Rapid rule-based machine translation between Dutch and Afrikaans

Pim Otte
Mendelcollege
Pim Mulierlaan 4
2024 BT Haarlem
5666@mendelcollege.nl

Francis M. Tyers
Dept. Lleng. i Sist. Inform.
Universitat d'Alacant
E-03070 Alacant
ftyers@dlsi.ua.es

Abstract

This paper describes the design, development and evaluation of a machine translation system between Dutch and Afrikaans developed over a period of around a month and a half. The system relies heavily on the re-use of existing publically available resources such as Wiktionary, Wikipedia and the Apertium machine translation platform. A method of translating compound words between the languages by means of left-to-right longest match lookup is also introduced and evaluated.

1 Introduction

Dutch is a West-Germanic language spoken by nearly 23 million people, mostly from the Netherlands and Flanders, the Dutch-speaking part of Belgium, and a minority living in former colonies of the Netherlands, such as Suriname, Aruba and the Netherlands Antilles. Dutch, as it is today, started developing in the 16th century in the major trade cities, such as Amsterdam and Antwerp (Shetter and Ham, 2002). Afrikaans is spoken by at least 5 million people, mainly in South Africa but also in Namibia. Afrikaans is a variety of Dutch that originates from that spoken by the Dutch colonists of the Cape Colony. In 1925 Afrikaans replaced Dutch as an official language in South Africa, to be the joint official language together with English (Donaldson, 1993). Currently, Afrikaans is one of the eleven national languages.

In this paper we will describe the development of `apertium-af-nl`, a bi-directional Afrikaans and Dutch machine-translation system based on the Apertium platform. As Afrikaans and Dutch are largely mutually intelligible, this machine translation system focuses on dissemination, the translation of text for the purpose of being post-edited and then being published.

This is not the first system to work with this language pair, van Huyssteen and Pilon (2009) describe a rule-based system to convert in a single direction from Dutch to Afrikaans. The reason we have chosen to work with a rule-based approach, instead of the ubiquitous corpus-based/statistical approach, is that the latter needs parallel corpora for the two languages. The only freely available Afrikaans–Dutch corpus, is KDE4[1], which is translated via English and domain specific. We feel that these corpora do not approach the quality required for the statistical approach, which makes the rule-based approach favourable.

The paper is laid out as follows: firstly, we will describe the reuse and creations of resources. We will then discuss several grammatical features of Afrikaans and Dutch and how these were treated in the machine-translation system. We will then present a section in which the system is evaluated. Finally, we will discuss the system and future work that could be done.

2 Method

The system is based on the Apertium machine translation platform.[2] The platform was originally aimed at the Romance languages of the Iberian peninsula, but has also been adapted for other, more distantly related, language pairs. The whole platform, both programs and data, are licensed under the Free Software Foundation's General Pub-

[1] `http://opus.lingfil.uu.se/KDE4.php`
[2] `http://www.apertium.org`

lic Licence[3] (GPL) and all the software and data for the 25 supported language pairs (and the other pairs being worked on) is available for download from the project website.

`apertium-af-nl` has been developed over the course of one and a half months. The vast majority of the work has been done by a Dutch secondary school student supervised by a PhD student. However, since for the latter this was not paid and for the former it was not for school, there were very few full 8-hour days of work during this period.

2.1 Existing resources

One existing resource was reused with very little modification: the morphological transducer for Afrikaans, which was created during a separate, currently dormant, project on English–Afrikaans machine translation. However, some changes were made. The structure of verb entries was wholly revised, and both infrequent words and words for which a translation could not be found, were removed.

2.2 Resources created

2.2.1 Dutch morphological transducer

There are a number of existing morphological analysers for Dutch (Bosch et al., 2007; Laureys et al., 2004; DePauw et al., 2004), some of which also function as morphological generators. Our decision to make a new morphological analyser was based on the following rationale:

- Licence: Neither the CELEX morphological database for Dutch (Laureys et al., 2004), nor the finite-state morphological transducer in the FLaVoR project (DePauw et al., 2004) are available under a free licence. As our objective is to publish and distribute the system described here, this made them unusable.
- Bidirectional: We wanted the dictionary to be able to be used for both morphological analysis and generation. Other analysers, for example the one described in Bosch et al. (2007) are only designed for analysis.
- Tagset: When creating a new machine translation system, it is convenient if the tags which represent the same or similar features are equivalent in the morphological analysers/generators for each of the languages, e.g. `<sg>` on both sides for 'Singular' instead of `<sg>` on one side and `ev` [4] on the other.

