:

- Annotation on the part-of-speech layer should be in excess of 98%
- The tag set below is based on:
 - EAGLES guidelines (<http://www.ilc.pi.cnr.it/EAGLES/home.html>);
 - *Corpus Gesproke Nederlands* tag set (lands.let.ru.nl/cgn/); and
 - Dissertation of Suléne Pilon:
 - Pilon, S. 2005. *Outomatiese Afrikaanse Woordsoortetikettering [Automatic Afrikaans Part-of-Speech tagging]*. Unpublished dissertation. Potchefstroom: North-West University.
- The general principles of the Text Encoding Initiative P5 Guidelines (TEI P5; <http://www.tei-c.org/Guidelines/>) should be followed when in doubt.

Technical specifications

- All documents and annotation layers should be saved in UTF-8 format.
- Depending on the annotation software, XML encoding should be used. This will be determined by the software developers; annotators should in general not be concerned about appropriate syntax for the XML encoding.

Authoritative references

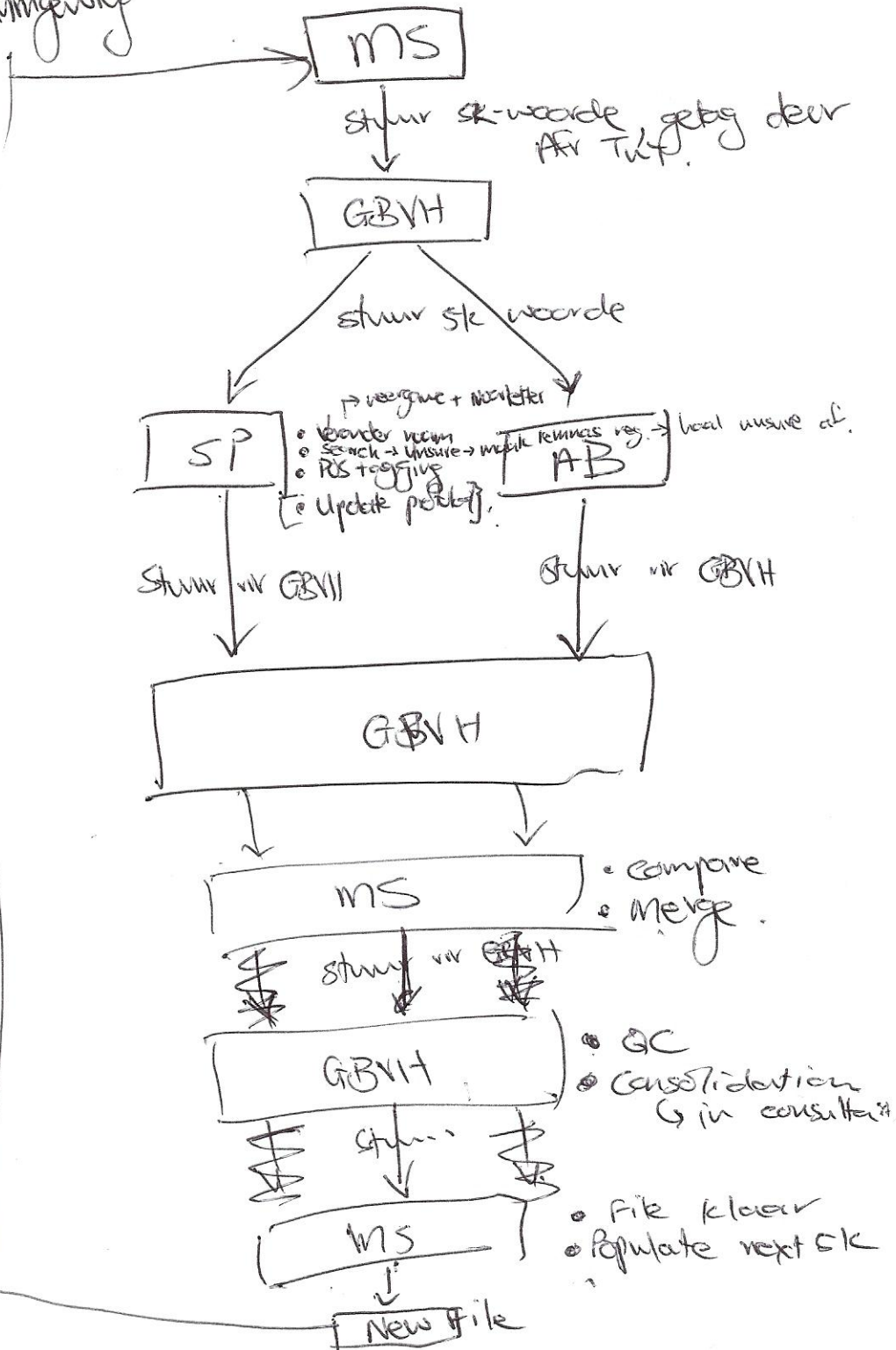
The following language-specific works of reference should be considered authoritative when in doubt:

- Odendal, F.F. & Gouws, R.H. (reds.). 2005. *Verklarende handwoordeboek van die Afrikaanse taal*. Vyfde uitgawe. Pinelands: Maskew Miller Longman.
- Taalkommissie. (samest.). 2009. *Afrikaanse Woordelys en Spelreëls*. Tiende uitgawe. Kaapstad: Pharos Woordeboeke.

Process

<Describe here step-by-step the process to be followed. Use number lists>

Naamgewing



File name convention:

- When you receive a file to be annotated, its name will typically be in the following format:
 - GOV-ZA.50000ParalleleEnWoorde.af.pos.part1.1.0.0.MS.2011-11-04

- When you save this file to your computer, increase the version number, and change your initials (no need to change the date). The file name should then be something like:
 - GOV-ZA.50000ParalleleEnWoorde.af.pos.part1.1.0.1.GBVH.2011-11-04

Other

- Remember to carefully take note of time spent on the project. Please make a distinction between **administration, training, meetings/discussions, annotation** and **other work** (e.g. if it is required of you to make changes to protocol, or annotate testing data, etc.).
- If you find any errors in the data that do not necessarily pertain to POS tagging but which you still want to flag to the project leader (e.g. editorial errors), mark it with **###** in the comment field
 - **###This is probably a contextual error.**

Tag set

For purposes of annotators, this tag set is described in Afrikaans, and is by and large taken over from Pilon (2005), as well as various other documents of CText, produced by Pilon.

Some notes and guidelines in the appendixes provide additional examples and practical guidelines.

Example	Value ¹	POS tag
Nouns		
<i>stoel</i>	soortnaam/enkelvoud/basis	NSE
<i>stoele</i>	soortnaam/meervoud/basis	NSM
<i>stoeltjie</i>	soortnaam/enkelvoud/diminutief	NSED
<i>stoeltjies</i>	soortnaam/meervoud/diminutief	NSMD
<i>Koos</i>	eienaam/enkelvoud/basis	NEE
<i>Kose</i>	eienaam/meervoud/basis	NEM
<i>Kosie</i>	eienaam/enkelvoud/diminutief	NEED
<i>Kosies</i>	eienaam/meervoud/diminutief	NEMD
<i>goud</i>	massanaam	NM
<i>trop</i>	versamelnaam/enkelvoud/basis	NVE
<i>troppe</i>	versamelnaam/meervoud/basis	NVM
<i>troppie</i>	versamelnaam/enkelvoud/diminutief	NVED
<i>troppies</i>	versamelnaam/meervoud/diminutief	NVMD
<i>teelepel</i>	maatnaam/enkelvoud/basis	NME
<i>teelepels</i>	maatnaam/meervoud/basis	NMM
<i>teelepeltjie</i>	maatnaam/enkelvoud/diminutief	NMED
<i>teelepeltjies</i>	maatnaam/meervoud/diminutief	NMMD
<i>liefde</i>	abstrak	NA
Verbs		
<i>Jy moet opskryf wat ek sê</i>	ongemarkeerd/hoof/skeibaar/oorganklik	VTHSG
<i>Jy moet die brief pos</i>	ongemarkeerd/hoof/onskeibaar/oorganklik	VTHOG
<i>Jy het die brief gepos</i>	gemarkeerd/hoof/onskeibaar/oorganklik	VVHOG
<i>Ek wil wegloop</i>	ongemarkeerd/hoof/skeibaar/onoorganklik	VTHSO

¹ Values are given in Afrikaans in order to make the abbreviations used for the POS tag more transparent.

