THE DUTCH LANGUAGE IN THE EUROPEAN INFORMATION SOCIETY

3.1 GENERAL FACTS

With about 23 million native speakers, Dutch is the 8th most widely spoken native language in the EU. It is the commonly used language in the Netherlands and the Flemish part (called Flanders) of Belgium and one of the official languages in Surinam, Aruba, Curacao and Sint-Maarten, where it is used by parts of the population. It is also spoken in the EU in France and Germany, and outside the EU in Brazil, Canada, Indonesia (Java and Bali), South Africa, and the United States. The official Dutch name for the language is *Nederlands*, though Dutch as spoken in Flanders is usually called *Vlaams* (‘Flemish’). This White Paper focuses on the situation of the Dutch language and LT for it in the Netherlands and Flanders, which together we will designate with the term ‘the Low Countries’.

In the Netherlands, Dutch is the common spoken and written language and the native language of the vast majority of the population. The Netherlands has one officially recognised minority language, Frisian, spoken in the province of Friesland (Frisia). There are several immigrant languages. No reliable figures on the number of speakers of immigrant languages are known. However, the Centraal Bureau voor de Statistiek (Statistics Netherlands) [6] does provide figures for immigrants by ethnicity (= nationality). For ethnicities from outside the Netherlands some 1.5 million are from Western origin, and for non-western origin the figures are: Morocco (Rif Berber, estimated at 75%, and (Moroccan) Arabic, estimated at 25%) 350k persons, Netherlands Antilles and Aruba (Papiamentu) 138k persons, Surinam (Dutch, Sranan, Guyanese Creole English, Hindustani, Javanese) 342k persons, Turkey (Turkish) 383k persons, and other non-western (various languages) 644k persons.

In Belgium, Dutch is, by law, the language of Flanders, and one of the two languages (next to French) of the Brussels region. Belgium also has a French-speaking region and a German-speaking region.

Dutch has a variety of dialects, including (in the Netherlands) Achterhoeks, Drents, Gronings, Limburgs, Sallands, Stellingwerfs, Twents, Veluws and Zeeuws, and in Flanders West-Vlaams, Antwerpser, Oost-Vlaams, Brabant and Limburgs. The orthography is standardised but there were some changes in the standard recently (1996 and 2006). The standard is obligatory in education and governmental publications. Some of the recently proposed changes have led to different interpretations of the standard by different publishers, causing small differences in spelling (e.g., the *Groene Boekje* [7]: actievoeren v. *Van Dale*: actie voeren), and some spelling changes were not accepted by all publishers [8], who spell certain words differently (esp. with regard to the
so-called tussen-n in compounds), in accordance with the so-called Witte Boekje [9]. Dutch orthography can be quite complicated for certain words and constructions, so complicated that every year the so-called Groot Dictee [10] is organised by the Netherlands and Flanders and broadcast on national TV. The Groot Dictee is so difficult that anyone scoring less than 30 errors in about 8 sentences can be considered an excellent speller! In general, all Dutch dialects in the Netherlands share the same core grammar, though some dialects exhibit differences in some syntactic constructions. There are several lexical differences between dialects, and especially between Dutch as spoken in the Netherlands and Dutch as spoken in Flanders, e.g., the word ajuan is only used in Flanders instead of the standard Dutch ui (‘onion’). There are also several words that are the same in Flanders and in the Netherlands but have a different meaning, e.g., middag (lit. ‘midday’) in the Netherlands means the period of the day from 14:00-17:00 hrs, while in Flanders it means the period of the day from 12:00-14:00. Flemish also uses many words originating from French, e.g., terms for car engine parts, while Dutch in the Netherlands uses more English or English-inspired words in this domain. This also sometimes has consequence for pronunciation, e.g., the words flat and tram are in use both in the Netherlands and in Flanders, they are borrowed from English but in Flanders the borrowing went via the French language, so that in Flanders these words are pronounced as fl[A]t and tr[A]m while in the Netherlands they are pronounced as fl[E]t and tr[E]m.

3.2 PARTICULARITIES OF THE DUTCH LANGUAGE

The Dutch language exhibits some specific characteristics, which contribute to the richness of the language by allowing the speakers to express ideas in a large variety of ways. One such particularity is that it is quite common to put non-subjects sentence-initially (much more common than in English).

Certain linguistic characteristics of Dutch are challenges for computational processing.

For example, consider the English sentence *The woman was going to the store every day*. In English, there are very limited possibilities to use a different word order in this sentence, but in the Dutch equivalent almost any phrase can be the initial phrase in the sentence:

- De vrouw ging elke dag naar de winkel.
- Elke dag ging de vrouw naar de winkel.
- Naar de winkel ging de vrouw elke dag.

Word order in Dutch is thus much freer than in English (but not as free as in German).

Also, the Dutch language is quite productive in creating new compounds, though the use and productivity of compounding is not as extreme as in German. Nevertheless, newly formed compounds occur frequently and are difficult to process for NLP technology.

Word order is relatively free in Dutch sentences.

Another characteristic of Dutch that makes processing difficult is formed by separable verb prefixes that can occur far from the verb in nested constructions like:

Hij stelde zich na mij een drankje aanboden te hebben en wij in gesprek geraakt waren aan ons voor.  
(He *introduced* himself after he offered me a drink and we started a conversation.)

The meaning of a verb containing such a separable prefix like voor, in or uit can very often not be derived