[3] `http://www.fsf.org/licensing/licenses/gpl.html`

Dutch

Etymology

hoofd- ("main, head") + *stad* ("city")

Pronunciation

Noun

hoofdstad m. (*pl* hoofdsteden, *dimin* hoofdstadje, *dimin pl* hoofdstadjes)

1. capital city

Figure 2: English language Wiktionary article for Dutch *hoofdstad* 'capital city' `http://en.wiktionary.org/wiki/hoofdstad`

The open categories (nouns, verbs, adjectives, adverbs) for the Dutch morphological analyser were extracted semi-automatically from Wiktionary,[5] a free online, multilingual dictionary that often includes inflectional information. On the English Wiktionary, in the case of Dutch nouns it often (although not always) gives the gender and the plural and diminutive forms (see for example Figure 2). The category `Dutch nouns` in the English Wiktionary has a total of 10,610 entries, while the corresponding category on the Dutch Wiktionary has 13,176 entries.

Closed categories were added by hand based on a grammar of Dutch (Shetter and Ham, 2002).

2.2.2 Bilingual dictionary

The bilingual dictionary has been developed in several ways. Exact matches from the dictionaries were automatically added to the bilingual dictionary. Proper names were added in the way that is described in Tyers and Pienaar (2008). After that cognates were added. There are several small common spelling differences between Afrikaans and Dutch. The bilingual dictionary was expanded further by categorising these spelling differences and automatically adding translations to the bilingual dictionary if the spelling difference was the only difference between the two words. Finally, some entries were done by hand. This included closed categories, but also words that frequently appeared in Wikipedia which were not yet in the bilingual dictionary.

[4] `ev` stands for *enkelvoud* 'singular' and is from the tagset of the Tadpole morphological analyser (http://ilk.uvt.nl/tadpole/).

[5] `http://www.wiktionary.org`

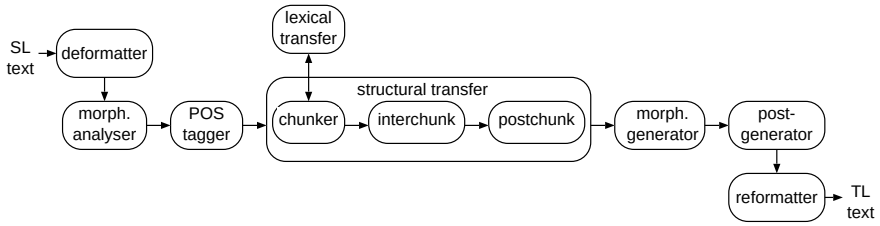

Figure 1: Modular architecture of the Apertium MT platform. The compound analysis and generation modules are included at the morphological analyser and morphological generator stages respectively.

2.3 Transfer rules

A total of 32 transfer rules have been written for the direction Afrikaans→Dutch and 15 for the direction Dutch→Afrikaans. Some of these transfer rules are discussed below.

2.3.1 Afrikaans to Dutch

Afrikaans hardly uses the word-attached genitive *s* (Donaldson, 1993). The word *se* is used to indicate possesion. Therefore, a transfer rule has been added to remove the *se* and instead make the preceding noun genitive. Note that in Dutch the genitive is not the preferred translation. A construction using the word *van* 'of' would be preferable, but would need restructuring of the phrase.

The verb *hê* 'have' is the only verb used in Afrikaans as auxiliary verb with a past participle. In Dutch the verbs *hebben* 'have' and *zijn* 'be' are both used, the latter mostly in cases of movement and a few exceptional cases, the former in all others. To handle this, two transfer rules have been added, to handle the patterns 'hê + past participle' and 'hê + nie + past participle + nie', which change the verb 'to have' into the verb 'to be', when the past participle is found in a list of verbs that go with 'zijn'.

