5

Language Documentation & Conservation Special Publication No. 21 (October 2020)
Interdisciplinary Approaches to Language Documentation
ed. by Susan D. Penfield, pp. 72-112
http://nflrc.hawaii.edu/ldc/ http://hdl.handle.net/10125/24944

Endangered Language Documentation: The Challenges of Interdisciplinary Research in Ethnobiology

Jonathan Amith

Abstract

In 2004, three national institutes jointly published *Facilitating interdisciplinary research*, a report that set standards for evaluating the interdisciplinarity of cross-disciplinary collaborations. Although endangered language documentation (ELD) projects often assemble multidisciplinary teams, the 2004 criteria, today followed by the NSF, create such a high bar for interdisciplinarity that it is probably better to evaluate the cross-disciplinary impact of ELD projects through a different criterion: that of service vs. science. According to this perspective, the cross-disciplinary goal of ELD projects should be to decrease reliance on outside provisioning of services while increasing their contribution to the research goals of external disciplines. This article first suggests that ELD projects should actively promote and evaluate the use of project results across disciplines, beginning with greater attention to the archiving process and issues of discoverability and transparency of data. It then explores the potential for the cross-disciplinary impact of ELD ethnobiological research, which has often simply asked taxonomists to identify collected material to species, a service that only marginally benefits biological research agendas. To promote scientific collaboration across disciplines, ELD ethnobiological projects are best designed if they contribute methodologically, substantially, and theoretically to biological research. This article concludes with a description of such an effort.

1. Introduction

In 2004 the National Academy of Sciences, National Academy of Engineering, and Institute of Medicine published a report entitled *Facilitating interdisciplinary research*, a report that a decade later continues to have a profound effect on National Science Foundation efforts to promote collaboration among disciplines. Indeed, in its "Introduction to interdisciplinary research," the NSF reproduces verbatim the 2004 definition of Interdisciplinary Research (IDR):[33]

> Interdisciplinary research is a mode of research by teams or individuals that *integrates* information, data, techniques, tools, perspectives, concepts, and/or theories from two or more

33 The relevant NSF website is at http://www.nsf.gov/od/iia/additional_resources/interdisciplinary_research/index.jsp, accessed Oct. 10, 2014.

> disciplines or bodies of specialized knowledge to advance fundamental understanding or to solve problems whose solutions are beyond the scope of a single discipline or area of research practice (emphasis added).

The key word in the preceding definition is"integrates," as it is precisely this integration that is taken to distinguish interdisciplinary from multidisciplinary studies. Figure 1, also from the 2004 report, schematizes the difference between these two terms by graphically distinguishing the optimal directionality of "post-project" research: interdisciplinary research has the potential to forge new disciplinary ventures. Paradoxically then, in the best of circumstances this positive result effectively redistricts academic disciplines and moves the goalposts for future challenges to disciplinary boundaries.

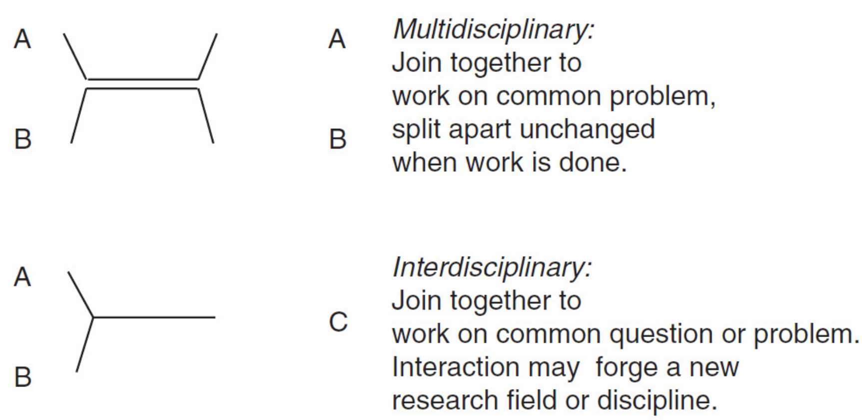

FIGURE 1.