<i>Die kind eet</i>	ongemarkeerd/hoof/onskeibaar/onoorganklik	VTHOO
<i>Sy het weggeloop/geëet</i>	gemarkeerd/hoof/onskeibaar/onoorganklik	VVHOO
<i>Ek is moeg</i>	ongemarkeerd/hoof/onskeibaar/koppel	VTHOK
<i>Ek was moeg</i>	gemarkeerd/hoof/onskeibaar/koppel	VVHOK
<i>Sy hou van wortelsap</i>	ongemarkeerd/hoof/onskeibaar/voorsetsel	VTHOV
<i>Sy het van wortelsap gehou</i>	gemarkeerd/hoof/onskeibaar/voorsetsel	VVHOV
<i>Jy kan my help</i>	teenwoordig/hulp/onskeibaar/modaliteit	VTUOM
<i>Jy kon my help</i>	verlede/hulp/onskeibaar/modaliteit	VVUOM
<i>Ek sal droom</i>	hulp/onskeibaar/tyd	VUOT
<i>Ek help haar leer</i>	teenwoordig/hulp/onskeibaar/aspek	VTUOA
<i>Ek het gehelp leer</i>	verlede/hulp/onskeibaar/aspek	VVUOA
<i>Daar word gedrink</i>	teenwoordig/hulp/onskeibaar/modus	VTUOP
<i>Daar is gedrink</i>	verlede/hulp/onskeibaar/modus	VVUOP
Adjectives		
<i>lelike meisie</i>	stellend/attributief	ASA
<i>leliker meisie</i>	vergroterend/attributief	AVA
<i>lelikste meisie</i>	oortreffend/attributief	AOA
<i>die meisie is lelik</i>	stellend/predikatief	ASP
<i>hierdie meisie is leliker</i>	vergroterend/predikatief	AVP
<i>hierdie een is die lelikste</i>	oortreffend/predikatief	AOP
Pronouns		
<i>hierdie</i>	aanwysend	PA
<i>niemand</i>	onbepaald	PO
<i>myne; myns insiens</i>	eerste/enkelvoud/besitlik	PEEB
<i>joune; jous insiens</i>	tweede/enkelvoud/besitlik	PTEB
<i>syne</i>	derde/ orphol/enkelvoud/besitlik	PDHEB
<i>hare</i>	derde/vroulik/enkelvoud/besitlik	PDVEB
<i>dit is ons boek; onses insiens</i>	eerste/meervoud/besitlik	PEMB
<i>dit is julle boek</i>	tweede/meervoud/besitlik	PTMB
<i>dit is hulle boek</i>	derde/meervoud/besitlik	PDMB
<i>wie</i>	vraend	PV
<i>ek sny myself</i>	eerste/enkelvoud/wederkerend	PEEW
<i>jy sny jouself</i>	tweede/enkelvoud/wederkerend	PTEW
<i>hy sny homself</i>	derde/ orphol/enkelvoud/wederkerend	PDHEW
<i>sy sny haarself</i>	derde/vroulik/enkelvoud/wederkerend	PDVEW
<i>dit hou sigself besig met...</i>	derde/onsydig/enkelvoud/wederkerend	PDOEW
<i>ons sny onself</i>	eerste/meervoud/wederkerend	PEMW
<i>julle sny julleself</i>	tweede/meervoud/wederkerend	PTMW
<i>hulle sny hulleself</i>	derde/meervoud/wederkerend	PDMW
<i>ek slaap</i>	eerste/enkelvoud/ongemarkeerd /persoonlik	PEENP
<i>jy slaap</i>	tweede/enkelvoud/ongemarkeerd /persoonlik	PTENP
<i>hy slaap</i>	derde/ orphol/enkelvoud/ongemarkeerd /persoonlik	PDHENP
<i>sy slaap</i>	derde/vroulik/enkelvoud/ongemarkeerd /persoonlik	PDVENP
<i>dit slaap</i>	derde/onsydig/enkelvoud/ongemarkeerd /persoonlik	PDOENP
<i>ons slaap</i>	eerste/meervoud/persoonlik	PEMP
<i>julle slaap</i>	tweede/meervoud /persoonlik	PTMP
<i>hulle slaap</i>	derde/meervoud/persoonlik	PDMP
<i>jy sien my</i>	eerste/enkelvoud/gemarkeerd/persoonlik	PEEDP
<i>ek sien jou</i>	tweede/enkelvoud/gemarkeerd /persoonlik	PTEDP
<i>ek sien hom</i>	derde/ orphol/enkelvoud/gemarkeerd /persoonlik	PDHEDP
<i>sy sien haar</i>	derde/vroulik/enkelvoud/gemarkeerd /persoonlik	PDVEDP
<i>ons help mekaar</i>	wederkerig	PW
<i>die man wat daar loop</i>	betreklik	PB
Determinors		
<i>die (der, des, den, ter, ten, 't)</i>	bepaald	LB

'n	onbepaald	LO
Adverbs		
<i>nogal</i>	stellend/graad	BS
<i>vinniger</i>	vergroterend/ wyse	BV
<i>gereeldste</i>	oortreffend/frekwensie	BO
Prepositions		
<i>op (op die stoep)</i>	voorsetsel	SVS
Conjunctives		
<i>En</i>	neweskikkend	KN
<i>Aangesien</i>	onderskikkend	KO
Number words		
<i>twee (twee meisies)</i>	hooftelwoord/adjektief/bepaald	THAB
<i>twee (daar was net twee)</i>	hooftelwoord/voornaamwoord/bepaald	THPB
<i>twee (twee is beter)</i>	hooftelwoord/naamwoord/bepaald	THNB
<i>twee (twee honderd)</i>	hooftelwoord/ orpho/bepaald	THBB
<i>horde</i>	hooftelwoord/adjektief/onbepaald	THAO
<i>hordes</i>	hooftelwoord/voornaamwoord/onbepaald	THPO
<i>meer (meer opwindend)</i>	hooftelwoord/ orpho/onbepaald	THBO
<i>eerste</i>	rangtelwoord/adjektief/bepaald	TRAB
<i>eerstes</i>	rangtelwoord/voornaamwoord/bepaald	TRPB
<i>eerstens</i>	rangtelwoord/ orpho/bepaald	TRBB
<i>laastes</i>	rangtelwoord/voornaamwoord/onbepaald	TRPO
<i>soveelste</i>	rangtelwoord/adjektief/onbepaald	TRAO
<i>laastens</i>	rangtelwoord/ orpho/onbepaald	TRBO
Expletives		
<i>Eina!</i>	Tussenwerpsel	W
Unique/Unspecified		
<i>"dis"</i>	dis	UDS
<i>daar's</i>	enklities	UE
<i>daar (Daar word beweer dat...)</i>	eks-daar	UXD
<i>en (Hy sit en orp)</i>	partikel/skakel	UPE
<i>se; van (Piet se boek; die boek van Piet)</i>	partikel/genitief	UPS
<i>so; soos; nes (vet soos 'n vark)</i>	partikel/vergelyking	UPV
<i>nie (Ek ken nie die antwoord nie)</i>	partikel/ontkenning	UPO
<i>van (een/twee/'n klomp van die orpho)</i>	partikel/deel	UPD
<i>so/wat (So/Wat 'n mooi dag!)</i>	partikel/betreklik	UPB
<i>te (Hier is te veel mense)</i>	partikel/graad	UPG
<i>te (besig om te werk; iets te ete/drinke/lese)</i>	partikel/infinitief	UPI
<i>oor; in; op; ens. (skryf dit teen die bord op)</i>	partikel/ww.	UPW
Residue		
<i>facto</i>	vreemdetaalwoord	RV
<i>a+b</i>	formule	RF
<i>%</i>	simbool	RS
<i>ABSA</i>	akroniem: letterklankwoord	RK
<i>mv</i>	afkorting	RA
<i>AA</i>	akroniem: letternaamwoord	RL
<i>???</i>	ongeklassifiseerd	RO
<i>on (dit is on)waarskynlik)</i>	prefiks	RPF
<i>e (die moord(e) is gepleeg deur)</i>	suffiks	RSF
<i>fakulteits- (die op die fakulteits- en universiteitsraad)</i>	woorddeel	RWD
Punctuation		
<i>.!?</i>	sinseinde	ZE

, : ;	sinmiddel	ZM
([{ “	links-parentese	ZPL
)] } ”	regs-parentese	ZPR

Changing the protocol

Since this protocol is considered to be a dynamic document, it is subject to change from time to time. For purposes of good governance, the following principles and procedures should be followed:

- The language expert (in consultation with the project leader) has the final say in changing this procedure document, and is the only authoritative person who might change the protocol.
- If an annotator (or other project member) wishes change a specification, s/he should log a request on Bugzilla, indicating the nature of the change request. The language expert then considers the request, and adds the change to the protocol document. The protocol document is saved with a new version number. All other annotators are then informed of the change, and requested to change previous annotations, where applicable.
- If an annotator wishes to add specific examples to the protocol, s/he sends an email to the language expert with the suggested examples. The language expert then adds the examples to this document, and when a significant number of examples have been added, saves the document with a new version number. All other annotators are then informed of the new version of the protocol document.

Future work

During quality control a few examples were noted that perhaps call for future work and subsequent new or additional part of speech tags. These examples are currently annotated with RO (ongeklassifiseerd) that relates to unclassified tokens.

The first couple of examples pertain to adjectives that are used independently (selfstandige gebruik van adjektiewe).

They are:

- *belangrikste* in the sentence “*Die belangrikste is dat ons sukses in die stryd teen misdaad ...*” (part 4).

A possible new tag **AON** (adjektief/oortreffend/naamwoord) could be considered.

- *eie* in the sentence “*... ten volle aanvaar en ons eie gemaak.*” (part 1).

A possible new tag **ASN** (adjektief/stellend/naamwoord) could be considered.

The second example refers to a *telwoord* that is used independently. At the moment the tag set caters only for a count word to function as a definite noun, THNB (telwoord/hoof/naamwoord/bepaald), but not as an indefinite noun as in the example below.

- *paar* in the sentence “*... om net 'n paar te noem.*” (toetstek).

A possible new tag **THNO** (telwoord/hoof/naamwoord/onbepaald) could be considered.

Roles and contact details

The following roles in this project can be identified:

- **Project leader:** Responsible for successful technical delivery of project.
 - Martin Puttkammer
martin.puttkammer@nwu.ac.za
- **Project manager:** Responsible for timeous delivery of project, within budget. If you suspect that you will not be able to deliver on a set milestone, contact the project manager.
 - Ané Bekker
ane.bekker@nwu.ac.za
- **Language expert:** Responsible for language specific requirements and specifications. This person also controls the protocol document.
 - Afrikaans: Gerhard van Huyssteen
gerhard@trifonius.co.za
082 562 4814
- **Annotator:** Responsible for annotating the corpus, according to the above procedures and principles.
 - Afrikaans: Suléne Pilon and Anneke Butler

Appendix

Nouns

Notes

- *Massanaam*, *maatnaam* en *versamelnaam* moet bykom by *Tipe* omdat alle naamwoorde in Afrikaans wat nie eiename is nie, nie soortname is nie. Hierdie kategorieë is belangrik omdat massaname (gewoonlik) nie meervoud en verkleining kan neem nie; versamelname en massaname word nie vas aan die volgende naamwoord geskryf nie.
- *Dimunitief* word bygevoeg onder *Graad*. EAGLES laat dit (en ander verbuigings) uit, aangesien dit suiwer orphologies is. *Meervoud* val ook hieronder, aangesien dit, vir Afrikaans, geen invloed op die etikette van die res van die sin het nie.
- “*Algemeen*” moet by *Genus* kom as teenpool vir *onsydig* in ’n twee-genus sisteem soos die van Deens en Nederlands – dit is egter nie relevant vir Afrikaans nie. Insgelyks word *vokatief* en *onverbuigbaar* by *Kasus* gesit vir byvoorbeeld Grieks, maar is onbelangrik vir Afrikaans. Ook *Definitiwiteit* is ’n kategorie wat volgens EAGLES belangrik is vir Deens, maar kan weggelaat word vir Afrikaans.
- *Telbaarheid* word uitgelaat aangesien dit presies ooreenstem met *Tipe* – slegs massanaamwoorde is ontelbaar. Dit is dus “dubbele werk” om ’n *Telbaarheid*-eienskap te hê.