Nouns in Afrikaans do not exhibit gender, where nouns in Dutch can be one of four genders, neuter, masculine, feminine or common. The definite article *het*, *die* in Dutch must agree with the noun it modifies. A number of transfer rules were written for patterns such as 'determiner + noun', 'determiner + adjective + noun', etc., which propagate the gender of the head noun to the article.

In Afrikaans, finite verbs do not agree in person and number with the subject of the sentence, where in Dutch they do. A rule was added which transfers the person and number of subject pronouns to the verb following them. This is a limited rule as it does not deal with non-pronominal subjects.

Negation in Afrikaans and Dutch differs in the use, in Afrikaans, of a negation scope marker *nie*. Translating from Afrikaans to Dutch this marker needs to be removed after 'nie' and also after other negatives such as *niemand* 'no one', *niks* 'nothing' and *geen* 'not'. Translating from Dutch to Afrikaans this marker needs to be added.

2.3.2 Compound words

Both Afrikaans and Dutch are languages in which words combine very productively into compounds. For example the words *infrastruktuurontwikkelingsplan* 'infrastructure development plan' and *lugmagbasis* 'air force base'. As it is impractical to introduce all compound words into the lexicons, compound word analysis is performed on all unknown words. The analysis process works longest-match left-to-right using the same transducer as is used for morphological analysis. This process only looks for compounds made up of just nouns, because they are more frequent than other compounds. Results are restricted by two special symbols which do not appear in the output, `compound-L` and `compound-R`. The `compound-L` symbol is used for forms that can only appear on the left side (e.g. surface form) of a compound, where `compound-R` is used for forms that can either appear in a compound, or end it. Epenthetics, that is linking letters that occur between compound words, are also taken care of heuristically in this way. For example the -s- in *ontwikkelingsplan*, the -en- in *pannenkoek* and the -e- in *paardebloem*. Notice that the epenthetic -e- is not productive in Dutch, that is, it is not used in new compounds.

There are some limitations to this method. For example: although both *macht-* and *machts-* can be analysed as an internal part of a compound, only one of them can be generated. Which one will be generated is decided based on the inflectional paradigm to which the word belongs.

2.3.3 Separable verbs

Another feature of Afrikaans and Dutch is separable verbs, for example the word *afslaan* 'to turn,

to decline, to stop'. This can appear in the following forms *afslaan*, *sla af*, *afgeslagen*. Additionally the two constituent parts of the verb in *sla af*, the verb itself *sla* and the particle *af* may be separated by a word or phrase, *Ik sla het aanbod af.* 'I decline the offer'.

The following cases are supported,

- Infinitive: afslaan → afslaan 'to turn'
- Participle: afgeslagen → afgeslaan 'turned'
- Non-separated: Ik sla af naar rechts. → Ek slaan af na regs 'I turned to the right'
- Subordinate: Toen ik de bal afsloeg → Toe ek die bal afslaan 'When I teed off the ball'

Verbs separated by a word or phrase are currently translated word-for-word, so the particle and verb are translated. This causes a problem when the verb is not constructed equally in Afrikaans and Dutch. Also, when one part of the verb, does not exist as a stand-alone verb, it is not recognised by the analyser. for example in *aankondig* 'to announce' *kondig* is not a word. Thus ... *kondig* ... *aan* cannot be analysed currently.

A module is under development to handle separable verbs, but is currently in the prototype stage.

There are currently 484 separable verbs defined in the bilingual dictionary. Of these, 439 are separable in both languages, 33 are separable in Afrikaans but not in Dutch, and 12 are separable in Dutch but not Afrikaans.

2.4 Current status

In terms of dictionary entries, the pair currently has 7,258 entries in the Afrikaans morphological dictionary, 7,048 in the Dutch morphological dictionary and 5,982 in the Bilingual dictionary.

3 Evaluation

The system was evaluated in five ways. The first was the coverage[6] of the system. The second was an evaluation of the compound analysis part of the system – new with respect to other Apertium language pairs. The third was the word error rate (WER) of the translations produced when comparing with a corrected sentence. The fourth was an analysis of the errors found by the second evaluation and finally a comparative evaluation with existing systems.