The report also lauds the benefits of a problem-solving approach to research and the incentive system that this creates: "IDR works best when it responds to a problem or process that exceeds the reach of any single discipline or investigator" (2004: 53). It works less well in environments such as those that predominate in most academic institutions (though not necessarily university affiliated laboratories, which often operate in collaboration with industrial partners), where the goals are concerned less with collective problem solving and more with an individual's academic standing within a department, college,or university.

The National Science Foundation, which sponsored the set of presentations that form the basis of the papers included in this special issue, implicitly recognizes the challenges of promoting interdisciplinary collaboration within the constructs of academic criteria by crafting initiatives such as CREATIV (**Creative Research Awards for Transformative Interdisciplinary Ventures) and its successor, INSPIRE (**Integrated NSF Support Promoting Interdisciplinary Research and Education), in which program officers are empowered to consider and grant awards outside of the normal panel review process. The stated reason for this approach is to avoid the perceived inherent conservatism of panels, adverse to high-risk ventures and oriented to academic, discipline-based evaluations of NSF proposals. Rather, these two programs seek to encourage bold, interdisciplinary proposals that, in the politic words of the INSPIRE program solicitation, "some may consider to be at a disadvantage in a standard NSF review process."

In addition to the criteria of *integration*, however, multidisciplinary and interdisciplinary projects can also be distinguished by the relative degree to which project participants from different disciplines either provide technical *services* to a colleague or are motivated to participate by the project's relevance to their own *scientific* research agendas. This tension between the service and scientific roles of project participants from distinct disciplines is not uncommon. For example, ethnobiological projects (a research topic that is here explored in detail) rely on the taxonomic expertise of systematists who often gain little more than a voucher specimen of a common species in the "gift for determination" exchange through which herbaria and entomological collections incrementally build up their inventories and biologists can obtain specimens delivered from areas they have not visited. In these situations the taxonomist is at the hub of an exchange of services: he or she provides an identification service to the ethnobiologist while serving as a conduit for specimens that enhance the collection of the host institution and may provide the specialist with vouchers from deficiently covered regions. Description of a new species or even a register that represents a significant range extension to a known species may of course constitute publishable data.[34] But even in the case of species new to science, collected inadvertently by a social scientist in the field, the contribution to the career of the gift-receiving taxonomist is minimal. Indeed, the taxonomist will often prefer to wait to describe the new species until he or she is developing a monographic treatment of the genus to which the new species belongs, a treatment that may take years, or never be completed. The discovery and description of a new species is seldom as significant an event in the career of a systematist as it is in the romanticized public perception of scientific discovery in natural history. Thus the primary foundation of ethnobiological projects, particularly in light of the definitions presented in Figure 1, is more multidisciplinary than interdisciplinary.

A discrepancy between the degree to which a project pursues research goals central to one discipline while relying on the services of another not only affects the degree to which interdisciplinary integration is problematic and deficient but it also affects the degree to which commitment to project goals is equal among all participants and, in regards to funding, the degree to which multiple disciplinary programs will dedicate resources to the joint project.Viewed from a perspective that my own research in ethnobiology has led me to consider: it is easy to imagine that an anthropologist or linguist documenting an endangered language will ask taxonomists to provide determinations to the species level of local biotaxa. It is less easy to imagine a biologist approaching an endangered language (EL) researcher with a request to provide a service of botanical field collections to enhance the systematic study of a given taxon or clade, or even the coverage of a given region. Striking a more equal balance in an EL ethnobiological documentation project between the benefits to the documentarian and to the biologist is not a trivial task.

---

34 For example, a field expedition I organized to document Indigenous (Yoloxóchitl Mixtec) knowledge of Hymenoptera (basically ants, bees, velvet ants, and wasps) resulted in the collection of a species in an area well beyond its previously known range and a short publication reporting this event (González et al. 2014).

Efforts to achieve such a goal are described in the following sections culminating in the extensive collaborative project described in §4.3.