Some guidelines

- Words like “emigrant(e)” and “onderwyser(s)” where the singular and plural form of the nouns are implied, are not tagged as nouns but as unspecified (“onspesifiseerd”) with the tag RO.
 - Tref onderskeid tussen:
 - Soortnaam: *stoel*
 - Kan meervoud neem
 - Kan *die* en *’n* vooraan sit
 - Eienaam: *Afrikaans*
 - Word met hoofletter geskryf
 - Massanaam: *goud*
 - Kan nie meervoud neem nie
 - Kan *die* vooraan sit, maar nie *’n* nie
 - Maatnaam: *koppie/emmer*
 - Kan gemeet word
 - Word nie vas aan volgende naamwoord geskryf nie: *koppie meel*
 - Versamelnaam: *trop/swerm/skool*
 - Dui versameling aan
 - Word nie vas aan volgende naamwoord geskryf nie: *skool visse*
 - Abstrak: *liefde/haat*
 - If it is PL or can take PL, it is a count noun (**NSE, NSM, NSED, NSMD**)
 - Words can be different nouns in different contexts, e.g. “gebruik” in the following sentences:
 - Maak *gebruik* (NA) van ... [make use of ...]
 - Die *gebruike* van ’n volk (NSE) ... [the customs of ...]
- The same applies for:

- *gees* as spirit = NA and *gees* as ghost = NSE
 - *gemeenskap* as community of people = NSE and *binne gemeenskap van goedere* as community of property = NA
 - *konstruksie* as the frame for a building = NSE and *konstruksie* as the building industry = NA
 - *raad* that means advice = NM and *raad* that means the board = NSE.
 - *'n reg* that means a right = NSE and *die reg* that means law = NA.
- Note that words that are inherently categorised as mass nouns or abstract nouns (e.g. *water* or *liefde*) can be count nouns in some contexts, and should then be tagged as count nouns.
 - Ek soek twee waters (=NSM)
 - Die liefde (=NSE) van my lewe.
- If it is a deverbal or deadjectival nominalisation, it is an abstract noun (NA)
 - *bestuur, bekruij*
 - *gelag, gelaggery, gespeel*
 - *inligting, werking*
 - If it can not take PL and is sensorically observable or quantifiable in some concrete way, it is a mass noun (NM)
 - If it can not take PL and is not sensorically observable or quantifiable in some concrete way, then it is a abstract noun (NA)
 - All school subjects/fields of study, etc. are NA – even when they are written with capital letters
 - Typical NA suffixes include (but are not limited to):
 - –skap
 - –logie
 - –asie
 - –kunde
 - If it quantifies a mass noun, then it is a measure name (NME, NMM, NMED, NMMD)
 - If it groups a count noun, then it is a collective noun (NVE, NVM, NVED, NVMD)
1. Can it (a) be preceded by the indefinite determinator *'n*; **and** (b) take PL?
 - a. Yes: NS
 - b. No: Go to next
 2. Is it a deverbal/deadjectival nominalisation?
 - a. Yes :NA
 - b. No: Go to table

3.

	PL	Sensory observable/concretely quantifiable
NS	+	+
NM	-	+
NA	-	-

- Email addresses and URL's are tagged as NEE, and marked as named entities on the secondary level.
- “gebeure”, “besonderhede” en “omstandighede” is NM, want dit kan waargeneem word. Sien lys in Tokenisation Protocol.
- Sommige naamwoorde kan 'n enkelvoud- of meervoudinterpretasie kry, byvoorbeeld “inkomste”, “onkoste”, “uitkomste”, “klagte”, “bylae”, “pianiste”, ens.

As dit onduidelik uit die konteks is, moet sulke gevalle by verstek as NSE geëtiketteer word.

Named entities

Words that are used in context to refer to a single, unique instantiation through convention, an institutionalised process or process, and of which the orthography is determined by law, or by some authority (Puttkammer, 2005).

- Tag by default any named entity (e.g. *Pieter Botha*) as **NEE/NEM/NEED/NEMD**, and also indicate on the secondary level that it is a named entity.
- In some cases (e.g. *Department van Kuns en Kultuur*) these ordinary words should be tagged as ordinary words (e.g. *Departement* should be tagged as **NSE**, and *en* should be tagged as **KN**), and on the secondary level indicated as named entities.
- English words that are part of proper nouns should be tagged as such (**NEE/NEM/NEED/NEMD**), and not as “other language”.
 - Question: Is it a word from another language AND used as a proper noun?
 - Tag as proper noun, and mark on secondary level as named entity (in other words, not necessary to indicate anywhere that it is from another language).
 - Question: Is it a word from Afrikaans AND used as a proper noun?
 - Tag as ordinary word, and mark on secondary level as named entity

Verbs

Notas

- Oorganklikheid:
 - ‘n Duimsuigtoets vir oorganklikheid is om te vra “S V wie/wat?”. As jy antwoord kry, is dit oorganklik?
 - As ‘n sin in die passief is, is die werkwoord per definisie oorganklik; parafrasering is “X V wie/wat?”
 - As die werkwoord oorgaan in ‘n bysin wat as “dit” geparafraseer kan word, is die werkwoord oorganklik (geparafraseer: “S V dit”)
 - As daar ‘n wederkerende (“hom”/”homself”) of wederkerige (“mekaar”) voornaamwoord is, is die werkwoord oorganklik
 - M.b.t. lang infinitiefsinne: parafraseer: “Om te V X”; as X ‘n waarde het, is die werkwoord oorganklik
 - Dit is lekker
 - Dit=POENP
 - Om te eet, is lekker
 - Eet=VTHOO
 - Om pap tee et, is lekker
 - Eet=VTHOG
 - Om te glo dat kennis mag is, is gewild
 - Om te glo dit, is gewild
 - Dus: glo = oorganklik
 - Voorsetselwerkwoorde is altyd oorganklik (VVHOV/VTHOV). Voorsetselwerkwoorde moet in praktyk by uitsondering as sodanig geëtiketteer word. Gebruik die volgende toetse:
 - Is hierdie werkwoord moontlik ‘n deeltjiewerkwoord?

- Ja: kan glad nie voorsetselwerkwoord wees nie
 - ??? 'n Skeibare werkwoord - (V(T|V)HS(G|O) – kan nooit 'n voorsetselwerkwoord wees nie.
 - Nee: Gaan na toets 2
- Is daar in die sin 'n voorsetsel met verpligte voorwerp ná die voorsetsel?
 - Nee: Hy appelleer teen die uitspraak (Jy kan net sê “Hy appelleer” en dan bly “appelleer” se betekenis behoue.)
 - Ja: Hy reken op haar (Jy kan nie net sê “Hy reken” nie – dit beteken dan iets anders.)
- Ander bevestigende toetse sluit in:
 - Daar is 'n besondere noue verwantskap is tussen die werkwoord en die spesifieke voorsetsel.
 - Die voorsetsel kan nie in die spesifieke sin met 'n ander voorsetsel vervang word nie.
 - Let op enkele uitsonderings:
 - wanhoop oor/aan
 - behoort tot/aan
 - Kan dikwels met 'n enkele Engelse word vertaal word
 - Hou van = like
 - Grens aan = borders
 - Voorsetselwerkwoorde kan nooit as iets anders geëtiketteer word nie
 - “hou” in “hou van” is altyd VTHOV
- Voorbeelde:
 - teer op
 - hou van
 - behoort tot/aan
 - wanhoop oor/aan
 - voldoen aan
 - grens aan
 - reken op
- “Frasewerkwoorde” kan genominaliseer word (inagneming), maar voorsetselstukke kan nie genominaliseer word nie (voor die hand liggend)
 - in ag neem
 - in staat stel
 - seker maak (seker=BS; maak=V - kyk binne konteks)
 - nodig hê
 - Let op: Die medisyne is nodig (nodig=ASP)
 - los skryf
- Werkwoorde mag slegs as skeibaar gemerk word wanneer dit 'n teenwoordigetyddeeltjiewerkwoord is, waarvan die dele in die onderhawige geval vas geskryf is. Let dus op dat as dit in die verlede tyd is, of as die twee dele van mekaar geskei is in die onderhawige geval, kan dit nooit as skeibaar geëtiketteer word nie.
 - Hy wil die trui aantrek (aantrek=VTHSG)
 - Hy trek die trui aan (trek=VTHOG; aan=UPW)
 - Hy het die trui aangetrek (aangetrek=VVHOG)
- Alle skeibare werkwoorde is deeltjiewerkwoorde, maar deeltjiewerkwoorde waarvan die dele los geskryf is, is nie skeibaar nie (want dit is klaar geskei).
- Ongemerkte werkwoorde wat in verledetydsinne voorkom en dus lyk soos teenwoordigetydwerkwoorde, word as ongemerkte werkwoorde geëtiketteer
 - Hy het die brief onderteken (onderteken=VTHOG)
 - Hy onderteken die brief (onderteken=VTHOG)

