

Developing infrastructure for team-based language documentation and description: The module-and-seminar model

Lev Michael
University of California, Berkeley
levmichael@berkeley.edu

1st International Conference on Language Documentation and Conservation

1 Introduction

- This talk motivates, describes, and evaluates the infrastructure developed for team-based language documentation and description (TBLDD) in the context of the Iquito Language Documentation Project (ILDP).
- Team-based linguistic documentation and description:
 - Confers significant advantages over the traditional one-language-one-linguist model,
 - But it also poses challenges for the fruitful integration and coordination of research activity.
- In the ILDP we sought to retain the advantages of TBLDD (see §2), while mitigating the difficulties it poses to research coordination and integration (see §3).
- The infrastructure we developed for doing so, which we call the Module-and-Seminar Model (MSM), consisted of:
 - The delimitation of a large set of relatively short research topics – the **modules**
 - The assignment of modules to individual researchers in accord with a **module schedule** prepared at the beginning of each field season
 - The presentation of each module at a daily **seminar**
 - The inclusion of modules in a bound volume at the end of each field season
 - Two **coordinators**, who were responsible for coordinating research activities in the group

1.1 Background on the ILDP

- The ILDP is a multi-year team-based project to document and describe Iquito, a highly endangered Zaparoan language of northern Peruvian Amazonia.
- Apart from standard documentary and descriptive goals, the project also sought to develop:
 - strategies for incorporating community members into the project in active research roles (discussed in Christine Beier's March 12 ICLDC talk)
 - frameworks to facilitate team-based linguistic fieldwork (the topic of this talk)
- Principal fieldwork was carried out between 2002 and 2006; short supplementary fieldwork trips and write-up of materials (dictionary, text collection, and grammar) are ongoing.
- Depending on the year, the project's research teams consisted of:

- 4-7 visiting linguists, who were graduate students from the University of Texas at Austin and *licenciatura* students from the *Universidad Nacional Mayor de San Marcos* (Lima, Peru); total of 12 students participated in the project, many for two or more years.
 - 2-3 community linguists, who were community members trained in aspects of language documentation and description
- Fieldwork by community linguists continued throughout the year, while teams of visiting linguists were in the community for two months during the North American academic summer.
 - ILDP funding:
 - 2001-2002 research and language revitalization activities were funded by grants from the Endangered Language Fund and Cabeceras Aid Project, and by grants to individual researchers from the University of Texas at Austin.
 - Fieldwork and language revitalization activities during 2003-2006 were funded by a Major Project grant from the Hans Rausing Language Endangered Languages Program.

2 Advantages of team-based language documentation

- By providing social, intellectual, and material infrastructure for student researchers, their entry into the field is facilitated, and their chances of success are enhanced. This is especially true in rustic and geographically remote fieldwork contexts like the Amazon Basin.
- Creates a scholarly community around a language, increasing the richness and quality of the resulting work. This could be especially significant in cases of underdocumented language families, where there might be no other linguists with relevant expertise.
- Permits greater specialization by individual linguists, and hence greater depth in description and analysis.
- Significantly increases the amount of work that can be carried out on a language during relatively short field stays, by increasing the total number of person hours devoted to the project.
- With the right infrastructure, permits relatively short-term involvement in the project (i.e. for a single field season), which is advantageous for attracting fieldworkers.

3 Lateral and longitudinal challenges in integrating and coordinating team-based language documentation and description

- Language documentation and description is an intrinsically holistic activity, requiring a high degree of integration and coordination among different facets of the activity.
- Team-based research, however, disrupts the de facto integration and coordination among the different facets of a research project that results from having only a single researcher working on a project.
- Thus, although team-based language documentation confers advantages over the single-linguist model, it also poses new challenges of integration and coordination.
- We directly experienced many of these challenges during the 2002 fieldwork season, our first fieldwork season (with four visiting linguists), and these experiences inspired our development of the module-and-seminar model.

3.1 The Lateral Integration Problem

- With multiple researchers working on documentation and description, each researcher discovers facts about the language and develops analyses useful to other researchers on the team.
- Without some efficient way of sharing this knowledge, there is the risk that:
 - There will be a great deal of redundant research, as multiple researchers carry out similar work to resolve (unknowingly) shared analytical and empirical issues.
 - Researchers will be ignorant of the results produced by other researchers that are relevant to their own work.