[6]Here coverage is defined as *naïve coverage*, that is for any given surface form at least one analysis is returned. This may not be complete.

Corpus	Tokens	Coverage
`af` Wikipedia	2,926,943	82.1% ± 0.8
`nl` Wikipedia	18,569,183	80.5% ± 0.7

Table 1: Naïve vocabulary coverage for the two morphological analysers.

Corpus	Corr. Seg.	Corr. Trans.
top-1,000	914	776
random-1,000	957	801

Table 2: Compound word accuracy in analysis and translation.

3.1 Coverage

Lexical coverage of the system is calculated over the Afrikaans and Dutch Wikipedias: Both corpora were split into four sections and coverage calculated over each of the sections in order to calculate the standard deviation.

The database dump of the Dutch Wikipedia was from the 1st November 2010, and that of the Afrikaans Wikipedia from the 31st July 2009. Both database dumps were stripped of formatting.

3.2 Compound words

In order to test the accuracy of the word compounding/decompounding strategy we tested two lists of words which received compound analyses from the Wikipedia. This test was only conducted in the Afrikaans→Dutch direction, but we expect similar results in the other direction. The first set of sentences was constructed by taking the 1,000 most frequent words which received a compound analysis from the corpus, the second was by taking a list of all the words and selecting 1,000 pseudo-randomly.[7] A total of 6,866 unknown words from the corpus received a compound analysis.

We include results for both correct segmentation (meaning the word was decompounded correctly) and correct translation (meaning the word was translated correctly). This allows us to take into account the *free ride* phenomenon, whereby an incorrect analysis may lead to a correct translation. There were 19 free rides in the top-1,000, and 5 free rides in the random-1,000.

[7]Using the Unix `unsort` program.

3.3 Quantitative

The translation quality was measured using word error rate (WER). This metric is based on the Levenshtein distance (Levenshtein, 1965) and was calculated for each of the sentences using the `apertium-eval-translator` tool.[8] A metric based on word error rate was chosen to be able to compare the system against systems based on similar technology, and to assess the usefulness of the system in a real setting, that is of translating for dissemination.

Four sets of 100 sentences were selected pseudo-randomly from Wikipedia.[9] The first two sets (C1, C3) contained no unknown words, whereas the second two sets could contain unknown words (C2, C4). This is to give an idea of the performance of the system in 'ideal' and 'realistic' settings.

For the Dutch to Afrikaans direction, the sentences were translated by the system, and then postedited by a native speaker. For the Afrikaans to Dutch direction, we took the reference translation, as postedited by the native speaker and used it as a source of Dutch to be translated to Afrikaans, then used the original Afrikaans sentence as a reference translation.

Confidence intervals were calculated through the bootstrap resampling method as described by Koehn (2004).

3.4 Qualitative

In order to inform ourselves of where the effort could be expanded in order to improve the system we undertook a qualitative evaluation by reviewing the translation errors from the Afrikaans to Dutch direction and categorising them as in Table 4. An example of each of the kind of error is found below. In all examples, the first sentence is Afrikaans, the second the Dutch machine translation, the third the post-editted Dutch and the fourth is the English translation of the sentence.

3.4.1 Unknown word

The example in (1) shows two errors caused by unknown words. The first error *Nystad* is a *free ride*, meaning that although it is an error it does not affect the final quality of the translation.

[8] `http://sourceforge.net/project/showfiles.php?group_id=143781&package_id=206517`; Version 1.0, 4th October 2006.

[9] The test corpora can be downloaded from *removed for review*

Error type	Count	% of total
Syntactic transfer	235	42.4
- Verb concordance	99	17.9
- Auxiliary verbs	13	2.3
- Relative pronoun	11	2.0
- Capitalisation	10	1.8
- Chunking error	9	1.6
- Other	93	16.8
Unknown word	147	26.5
Disambiguation	106	19.1
Morphology	28	5.1
Polysemy	23	4.2
Multiword	6	1.1
Compounding	6	1.1
Separable verb	3	0.5
Total	554	100

Table 4: Contribution to total error by type. Syntactic transfer errors are split into further categories.