Before proceeding, however, it is important to note that endangered language documentation may also be at the "service" end of the spectrum in cross-disciplinary projects.[35] One service, to biologists, has already been mentioned. Through community-based collaboration, ethnobiological projects are able to extensively collect flora and fauna, often in areas that are poorly explored. Herbaria and museum collections may thus be built up at a relatively low cost and new geographical references and species are often discovered.[36] Traditional ecological knowledge, such as observations on habitat and behaviour and local uses of flora and fauna, constitute important additions to human knowledge of the biosphere.

The most common "service" of documentation, however, is one that relates back to Himmelman's (1998) early distinction between documentary and descriptive linguistics: the creation of a representative corpus of "the linguistic practices characteristic for a given speech community." "In this view," Himmelmann notes, "language documentation may be characterized as radically expanded text collection" and is "uncompromisingly data-driven" (1998:165). One basic question to be asked, then, is the following: If documentation is to be considered a data acquisition and provisioning activity, how easily can the data provided in documentation efforts be mined for descriptive and theoretical work in linguistics and related disciplines? For example, although natural speech corpora created by documentarians could provide a basis for a corpus-based approach to research, it is not clear how often either the developer of an EL corpus or linguists who pursue corpus-based studies have used EL corpora for statistically-based lexicosemantic or morphosyntactic research, such as attempting to discover frequency patterns (e.g., collocations, morphological co-occurrences) within the textual transcriptions.

The potential for archived EL corpora to inform theoretical work in linguistics is an issue that has been addressed in at least one project. An NSF grant to Douglas Whalen, in which I was invited to participate, explored whether archived EL documentation could yield an adequate phonetic description of the language in question. Among the questions asked was whether a forced alignment system developed for a major language could be bootstrapped for use on an endangered language, providing a close enough approximation to segmental boundaries that would, in turn, facilitate phonetic analysis. A second question was whether the phonetics of natural (not elicited) speech (the type of speech most extensively collected in documentation projects) could yield material that would correspond to an accurate phonetic description of the targeted language. Traditionally, such descriptions have relied on elicited, careful speech. The answers to these questions are provided in a series of articles (Christian DiCanio was the senior author in all; see references).

My experience in this and other projects that attempt to use EL corpora of "the linguistic practices characteristic for a given speech community" has made

35 Cross-disciplinary, or across disciplines, is used as a neutral term, implying neither multidisciplinarity nor interdisciplinarity as described in the paragraphs above.

36 Amith'sethnobotanical project in the Sierra Nororiental de Puebla has yielded over 125 new state plant registers and his collaboration on Cerambycidae with Steve Lingafelter (Amith&Lingafelter 2016) has documented 13 new state registers in Guerrero and Puebla. Ongoing work with Kevin Williams on Mutillidae has discovered 11 species new to science.

it clear that creating a fit between the EL data and natural language data-processing tools for descriptive or theoretical linguistic study is not a trivial matter. Deficiencies in field data collection are not easily remedied and often require labor-intensive rectification or enhancement of the original deposit. Recognizing the need to better ensure that archived EL materials be adequate for studies similar to his, Whalen suggested the need to establish a "definition of phonetic norms not only for endangered language documentation, but for documentation of any language."[37] The lack of such norms, covering initial documentation and final accession and archiving, can only adversely affect the ease of using primary documentation materials for future work in descriptive and theoretical linguistics.

This introduction has briefly presented two perspectives on cross-disciplinary efforts. The first focuses on the integration of disciplines and the potential for lasting impact on future collaborative activities as a distinguishing feature of interdisciplinarity. Implicitly, the research questions addressed in such ventures must be of significance not only to the individual participants but also to the disciplines that they represent. When the NSF establishes interdisciplinary funding initiatives, as they are increasingly doing, the officers of the different programs, despite increased discretionary powers, are in effect responsible for ensuring that the proposal presented is relevant, particularly theoretically relevant, to the discipline and program goals that they represent.

