Towards a journal for linked dictionaries of minor languages

Summary

We are planning to start an open-access electronic journal that publishes dictionaries of minor languages in database format, for which we need funding for a start-up phase of three years (for one scientist who coordinates the work in this initial phase). Collecting the words of as many languages around the world as possible is an important task of comparative linguistics, but traditional print publication of dictionaries of minor languages has become quite impossible. Existing online dictionaries are not regular refereed publications and thus do not contribute to career-building, and they are often made available outside a stable institutional context. Moreover, their entries are usually listed on HTML pages that emulate the previous technology (paper pages), rather than exploiting the possibilities of electronic publication (database publication, using the principles of Linked Data). The technical prerequisites for a database publication of dictionaries already exist at our institutions, and many researchers have dictionaries that they would be interested in publishing but have no means to do so. Thus, all we need is to set up an editorial board, establish a workflow, publish a number of seed dictionaries and advertise the journal among the community of minor language researchers.

The text below contains relevant passages from an application to the DFG ("Open-access publication of linked dictionaries for cross-linguistic comparison")

1. Why we need electronic open-access dictionaries

Like many other media, dictionaries have tended to move into the internet, and many people nowadays use online dictionaries for everyday purposes (e.g. leo.org, which has popular online dictionaries for translation between German and a number of widely spoken languages). Scientific, high-profile dictionaries like the OED or the Grimm Deutsches Wörterbuch are also increasingly used in their electronic version.

However, scientific dictionaries of small languages that have little or no practical use outside their speech communities have not often been published electronically so far, even though there are no major technical obstacles to this, and even though this would be highly beneficial for science in a wide variety of ways. In the following, we list a few reasons why online dictionaries of minor languages are very important for linguistic science world-wide:

– Minor languages are rapidly disappearing on all continents, but at the same time, interest (both scientific interest and wider public interest) in them has been rising over the last two decades.

– Due to language documentation efforts such as DoBeS (Volkswagen Foundation), HRELP[1] (Arcadia/Lisbet Rausing Charitable Fund), and DEL[2] (U.S. National

---

[1] http://www.hrelp.org/
[2] http://www.neh.gov/grants/preservation/documenting-endangered-languages

Endowment of the Humanities) as well as similar smaller-scale activities elsewhere, there is now a lot more data on minor languages from around the world. Many dozens of linguists have gathered new data on little-studied languages in modern formats.

– There is no alternative to electronic open-access publication of minor-language dictionaries. Dictionaries of such languages used to be published as print books (e.g. Winter 2003), but such publication is now very difficult or impossible. Major publishers find no global markets for minor-language dictionaries, and smaller, local publishers can publish dictionaries only for local purposes, so that they will hardly be accessible to the scientific community world-wide. (It is true that especially in the wealthier countries, there is sometimes funding for dictionaries for small communities interested in revitalizing their languages, e.g. Liljeblad & Fowler 2011. But this does not work as a general solution, and especially in these countries, people are increasingly interested in online dictionaries.)

– Early-career researchers in language documentation and description thus have no means for prestigious peer-reviewed publication of their research on the lexical resources of their language. Language documentation involves the collection of lexical data as a necessary prerequisite, and other researchers would like to have access to this information, but unlike work on grammatical structures, dictionaries cannot be published at present.

– Lexical data from a wide range of languages have increasingly been used in large-scale comparative work such as the research in the ASJP framework (Brown et al. 2008) or the work by Russell Gray and associates on Austronesian and other language families (e.g. http://language.psy.auckland.ac.nz/austronesian/). There is now a sizable community of researchers (also scholars coming from disciplines other than linguistics) who want to work on lexical comparative data, but are hampered by the lack of relevant comparable data.

– Electronic open-access publication, unlike traditional book publication, makes it easy to publish subsequent editions that supersede earlier editions. If there is a market for a first edition of a print dictionary, there will rarely be a market for a second edition, but there is no significant obstacle to electronic publication of dictionaries of the same language by the same author. This will make dictionary publication attractive to many scholars who are engaged in long-term (sometimes life-long) work on a language: They can now publish a dictionary that they know need not be their final word on the language.