- Die brief is onderteken (is=VVUOP, onderteken=VTHOG)
- Wanneer “is” en “wees” hoofwerkwoorde is (i.e. koppelwerkwoorde) is, word dit hanteer as ongemarkeerd, en beide is dan VTHOK
 - Die kinders is soet in die kerk (is=VTHOK)
 - Die kinders sal soet wees in die kerk (wees=VTHOK)
- “is” en “word” kan koppelwerkwoorde of hulpwerkwoorde van vorm (VVUOP/VTUOP) wees. Let op dat “wees” moontlik ook ‘n hulpwerkwoord van vorm kan wees wanneer die passiewe sin in die toekomende tyd uitgedruk word, bv. Teen môre sal die huis klaar geverf (VVHOG) wees (VTUOP).)
 - As dit met ‘n deelwoord voorkom, is “is” ‘n VVUOP en “word” ‘n VTUOP.
 - is/word (=VVUOP/VTUOP) geregistreer/geskiet/gebyt/gebore (=VVHOG)
 - is (=VVUOP) getroud/geskei (=VVHOG)
 - is/word (=VVUOP/VTUOP) bederf/hersien/ervaar/onthou/vergeet (=VTHOG)
 - Let op: ‘n deelwoord is nie altyd ‘n VTHOG of VVHOG nie! In sommige konstruksies kan dit ook ASA of ASP wees. Wanneer ‘n deelwoord saam met “was” of “wees” in ‘n sin voorkom, is die deelwoord altyd ‘n ASP.
 - As dit die enigste werkwoord in die sin is, is “is/word” ‘n koppelwerkwoord (VTHOK).
 - Hy is mooi. (is = VTHOK)
 - Hy word ‘n onderwyser. (word = VTHOK)
 - As dit saam met ‘n sterk verlede deelwoord voorkom, is “is” ‘n koppelwerkwoord (VTHOK).
 - My hart is gebroke. (is = VTHOK, gebroke = ASP)
 - Hy is betrokke. (is = VTHOK, betrokke = ASP)
 - As “is/was/word/wees” saam met ‘n adjektief (ASP) voorkom, is dit ‘n koppelwerkwoord (VTHOK/VVHOK).
 - Sy is oorlede. (is = VTHOK, oorlede = ASP)
 - Sy was gesond. (was = VVHOK, gesond = ASP)
- “het” en “hê” word hier hanteer as ongemarkeerd, en beide is VTHOG
 - Ek het ‘n kar
 - Ek wil ‘n kar hê
- As daar ‘n “vir” as deel van die voorwerp is (“X glo vir Susan”), is “glo” oorganklik, want dit is gelyk aan “X glo Susan”

Some guidelines

- Belangrik vir etiketstel:
 - Tyd: teenwoordig/toekomend/verlede
 - Status: hoof/hulp
 - Skeibaarheid: *ek skryf by die skool in* vs. *ek moet by die skool inskryf*
 - Tipe:
 - oorganklik
 - slegs hoofwerkwoord
 - onoorganklik
 - slegs hoofwerkwoord
 - koppel
 - slegs hoofwerkwoord
 - *bly*
 - *gaan*
 - *heet*

- *klink*
- *smaak*
- *voel*
- *word*
- *lyk*
- *ruik*
- voorsetsel
 - slegs hoofwerkwoord
- deeltjie
 - slegs hoofwerkwoord
- skakelwerkwoord
- hulpwerkwoord van modaliteit
- hulpwerkwoord van modaliteit
 - *sal*
 - *het*
 - *wil*
 - *wou*
 - *sou*
- aspek
 - sien lyste
- modus
 - sien lyste
- omskrywingswerkwoord
 - *is*
 - *word*

Some problematic cases/examples

- Hy plaas die onderneming onder administrasie
 - Plaas=VTHOG
 - Onder=SVS
 - Administrasie=NA
- *BTW ingesluit* (hanteer asof “BTW is ingesluit deur die maatskappy”)
 - BTW = RL
 - Ingesluit = VVHOG

Adjektiewe

Notas

- Let op na die volgende met betrekking tot deelwoorde:
 - Wanneer ‘n deelwoord saam met “was” of “wees” gebruik word, is die deelwoord ‘n adjektief en nie ‘n werkwoord nie.
 - Hy was getroud. (was=VVHOK, getroud=ASP)
 - Sy wil nie gesteur wees nie. (wees=VTHOK, gesteur=ASP)
 - Sterk verlede deelwoorde word altyd as adjektiewe hanteer.
 - Die besope man is gearresteer. (besope=ASA)
 - Sy wil nie betrokke raak nie. (raak=VTHOK, betrokke=ASP)
 - Sien ook relevante besprekings onder “Verbs”.

Bywoorde

Notas

- 'n Bywoord van Skakeling voeg nie twee sinne aan mekaar nie.
 - Dit stel jou dus in staat om... (dus = BSS)
 - Jy het nie geslaap nie, dus is jy moeg (dus = KO)
- “ook” is altyd BSS
- Wanneer “meer” en “mees” gebruik word in trappe van vergelyking (m.a.w. as die lemma “baie” is), is dit 'n bywoord.

Voorsetsels

Notas

- Infinitiefkonstruksie (om...te): om = SVS, te = UPI
- “ongeag” is meestal 'n voorsetsel
- “ooreenkomstig” is 'n voorsetsel as dit ‘volgens’ beteken (*Ooreenkomstig sy insig word 'n man geprys*). Andersins is dit 'n A of B (*'n Afskrif ooreenkomstig die oorspronklike*).
- Die onderstaande woorde is meestal (lees: amper altyd) voorsetsels
 - ooreenstemmend
 - ooreenkomstig
 - volgens
 - sonder
 - ongeag
 - ten
 - weens
- Sommige voorsetsels lyk soos deelwoorde, maar word as voorsetsels gemerk.
 - insluitende
 - gedurende
 -

Konjunkte

Notas

- “en” is altyd KN, behalwe in die verbalehendiadiskonstruksie (sit en slaap) waar dit dan geëtiketteer word as UPE.

- “of” is KN wanneer dit in Engels met “or” vertaal kan word, maar KO wanneer dit met “if” in Engels vertaal kan word.
- “daarom” is meestal ’n voegwoord (KO). Let op as “daarom” as ’n voorsetsel gebruik word in “die lint is daarom gebind” (dan SVS).
- Merk “behalwe” deurgaans as a voegwoord (KO) al dui die HAT en WAT ook aan dat dit ’n voorsetsel kan wees.

Voornaamwoorde

Notas

- “wie”, “wat”, “waar”, “wanneer” en “watter” is PB of PV, afhange van die sinsoort.
 - Let op dat HAT “waar” verkeerdelik aandui as voegwoord.
- Wanneer die betreklike voornaamwoord “wat” met ’n voorsetsel verbind (*waarin, waarop, waarlangs*) word die resulterende woord as PB geëtiketteer. Dit geld ook *waarheen*.
- elke = PO

Telwoorde

Notas

- enkele (wat ’n paartjies beteken – bv. “enkele studente het beter presteer”) = THAO
- enkele (wat een beteken – bv. “daar is slegs ’n enkele tienrandnoot in my beursie”) = THAB
- *albei* = THAB
- *alle* = THAO
- *geen* = THAB (en tweede ontkenningwoord is UPO)

Consistency checks

- The following words should be marked as SVS:
 - *insluitende/insluitend*
 - *gedurende/gedurende*
 - *ongeag*
 - *bykomend*
 - *tydens*
- Make “geen” THAB and the second “nie” UPO.
- Search for “dat” and make sure that the verb before “dat” is marked as V(V|T)HOG
- Kontroleer dat die volgende voegwoorde (KO) is:

- Sodra
- Nadat
-

Fixed expressions

- Afrikaans has a large number of fixed expressions where relicts of inflection processes have been fossilised. In such cases, **all the words in the archaic fixed expression are left unchanged** (i.e. treated as lemmas).
 - *ter* *ter*
 - *harte* *harte*
 - *neem* *neem*
- The following examples are taken from, amongst others, Botha, JP & Van Aardt, JMH. 1978. *Afrikaans vir die praktyk*. Johannesburg: McGraw-Hill.
- Words in archaic fixed expression should be tagged with the closest possible POS tag, while every word should also be marked as fixed expression. See the lists in the appendix.
 - *in* **SVS**
 - *der* **LB**
 - *ewigheid* **NA**

Old genitive

<i>allersins</i>	<i>enigsins</i>	<i>smiddags</i>
<i>brood ter smarte</i>	<i>geensins</i>	<i>smorens</i>
<i>deesdaags</i>	<i>goedsmoeds</i>	<i>snags</i>
<i>desgelyks</i>	<i>grotendeels</i>	<i>soggens</i>
<i>desnoods</i>	<i>kind des doods</i>	<i>steen des aanstoots</i>
<i>desondanks</i>	<i>kwelling des geestes</i>	<i>stem des roepende</i>
<i>destyds</i>	<i>myns insiens</i>	<i>sweet des aanskyns</i>
<i>eersdaags</i>	<i>onses insiens</i>	<i>syns insiens</i>
<i>engel des ligs</i>	<i>saans</i>	<i>syns weegs</i>
<i>Gods water</i>		

- Note that the partitive genitive **-s** should be removed in cases like:
 - *iets goeds* *iets goed*
 - *niks nuuts* *niks nuut*

Old dative

<i>agterweë</i>		
<i>byderhand</i>	<i>by magte</i>	<i>by monde van</i>
<i>by wyse van</i>	<i>by ontvangs van</i>	
<i>dienaangaande</i>	<i>deser dae</i>	
<i>in der ewigheid</i>	<i>in der mine</i>	<i>in dier voege</i>
<i>in deser voege</i>		
<i>met dien verstande</i>	<i>met referte tot</i>	<i>mettertyd</i>
<i>metterwoon</i>		
<i>onder andere</i>	<i>op heter daad</i>	<i>onder meer</i>
<i>sedertdien</i>	<i>sindsdien</i>	
<i>te berde bring</i>	<i>te dien effekte</i>	<i>te dien einde</i>
<i>te gelegener tyd</i>	<i>te goeder trou</i>	<i>te lande</i>

<i>te velde trek</i>	<i>ten aanhore van</i>	<i>ten bate van</i>
<i>ten behoefte van</i>	<i>ten beste</i>	<i>ten dele</i>
<i>ten eerste, tweede, ens.</i>	<i>ten einde</i>	<i>ten enemale</i>
<i>ten gevolge van</i>	<i>ten/te gronde gaan</i>	<i>ten gunste van</i>
<i>ten huise van</i>	<i>ten koste van</i>	<i>ten laste</i>
<i>ten minste</i>	<i>ten nadele van</i>	<i>ten onregte</i>
<i>ten opsigte van</i>	<i>ten regte</i>	<i>ten seerste</i>
<i>ten slotte</i>	<i>ten spyte van</i>	<i>ten strengste</i>
<i>ten volle</i>	<i>ten voordele van</i>	<i>ter aarde bestel</i>
<i>ter bekwamer tyd</i>	<i>ter dood (veroordeel)</i>	<i>ter elfder ure</i>
<i>ter ere van</i>	<i>ter harte neem</i>	<i>ter insae</i>
<i>ter insage</i>	<i>ternouernood</i>	<i>ter ore kom</i>
<i>ter perse gaan</i>	<i>ter plaatse</i>	<i>(wel) ter ruste</i>
<i>ter sake</i>	<i>ter sprake</i>	<i>ter tafel</i>
<i>wel ter tale</i>	<i>ter verduideliking</i>	<i>ter wille van</i>
<i>uit hoofde van</i>		
<i>van aller harte</i>	<i>van harte (geluk)</i>	<i>van nature</i>

NCHLT-TOETSTEKS

Laai die elektroniese aansoekvorm af of kry dit by jou naaste SAID-kantoor.