Second, lack of cross-disciplinary relevance can also be interpreted in relation to the science/service polarity mentioned above. When a collaborator provides a service to a cross-disciplinary project that is not particularly well integrated with that contributor's scientific research agenda but rather relies on his or her discipline-specific skills, it is unlikely to stimulate resource commitment beyond the principle discipline that drives the project. In addition, when a principle justification for EL documentation is the utility of the material for contemporary or future descriptive and theoretical work, it is fair to ask how best to gather, process, and archive the primary documentation materials so that they can fulfill this goal. This question needs to be posed, and the challenges of using the results of EL documentation projects for work in theoretical and descriptive linguistics, particularly in projects that utilize natural language processing of corpus material, need to be addressed.

Nevertheless, despite a need to discuss the efficiency of EL documentation as a data/service provider for theoretically-based linguistic research, this essay will be limited to an exploration of cross-disciplinary ethnobiological research, research that represents what is basically the inverse situation in which EL documentation requires, at a minimum, significant service provisioning by biologists (although, as noted above, the ethnobiologist also provides a service in the voucher sent as "gift for determination" and the traditional ecological knowledge that may accompany the collection and label). Yet for a sustainable long-term effort, the heavy service needs of the documentarian should be complemented by either integrating the social science research agenda with that of biologists or offering greater services, such as more extensive collections, to these same biologists. The present essay explores several paths to a more integrated and mutually beneficial relation between language documentation and biology: a focus on biosemantics

37 Abstract of proposal 0966411, "From Endangered Language Documentation to Phonetic Documentation" at http://www.nsf.gov/awardsearch/showAward?AWD_ID=0966411&HistoricalAwards=false, accessed Oct. 22, 2014.

and classification accompanied both by extensive biological fieldwork in particular families of organisms, both flora and fauna, of interest to collaborating biologists (§4.2), and by molecular analysis of local inventories of biotaxa that will simplify and accelerate identification of vouchers to species and will create a permanent material resource of use across many disciplines (§4.3).

The following section (§2) explores the complexity of Indigenous nomenclature and classification of local flora and fauna. This complexity reflects a wide range of factors, three of which are explored in §2: (a) intricate patterns of correspondence between Indigenous and Western classificatory schemes; (b) intracommunity and speaker-specific variation, particularly in the criteria of category membership; and (c) complicated lexicosemantic relations among Indigenous terms for biotaxa.

This complexity, I suggest, presents unique challenges to cognitively focused ethnobiological research, research primarily concerned with the naming and classification of the natural environment. The argument presented below is that these challenges can best be met by a mutually beneficial collaboration between linguists or anthropologists working on language documentation and biologists interested in a thorough documentation of local flora and fauna. Language documentation, particularly the compilation and analysis of a corpus of conversations about local flora and fauna (as well as related topics in material culture, food, and medicine) offers a seldom exploited tool for understanding how native speakers communicate about the natural environment, often despite significant differences in nomenclature and classification of biotaxa among conversation participants. Extensive collection of flora and fauna is both necessary to fully understand Indigenous nomenclature and classification and of potential interest to taxonomists and systematists. §3 examines the potential role of language documentation, along with an emerging strategy to engage native communities in documenting their own language and culture, in ethnobiological research. Finally, §4 explores the potential for mutually beneficial collaboration between those engaged in language documentation projects and a wide range of biologists, including molecular biologists.

2. Classification

Ethnobiological projects most often target a single community (although the local settlement pattern may be disperse), a strategy that reflects both logistical and research considerations. Self-contained multisited projects are rare (but cf. Turner 1973 and, particularly, Martin 1996, for outstanding comparative studies) and most comparative analyses of traditional ecological knowledge rely on accessing the results of several single-sited efforts developed by various researchers over a long period of time. Even when limited to a single community, however, documentation projects that develop the Boasian core of resources—corpus, lexicon, and grammar—are particularly challenged by biosemantics: the nomenclature and classification of local flora and fauna. At a very basic level, difficulties emerge in determining the lexicosemantics of words and combinations of words that reference the domain of natural history. For example, even when the native term has a single prototypical referent in a localized ecosystem, the meaning of this term is often easily expanded (and by different speakers in different patterns) to include other referents, as occurs when a speaker encounters related species from ecosystems different from those of his or her home environment. Moreover, difficulties in translating or glossing Indigenous terms to a Western language are compounded if one considers that biosemantic terminology often references a category or classificatory scheme and not a single referent. Viewed as a classificatory process, ethnographic research on biosemantics quickly shifts from determining the referent of a given term to understandinga divergent set of native opinions about the internal structure and often indeterminate boundary limits of categorical divisions of the natural environment. Time and time again speakers may differ not so much (or not only) in their identification of what might be considered the prototypical referent of a given term but rather in their varied cognitive organization of the environment: categorical strategies for structuring, limiting, and expanding the set of denotata covered by a given term.