(1) Hierdie besetting is in 1721 met die Verdrag van Nystad erken.
Deze bezetting is in 1721 met het Verdrag van **Nystad *erken*.
Deze bezetting is in 1721 met het Verdrag van *Nystad erkend*.
'This occupation has been acknowledged in 1721 with the Treaty of Nystad.'

The unknown words are marked with asterisk.

3.4.2 Morphology

Most errors in the morphological analyser were caused by a flaw in the automatic extraction process. The example in (2) shows a morphological error due to gender. The country *DDR* 'GDR' is feminine, which should go with the determiner 'de'. However, because it is marked as neuter in the morphological analyser, it is translated with 'het'. The vast majority of countries are in fact neuter, but *DDR* is not.

(2) In die DDR volg Erich Honecker Walter Ulbricht as partyleier op.
In *het* DDR volgen *Erich *Honecker *Walter *Ulbricht dan *partyleier op.
In *de* DDR volgt Erich Honecker Walter Ulbricht als partijleider op.
'In the DDR Erich Honecker succeeds Walter Ulbricht as party leader.

Errors of this type could be fixed with a more thorough revision of the morphological analyser.

Dir.	System	C1	C2	C3	C4
af-nl	Apertium	16.625 ± 1.465	23.405 ± 1.235	15.225 ± 1.735	22.195 ± 2.515
	Google	**9.485 ± 1.115**	**10.575 ± 1.795**	**7.63 ± 1.45**	**12.185 ± 1.545**
nl-af	Apertium	**15.435 ± 1.885**	**21.72 ± 1.06**	**18.375 ± 2.785**	**24.975 ± 2.075**
	Google	21.81 ± 1.72	25.71 ± 1.22	24.31 ± 3.22	30.965 ± 2.385

Table 3: Accuracy for the test corpora for the two systems as measured by Word Error Rate with 95% confidence interval.

3.4.3 Disambiguation

One of the biggest disambiguation problems for Afrikaans is distinguishing between short infinitive and present tense, which are morphologically the same. In example (3), in the Afrikaans sentence, the verb *volg* 'follow' could be present tense or infinitive. It has been tagged as infinitive, where present tense is the correct option.

(3) Hier volg 'n lys van hoofstede.
Hier *volgen* een lijst van hoofdsteden.
Hier *volgt* een lijst van hoofdsteden.
'Here follows a list of capital cities.'

Distinguishing between these two analyses is a difficult problem for a bigram part-of-speech tagger.

3.4.4 Multiword

Example (4) is causing problems because it is hard, if not impossible, to catch the meaning of the Afrikaans *dwarsoor* in one Dutch word. An appropriate multiword has solved the initial problem, but this causes additional issues with the article of *wereld* 'world' as that is included in the phrase *over de hele* 'all over the'.

(4) Duitse argitekte pak projekte dwarsoor die wêreld aan.
Duitse architecten pakken projecten *over de hele de* wereld aan.
Duitse architecten pakken projecten *over de hele* wereld aan
'German architects are taking on projects all over the world.'

3.4.5 Syntactic transfer

In (5) the singular verb does not match the plural subject, the noun *vrouwen* 'women'. This could be solved by identifying the subject of the sentence and matching the plurality of the verb with it.

(5) Die belangrikste rol wat die vroue egter in die stryd teen apartheid gespeel het, ...
De belangrijkste rol wat de vrouwen echter in de strijd tegen apartheid gespeeld *heeft*, ...
De belangrijkste rol die de vrouwen echter in de strijd tegen apartheid gespeeld *hebben*, ...
'The most important part that women played in the struggle against apartheid, ...'

Afrikaans uses the verb *hê* 'have' with all past participles, whereas Dutch uses the verb *zijn* 'be' in cases of, amongst others, verbs that imply movement. This could be fixed by tracking the auxiliary verb in a sentence and alter it if the past participle is in a list of movement verbs.

(6) Die sand het dan saam met die water weggespoel.
Het zand *heeft* dan samen met het water weggespoeld.
Het zand *is* dan samen met het water weggespoeld.
'The sand was washed away along with the water.'