Die aansoekvorm moet deur die aansoeker self ingevul word en nie deur 'n derde party nie, anders sal die aansoek nie oorweeg word nie.

Raadpleeg die aansoekgids om jou deur die lisensiëringsproses te help.

Doen die volgende in die afdeling Client Types:

- merk die blokkie Manufacturer for AGOA onder registration.

Verstrek die volgende in die afdeling Manufacturer for AGOA Particulars:

- jou onderneming se handelsnaam
- die plek waar die onderneming fisies geleë is (bv. Plot 10)
- die straatnaam en -nommer van die onderneming
- die voorstad waarin die onderneming geleë is
- die stad of dorp waar die onderneming geleë is
- die poskode van die voorstad waarin die onderneming geleë is.

Die afdeling Authority to Apply moet deur die alleeneienaar, direkteur, vennote, lid, trustee, of persoon wat die besigheid bestuur ingevul word wanneer daar aansoek gedoen word om as 'n doeane-en-aksyns spesiale vervaardiger vir AGOA geregistreer te word.

Die naam van die aansoeker of die geregistreerde naam van die entiteit, en nie die handelsnaam nie, moet in hierdie veld ingevul word.

waar 'n asterisk (*) voorkom, skrap dit wat nie van toepassing is nie.

In die afdeling Declaration moet jy die onderstaande inligting verstrek:

- die voorletters en van van die persoon wat om lisensiëring aansoek doen
- die aansoeker se status, bv. direkteur, bestuurder, eienaar
- die aansoeker moet die aansoekvorm onderteken
- die datum waarop die aansoek onderteken is
- die plek waar die aansoek onderteken is.

Vul die aansoekvorm in en handig dit by die naaste SAID-doeane-en-aksynskantoor in.

Heg oorspronklike of gewaarmerkte kopieë van die volgende dokumente by die aansoekvorm aan:

- die onderneming se registrasiesertifikaat, soos uitgereik deur die registrateur van maatskappye of, in die geval van 'n trust, deur die meester van die hooggeregshof
- besluit, toestemming, of enige ander toepaslike magtiging
- 'n Permit van die Departement van Handel en Nywerheid - indien van toepassing
- formules - indien daar ingevolge kortingsitem 607.04.10(02), 607.04.10(03), 607.04.10(09) en 607.04.10(12) aansoek gedoen word
- identiteitsdokumente - in die geval van 'n Suid-Afrikaanse burger

Paspoortdokumente - in die geval van:

- 'n nie-inwonende individu ('n nie-ingesetene)
- 'n beslote korporasie en trust: alle lede, vennote of trustees
- 'n maatskappy: alle direkteure, insluitende die besturende direkteur en die finansiële direkteur.

Crime Stop is 'n reeks oproepsentrums wat by die publiek inligting oor misdadigers en hulle bedrywighede insamel. Lede van die publiek word gehelp deur oproepsentrumpersoneel wat goed opgelei is in ondersoekende onderhoudvoering. Dit stel jou in staat om die Suid-Afrikaanse Polisie (SAPD) in 'n veilige omgewing, en sonder enige bedreiging, oor kriminele bedrywighede in te lig. Jy kan vra om anonimiteit te bly (dit wil sê om nie te sê wie jy is nie.) Om te waarborg dat informante se identiteit nie bekend sal word nie, word opsporingsmeganismes en inbeller-identifikasietegnologie nie by Crime Stop toegelaat nie. Die inligting wat jy verstrek, word aan speurders oorgedra.

Sekere Crime Stop-oproepsentrums verbind informante ook regstreeks met ondersoekbeamptes deur die telefoonkonferensiestelsel te gebruik. Dit stel die ondersoekbeampte in staat om die informant regstreeks te ondervra.

Let wel: dit word slegs gedoen indien jy daartoe instem.

Jy kan kriminele bedrywighede ook elektronies op die Suid-Afrikaanse Polisie (SAPD) se webwerf aanmeld. Jou anonimiteit word ook hier gewaarborg, behalwe as jy aandui dat dit nie gedoen hoef te word nie.

Bel die Crime Stop-nommer: 08600 10 111.

Die agent by die oproepsentrum sal sekere vrae aan jou stel. Indien jou inligting bruikbaar is, sal hy of sy 'n unieke kodenommer aan jou verstrek.

Indien die inligting wat jy verstrek nie bruikbaar genoeg is nie, sal die agent aan jou verduidelik waarom dit nog nie bruikbaar genoeg is nie en jou raad gee oor hoe om bykomende inligting te bekom.

Bruikbare inligting word gekoppel aan die kodenommer wat aan jou verstrek is.

Jy sal gevra word om Crime Stop binne 'n gegewe tyd weer by 08600 10 111 te bel om te hoor of die inligting positief was en of jy vir 'n beloning kwalifiseer (mits jy vooraf om 'n beloning aansoek gedoen het).

In die toekoms sal die agent die saak slegs met jou bespreek indien jy die korrekte kodenommer verstrek, en mits jou identiteit ooreenstem met die 'profiel' wat op rekord is.

Die agent sal die oproep onmiddellik beëindig indien hy of sy vermoed dat jy nie die persoon is wat die oorspronklike oproep gemaak het nie. Hierdie is 'n beskermingsmaatreël, en is nie onderhandelbaar nie.

Indien jy 'n misdaad elektronies aanmeld, moet jy die inligting op die terugvoervorm op die SAPD-webwerf verstrek. 'n Verwysingsnommer sal op die skerm verskyn sodra jy die vorm ingedien het.

Crime Stop sal die inligting aan die betrokke ondersoekende oordra om op te volg, met die versoek om binne 24 uur terug te rapporteer en so gou moontlik oor die afloop van die ondersoek verslag te doen.

Indien jy vir 'n beloning in aanmerking kom (en jy wel daarvoor aansoek gedoen het), sal die ondersoekbeampte die bedrag van die beloning vasstel en die eis ooreenkomstig National Information 2/2001 (Nasionale Inligting 2/2001) verwerk.

Sodra die beloning gereed is, sal Crime Stop met jou in verbinding tree om af te spreek oor hoe die betaling sal geskied.

Ek staan vanaand hier voor u, 20 jaar nadat President Nelson Rolihlahla Mandela as 'n vry man uit die gevangenis gestap het.

Ons het juis hierdie dag gekies om hierdie Gesamentlike Sitting van die Parlement byeen te roep om die Staatsrede te lewer en sodoende 'n waterskeidingsoomblik wat ons land se geskiedenis vir ewig verander het, te gedenk.

Madiba se vrylating was die uitkoms van Suid-Afrikaners se onvermoeide stryd.

Voormalige politieke gevangenes en veterane wat hier saam met ons is, was deel van daardie proses en kan getuig daarvan.

U sal natuurlik onthou dat die massas in hierdie land, in hul onderskeie formasies, met vasberadenheid gereageer het op die oproep om die land onregeerbaar en apartheid onwerkbaar te maak.

Ons vier hierdie dag saam met voormalige politieke gevangenes wat ons spesiaal uitgenooi het om hier by ons aan te sluit.

Ons verwelkom in die besonder diegene wat uit die buiteland gekom het om vandag hier te wees: Helene Pastoors, Michael Dingake van Botswana, mnr Andimba Toivo ya Toivo van Swapo in Namibië.

Dit is ook ons voorreg om die regsman van die Rivonia Hoogverraadverhoor - Lord Joel Joffe, wat tans in Londen gebaseer is, en Regter Arthur Chaskalson hier by ons te hê.

Ons onthou en vereer vandag ook mnr Harry Schwarz, wat verlede week oorlede is.

Hy was onder andere ook 'n lid van die Rivonia-regspan.

Ons bedank graag ons vriende en comrades in die internasionale gemeenskap vir hul stryd saam met ons om ons vryheid te wen.

'n Spesiale woord van welkom aan die Madiba-familie.

Hulle is 'n simbool van die opofferings wat baie moes maak wat soveel weens apartheid moes trotseer.

Ons groet die leiers van die regerende party en sy Alliansie-vennote, vir wie hierdie 'n besonder spesiale geleentheid is.

Landsgenote en vriende,

Op hierdie spesiale dag moet ons ook die bydrae erken van diegene in die leierskorps van die Nasionale Party, wat uiteindelik besef het dat apartheid geen toekoms sou hê nie.

Laat ek spesifiek die rol van voormalige President PW Botha hier noem.

Dit was hy wat die bal aan die rol gesit het met gesprekke oor die moontlike vrylating van politieke gevangenes.

President Botha het met die voormalige Minister van Justisie, mnr Kobie Coetzee, saamgewerk, wat op sy beurt hierin bygestaan is deur dr Neil Barnard en mnr Mike Louw.