Each of the following three subsections addresses factors that contribute to the complexity of describing (in effect, translating) native terms for biotaxa: (§2.1) mismatches between Indigenous and Western taxonomic categories; (§2.2) speaker-specific variation; and (§2.2) complex lexicosemantic relations among terms ostensibly within a single classificatory group.

2.1 The relationship of Indigenous and Western classificatory schemes The relationship of a given native term to a subset of the natural environment may vary greatly. One of the simplest relations is a single native lemma, well-known throughout the community, that references a cohesive unit, be it (1) a unique terminal taxon or (2) a high-level node in what, viewed from a Western scientific perspective, is a monophyletic taxonomic structure. Examples of the latter that I have encountered include words for 'earwigs' (the order Dermaptera) and the parasitic plant known as 'dodders' (genus *Cuscuta*, now in the Convolvulaceae family). The Balsas Nahuatl term *mēmēya* (from the verb *mēya* 'to issue or flow forth [a liquid]') refers to the latex that characterizes one of four subgenera now included in the genus *Euphorbia* (family Euphorbiaceae) but previously grouped into a separate genus, *Chamaesyce* (commonly called 'sandmats').[38] In cases in which the node referenced by an Indigenous term is designated by a lexical item in Western nomenclature, the definition of the Indigenous term is *almost* a matter of simple equivalence: *sakapahli* = genus *Cuscuta* spp., or *mēmēya* = subgenus *Chamaesyce*. Strictly speaking, the native term may be *translated* as the name for a node in Western taxonomy only if, like this scientific name, it covers all locally present lower-level descendants from the designated node and does not extend to taxa below a different taxonomic node. That is, it must be locally monophyletic as viewed from a Western perspective. In all three preceding cases (earwigs, dodders, and sandmats) this appears to be the case. The morphology of the referents is relatively unique, salient, and easily delimited, although in these examples each Indigenous term corresponds to a different level within Western scientific taxonomy (order, genus, and subgenus).[39]Such direct correspondences between taxonomy

38 For the phylogeny of Chamaesyce see Yang et al. (2012).

39 For a summary representation of the nature of the correspondence between a native term and Western taxonomic categories and general classes (e.g., by sex, developmental stage) see Ellen (1993:69), Table 1. Nevertheless while the native terminal category may intersect Western taxonomy at a range of nodes, this does not mean that the relation is monophyletic (includes all regionally extant descendents of that node). To the extent that an Indigenous term does not reference a Western monophylectic category, a simple translation of an Indigenous term to a Western taxonomic node is inaccurate. In such cases the nature of classificatory overlap is a topic for empirical research on category structure and boundaries.

and nomenclature in two cultural systems are, however, rare. More common are imperfect equations. Some of the most frequent categorical relations between cultures are described below. All of these represent problems of simple L1= L2 translation.

