

# TRANSCRIPTION STANDARD

Semantics and Spoken Language  
Department of Linguistics  
Göteborg University

Version 6

Joakim Nivre

## 1. Introduction

This document<sup>1</sup> describes the standard for machine-readable transcriptions of spoken language used within the research program Semantics and Spoken Language at the Department of Linguistics, Göteborg University.<sup>2</sup> The document is organized as follows. In section 2, we describe the basic format requirements of transcription files. Section 3 is devoted to the transcription of utterances, or contributions, while sections 4–7 deal with the coding of various types of background information through the use of so-called information lines. In section 8, finally, we give instructions for the composition of the transcription header.

## 2. Basic Format Requirements

### 2.1 Global Structure

Every transcription is divided into two consecutive parts:

- Header
- Body

---

<sup>1</sup> This document is the result of a collective activity involving many people over a number of years. Important contributions have been made by Jens Allwood, Elisabeth Ahlsén, Leif Grönqvist, Johan Hagman, Staffan Larsson, and Sylvana Sofkova.

<sup>2</sup> Activities within this program have been funded by The Swedish Council for Research in the Humanities and Social Sciences (HSFR) through the project Semantics and Spoken Language and by The Bank of Sweden Tercentenary Foundation through the project Spoken Language and Social Activity.