Towards a citation standard

Jeff Good, MPI EVA (good@eva.mpg.de)

Functions of citations

- Location of the resource

- Indication of type of resource

- Credit for ideas and research

- Signaling prestige

Functions of citations

- Which work is more prestigious?

Bender, Emily M., Dan Flickinger, and Stephan Oepen. 2002. The Grammar Matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In *Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics*, 8–14, Taipei, Taiwan.

Dixon, R. M.W. 1979. Ergativity. *Language* 55:59–138.

Desirable features

- In presumed order of importance

    - Unambiguous indication of how to access the source

    - Clear indication of the resource type

    - Clear indication of who created the intellectual content

    - Some indication of the quality of the cited work

Location

- Where the material is archivally housed

- If possible, how to access the material via the internet (probably means giving a URI)

- How to access the material within the context of the archive itself (probably means giving a resource identifier)

- Should we encourage all identifiers to be URI's?

Location

- We also need to make sure people can not only locate the cited resource but that they can see what the resource looked like when it was cited

- If someone updates their texts in the archive, the reader needs to know what version of the texts the reference is based on

- This would seem to mean we need some kind of versioning system

Location

- Of course, people need to not only locate the resource in question—they have to figure out where to look in the resource for the cited content

- What is the digital equivalent of page numbers?

Type

- We need a signal that a resource is a digital resource

  - URI's would do this

- We want to know the type of digital resource: audio recording, annotated text, etc.

  - Maybe archives could have collection sets, e.g., *AILLA Annotated Texts*, *AILLA Audio Recordings*, *AILLA Video Recordings*

Credit

- Here, we want to make sure we facilitate current thinking in how to assign credit to speakers as well as linguists

- In some cases credit is easy: You list a person's name at the beginning of a reference

- Sometimes too many people may be involved to allow this—but this is a problem for paper resources as well

Credit

- But what about aggregated data?

- A scenario: Someone mines all the interlinear text found at PARADISEC and uses it to draw a new genetic tree for the languages represented by the archive

- How do we cite this? There could be (someday, we hope) hundreds of authors of the texts and thousands of texts

- And we need to know what the collection was right when it was sampled

Credit

- This isn't quite as science fictional as it sounds

- Though it's not an archive, the *World Atlas of Language Structures* database has been used in aggregate form like this

- Current practice is to site the WALS edited book...

- ...but what if there is no book? And is that really good practice anyway?

Prestige

- *Language* is prestigious; *OUP* is prestigious

- *LINCOM* is not prestigious. "Unpublished ms." is not prestigious

- Not all archived materials are equally good or valuable

- While a citation can't say everything about a resource, it should say something about its likely quality

Prestige

- Two possible routes

  - Archives have different categories of collections: *ELAR Deposits, ELAR Gold, ELAR Platinum, ELAR Not Deleted*

  - A "journal" with short reviews of high-quality deposits (is there a model for this elsewhere?)

Science fiction

- Could we move to a model where, when citation is needed, an archive gives the researcher a URI specific to the relevant "search"?

- That URI would accompany a standard citation

- It would lead the reader to the right part of a resource and show the reader what the resource looked like when the citation was made