Another issue is relative pronouns. Afrikaans always uses the word *wat*, where the equivalent Dutch word depends on the antecedent. In Dutch *wat* is used when i.e. the antecedent is an entire sentence. In this case (7) the antecedent is *formules*, for which the appropriate relative pronoun is *die*.

(7) Pi kom voor in baie formules in meetkunde wat sirkels en sfere betrek.
Pi komen voor in vele *formules in meetkunde *wat* cirkels en *sfere betrekken.
Pi komt voor in vele formules in meetkunde *die* cirkels en bollen betrekken.
'Pi appears in many formulas in geometry which concern circles and spheres.'

Capitalisation is generally straightforward. An exception is when a sentence starts with an apostrophe in one language and does not start with that in the other. The Afrikaans indefinite article is

'n, which cannot be capitalised. Therefore in (8) the translation has a capitalisation error. The word *een* should be capitalised, while *Pet* should not be. In Apertium, changes in word case are performed in the syntactic transfer stage, thus this could be solved by altering the set of transfer rules.

(8) 'n Pet vorm ook deel van die uniform.
een Pet vormen ook deel van het uniform.
Een pet vormt ook deel van het uniform.
'A cap is also part of the uniform.'

Apertium uses fixed length chunks for transfer. In example (9) there is an error due to this: *preciese* 'exact, precise' is an adjective modifying *grens* 'border'. While there is a pattern 'adj cc adj noun', there is no pattern 'adj adj cc adj noun'. This causes the chunker to put 'preciese' in a seperate chunk, which results in the predicative form, rather than the attributive.

(9) Daar is geen presiese geografiese of geologiese grens tussen Europa en Asië nie.
Daar is geen *precies* geografische of geologische grens tussen Europa en Azië niet.
Er is geen *precieze* geografische of geologische grens tussen Europa en Azië.
'There is no exact geographical or geological border between Europe and Asia.'

This error could be fixed by adding the aforementioned pattern.

Example (10) is one of those that was included in the 'other' category of syntactic transfer errors. The words *om te* come before infinitives in both Afrikaans and Dutch, much like *to* in English. However, the behaviour is not identical in Afrikaans as in Dutch.

(10) Jy kan aan Wikipedia meewerk sonder om enige besprekingsblaaie te lees.
Jij kunt aan Wikipedia *meewerk zonder *om* enig *besprekingsblaaie *te* lezen.
Jij kunt aan Wikipedia meewerken zonder enige besprekingsbladen te lezen.
'You can work on Wikipedia without reading any talk pages.'

Dutch cannot have *om te* after a preposition, in this case 'zonder' (without). A simple transfer rule could fix this for the case that *om te* is next to each other. However, in the case that it is seperated it is harder to solve.

3.4.6 Polysemy

The sentence in (11) has an error due to polysemy. The Afrikaans *algemene*, here as an attributive adjective, can be translated into Dutch as either *algemeen* or *voorkomend* (the former means 'general', the latter 'common' in English). While the Afrikaans word *algemeen* is used for both of these, they have a distinct meaning in Dutch.

(11) Sink is die vierde mees algemene metaal in gebruik.
Zink is de vierde meest *algemene* metaal in gebruik.
Zink is het op drie na meest *voorkomende* metaal in gebruik.
'Zinc is the fourth most common metal in use.'

Choosing the correct translation would require a module for lexical selection. However, it might also be worth changing the default translation.

3.4.7 Compounding

The error in example (12) is due to a specific rule in Dutch to do with compounds, *klinkerbotsing* – which also exists in Afrikaans as *vokaalopeenhoping*. If a compound is built-up from two words as such that the two vowels around the splitting point constitute a sound on their own, which means the word could be mispronounced, a hyphen should be used to distinguish the different parts of the compound.

(12) Die motornywerheid is die ekonomiese basis van Oshawa, ...
De *autoindustrie* is de economische basis van *Oshawa, ...
De *auto-industrie* is de economische basis van Oshawa, ...
'The car industry is the economic base of Oshawa, ...'

