

Towards Processable Afrikaans

Laurette Pretorius¹ and Rolf Schwitter²

¹ School of Computing, University of South Africa and Meraka Institute, CSIR,
Pretoria, South Africa, LPretorius@csir.co.za

² Centre for Language Technology, Macquarie University,
Sydney 2109 NSW, Australia, Rolf.Schwitter@mq.edu.au

Abstract. In this paper we discuss a number of structural problems that we face when we design a machine-oriented controlled natural language for Afrikaans taking the underlying principles of Attempto Controlled English (ACE) and Processable English (PENG) as a starting point.

1 Introduction

Machine-oriented controlled natural languages are well-defined subsets of natural languages that can be translated unambiguously into a formal language (and sometimes vice versa) and be used for automated reasoning. Over the last decade, a number of machine-oriented controlled natural languages have been designed and used for specification purposes, knowledge acquisition and knowledge representation, and as interface languages to the Semantic Web – among them Attempto Controlled English (ACE) [6], Boeing’s Computer-Processable Language (CPL) [2], Common Logic Controlled English (CLCE) [9], and Processable English (PENG) [7]. All these controlled natural languages are based on English, but little work has been done in this domain for other languages. In this paper, we look at Afrikaans and work towards a machine-oriented controlled natural language for Afrikaans, based on PENG. Afrikaans is an Indo-European language derived from 17th century Dutch and mainly spoken in South Africa and Namibia. It is estimated that Afrikaans is the home language of 6.45 million people and the second or third language for 6.75 million people.³ In the following, we will discuss how a subset of ACE and PENG is related to Afrikaans and identify a number of structural problems that we face when we intend to translate Afrikaans automatically into a first-order notation.

In addition to English and Afrikaans, South Africa has nine other official languages that belong to the Bantu language family, and differ significantly from English and Afrikaans. Investigating the development of controlled natural languages for the South African Bantu languages constitutes a future challenge.

2 ACE and PENG

ACE and PENG are machine-oriented controlled natural languages that can be used as specification and knowledge representation languages. They are tractable

³ <http://www.southafrica.info/about/people/language.htm>

subsets of English that can be unambiguously translated by a machine into first-order logic. A subset of the grammar of these two languages covers declarative sentences and questions. Their vocabularies consist of predefined function words, some predefined phrases, and content words. In a nutshell, an ACE or PENG text consists of a sequence of declarative sentences that are linked by approved forms of intra- and intersentential anaphoric expressions.

2.1 Simple Sentences

Simple sentences in ACE and PENG have the following general structure:

subject + verb + [complements] + { adjuncts }

Every sentence has a subject and a verb. Complements are required for transitive verbs and ditransitive verbs but adjuncts are always optional. For example, the following sentence consists of a subject that has the form of a noun phrase *a customer*, a transitive verb *inserts* that takes the noun phrase *a card* as a complement, and an adjunct in form of a prepositional phrase *into the slot*:

1. A customer inserts a card into the slot.⁴

Not only the verb can be described in more detail as illustrated by the prepositional modifier in (1) but also the nouns. For example in (2):

2. John's customer who is new inserts a valid card of Mary into a slot A.

the noun *customer* is further described with the help of a possessive noun and a relative clause, the noun *card* via an adjective and an *of*-prepositional phrase, and finally the noun *slot* via a name as apposition.

2.2 Complex/Composite Sentences

Complex sentences (or composite sentences in "ACE-speak") are built recursively from simpler sentences through coordination (3 and 4), subordination (5), quantification (6), and negation (7-9):

3. A customer inserts a card and enters a code.
4. A customer inserts a card or enters a code.
5. If a card is valid then a customer inserts it.
6. Every customer inserts a card.
7. A customer does not insert a card.
8. No customer inserts a card.
9. It is not the case that a customer inserts a card.

Note that this is only a small subset of complex sentences that are approved by these two controlled natural languages (for details see [6]).

⁴ The examples have been taken from the specification of an ATM [5].

2.3 Questions

ACE and PENG support *yes/no*-questions (10) and *wh*-questions (11):

10. Does the customer insert the card into the slot?
11. Where does the customer insert the card?

Yes/no-questions can be used to investigate whether or not a specific situation is true, and *wh*-questions interrogate different aspects of a situation.

3 Afrikaans Specifics

Afrikaans and English have co-existed in South Africa for many years and the main differences between the two languages are well documented (see, for example, [1]). We focus on the following five main aspects of Afrikaans in which the language differs from English and consequently from ACE and PENG.

3.1 Particle Verbs

Particle verbs (phrasal verbs) (Afrikaans: partikelwerkwoorde) are common in Afrikaans, as also in English. The verb ‘indruk’ (insert) is one example. In Afrikaans, the verb part of the particle verb behaves like an ordinary central verb in terms of placement in the sentence. The complement/particle has varying positions according to the construction [3, p.104]. For example, in main and simple sentences without an auxiliary verb such as in (13):

12. A customer inserts a card.
13. 'n Klant **druk** 'n kaart **in**.

the short complement/particle stands at the end of the sentences but in subordinate clauses and after auxiliary verbs it is prefixed to the verb.