There are currently a number of places with resources for comparative lexical research: Apart from etymological dictionaries (as published by the Tower of Babel etymological database project, http://starling.rinet.ru/), the following major resources exist:

– Intercontinental Dictionary Series, published by the Max Planck Institute for Evolutionary Anthropology: about 1300 words for over 200 languages (http://lingweb.eva.mpg.de/ids/), collected since the 1980s by Mary Ritchie Key and later by Bernard Comrie

– The ASJP database: 40 words in reduced transcription from over 5000 languages, put together by Søren Wichmann and colleagues the Max Planck Institute for Evolutionary Anthropology (http://cldbs.eva.mpg.de/asjp/)

– Austronesian Basic Vocabulary Database, by Simon Greenhill, Robert Blust and Russell Gray (http://language.psy.auckland.ac.nz/austronesian/): about 200 words from 979 Austronesian languages

– LEGO wordlists, collected by Helen Aristar-Dry, Anthony Aristar, and Jeff Good (http://legodev.linguistlist.org/about)

However, none of these can be said to be collections of dictionaries. The latter three mainly collate lexical information from a variety of sources, and the former is not a real citable publication either. But most crucially, they provide only very basic lexical information, rarely going beyond a single string and a simple counterpart in English. Thus, they are collections of word lists rather than dictionaries.[3]

Various individual dictionaries of minor languages have been made available online, for example:

– Araki dictionary: http://alex.francois.free.fr/AF-Araki.htm
– Archi dictionary: http://www.smg.surrey.ac.uk/archi/linguists/
– Bambara dictionary: http://www.bambara.org/lexique/lexicon/main.htm[4]
– Gamilaraay web dictionary:
http://coombs.anu.edu.au/WWWVLPages/AborigPages/LANG/GAMDICT/GAMDICT.HTM
– Passamaquoddy/Maliseet dictionary: http://pmportal.org/browse-dictionary
– Yucatec Maya dictionary:
http://193.175.207.216/yuclex/diccionario/dic_index.html

But these use a wide variety of formats and are not readily comparable, and they are published without peer review (thus without giving their authors the usual scientific recognition) and without a clear perspective of permanence. Once their authors lose funding or retire, they could quickly disappear. Moreover, in many cases it is not clear how they should be cited, and how their authors would list them on their CV, so they do not contribute to career-building.

The only larger project that has made available a collection of online dictionaries of minor languages is Swarthmore College's "Talking Dictionaries" (http://talkingdictionary.swarthmore.edu/, coordinated by David Harrison). However, the entries in these dictionaries (currently there are over 20) are not linked to each other or to anything else, and the publication is not peer-reviewed (as far as we know).

---

[3] A non-academic project that uses principles quite similar to ours is OmegaWiki (http://www.omegawiki.org/). This is of course primarily concerned with the bigger languages. But we will be able to link our lexical data to OmegaWiki's.
4 Both the Araki and the Bambara dictionaries are based on LexiquePro, a tool for producing electronic dictionaries (see http://www.lexiquepro.com/). LexiquePro is very useful, but the resulting electronic dictionaries are published as simple HTML pages, not using the principles of Linked Data.

Most of the existing online dictionaries of minor languages simply emulate the structure of the previous technology (paper pages), by simply listing the words on HTML pages. This leads to a familiar look, but it means that the usability of the dictionaries is severely restricted (one cannot readily sort the entries by different criteria, one cannot export them in database format, one cannot link to individual data points).

Thus, we maintain that electronic dictionaries need to be published in the same way as other scientific contributions: with peer review and peer selection, in a journal that regularly accepts submissions and stores them and makes them accessible indefinitely. Moreover, dictionaries of minor languages need to be published in such a way that the data can be readily used by comparative linguists as well, i.e. in a database format, with data that can be easily exported, as in the *World Loanword Database* (Haspelmath & Tadmor 2009b). So far, nothing of this sort exists, as far as we know, even though it is an urgent need of the community of linguists working on minor languages.

2. Goals of our project

The goal of this project is to set up and establish an electronic journal that publishes open-access dictionaries of less widely spoken, little-known languages[5] for the purpose of scientific research on these languages (i.e. "scientific dictionaries"). In addition, the dictionaries can have various features serving the needs of speaker communities.

The dictionary journal[6] should be open to linguists around the world, and it should aim to become a world leader. It should not be the only dictionary journal: We hope that others will follow our initiative, because we see a lot of demand for peer-reviewed dictionary publication. But as the first dictionary journal, it will set standards, and we will try to preserve its status as the most prestigious outlet for dictionary publication.