3.4.8 Separable verb

Example (13) demonstrates the problem with seperable verbs. The Afrikaans *ruk ... hand uit* corresponds with the Dutch expression *loopt ... uit de hand*. However, *ruk* 'to pull' in itself could never be translated as *lopen* 'to walk'. Note that 'uit de hand lopen' technically is not a seperable verb, but it poses the exact same problem as one.

(13) Die situasie ruk deur massabetogings hand uit.

De situatie **ruk* door *massabetogings *hand uit*.
De situatie *loopt* door massabetogingen *uit de hand*.
'The situation got out of control because of mass protests.'

3.5 Comparative

We compared our system to the other available MT system for Afrikaans to Dutch and Dutch to Afrikaans, Google Translate[10], a popular web-based statistical machine translation system. The evaluation was performed in the same way, the test corpora were translated with Google, and then post-edited.

For Afrikaans to Dutch, Google substantially outperforms the prototype Apertium system, with error rates reduced by a half. For Dutch to Afrikaans, the Apertium system performs better, although this could be due to the method used for testing the Dutch to Afrikaans direction favours more literal translations. E.g. it does not rely on post-edition. Another possible explanation could be that there are substantially bigger monolingual corpora for Dutch than for Afrikaans for building language models.

4 Discussion

We have presented a bi-directional rule-based machine translation between Dutch and Afrikaans, two closely-related Germanic languages. The system gives promising results, and offers an improvement in translation quality in the Dutch to Afrikaans direction over another publically available system, but does not offer any improvement in translation quality in the Afrikaans to Dutch direction.

We have shown that the development of an RBMT system between closely-related languages does not necessarily take a long time, and can be carried out by people with little formal training, and that the resulting system provides comparable results, in one direction at least, with a leading corpus-based machine translation system.

4.1 Future work

The three biggest issues in the system come from lack of dictionary coverage, poor morphological disambiguation and insufficient syntactic transfer. Thus these areas are ones that we intend to concentrate on. In addition, false friends have not specifically been looked at. We could review the list of false friends in (van Huyssteen and Pilon, 2009) to see if any translations could be improved.

[10] `http://translate.google.com/`

Acknowledgements

Development partially supported by the Google Code-in, a contest to introduce pre-university students to contributing to open-source software. Support also received from the Spanish Ministry of Science and Innovation through project TIN2009-14009-C02-01. Thanks also to Friedel Wolff.

References

Van den Bosch, A., Busser, G.J., Daelemans, W., and Canisius, S. 2007. An efficient memory-based morphosyntactic tagger and parser for Dutch. *Selected Papers of the 17th Computational Linguistics in the Netherlands Meeting* 99–114.

De Pauw, G., Laureys, T., Daelemans, W., and Van Hamme, H. 2004. A Comparison of Two Different Approaches to Morphological Analysis of Dutch. *Proceedings of the Workshop of the ACL Special Interest Group on Computational Phonology (SIGPHON), ACL2004*

Donaldson, Bruce C. 1993. *A grammar of Afrikaans*. Walter de Gruyter, Berlin

Koehn, P. 2004. Statistical significance tests for machine translation evaluation. *Proceedings of the Conference on Empirical Methods in Natural Language Processing* 388–395.

Laureys, T., De Pauw, G., Van hamme, H., Daelemans, W., and Van Compernolle, D. 2004. Evaluation and adaptation of the Celex Dutch morphological database. *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)* 1247–1250.

Levenshtein, Vladimir. 1965. Binary codes capable of correcting deletions, insertions, and reversals. *Doklady Akademii Nauk SSSR* 845–848.

Tyers, F. M. and Pienaar, J. A. 2008. Extracting bilingual word pairs from Wikipedia. *Proceedings of the SALTMIL Workshop at the Language Resources and Evaluation Conference*, LREC2008 19–22.

Shetter, William Z. and Ham, Esther. 2002. *Dutch: An Essential Grammar*, 9th edition. Routledge, Oxford.

van Huyssteen, Gerhard and Pilon, Suléne. 2009. Rule-based Conversion of Closely-related Languages: A Dutch-to-Afrikaans Convertor. *Proceedings of the 2009 Conference of the Pattern Recognition Association of South Africa*, Stellenbosch, SA 23–28.