3.2 The Preterite or Past Tense

The preterite of the majority of Afrikaans verbs is formed by the auxiliary verb ‘het’ + the past participle of the verb. Moreover, in the presence of complements and/or adjuncts the auxiliary verb ‘het’ is separated from the past participle, with the past participle occurring at the end of the sentence. For example, for the verb ‘indruk’ (insert), we have (15):

14. A customer inserted a card.
15. 'n Klant **het** 'n kaart **ingedruk**.

3.3 Subordinate Clauses

Similar to English simple sentences in Afrikaans also follow the general structure:

subject + verb + complements + adjuncts (S+V+C+A)

However, in subordinate clauses the structure is usually S+C+A+V. So composite sentences that consist of a main sentence and a subordinate clause such as in (17) have the following structure [1, pp.46-47] where () indicates optionality:

$$\underbrace{S + V(+C + A)}_{main} + \underbrace{S(+C + A) + V}_{subordinate}$$

16. John's customer who is new inserts a valid card of Mary into a slot A.
17. John se klant wat nuut is druk 'n geldige kaart van Mary in 'n gleuf A in.

In composite sentences that consist of a conditional subordinate clause followed by the main sentence such as in (19):

18. If a card is valid then a customer inserts it.
19. As 'n kaart geldig is dan druk 'n klant dit in.

the main sentence has a V+S+O structure [3, p.222].

3.4 Negation

One of the most complex issues in Afrikaans is its double negation [1, pp.54-58]. Although double negation is the 'rule' there are constructions in which not all negations are doubled. As a guideline, negation in the main sentence is usually doubled, even when they are composed, for example in (21):

20. It is not the case that a customer inserts a card.
21. Dit is **nie** die geval dat 'n klant 'n kaart indruk **nie**.
22. It is not the case that a customer does not insert a card.
23. Dit is **nie** die geval dat 'n klant **nie** 'n kaart indruk **nie**.

However, in subordinate clauses that form part of a main sentence that already has negation such as in (23), negation is usually not doubled.

3.5 Questions

Both *yes/no*-questions and *wh*-questions in Afrikaans are obtained from declarative sentences by means of inversion [4, p.186], for example:

24. Does a customer insert a card?
25. Druk 'n klant 'n kaart in?
26. What does a customer insert?
27. Wat druk 'n klant in?
28. Where does a customer insert a card?
29. Waar druk 'n klant 'n kaart in?

In particular, if a declarative sentence has the structure S+V+(C+A) then an associated *yes/no*-question will have the structure V+S+(C+A) while an associated *wh*-question has either a WH+V+S(+A) or a WH+V(+S)+A structure.

4 First experiment

The experiment was done with PENG Light [8], for which a systematically compiled suite of approximately two hundred test sentences had been developed. As a first step we translated selected test sentences into Afrikaans to ascertain to what extent PENG Light is suitable for these sentences. The PENG Light lexicon was extended to include the Afrikaans words. We focussed on particle verbs, the past tense and double negation, which all constitute forms of long-distance dependencies. In all cases the Afrikaans constructions could be accommodated by modifications to the PENG grammar. It remains to consider to subordinate clauses, *yes/no* questions and *wh*-questions, and to run PENG Light on the full, translated suite of test sentences.

5 Conclusions

We compared a common subset of ACE and PENG with Afrikaans and identified five structural aspects in which Afrikaans differs from these two machine-oriented controlled natural languages: (1) particle verbs are frequent in Afrikaans and their treatment requires a mechanism that deals with long distance dependencies if these verbs occur in a main sentence; (2) the past tense in Afrikaans is mainly formed by the auxiliary verb ‘het’ + the past participle of the verb, with the past participle occurring at the end of the sentence; (3) verbs in subordinate clauses occupy the final position; (4) double negation appears in main sentences but not necessarily in subordinate clauses that depend on the main sentence; (5) questions are built via inversion and their forms are in general simpler than in English. PENG Light was successfully modified and applied to Afrikaans sentences with particle verbs, the past tense and double negation, and will now be extended to cater for subordinate clauses, *yes/no* questions and *wh*-questions.

References

1. Carstens, W.A.M.: Norme vir Afrikaans. J.L. Van Schaik (1994)
2. Clark, P., Harrison, P., Jenkins, T., Thompson, T., Wojcik, R.: Acquiring and Using World Knowledge Using a Restricted Subset of English. In: Proceedings of FLAIRS05, pp. 506–511 (2005)
3. De Villiers, M.: Afrikaans Grammatika vir Volwassenes. Nasou Beperk (1983)
4. Du Toit, P.J.: Taalleer vir Onderwyser en Student. Academica (1982)
5. Fuchs, N.E., Höfler, S., Kaljurand, K., Rinaldi, F., Schneider, G.: Attempto Controlled English: A Knowledge Representation Language Readable by Humans and Machines. In: Reasoning Web: First International Summer School 2005, LNCS 3564, pp. 213–250 (2005)
6. Fuchs, N.E., Kaljurand, K., Kuhn, T.: Attempto Controlled English for Knowledge Representation. In: Reasoning Web, LNCS 5224, pp. 10–124 (2008)
7. Schwitter, R., Tilbrook, M.: Meaningful Web Annotations for Humans and Machines using Controlled Natural Language. In: Expert Systems, 25(3), pp. 253–267 (2008)
8. Schwitter, R.: Working for Two: A Bidirectional Grammar for a Controlled Natural Language. In: Proceedings of AI 2008, LNAI 5360, pp. 168–179 (2008)
9. Sowa, J.F.: Common Logic Controlled English. Draft, 24 February 2004, <http://www.jfsowa.com/clce/specs.htm> (2004)