The primary goal of the dictionary journal is to publish new dictionaries from recent research that have not been published so far. This does not exclude the possibility of publishing updated editions of dictionaries by authors who have published a previous dictionary in a rudimentary form (e.g. on a self-programmed website, in a working paper, or through a local publisher with limited distribution). But in this project, we do not aim to digitize and make accessible published legacy dictionary materials by other authors, even though this is of course also a worthwhile task.[7]

---

[5] We sometimes call these languages "minor languages" for convenience. We have no strict definition of these languages, and we would not exclude publishing dictionaries of languages with many speakers. However, when a language is an official language of a country or otherwise has a lot of institutional support, the considerations leading us to propose this journal do not apply in the same way.

[6] The word "journal" typically evokes a collection of prose research papers. But dictionaries are databases, not prose documents, and we will publish them as databases, not as text or PDF files. Nevertheless, we use the term "journal" because we are aiming for a serial publication with a uniform editorial structure that will exist for an indefinite time.

[7] Two retrodigitizing projects that we are particularly familiar with are Michael Cysouw's "Quantitative language comparison" (http://www.qlc.sprachwiss.uni-muenchen.de/about/index.html)

The dictionaries that we aim to publish are comprehensive word collections of individual languages, minimally with translation into English. We will not consider highly specialized dictionaries such as etymological dictionaries, spelling dictionaries or pronunciation dictionaries. Dictionaries dealing with particular topical areas (e.g. dictionary of fishing terminology) may be acceptable, but this needs to be discussed in the course of the project. All dictionaries will have to include minimal grammatical information such as word-class and (where relevant) inflection class and other lexical properties that reflect irregular grammatical behaviour, but we do not require detailed syntactic information, as this would be a different project. If, however, certain authors want to include more detailed information on the syntactic behaviour of the corresponding lexical items, we will encourage and help them to do so. Thus, the dictionaries will be fairly diverse, depending on the author's particular interests, though their presentation will be as uniform as possible, and they will all be linked among each other via the English glosses/definitions.

We hope to publish about 10 dictionaries in the first year, 25 in the second year, and 35 in the third year. We realize that estimating the effort that will have to go into such a journal is not easy, but we are confident that we have all the prerequisites for launching such a new venture.

Apart from serving world-wide and comparative linguistics, this project is also of great importance for the communities of speakers who are often struggling to preserve, maintain or revitalize their languages (cf. Corris et al. 2004, Mosel 2011, Ogilvie 2011). The disappearance of small languages is widely regarded as an important cultural and social issue, and many different players around the world are making efforts to strengthen speaker communities of minor languages. While our primary interest is in comparative linguistics, we have good connections to scholars and activists involved in language revitalization efforts. The needs of speaker communities will always be present during our work as well. Since our publication is open-access, there is no obstacle for use of our dictionaries by lay people. When desired, we will of course include translations of all words not only into English, but also into another widely spoken language that the members of the speaker communities may be more familiar with (Spanish, Russian, Chinese, etc.). We will make sure that the dictionaries are published in such a way that a special app for portable devices can easily be created, though creating such an app is not currently part of our plans.

We anticipate that the possibility of easy and prestigious dictionary publication will have a significant impact on the future directions of linguists' work. So far, published grammars tend to carry more prestige than published dictionaries, but although we are primarily grammarians ourselves, we feel that this is not justified. Producing a good dictionary requires the same kind of sophisticated analytical work as producing a grammar, and it should get the same attention from linguists working on minor languages.

---

and the Mon-Khmer Languages Project by Paul Sidwell and Doug Cooper (http://sealang.net/monkhmer/). The LEGO project (http://legodev.linguistlist.org/) has also been mostly concerned with retrofitting existing word lists and dictionaries.

The dictionaries will be open-access publications, probably using a CC-BY licence,[8] which puts minimal restrictions on the use of the dictionaries: All that is required is attribution (i.e. proper citation), as in all scientific contexts. Thus, users will be allowed to download the dictionaries and use them in their own preferred formats. We may also provide an export option as a linear alphabetical text file (to yield a format that is familiar to all older users), but this is explicitly intended as a derived format. The dictionaries are primarily electronic databases and are published as such, linked to other dictionaries.

