



All papers ▾

Go!

Computer Science > Computation and
Language

A Formal Framework for Linguistic Annotation (revised version)

[Steven Bird](#), [Mark Liberman](#)

(Submitted on 26 Oct 2000)

`Linguistic annotation' covers any descriptive or analytic notations applied to raw language data. The basic data may be in the form of time functions - audio, video and/or physiological recordings - or it may be textual. The added notations may include transcriptions of all sorts (from phonetic features to discourse structures), part-of-speech and sense tagging, syntactic analysis, `named entity' identification, co-reference annotation, and so on. While there are several ongoing efforts to provide formats and tools for such annotations and to publish annotated linguistic databases, the lack of widely accepted standards is becoming a critical problem. Proposed standards, to the extent they exist, have focused on file formats. This paper focuses instead on the logical structure of linguistic annotations. We survey a wide variety of

Download:

- [PDF](#)
- [PostScript](#)
- [Other formats](#)

Current browse context:

cs

[< prev](#) | [next >](#)

[new](#) | [recent](#) | [0010](#)

References & Citations

- [NASA ADS](#)

DBLP - CS Bibliography

[listing](#) | [bibtex](#)

[Steven Bird](#)

[Mark Liberman](#)

Bookmark([what is this?](#))



existing annotation formats and demonstrate a common conceptual core, the annotation graph. This provides a formal framework for constructing, maintaining and searching linguistic annotations, while remaining consistent with many alternative data structures and file formats.

Comments: 29 pages, 20 figures, to appear in Speech Communication, An earlier version appeared as [cs.CL/9903003](https://arxiv.org/abs/cs.CL/9903003)

Subjects: **Computation and Language (cs.CL)**; Databases (cs.DB); Data Structures and Algorithms (cs.DS)

ACM classes: A.1; E.2; H.2.1; H.3.3; H.3.7; I.2.7

Cite as: [arXiv:cs/0010033 \[cs.CL\]](https://arxiv.org/abs/cs.CL/0010033)
(or [arXiv:cs/0010033v1 \[cs.CL\]](https://arxiv.org/abs/cs.CL/0010033v1) for this version)

Submission history

From: Steven Bird [[view email](#)]

[v1] Thu, 26 Oct 2000 17:42:30 GMT (100kb)

[Which authors of this paper are endorsers?](#) | [Disable MathJax \(What is MathJax?\)](#)

Link back to: [arXiv](#), [form interface](#), [contact](#).