



Reports from the SCARRIE project  
Editor: Anna Sågvall Hein

## An Error Database of Swedish

Olga Wedbjer Rambell  
Bengt Dahlqvist  
Erik Tjong Kim Sang  
Nils Hein

## **Preface by the editor**

SCARRIE is short for *Scandinavian Proof-reading Tools*. It is the name of a project within the EU TELEMATICS APPLICATIONS Programme. The project run for 30 months starting in November 1996.

*"The SCARRIE project aims at the development of a high-quality proof-reading tool for the Scandinavian publishing industry. A concrete concrete result of the project will be a carefully evaluated demonstrator, the SCARRIE pilot, designed to meet the needs formulated by a user group consisting of representatives for Danish, Norwegian and Swedish newspapers and publishing houses. "* (from LE3-4239 SCARRIE Project Programme. Annex I.)

The SCARRIE demonstrator for Swedish was developed by the Department of Linguistics at Uppsala university. Fundamental user input was provided by two Swedish newspapers, Svenska Dagbladet and Upsala Nya Tidning. We gratefully acknowledge their contributions. See further <http://www.scarrie.com/> for general information about the project and <http://stp.ling.uu.se/~ljo/scarrie-pub/> for a test version of the resulting Swedish SCARRIE pilot.

The achievements made in the SCARRIE project were continuously delivered to the European Commission via the project officer. When the project was still running the availability of most of the deliverables was restricted. Now the project is closed, and we make some of them available to a larger community. They all concern the development of the Swedish SCARRIE pilot.

A language checker will not cope with all kinds of errors, and based on large-scale error collection and error analysis a target set of error types was defined. The error types were classified in accordance with a fine-grained error typology. In this paper the error database will be presented. The delivery of the report was made in January 1998.

For reports on other aspects of the SCARRIE project, see Working Papers in Computational Linguistics & Language Engineering No. 3 - 4 and No. 6 - 13.