After the three-year period of funding by the DFG, we trust that the workflows will be so standardized and automatized that further costs will be quite small, so that the journal can be run without a full staff member coordinating it. It will not require more efforts than editing any other linguistics journal, and we trust that we will find someone else to take over editorship in case we may not be able to continue at some point. We have no plans of charging APCs (author processing charges) at any point in the future. Our journal should continue to be an activity of academics primarily for academics, and it should be funded by the usual funders who fund scientific research.

Due to the general nature of funding for innovative scientific efforts, there are currently no specific funding promises for the technical service that is required by the journal beyond 2016. In line with this, we are not asking the DFG to make any commitments for the longer term. However, since the journal is fully integrated into the Cross-Linguistic Linked Data (CLLD) set of databases, which includes a number of other important and prestigious databases (*WALS Online*, Glottolog/Langdoc, *APiCS* Online, World Loanword Database, eWAVE, ValPaL, ASJP), we have no doubt that funding options will exist: The resources created over the years within the CLLD framework are too big to disappear, as there are too many stakeholders. One concrete possibility that we have discussed is to integrate the databases into The Language Archive,[9] which has long-term funding from the Max Planck Society (MPG), the BBAW (Berlin-Brandenburgische Akademie der Wissenschaften), and the KNAW (Royal Dutch Academy of Sciences).

3. Procedures

The work in this project has three main aspects: A technical aspect (online publication of lexical databases using the linked-data approach), a conceptual aspect (the structure of the lexical databases), and a management aspect (coordinating the editorial process).

3.1. Technical prerequisites

3.1.1. General. The technical infrastructure is the responsibility of Robert Forkel, who is employed as a database manager, programmer and linked-data expert by the four-year project Cross-Linguistic Linked Data (CLLD, 2013-2016), funded by the Max Planck Society (PI Martin Haspelmath).

---

[8] http://creativecommons.org/licenses/by/2.0/, see Shieber (2012) for justification.
[9] http://tla.mpi.nl/

Robert Forkel has extensive experience in programming online databases for publishing cross-linguistic datasets: *WALS Online* (since 2008) and the *World Loanword Database* (since 2010) were created by him. During the four years of the CLLD project, he will create similar published database sites for the *Atlas of Pidgin and Creole Language Structures* (Michaelis et al. 2013, *APiCS Online*), the *ASJP* lexical database (Brown et al. 2008 etc.), and the *Intercontinental Dictionary Series* (Comrie et al.), and he will ensure that other databases published by MPI-EVA (Glottolog/Langdoc, Nordhoff et al. 2012; *eWAVE*, Kortmann & Lunkenheimer 2011; and *ValPaL*, Hartmann et al. 2013) will remain compatible and linkable with the databases created by him.

Creating and curating the dictionary journal was part of the CLLD project tasks from the very beginning, so the technical infrastructure is taken care of by this project. No DFG funding is required for this part of the project. Data storage and hosting is ensured via a GWDG server owned by the Max Planck Institute for Evolutionary Anthropology.

The dictionaries will be published in such a way that the browsing function of print dictionaries will be preserved, by showing a table of words sorted by some criterion such as alphabetical order. (That is, users will not just be confronted with a simple search field as in leo.org and in other online dictionaries.) In fact, the browsing function will be enhanced, because browsing will also be possible by other important fields such as word-class or semantic field. Thus, electronic dictionaries are superior to page-based print dictionaries not only due to their searchability, but also due to their enhanced browsability.

3.1.2. The Linked Data approach.
In short, Linked Data means "serving data the web way". It can be described as a data access protocol that is as much as possible in line with HTTP (the Hypertext Transfer Protocol) based on four principles (Bizer et al. 2009):

1. Use URIs to identify things.
2. Use HTTP URIs so that these things can be referred to and looked up ("dereferenced") by people and user agents.
3. Provide useful information about the thing when its URI is dereferenced, using standard formats such as RDF/XML.
4. Include links to other, related URIs in the data provided to improve discovery of other related information on the Web.

In particular the principle of using URIs to identify all resources is essential for publishing data in a distributed web of data, because it combines the property of global uniqueness for identifiers with a standard resolution and access mechanism. Thus, Linked Data allows for a distributed development of data resources (with a well-defined method for merging) and of data consumers, i.e. tools which make use of the data.

We also expect Linked Data to eventually supersede protocols like OAI-PMH (currently used by OLAC[10]), which go against the grain of the web by not using techniques like URIs and principles like REST (representational state transfer) to make resources available. Since these can readily be replaced by Linked Data, a consolidation of data (and metadata) access methods is in reach.