One common classificatory pattern is the 'extension'[40] of a term denoting a fairly cohesive set of organisms to others that share morphological, behavioral, or even functional similarities with the prototypical group.Anexample of the apparent extension of a category that is generally well recognized and firmly bounded occurred with velvet ants (family Mutillidae) in the Balsas valley (Nahuatl). Over 150 collections of Mutillildae have been consistently named with a single term, ītskwin tiōpixki (lit. 'the priest's dog'). Most collections were of several species in the genus *Pseudometheca*, all of which were small and greyish black. The prototypical referent, however, seems to be the more strikingly marked large black-and-orange mutillids in the genus *Dasymutilla*. Once, however, a very knowledgeable Nahuatl-speaking consultant found a spider wasp (*Psorthaspis formosa,* family Pompilidae) that she categorized as ītskwin tiōpixki along with the dozens of Mutillidae that we had already collected together. At the time she commented on the fact that she had never seen an ītskwin tiōpixki with wings, a comment that suggests that a common (though not necessary and sufficient) characteristic of ītskwin tiōpixki is winglessness. This comment makes sense given that the prototypical referents of ītskwin tiōpixki are velvet ants, a family of insects in which females are diurnal, terrestrial, and wingless, while males are winged and nocturnal (and thus seldom seen or noticed). The Pompilidae collected was a non-prototypical ītskwin tiōpixkiin having wings, yet similar enough in body form and marking to merit inclusion in this category. Its designation as an ītskwin tiōpixki, however, was probably ad hoc, a spot decision based on general morphological features despite the noted unusualness of the wings. No Nahuatl-speaking consultant ever mentioned the possibility of a winged ītskwin tiōpixki and most Pompilidae seen were large and metallic black or blue. They were generally considered a type of wasp.[41]

Another example involves an Indigenous category that is polyphyletic from a Western perspective. I have often found that the nomenclature and classification of both mantids (family Mantidae) and stick insects (order Phasmatodea) is as straightforward as the examples given in the previous paragraph: an Indigenous term corresponds to a named node in Western taxonomy and all descendants readily found in the local ecosystem.[42] In the Mixtec community of Yoloxóchitl, however, one term ($ko^{1}li^{4}li^{4}$) covers insects from both the Mantidae family and the Phasmatodea order.[43] But twice during collection trips an assassin bug (family Reduviidae) of the subfamily Emesinae ('thread-legged bugs') was encountered. Both times all consultants present designated them as $ko^{1}li^{4}li^{4}$, a term otherwise exclusively reserved for the jointly categorized mantids and stick insects.

40 Extended ranges of terms are noted in many of the classic works by Berlin, Hunn, Breedlove, and Laughlin, and others.

41 In Yoloxóchitl, however, several male Mutillidae were caught and consultants invariably identified them by the Mixtec name for velvet ants: ndi3ka'3a3 (lit., 'panther', Pantheraoncahernandesii) followed by nda3yu4 ('soup') or ndu3chi4 ('beans') as depending on their markings mutillids are said to be omens of what one will find to eat upon returning home: brightly colored mutillids foretell soup, darkly colored mutillids foretell beans. The association of mutillids with omens seems to be a common Mesoamerican belief (e.g., it also occurs in Totonac [David Beck, personal communication], Triqui, and Mazatec).

42 In Yoloxóchitl Mixtec one term covers insects from the Mantidae family and the Phasmatodea order. Given that all mantises collected to date have been in the family it is not certain whether other insects within the order Mantodea would also be classified by the same name, though this probably would be the case. At any rate, viewed together the Mantodea and Phasmatodea orders are not monophylectic; they do not represent all descendents from a common higher-level node. In Balsas Nahuatl the two orders are distinctly named.

43 Some speakers pronounce this as ko1ko1li4li4, or ko14li4li4; tones in Yoloxóchitl Mixtec range from low (1) to high (4), with additional rising, falling, and complex tones.

