

COLLECTION PUBLISHED SOURCES*

Joseph Grimes

This work is produced by OpenStax-CNX and licensed under the
Creative Commons Attribution License 2.0[†]

Abstract

Published sources that cover more than one speech variety are Collection metadata.



Figure 1: Published source from the Collection panel of JG-SulSel12.

The **word lists** that make up a collection come from a variety of sources:

- Published word lists or dictionaries
- Word lists you and your associates have collected
- Word lists other linguists have passed on to you
- Notes on languages made by nonlinguists
- Documents in some of the languages

Language surveys may cover areas where little is known about what languages are spoken there (yes, there are still parts of the world where this is true). Even in such situations, it is profitable to include word lists from well documented languages of the region as well as the lists the survey picks up. The better documented languages might or might not be related to some of the languages included in the survey. You want to know that, in order to tie the survey data in with comparative work already done.

Published sources cover, of course, anything modern libraries can catalogue with ease that contains data you use or discussions of comparative results:

- Books
- Articles
- Microform documents
- Archived paper documents

*Version 1.3: May 27, 2006 12:23 am -0500

[†]<http://creativecommons.org/licenses/by/2.0/>

- Electronic documents (some, not all)

Looked at another way, anything you can write a bibliographic entry for that would be acceptable in a refereed linguistics journal can be considered a published source.

What about Web pages or computer files as published sources? That's still a problem. There isn't enough experience among linguists for a consensus to form. Web pages come and go all the time. Nobody likes to be confronted with the dread Error 404 "Page cannot be found" message when all the World Wide Web's technology draws a blank on getting a page you asked for.

Permanence is the issue when deciding what can be cited. The Web page you are looking at in this tutorial has been designed to be accessible fifty or more years from now. You can see several acceptable ways to cite it by clicking on "Cite this content" at the bottom of this page. A good rule of thumb is

NOTE: A Web page to which a well funded university, museum, archive, or library is committed for the long haul is citable. Anything that is posted on a personal Web page, even that of a reputable linguist, is probably not citable.

What you want is to archive your material on a Web site that is committed to permanence. Make arrangements with one of the linguistic archives, on the order of AILLA¹, DOBES², or PARADISEC³. Or find a university or museum library committed to the preservation of electronic data. Some national archives may be set up as repositories for scientific data.

Keep in mind that while "endangered languages" are very much on people's minds, far fewer people are conscious of the problem of "endangered data."

Another question that comes up is whether a published source should be listed as metadata for the collection as a whole or as metadata for one of the speech varieties. The answer depends upon its scope. In one collection, two comparative works are given full references under the collection, because neither is restricted to a single variety in the collection. Then under the specific speech varieties, in the published source field, all that is needed is the author-year type of citation: *Savage 1986, Reid 1971*.

¹<http://www.ailla.utexas.org/>

²http://www.mpi.nl/DOBES/archive_info/

³<http://paradisec.org.au/>