---

[10] OLAC: http://www.language-archives.org/

Last but not least, Linked Data describes a minimal protocol which data publications must adhere to in order to remain usable. Additionally, this protocol is trivial to implement for static data, thus it provides an exit strategy (or at least a strategy to scale down efforts) in case resources can no longer be kept dynamic (i.e. updatable). While this would obviously not be the ideal state for data resources to be in, it still allows for more credible claims of persistence of data publications.

The fact that this approach is gaining traction is underlined by the fact that the European Commission is sponsoring multiple Linked Data projects,[11] and that CrossRef has announced Linked-Data-friendly DOIs.[12]

Also from the point of view of persistence, our Linked Data approach has advantages: It is a technology to keep data available and accessible at a minimum effort. We can just dump the requirement for resources to be updatable and host the data as simple files on any webserver. Reconstructing a proper database from these is straightforward.

3.2. Conceptual foundations

The most important conceptual task that we will tackle in this project is to set up a comprehensive data model for linked dictionaries of minor languages. Since we expect our journal to become influential, it may well be that our choices will become standard practice in lexicography of minor languages (in much the same way as the Leipzig Glossing Rules have become standard for interlinear morpheme-by-morpheme translations).[13] Of course, we will be able to build on a lot of previous experience and a lot of earlier literature on lexicography (e.g. Kirkness 2005, Atkins & Rundell 2008, Svensén 2009, Granger & Paquot 2012, to name but a few). But much of the previous literature discusses printed dictionaries of big languages, so only part of it is relevant for our concerns. Existing published electronic dictionaries of minor languages do not follow any kind of uniform data model. Unpublished dictionaries often follow models suggested by software tools like Toolbox, FLEx, LexiquePro and LEXUS. We will of course make sure that our data model is interoperable with existing formats such as LMF and LIFT and we will make full use of the results of the RELISH project (Aristar-Dry et al. 2012).

All the dictionaries that we will publish will contain a lot of information that is comparable across languages, but they can of course also contain some types of information that are specific to individual languages.

Dictionaries are sets of lexical entries, each of which contains information on the lemma in several fields. Some of these fields will be applicable (in principle) to all languages, especially the following (the fields with an asterisk are obligatory).

*lemma label
alternative form
pronunciation
*word-class
*English gloss/definition

---

[11] http://latc-project.eu/, http://planet-data.eu/about

[12] http://www.crossref.org/CrossTech/linked_data/

[13] A valuable paper that investigates earlier dictionaries from such a point of view is Bell & Bird (2000).

meaning paraphrase in the same language
semantic domain
general comment
links to translated example sentences
links to related entries[14]
etymological origin (e.g. loanwords, or even proto-language forms)
gloss in another major language

Other fields will be relevant only to some languages, e.g. gender, inflection class, special inflected forms, information on dialectal variants, and so on. It will be our task to determine the best set of database fields that allows both for comparison across languages and for expression of language-particular peculiarities. We will also need to decide how to deal with multiple senses and homonyms.

In addition to the lexical entries, the databases will contain a substantial number of glossed and translated example sentences showing the lexical entry (or a particular sense of an entry) in context. These will be stored in a separate database table, so that they can be linked to multiple entries. Glossed example sentences have become an indispensable component of descriptive grammars of minor languages. By contrast, for example sentences in dictionaries, glossing has so far not been practiced, but apparently only because of the restrictions of the traditional page-based format. Most active fieldworkers nowadays have large sets of glossed sentences, and these can be easily used in dictionaries as well. Unglossed example sentences are of limited use for minor-language dictionaries, as they can be fully appreciated only by readers who know the language quite well already.

Other important decisions relate to the kind of submissions that we want to accept for the journal. We will probably want to require that at least 1000 entries are represented in the dictionary, covering at least the core vocabulary,[15] and we will want to require that at least 10% of the entries are linked to glossed example sentences. We will allow the inclusion of pictures and sound files when they are used consistently and judiciously. But we also find it important that dictionary authors should have a substantial amount of freedom in focusing on those aspects of the language that interest them most. Some authors will be highly interested in syntax and will make sure to include full valency information for all verbs, for example. Others will be highly interested in traditional material culture and will invest heavily in description of words for traditional artifacts or special cultural practices.