3.2 The Longitudinal Integration Problem

- Another set of integration problems was posed by the organization of the project around relatively short field seasons (two months per year) and the variation in the year-to-year constitution of the visiting research team.
- These circumstances meant that:
 - Each year (potentially – and in fact) new researchers joined the team with no prior experience with the language, necessitating that the advances of previous years be available in some form to these researchers.
 - Even researchers from previous years stood in danger of losing a great deal of knowledge and partial analyses in the gap between field seasons.

3.3 The Lateral and Longitudinal Coordination Problem

- Another potential issue for TBLDD is coordination among different researchers regarding research topics.
- Without such coordination, it is perfectly possible with multiple researchers:
 - for two or more people to end up working on the same topic
 - for significant gaps to arise in documentation and description because of:
 - * chance
 - * avoidance of unpopular topics (note that the social dynamics of groups favor shirking of unpleasant tasks)
- This issue of coordination is relevant both within a given field season, and across multiple ones, as the coverage of topics develops in both range and depth.

4 The Module-and-Seminar Model

- In 2003, the second year of the project, we developed a research framework to address the integration and coordination issues just sketched: the **module-and-seminar model**.
- At the core of this model were:
 - sets of weekly research topics that were assigned to each researcher, upon which they wrote reports – the **modules**.
 - a two-hour **seminar** at the end of most work days, at which one module was presented by the researcher to which it was assigned, to the entire team for discussion.

- The topics were assigned to the researchers on the basis of a master **module schedule** prepared by the **project coordinators**, who also consulted with the researchers as they worked on their modules.
- The final version of each module was included in a summary volume with the texts and pedagogical materials produced that year, plus the latest version of the dictionary.
- This framework addressed the integration and coordination issues described above by:
 - Providing a regular venue for the sharing of analyses, so that the team as a whole could contribute to the research topics of individuals and be apprised of individual advances
 - Creating an immediate written record of work for reference during that field season, and via the summary volume, for reference by researchers in subsequent years
 - Via the module schedule, assuring even coverage of research topics, while avoiding overlaps

4.1 Project coordinators

- In the ILDP, there were two project coordinators, each with different domains of responsibility:
 - texts and pedagogical materials (Christine Beier)
 - dictionary and grammatical description (Lev Michael)
- The two project coordinators remained the same over the course of the project, despite significant changes in the constitution of the remainder of the research team.
- The coordinators constituted an important institutional memory for the project, significantly ameliorating the longitudinal issues described above.
- The grammatical description coordinator regularly consulted with researchers as they wrote their modules.
- Significantly, both coordinators were graduate students at the time, so their efficacy was not especially tied to greater expertise than the other team members, but simply their longer-term participation in the project.

4.2 The module schedule

- At the beginning of each field season, the grammatical description coordinator developed a module schedule in conjunction with individual researchers.
- Efforts were made to balance project needs with individual interests and skills in the preparation of the schedule.
- The schedule typically included:
 - 5-6 weeks of modules
 - 1-2 weeks of texts (typically front-loaded at the beginning of the fieldwork season as a way of immersing newcomers in the language)
 - 1 week of work on pedagogical materials
- Modules were normally assigned to each individual researcher for a one week period, who presented their results at the end of the one week period during the daily seminar.
- The seminar dates were assigned in a staggered fashion so that at any given time, each team member was working on a single module (or texts or pedagogical materials), which ‘came due’ and was presented at the daily seminar on different days.
- Page 7 of this handout is a copy of the table of contents of the summary volume of modules for 2004, and serves to give a sense of the number of modules produced in a field season and the nature and scope of module topics.

4.3 The seminar

- The number of seminars during each week corresponded roughly to the number of researchers on the team, with each researcher being responsible for presenting his or her work on a fixed day of the week.
- Seminars began at 4 pm each day and lasted until about 6pm (when the mosquitoes became intolerable).
- The researcher provided printed drafts of the module, which s/he then presented to the group for discussion.
- Following the seminar, the presenting researcher and the descriptive grammar coordinator met to discuss final amendments to the module.