Hulle het 'n sleutelrol gespeel in die proses wat tot Madiba se vrylating gelei het.

Agbare landsgenote,

Suid-Afrika moet nog ten volle erkenning gee aan die deurslaggewende rol van die voormalige President van die ANC, Comrade Oliver Tambo, wat die grondslag gelê het vir hierdie land om 'n uitmuntende voorbeeld van demokrasie en vryheid te word.

Dit is te danke aan sy uitstaande leierskap, versierendheid en duidelike visie dat hy die ANC kon lei om die strewe na 'n onderhandelde skikking te verskerp.

Sy wysheid het ook duidelik geblyk uit die Harare Deklarasie, wat hy geskryf en onvermoeid bepleit het.

Hierdeur is die basis gelê vir die historiese aankondigings wat President FW de Klerk 20 jaar gelede gemaak het.

President de Klerk het hierdeur groot moed en besliste leierskap aan die dag gelê.

Op hierdie groot dag wil ek ook die rol van wyle me Helen Suzman uitlig.

Sy was baie lank die enigste stem in die Parlement wat haar onverpoosd vir verandering beywer het.

Ons erken ook die rol van die leier van die Inkatha Vryheidsparty, Inkosi Mangosuthu Buthelezi, wat Madiba se vrylating geëis het, asook die vrylating van ander politieke gevangenes en die terugkeer van uitgewekenes.

Ons herbevestig ons innige dankbaarheid teenoor die internasionale gemeenskap vir hul onwrikbare steun vir ons stryd.

Hierdie oomblikke in ons geskiedenis demonstreer ons vermoë om saam te werk, selfs onder moeilike omstandighede, en om die land se belange bo alle ander oorwegings te stel.

Deur saam te werk, kan ons meer bereik.

Agbare lede,

Gedurende die loop van hierdie jaar sal ons die eeufeesviering van die stigting van die Unie van Suid-Afrika vier, wat in 1910 tot stand gekom het

Dit het tot 'n verenigde staat gelei.

Van besondere belang is dat die uitsluiting van swart mense van hierdie Unie een van die hoofredes was vir die stigting van die African National Congress in 1912.

Wanneer ons later vanjaar hierdie eeufees vier, moet ons besin oor hoe ver ons reeds as 'n land gekom het.

Agbare lede,

Ons onthou vandag die woorde van Madiba ten tyde van sy vrylating, en ek haal aan:

Ek staan voor julle, nie as 'n profeet nie, maar as 'n nederige dienaar van julle, die mense.

Julle onvermoeide en heldhaftige opofferings het dit vir my moontlik gemaak om vandag hier te wees.

Daarom plaas ek die oorblywende jare van my lewe in julle hande.

Hierdie woorde inspireer ons om nie te rus totdat ons die ideale bereik het van 'n samelewing wat vry is van armoede en ontneming nie.

Gedurende die twee dekades sedert Madiba se vrylating het ons land fundamentele veranderinge ondergaan.

President Mandela het hierdie land verenig tot 'n nie-seksistiese, nie-rassige, demokratiese en welvarende Suid-Afrika as doelwit.

Waar ons vandag Madiba se vrylating vier, moet ons onself weer daartoe herverbind om 'n beter toekoms vir alle Suid-Afrikans, swart en wit, te bou.

Kom ons streef opnuut na die ideaal waarvoor Madiba sy hele lewe lank geveg het - die ideaal van 'n demokratiese en vrye samelewing, waarin alle mense in harmonie en met gelyke geleenthede saamleef.

Agbare lede,

Ons het hierdie gesamentlike sitting in die aand byeengeroep sodat die meerderheid van ons land se mense, werkers en skoolkinders, ook deel van die geleentheid kan wees.

Ons is beïndruk deur die entoesiasme van die jeug oor hierdie geleentheid.

Tweehonderd ses-en-sestig kinders van al die provinsies het deelgeneem aan die pre-Staatsrede-debat oor die rol van die jeug in die stryd teen armoede.

Baie geluk aan die algehele wenner, Charlotte Le Fleur van Worcester Sekondêre Skool, asook aan al die deelnemers vir hul harde werk.

Landsgenote en vriende,

Ons ontmoet hier teen die agtergrond van 'n wêreldwye ekonomiese krisis.

Verlede jaar het ons ons eerste resessie in 17 jaar ervaar. Die krisis het ons ekonomie ongeveer 900 000 werksgeleenthede gekos.

Baie mense wat hul werk verloor het, was die broodwinners van arm families.

In Februarie verlede jaar het die regering, sakelui, arbeidsektor en gemeenskapsverteenwoordigers ooreengekom op 'n pakket van maatreëls om die skaal en impak van die krisis te bekamp.

Ons het talle van hierdie maatreëls reeds in werking gestel.

Ons het besliste teen-resessiebesteding deur die regering geïmplementeer, veral op infrastruktuur.

Ten einde 'n veiligheidskussing vir armes te skep, het ons die verhogings in maatskaplike toelaes vervoeg, en die Kindersorgtoelaag uitgebrei na kinders ouer as 14 jaar.

Gedurende die volgende drie jaar sal 'n addisionele twee miljoen kinders uit arm gesinne en huishoudings tussen die ouderdom van 15 en 18 jaar, by die Kindersorgtoelaag baat vind.

Die Nywerheidsontwikkelingskorporasie het R6 biljoen bewillig om maatskappye in nood by te staan.

Die Regering het 'n "opleidingsafleggingskema" ingestel om aan werkers die opsie van 'n opleidingstydperk, in plaas van aflegging, te bied.

Hierdie pogings is deur ons Openbare Werke-program versterk.

Die nasie sal onthou dat ek tydens die 2009-Staatsrede aangekondig het dat die Uitgebreide Openbare Werke-program teen Desember 2009 500 000 werksgeleenthede sou skep.

Ek wil beklemtoon dat dit nie poste in die hoofstroom-ekonomie is nie. Hierdie is werksgeleenthede wat geskep is om werklose mense van 'n inkomste, ervaring en opleidingsgeleenthede te voorsien.

Agbare lede, mede-Suid-Afrikaners

Ons kondig graag aan dat ons teen die einde van Desember reeds meer as 480 000 openbare werksgeleenthede geskep het, met ander woorde, 97% van die doelwit wat ons gestel het.

Hierdie poste is in areas soos konstruksie, tuis- en gemeenskapsgebaseerde sorg en omgewingsprojekte.

Ons het 'n aantal terreine geïdentifiseer waar ons in die toekoms verder gaan verbeter, soos om meer arbeidsintensiewe projekte te verseker.

Ons weet egter dat hierdie en ander maatreëls nie die uitwerkings van die resessie heeltemal kan temper nie.

Ons is dankbaar vir die gees van familie, gemeenskap en vrywillige werk wat soveel mense inspireer om diegene wat die swaarste deur die krisis getref word, deur hierdie moeilike tye te help.

Agbare lede,

Ekonomiese aanwysers dui daarop dat ons nou deur die ergste van die ekonomiese resessie is.

Ekonomiese aktiwiteit is aan die toeneem in Suid-Afrika, en ons verwag groei in die toekoms.

Die arbeidstatistiek wat Dinsdag vrygestel is, dui daarop dat die ekonomie tans werksgeleenthede skep eerder as om dit in te boet.

Dit is egter te gou om seker te wees oor die tempo van hierdie herstelproses.

Die Regering sal dus nog nie sy ondersteuningsmaatreëls opskort nie.

Dit is nou die tyd om die basis te lê vir sterker groei in die toekoms en vir groei wat tot meer werksgeleenthede lei.

Ons langtermyn-infrastruktuurprogram sal ons help om vinniger te groei.

Ons onderwys- en vaardigheidsprogramme sal ons produktiwiteit en mededingendheid verhoog.

Ons Nywerheidsbeleids-aksieplan en ons nuwe fokus op groen werksgeleenthede sal sterker en meer arbeids-absorberende nywerhede bou.

Ons landelike ontwikkelingsprogram sal landelike produktiwiteit verhoog en die lewens van die mense in landelike gebiede verbeter.

Ons kapitaalinvesteringsprogram lê ten grondslag aan ons strategie vir ekonomiese herstel en groei.

Oor die volgende drie jaar sal die regering R846 biljoen aan openbare infrastruktuur spandeer.

Op die terrein van vervoer, sal ons ons padnetwerk in stand hou en uitbrei.

Ons sal verseker dat ons spoornetwerk betroubaar en mededingend is en beter met ons seehawens geïntegreer is.

Om betroubare kragtoevoer te verseker, het ons 'n Inter-Ministeriële Komitee oor Energie ingestel, om 'n 20-jaar- geïntegreerde hulpbronplan te ontwikkel.

Hierdie komitee sal, onder andere, kyk na die deelname van onafhanklike kragprodusente, en arm mense teen stygende elektrisiteitspryse beskerm.

Ons sal 'n onafhanklike stelseloperateur instel, wat afsonderlik van Eskom Beherend sal funksioneer.

Eskom sal voortgaan om addisionele opwekkingskapasiteit te bou en die instandhouding van sy kragstasies te verbeter.

Ten einde 'n inklusiewe ekonomie te verseker wat groei en ontwikkeling bevorder, het ons die Adviesraad oor Breë Swart Ekonomiese Bemagtiging in die lewe geroep, onder voorsitterskap van die President.

Beleidsverandering moet fokus op ingrypings om werkseleenthede vir jongmense te skep.

Werkloosheidskoerse vir jongmense is aansienlik hoër as die gemiddeld.

Voorstelle sal ingedien word om die koste van die indiensneming van jonger werkers te subsidieer ten einde ondernemings aan te moedig om onervare personeel in diens te neem.

'n Verdere uitbreiding van openbare-werkskeppingsprogramme is ook reeds aan die gang.

Dit sluit in plaaslike-infrastruktuur- en tersiêre projekte en geletterdheidsprojekte, tuisgebaseerde versorging, skoolinstandhouding en vroeëkindontwikkelings-inisiatiewe.

Verlede jaar het ons die Nasionale Jeugontwikkelingsagentskap in die lewe geroep.