Finally, we will need to develop a way of comparing the meanings across the languages. Minimally, they can of course be compared via the English glosses. (Maximally, one would develop a sophisticated semantic metalanguage that is neutral with respect to semantic differences between languages, but this is beyond the scope of this project.) If the words could only be compared via the English glosses, this would cover a large segment of what can be compared at all, but we can improve on this: For example, we can make sure that the English glosses use a

---

[14] Since the dictionaries are primarily lexical databases, alphabetic sorting plays a limited role, which means that lemmas can be longer expressions such as *to take part,* or *to keep pace.* There is no need to turn these into sublemmas of either *take* or *part*, because they can simply be linked to both.

[15] Thus, we will probably make it a requirement that all words on the Swadesh-200 list and on the Leipzig-Jakarta list (Tadmor & Haspelmath 2009a) are represented.

uniform style (e.g. either ‚the moon' or ‚moon', and either ‚to take' or ‚take'), and we could require that certain frequent homonyms (*hide*, *like*, *bark*, etc.) are resolved. For dictionaries that include lexical information that has been compiled for a certain aspect of semantic typology, we will integrate the stimulus material that has been used. These data may be linked to in other dictionaries that refer to the same stimuli.

3.3. Coordination of the editorial process

Starting a completely new journal of a novel kind requires some heavy investment at the beginning, so we need a full position for a scientist to coordinate the editorial process.

In a first step, we need to identify four or five high-quality dictionaries by authors from different institutions that we know well, on languages in different parts of the world. These will be published as "seeds", using a non-standardized method of quality control. Once these initial dictionaries are published (we estimate that this will be six months after the beginning of the project), we will publicize the dictionary journal widely, and then we expect to get unsolicited submissions.

In order to establish the dictionary journal's scientific credentials, we will set up an Editorial Board of about 20 distinguished (and some junior) scientists with an interest in research on minor languages and some experience in lexicography. We are very well connected to this group of linguists, so this will be easy.

The coordinator will establish the contacts with the seed dictionary authors, and later with the dictionary submitters, and will coordinate the data transfer between the authors and the programmer. This will get less and less complicated with each new dictionary, as we gather more and more experience.

Once submissions come in, another important task of the coordinator will be to contact two reviewers and ask them to review the dictionary, supplying the database in a suitable format that allows for easy browsing and annotation. One of them should be an expert of the language family or area, while the other one can be a linguist with general expertise or expertise in lexicography. The Editorial Board members will be particularly (but not exclusively) relied on for this task.

All procedures of the editorial workflow will be documented in detail, and the most important parts will be published on the journal website, to make the process as transparent as possible, and to facilitate portability of the journal in case it needs to be taken over by a different editorial team.

We will investigate the range of import formats that are practical both for dictionary creators and for the journal editors. Thus, since many dictionary creators use SIL's Toolbox, LexiquePro and FLEx, the export formats of these programmes should probably be given special support. But we do not regard it as our task to accommodate idiosyncratic file formats and idiosyncratic database structures: Dictionary authors who want to submit their dictionary are required to comply with our Guidelines very closely on first submission. We will require different database tables for lexical entries and example sentences, full description of all nonstandard database fields, full description of the controlled values, Unicode encoding of

characters, and so on.

The project will consist of three phases: a preparatory phase (first six months), the primary phase of the two years in the middle (month 7 through 30), and the final phase.

During the preparatory phase, we will integrate the seed dictionaries into the database and web application, and we will discuss the conceptual issues. This will be the most intensive period for the PIs. During the spring of 2014 (ideally after three or four months), we will organize a workshop with 10 outside participants (potential members of the editorial board) to discuss our proposals. At the end of the preparatory phase, the journal will be launched officially with a call for contributions.

During the primary phase, we will publish more and more dictionaries and solve remaining technical and conceptual issues.

During the final phase, we will ensure that the transition to the post-DFG operation of the journal will be smooth. We will document the workflow of the publication process in full detail.