4.4 The life cycle of a module

- **Days 1-4:** Mainly elicitation and data harvesting from texts with a small portion of analysis per day (8-12 hour work days); consultation with descriptive grammar coordinator
- **Day 5-6:** Work shifts to roughly 25% elicitation and data harvesting, 50% analysis, and 25% write up
- **Day 7:** Work shifts to 75% write up, 25% mixture of analysis, last minute data collection, and consultation with descriptive grammar coordinator
- **Day 7, 4pm:** 2-hour presentation of module at the daily seminar followed by group discussion; post-seminar consultation with coordinator
- **Day 8/Day 1:** Incorporate comments from seminar into a final version of the module, start on new research topic

5 Evaluating the Module-and-Seminar Model

5.1 Limitations and Weakness of the Model

- This model depends a great deal on the continuity of the project's coordinators. Although the group of student researchers can vary considerably from year to year, this flexibility depends in good part on the institutional and analytical memory of the coordinators.
- We found that at a 2:1 ratio of non-coordinators to coordinators (a total team size of 6) or less, the visiting linguist team functioned well. Above that limit, the ability of the coordinators to carry out research was compromised.
- This methodology requires significant material and human resources, including:
 - The ability to print modules in the field, which requires access to computers and printers, and the electricity to run them
 - Space in which to meet for seminars
 - Sufficient access to consultants for each team member to work for an average of 4 hours a day.
- As the project progressed, and linguistic description and analysis of the language became increasingly detailed, it became more difficult for new team members to quickly get up to speed and contribute to the project.

5.2 Successes

- The integration and coordination problems experienced in the first year of team-based fieldwork in the ILDP (2002) were mainly solved by the adoption of module-and-seminar model in 2003-2006.
- This model made it possible to bring in students for even a single field season and have them make significant contributions.
- The model encouraged the continuous, incremental production of documentation materials.
- The team-based framework appears to have been successful in stimulating a great deal of work on the language:
 - Three *tesis de licenciatura* (Huamancayo 2005, Sullón 2005, Bautista 2008), five MA theses (Brown 2004, Lai 2005, DeJong 2005, Anderson 2006, Hallett 2006), and two dissertations (Lai 2009, Anderson in prep.)
- This model has been successful in bringing students into Amazonian fieldwork. The majority of the students would probably never have carried out fieldwork in the Amazon Basin without the support the team-based project provided (including the two who are currently completing dissertations on Iquito).
- Because of the general descriptive base provided by the project, student theses and dissertations achieved considerable depth on narrower topics:
 - Lai (2009) presents a very detailed theoretically-informed study of the semantics and pragmatics of tense, aspect, and mood in Iquito.
 - Anderson (in prep) focus on a typologically unusual means for marking reality status in Iquito that presents interesting challenges for morphological typology.
- The integrated team-based model has been successful in creating a community of active Iquito researchers (e.g. four of us are currently collaborating on a paper on the Iquito irrealis construction, and on the full descriptive grammar of the language).
- As developed, the model appears to be transposable to other projects and contexts.

6 References

- Anderson, Cynthia. in prep. The Iquito Irrealis Construction. PhD dissertation, University of Texas at Austin.
- Anderson, Cynthia. 2006. Adverbs and Phrase Structure in Iquito. MA thesis, University of Texas at Austin.
- Bautista, Sisi. 2008. Algunos aspectos discursivos de la coherencia funcional en los textos narrativos iquito. Tesis de Licenciado, Universidad Nacional Mayor de San Marcos.
- Brown, Mark. 2004. Topics in Iquito syntax : word order, possession, and nominal discontinuity. MA thesis, University of Texas at Austin.
- de Jong, Lynda. 2005. The Iquito Clause: Simple and Multi-Verb Constructions. MA thesis, University of Texas at Austin.
- Hallett, Taryne. 2006. A paradigm of event modality: The Iquito continuum. MA thesis, University of Texas at Austin.
- Huamancayo, Edinson. 2005. La Formación de Palabras Mediante la Derivación en Iquito. Tesis de Licenciado, Universidad Nacional Mayor de San Marcos.
- Lai, I-wen. 2005. Fundamental Aspects of the Iquito Language. MA thesis, University of Texas at Austin.
- Lai, I-Wen. 2009. Time in the Iquito Language. PhD dissertation, University of Texas at Austin.
- Sullón, Karina. 2005. El acento y tono en la lengua Iquito. Tesis de Licenciado, Universidad Nacional Mayor de San Marcos.