Hearing on Spelling, Flemish Parliament, Monday February 20, 2006.

Proposition: The commotion caused by the Dutch language spelling revision is boosted by the Media's fundamental lack of an adequate spelling checker.

The Media publish articles in their newspapers and on their websites in which spelling errors show up very frequently. The majority of these errors are not related to the spelling revision of October 2005. Some of these errors are caused by two different, important schools of thought about spelling rules, i.e. the Van Dale dictionary versus the Groene Boekje (the Little Green Book) 1995.

However, we cannot put all the blame on the Media for the appearance of so many incorrect ways of spelling in their publications. Instead, we need to analyse the way in which they check their spelling.

Usually, spelling checkers that come with the editorial layout systems are not capable of correctly determining spelling errors. These spelling checkers pass a substantial part of the errors in texts. As the length of words increases so increases the chance of passing the errors.

It is impossible for most developers of spelling checkers to develop a lexicon that covers a major part of the wealth of Dutch words. Instead, they put together a limited dictionary of word *stems* (lemmata). They then create artificial words by combinations of these *stems*. In mathematics such combinations are called permutations. Many of these permutations are actually non-existent or incorrectly spelled words. Because of the large number of combinations resulting from the method of permutations, many spelling errors are not detected.

Earlier tests we conducted showed that 40% to 50% of the words with incorrect *intermediate* letters, or other small deformations, were passed by most of the spelling checkers. For the Dutch language one could think of words like *mistanden, bolletjesliker, briefbus, straatgids, rechtzaak*, etc.

The spelling revision of October 2005 has caused a new category of problems for spelling checkers, particularly as far as the frequency of appearance is concerned: the spelling has crossed the word boundary. These words now contain a space: *24 uursdoktersdienst, 18 holesgolfbaan*, etc. These combinations are chopped up by spelling checkers and are not checked while keeping in mind their interconnectedness. Formerly, this often happened too. An error like *gebruik maken* would then be missed.

*TALO has developed a new generation of spelling algorithms and processes that do detect the above mentioned errors. *TALO is now working with developers in the graphics field to set up the appropriate language technology. However, often we have to deal with editorial systems that do not have the functions needed to treat the language in accordance with the appropriate norms. Based on the methodologies supplied by the manufacturers it is impossible to select more than one word (spell check). Developers need to find other methods in order to be able to add an adequate spelling checker to such an editorial system. The interface technique offered by the manufacturers is where the problem is.

Never before has the Taalunie thought about including this aspect of spelling checking in its judgment. However, from conversations it appears to have become aware of how much of an impediment to the correct use of the language the functionality of spelling checkers can be.

*TALO is making a contribution to spelling checker technology, also based on the experience it gained with the German spelling revision of 1996.

Dr. J.C. Woestenburg, *TALO b.v.