References

APiCS Online: Michaelis, Susanne & Maurer, Philippe & Haspelmath, Martin & Huber, Magnus (eds.) 2013. *The Atlas of Pidgin and Creole Language Structures Online.* Leipzig: Max Planck Institute for Evolutionry Anthropology. http://apics-online.info (to appear)

Aristar-Dry, H., S. Drude, J. Gippert, I. Nevskaya and M. Windhouwer. 2012. Rendering Endangered Lexicons Interoperable through Standards Harmonization: the RELISH Project. Presented at ELRA Conference on Language Resources and Evaluation Istanbul. Available at http://pubman.mpdl.mpg.de/pubman/item/escidoc:1480877:2/component/escidoc:1480876/RELISH-contribution-final.pdf

ASJP: Automated Similarity Judgment Program (http://email.eva.mpg.de/~wichmann/ASJPHomePage.htm), see Brown et al. 2008

Atkins, Beryl T. Sue & Michael Rundell. 2008. *The Oxford guide to practical lexicography*. Oxford: Oxford Univ. Press.

Bell, John & Steven Bird. 2000. A preliminary study of the structure of lexicon entries. Presented at the Workshop on Web-based Language Documentation and Description. Available at http://www.ldc.upenn.edu/exploration/expl2000/papers/bell/bell.html

Bizer, Christian & Tom Heath & Tim Berners-Lee. 2009. Linked Data - The Story So Far. In Heath, T., Hepp, M., and Bizer, C. (eds.). Special Issue on Linked Data, *International Journal on Semantic Web and Information Systems (IJSWIS)*. <http://linkeddata.org/docs/ijswis-special-issue>

Brown, Cecil H., Eric W. Holman, Søren Wichmann, and Viveka Velupillai. 2008. Automated classification of the world's languages: A description of the method and preliminary results. *STUF – Language Typology and Universals* 61.4: 285-308.

Corris, Miriam; Christopher Manning; Susan Poetsch; Jane Simpson. 2004. How Useful and Usable are Dictionaries for Speakers of Australian Indigenous languages? *International Journal of Lexicography* 17, 1: 33-68.

eWAVE: Kortmann & Lunkenheimer 2011

Glottolog/Langdoc: Nordhoff, Sebastian & Hammarström, Harald & Haspelmath, Martin (eds.) 2012. *Glottolog/Langdoc*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://glottolog.org

Granger, Sylviane & Magali Paquot. 2012. *Electronic lexicography*. Oxford: Oxford University Press.

Hartmann, Iren, Juliane Kabisch-Lindenlaub & Christian Marschke. 2010. *Hocak Teaching Materials, Volume 1: Elements of grammar, learner's dictionary*. (Ed.) Johannes Helmbrecht & Christian Lehmann. Albany: State University of New York Press.

Hartmann, Iren, Martin Haspelmath & Bradley Taylor (eds.). 2013. *Valency Patterns Leipzig: A database of valency information in 35 languages around the world*. Leipzig: MPI for Evolutionary Anthropology. http://valpal.info (to appear)

Haspelmath, Martin & Uri Tadmor (eds.). 2009a. *Loanwords in the World's Languages: A Comparative Handbook*. Berlin: De Gruyter Mouton.

Haspelmath, Martin & Uri Tadmor (eds.). 2009b. *World Loanword Database*. Munich: Max Planck Digital Library. http://wold.livingsources.org/.

Kirkness, Alan. 2005. Lexicography. In: *The Handbook of Applied Linguistics*. Davies, Alan and Catherine Elder (Eds). Oxford: Blackwell Publishing.

Kortmann, Bernd & Lunkenheimer, Kerstin (eds.). 2011. *The electronic World Atlas of Varieties of English*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Liljeblad, Sven S. 2011. *Northern Paiute-Bannock Dictionary*. University of Utah Press.

Mosel, Ulrike. 2011. Lexicography in endangered language communities. In Peter K. Austin & Julia Sallabank (eds.), *The Cambridge Handbook of Endangered Languages*, 337–353. Cambridge: Cambridge University Press.

Ogilvie, Sarah. 2011. Linguistics, Lexicography, and the Revitalization of Endangered Languages. *International Journal of Lexicography* 24(4). 389–404. doi:10.1093/ijl/ecr019 (2 May, 2013).

Shieber, Stuart M. 2012. The case for the journal's use of a CC-BY license. *Journal of Language Modelling* 0(1). 5–8. (2 May, 2013).

Svensén, Bo. 2009. *A handbook of lexicography: the theory and practice of dictionary-making*. Cambridge: Cambridge University Press.

ValPaL: Hartmann et al. 2013

WALS Online: Dryer, Matthew S. & Haspelmath, Martin. 2011. *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library.

World Loanword Database: Haspelmath & Tadmor 2009b

Winter, Werner. 2003. *A Bantawa dictionary*. Berlin: Mouton de Gruyter.