Ons het die Agentskap aangesê om vinniger te werk om sy strukture regdeur die land te vestig sodat dit ons kan help om jeugontwikkelingsprogramme in die regering as hoofstroomprojekte te loods.

Agbare lede,

Toe hierdie administrasie verlede jaar aan bewind gekom het, het ons onself daartoe verbind om harder te werk om 'n sterk ontwikkelingstaat te bou.

Ons het gesê dat dit 'n staat sou wees wat op die mense se behoeftes en aspirasies reageer, en wat beter en vinniger presteer.

Hierdie jaar, 2010, sal 'n jaar van aksie wees.

Die onderskeidende kenmerk van hierdie administrasie sal wees dat dit weet waar mense bly, hulle behoeftes verstaan en vinniger daarop reageer.

Die Regering moet vinniger, harder en slimmer werk.

Ons sal van die Uitvoerende Gesag en die Staatsdiens verwag om aan hierdie visie te voldoen.

Ons bou 'n staat wat op effektiewe werkverrigting fokus. Dit sal bereik word deur beplanning asook prestasie-monitering en -evaluering te verbeter.

Ons moet ook geslagsgelykheidsmaatreëls in die Regering se Aksieprogram integreer.

Hierdie optrede sal verseker dat vroue, kinders en persone met gestremdhede toegang tot ontwikkelingsgeleenthede kry.

Ons kondig graag aan dat ons in die Regering dinge op 'n nuwe manier gaan doen.

Departemente se werk sal gemeet word deur uitkomste, wat op grond van ons prestasie-monitering- en-evalueringstelsel ontwikkel word.

Die ministers wat vir 'n bepaalde uitkoms verantwoordelik is, sal 'n gedetailleerde leweringsooreenkoms met die president onderteken.

Hierdie ooreenkoms sal duidelik uiteensit wat gedoen moet word, hoe, deur wie, binne watter tydperk en met behulp van watter maatreëls en hulpbronne.

Soos u weet, is ons verbind tot vyf prioriteite:

- onderwys, gesondheid, landelike ontwikkeling en grondhervorming, skepping van ordentlike werkseleenthede en die bestryding van misdaad.

Daarbenewens sal ons werk om die effektiwiteit van plaaslike regering, infrastruktuurontwikkeling en menslike nedersettings te verbeter.

Ons sal 'n aantal sleutelaktiwiteite onderneem ter bereiking van hierdie uitkomste.

Ons het onderwys en vaardigheidsontwikkeling as die middelpunt van hierdie regering se beleide gestel.

In ons 2010-program wil ons die vermoë van kinders om in die grondslagjare te lees, skryf en tel, verbeter.

As ons dit nie doen nie, sal ons nie die gehalte van die onderwys kan verbeter nie.

Ons onderwysteikens is eenvoudig, maar van kritiese belang.

Ons wil hê dat leerders en onderwysers betyds in die skool, in die klas moet wees, besig om sewe uur per dag te leer en te onderrig.

Ons sal onderwysers help deur gedetailleerde daaglikse lesplanne te voorsien.

Ons sal werkboeke wat maklik is om te gebruik in al 11 tale aan studente voorsien.

Van vanjaar af sal alle leerders in graad 3, 6 en 9 geletterdheids- en gesyferdheidstoetse skryf, wat onafhanklik gemodereer word.

Ons mikpunt is om die slaagsyfer vir hierdie toetse van die huidige gemiddeld van tussen 35 en 40% tot minstens 60% teen 2014 te verhoog.

Die uitslae sal aan die ouers gestuur word om vordering te monitor.

Daarbenewens sal elk van ons 27 000 skole deur amptenare van die Departement van Basiese Onderwys geassesseer word.

Dit sal in 'n ouditeerbare skriftelike verslag opgeteken word.

Ons beoog om die aantal matriekstudente met universiteitstoelating teen 2014 tot 175 000 per jaar te verhoog.

Ons spoor ouers aan om met ons saam te werk om 'n sukses hiervan te maak.

Ons verwelkom verlede maand se verklaring deur die drie onderwysersvakbonde, NAPTOSA, SADTU en SAOU, waarin hulle hulle vanaf die begin van 2010 herverbind het tot die Kwaliteit Leer- en Onderrigveldtog.

Agbare lede,

Ons moet in ons jeug belê ten einde 'n vaardige en bekwame werksmag te verseker wat groei en werkskepping ondersteun.

Ons beplan dus om die opleiding van 16 tot 25-jarige persone te verbeter in verdere onderwys-en-opleidingsfasiliteite.

Dit sal ons in staat stel om aan diegene wat nie vir universiteit kwalifiseer nie, nog 'n kans op geleerdheid te bied.

Ons werk saam met hoëronderwysinstellings om te verseker dat studente wat daarvoor kwalifiseer, deur die Nasionale Finansiële Hulpskema vir Studente finansiële bystand ontvang.

Ons het ook ambisieuse doelwitte gestel vir vaardigheidsontwikkeling om addisionele ingenieurs en tegnisi te kweek en om die aantal gekwalifiseerde Wiskunde- en Wetenskaponderwysers te vermeerder.

Ons moet verseker dat meer jongmense leerderskappe in die privaat- en openbare sektor betree.

Nog 'n sleuteluitkoms is om 'n lang en gesonde lewe vir alle Suid-Afrikaners te verseker.

Ons sal voortgaan om ons gesondheidsorgstelsel te verbeter.

Dit sluit in die bou en opgradering van hospitale en klinieke en die verdere verbetering van die werksomstandighede van gesondheidswerkers.

Ons het 'n vennootskap met die Ontwikkelingsbank van Suider-Afrika (DBSA) gesluit om die funksionaliteit van openbare hospitale en hul distrikskantore te verbeter.

Ons werk ook saam met die DBSA en die Nywerheidsontwikkelingskorporasie in 'n publiek-private vennootskapsprogramme om hospitale te verbeter en finansiering vir projekte te verskaf.

Agbare lede,

Ons moet die feit in die gesig staar dat lewensverwagting by geboorte van 60 jaar in 1994, tot net onder 50 jaar vandag, gedaal het.

Ons doen dus stappe om sterftesyfers onder moeders te verlaag, om nuwe MIV-infeksies te verminder en om MIV en tuberkulose effektief te behandel.

Ons sal ook babasterftesyfers deur 'n massiewe immuniseringsprogram verlaag.

Ons sal gesondheidsprogramme in skole herinstel.

Ons sal al die ondernemings wat op Wêreldvigsdag met betrekking tot die voorkoming van nuwe MIV-infeksies en die behandeling daarvan gemaak is, implementeer.

Intensiewe stappe word geneem om te verseker dat hierdie werk op skedule bly.

Ons sal ook voortgaan met ons voorbereidings vir die instelling van 'n nasionale gesondheidsorgversekeringstelsel.

Mede-Suid-Afrikaners,

Ons werk hard om te verseker dat almal in Suid-Afrika veilig voel en veilig is.

Ons sal ons werk uitbrei om ernstige en geweldsmisdade te verminder en te verseker dat die regstelsel doeltreffend funksioneer.

Ons implementeer planne om die aantal polisiemanne en -vroue oor die volgende 10 jaar met 10% te vermeerder.

Ons het die stryd teen kappings, besigheids- en huisrooftogte asook kontakmisdade soos moord, verkragting en aanranding, as ons topprioriteite geïdentifiseer.

Ons het almal 'n rol om te speel.

Laat ons aan gemeenskapsveiligheidsforums deelneem.

Kom ons hou op om gesteelde goedere te koop.

Laat ons altyd gereed wees om die polisie van inligting oor misdaadaktiwiteite te voorsien.

Deur saam te werk sal ons misdaad verslaan.

Landsgenote en agbare gaste,

Die plaaslike regering moet werk.

Munisipaliteite moet die voorsiening van behuising, water, sanitasie, elektrisiteit, afvalbestuur en paaie verbeter.

Ons het verlede jaar met burgemeesters en munisipale bestuurders vergader.

Dit het waardevolle insig in die uitdagings op die terrein van plaaslike regering gebied.

Ons het ook verskeie gemeenskappe en munisipaliteite besoek, insluitende Balfour in Mpumalanga en Thembisa in Gauteng.

Na afloop van die Balfour-besoek het ons 'n ministeriële span, bestaande uit nege lede, gestuur om die gebied te besoek en aandag te skenk aan die kwessies wat deur die gemeenskap geopper is.

'n Hele aantal sake het reeds aandag geniet.

Ek het die ministers beveel om die uitstaande aangeleenthede ook op te los.

Ons wil herbeklemtoon dat geen griewe ooit geweld en die vernietiging van eiendom kan regverdig nie.

Ons het wetstoepassingsagentskappe beveel om strenger op te tree teen wetteloosheid in Balfour en ander gebiede.

In Desember 2009 het die Kabinet 'n omkeerstrategie vir die plaaslike regering goedgekeur.

Dit sal verseker dat die plaaslike regering oor die korrekte bestuurs-, administratiewe en tegniese vaardighede beskik.

Laat ons gedurende hierdie jaar van aksie saamwerk om plaaslike regering almal se besigheid te maak.

Ons werk daaraan om goed geleë informele nedersettings op te gradeer en om teen 2014 aan minstens 500 000 huishoudings behoorlike dienste en grondbesit te verskaf.

Ons beplan om meer as 6 000 hektaar goed geleë openbare grond vir lae-inkomste en bekostigbare behuising opsy te sit.

'n Nuwe sleuteliniatief sal wees om mense te akkommodeer wie se salarisse te hoog is om regeringsubsidies te ontvang, maar wat te min verdien om vir 'n normale bankverbandening te kwalifiseer.

Ons sal 'n waarborgfonds van R1 biljoen opsy sit as aansporingsmaatreël om die private bank en behuisingsektor aan te moedig om nuwe produkte te ontwikkel om in hierdie vraag na behuising te voorsien.

Landsgenote,

Verlede jaar het ons gesê dat mense op die platteland ook die reg het om elektrisiteit, water, spoeltoilette en paaie te hê.

Ons het gesê hulle moet sportsentrums en beter inkopiesentrums hê soos dié in die stede.

In hierdie verband het ons in Augustus verlede jaar die eerste loodsterrein van die Omvattende Landelike-ontwikkelingsprogram in Giyani, Limpopo, geloods.

Sedertdien is 231 huise gebou.

Vordering is ook gemaak met die voorsiening van infrastruktuur om landbou-ontwikkeling en opleiding vir gemeenskapsleiers te ondersteun.

Toegang tot gesondheids- en onderwysgeriewe het verbeter.

Ons implementeer tans soortgelyke programme in sewe terreine regoor die land, tot voordeel van 21 wyke.

Ons beoog om teen 2014 terreine in 160 wyke te hê.

Ons mikpunt is dat 60% van huishoudings in hierdie terreine teen 2014 uit eie produksie in hul voedselbehoefte sal voorsien.

Tree vir tree sal ons dit maak, soos die spreekwoord lui, landsgenote, groot dinge begin met klein treetjies.

Ons wil ook grondhervorming en landbou-ondersteuningsprogramme beter integreer.

Ons sukses op hierdie terrein sal gemeet word deur die toename in die aantal kleinskaalse boere wat ekonomies lewensvatbaar word.

Agbare Speaker en Voorsitter van die NROP,

Ons is nie 'n waterryke land nie.

Maar steeds verloor ons baie water deur lekkende pype en onvoldoende infrastruktuur.

Ons sal maatreëls instel om ons waterverlies teen 2014 met die helfte te verminder.

Agbare lede,

As deel van ons pogings om groter ekonomiese groei aan te moedig, doen ons stappe om die koste verbonde aan kommunikasie te besnoei.

Die Suid-Afrikaanse publiek kan uitsien na 'n selfs verdere verlaging in breëband-, selfoon-, landlyn- en openbare telefoon-tariewe.

Ons sal stappe doen om breëbandspoed te verhoog en 'n hoë standaard van internet-dienste te verseker wat in ooreenstemming met internasionale norme is.

Mede-Suid-Afrikaners

Hierdie regering sal verseker dat ons omgewingsbates en natuurlike hulpbronne goed beskerm en deurentyd verbeter word.

Saam met Brasilië, Indië en China, en in samewerking met die Verenigde State, wat die ontwikkelde wêreld verteenwoordig, het ons 'n wesenlike bydrae gelewer tot die ooreenkoms wat in Desember verlede jaar by die Klimaatberaad in Kopenhagen aanvaar is.

Alhoewel dit nie ver genoeg gaan nie, is dit tog 'n belangrike stap in die regte rigting, aangesien dit alle lande daartoe verbind om op klimaatsverandering te reageer.

Ons sal harder met ons internasionale eweknieë moet saamwerk om 'n wettig bindende verdrag te beding.

As Suid-Afrika, het ons ons vrywillig verbind tot spesifieke uitlaatgas-verlagingsteikens, en ons sal aanhou werk aan ons langtermynstrategie om klimaatsverandering teen te werk.

Agbare lede,

Ons sal harder werk om Suid-Afrika se belange wêreldwyd te bevorder.

Ons sal pogings ondersteun om die politieke en ekonomiese integrasie van die Suider-Afrikaanse Ontwikkelingsgemeenskap- (SAOG) streek te bespoedig en intra-streekhandel en beleggings te bevorder.

Suid-Afrika gaan voort om 'n toonaangewende rol te speel in kontinentale pogings om die Afrika Unie en sy organe te versterk en na eenheid te streef.

Ons sal ons energie daarop toespits om die Nuwe Vennootskap vir Afrika se Ontwikkeling (Nepad) te laat herleef, as 'n strategie vir ekonomiese ontwikkeling op die vasteland.

Mede-Suid-Afrikaners,

Die Staatsdiens moet reageer op die beroep om hierdie termyn deur vinniger optrede en beter prestasie deur die staat te laat kenmerk.

Ons vereis uitnemendheid en harde werk.

Ons benodig staatsamptenare wat toegewyd en bekwaam is en wat na landsburgers se behoeftes omsien.

Die Regering werk reeds aan die ontwikkeling en implementering van 'n staatsdiens-ontwikkelingsprogram, wat die norme en standaarde vir staatsamptenare in alle sferes sal bepaal.

Agbare Lede,

Ons sit ons pogings voort om korrupsie en bedrog in verkrygings- en tenderprosesse uit te wis, asook in aansoeke om bestuurderslisensies, maatskaplike toelaes en identiteitsdokumente, om net 'n paar te noem.

Ons is tevrede met die vordering van die regering op sommige terreine.

Hierdie week het ons 32 687 onregmatige maatskapliketoelaagbetalings opgeskort, ter waarde van R180 miljoen.

Ons Inter-Ministeriële Komitee op Korrupsie ondersoek maniere om korrupsie doelgerig die nek in te slaan.

Deur saam te werk kan ons meer doen.

Landsgenote,

Soos u weet, het ons die Presidensiële Klagtelyn ingestel om die regering en die Presidensie meer toeganklik vir die publiek te maak en om te help om blokkasies in dienslewering uit die weg te ruim.

Die klagtelyn verteenwoordig ons vasberadenheid om dinge anders in die regering te doen.

Dit het 'n verskil aan baie Suid-Afrikaners se lewens gemaak.

Ons kan mev Buziwe Ngaleka van Mount Frere hier noem, wie se oproep oor haar oorlede man se pensioen die eerste oproep was wat ons op die eerste dag van die klagtelyn geneem het.

Sy is vanaand hier saam met ons.

Ons het ook mnr Nkululeko Cele hier by ons, wat bygestaan is om sy identiteitsdokument te kry sodat hy by die Tshwane Universiteit van Tegnologie kon inskryf.

Dit is net twee van die talle suksesstories.

Uit hierdie en ander voorbeelde het ons swakhede geïdentifiseer wat deur die onderskeie regeringsfere reggestel moet word.

Deur die Speaker het ons 'n veelparty-afvaardiging van die Parlement uitgenooi om die oproepsentrum te besoek, sodat LP's eerstehandse ervaring kan kry van die werk wat gedoen word.

Landsgenote en vriende,

Ek het 'n uiteensetting gegee van die hoofelemente van ons planne vir 2010, ons kollektiewe verbintenis as regering teenoor die mense van Suid-Afrika.

Die Staatsrede bied 'n breë oorsig oor ons aksieplan.

Ministers sal die details in hul onderskeie begrotingsredes verskaf.

Agbare lede, mede-Suid-Afrikaners,

In November vanjaar vier ons die 150ste herdenking van die aankoms van die Indiërs in Suid-Afrika.

Dit bied 'n geleentheid om erkenning te gee aan die belangrike bydrae van die Indiese gemeenskap op die gebied van arbeid, die sakesektor, wetenskap, sport, godsdiens, kuns en die bereiking en konsolidasie van ons demokrasie.

Landsgenote en vriende,

Ek wil weer van hierdie geleentheid gebruik maak om ons almal se diepe meegevoel te betuig teenoor die Regering en die mense van Haiti met die enorme tragedie wat hulle getref het.

Ons is bly dat ons reddingspanne hulle kon gaan bystaan.

Ek wil graag besondere erkenning verleen aan een Suid-Afrikaner wat nooit huiwer om in tye van nood en rampe te help nie, en ons help om die visie van 'n sorgsame gemeenskap te bevorder.

Ons verwelkom vandag dr Imtiaz Sooliman van Gift of the Givers in hierdie huis.

Mede-Suid-Afrikaners,

Om as gasheerland vir die FIFA Wêreldbeker op te tree, maak van 2010 voorwaar 'n jaar van aksie.

Ons beplan reeds jare lank aan hierdie Wêreldbeker.

Daar is nog net drie maande oor voor hierdie geskiedkundige gebeurtenis plaasvind.

En ons is vasberade om 'n sukses daarvan te maak.

Die nodige infrastruktuur, sekuriteit en logistieke reëlins is alles gereed om 'n suksesvolle toernooi te verseker.

As 'n nasie is ons groot dankbaarheid verskuldig teenoor die 2010 Plaaslike Reëlinskomitee vir hul uitstekende poging.

Ons wens die voorsitter van die komitee, Irvin Khoza, die Hoof Uitvoerende Beampte, Danny Jordaan, en die afrigter van Bafana Bafana, Carlos Alberto Parreira, alles van die beste toe vir die volgende paar maande.

President Mandela het 'n sentrale rol gespeel om die land te help om die gasheerrechte vir hierdie groot geleentheid te wen.

Ons moet dus tot sy eer 'n sukses van die Wêreldbeker maak.

Landsgenote, kom ons skaar ons ook agter ons nasionale span, die Bafana Bafana.

Die belangrikste is, wees in besit van julle kaartjies, landsgenote!

Kom ons koop almal betyds kaartjies sodat ons die wedstryde kan bywoon.

Mede-Suid-Afrikaners,

Wanneer ons vandag Madiba se vrylating gedenk, herverbind ons onself tot versoening, nasionale eenheid, nie-rassigheid en om 'n beter toekoms vir ons as Suid-Afrikaners, swart en wit, te bou.

Ons word gelei deur dit wat Madiba in die beskuldigdebank gesê het, en ek haal aan:

Gedurende my lewe het ek myself toegewy aan hierdie vryheidstryd van die mense van Afrika.

Ek het gestry teen blanke oorheersing en ek het gestry teen swart oorheersing.

Ek het die ideaal gekoester van 'n demokratiese en vrye samelewing, waarin almal in harmonie saamleef met gelyke geleenthede.

Dit is 'n ideaal waarvoor ek hoop om te lewe en te bereik.

Maar, as dit nodig sou wees, is dit 'n ideaal waarvoor ek bereid is om te sterf.

Besiel deur ons ikoon Madiba, is dit nou vir my 'n eer om hierdie 2010-Staatsrede aan al ons helde en heldinne, besonge en onbesonge, bekend en onbekend, op te dra.

Kom ons werk saam om hierdie jaar van aksie 'n suksesvolle jaar vir ons land te maak.