In the preceding two cases there seems to be no evidence of mimicry.[44] The inclusion in a group corresponding to a recognized taxonomic node of species from an unrelated group of insects (in the two examples, Pompilidae and Emesinae) is based on superficial morphological similarity not related to protective mechanisms, such as Batesian and, less commonly, Müllerian mimicry. As expected, however, the inclusion of mimicsas extensions to unrelated groups is not uncommon in native classificatory systems. Thus the Yoloxóchitl Mixtec name *yo³ko² lu³tu³* (lit., 'wasp narrow') is prototypically *Parachartergus apicalis*, a swarming, aggressive wasp with white-tipped wings.[45]According to several consultants, the Mixtec name refers to the narrowing of the eyes of a person who has been attacked by these insects, though it is perhaps more likely that the name refers to the tapered form of their nests.[46] Yet *Mischocyttarus deceptus*, also with a white-tipped wing, is also invariably labeled a *yo³ko² lu³tu³*. As *P. apicalis* is aggressive and *M. deceptus* is not, it is clear from the behavioral description that the former is the prototypical referent, as this species manifests both the morphological *and* behavioral characteristics that speakers associate with wasps identified by the term *yo³ko² lu³tu³*.[47] A similar case is that of another aggressive wasp, *yo³ko² ndia'¹⁴na³*(lit., 'mask wasp'), which has been used by native Yoloxóchitl Mixtec speakers to designate three species: *Synoeca septentrionalis*, *Montezumia mexicana*, and *Montezumia azteca.* The name refers to the nest of the first, *S.septentrionalis*, a "mask-like" flattish structure pasted onto tree trunks. The two *Montezumia* are superficially similar to *S. septentrionalis* though they are non-aggressive and make nests that do not resemble masks.

In these two cases (involving *Parachartergus apicalis* and *Synoeca septentrionalls*) the expected insect behavior or meaning of the name indicates which of the collected species is the prototype. In another instance, still pending further study, a

---

44 The effect of mimicry on the relationship of Indigenous and Western scientific categories is further discussed in Amith & Lingafelter (2016), where it is noted that Batesian mimicry of certain Cerambycidae creates a situation in which the Indigenous term references a category that from a Western scientific perspective is paraphyletic.

45 On the aggressivity of P. apicalis, see O'Donnell &Hunt (2013), one of two swarm-founding wasps discussed, including another of the Agelaia genus. Mimicry is well described in another article (O'Donnell &Joyce (1999: 502)):

> Swarm-founding Neotropical Vespidae (tribe Epiponini) often have large, aggressively defended colonies (hundreds to ten thousands of adults), and are often mimicked by other eusocial wasps. In contrast, the largely Neotropical genus Mischocyttarusis characterized by species with independently founded small colonies (several dozen adults) of relatively non-aggressive wasps. Many Mischocyttarusspecies apparently mimic other vespid wasps (Richards 1978). At least some Mischocyttarusspecies wasps are capable of stinging humans. However, Mischocyttaruswasps are often reluctant to sting humans even in nest defense, and individuals of most species we have observed remain immobile or even flee their nest when disturbed (S.O'D[onnell]. and F. J. J[oyce]., pers. obs.). Because they possess stings but are relatively docile, Mischocyttarus species may mimic other Vespidae through a combination of Batesian and Müllerian processes.

The docile nature of Mischocyttarus is captured by the name for M. rufidens and, less often, M. mexicanusin Oapan (Nahuatl): īxtēmpā´ya, literally 'fuzzy eyed' in reference to the fact that one can position one's hand close to their nest without suffering attack, allegedly due to their poor sight. In Yoloxóchitl one consultant also described a yo3ko2 ma'3a4, literally 'wasp docile' (the word ma'3a4 is also used to mean 'cuckold') characterized by the fact that "one can take its nest a way without being stung." This wasp has not been collected but the described behavior suggests a Mischocyttarus. Finally, James Carpenter (personalcommunication) pointed to the original description of Polybiadecepta (syn. M. deceptus) and noted: "There are several species of Mischocyttarusthat mimic species in the Parachartergusapicalis species group, including deceptus and socialis; the other species is imitator. Fox did not give an etymology in his description of Polybiadecepta [note: Synonym of M. deceptus] but did remark on the similarity in coloration to Parachartergusapicalis. " Indeed, in the original article (Fox 1895:269–70), in describing the then new species P. decepta, William J. Fox notes: "Its similarity in color to Chartergusapicalis [Note: now a synonym of Paracha rtergusapicalis] is really remarkable."

46 For the nest of P. apicalis, see Dejean (1998)

47 Rosch & Mervis (1975:575) definite prototypicality as follows: "members of a category come to be viewed as prototypical of the category as a whole in proportion to the extent to which they bear a family resemblance to (have attributes which overlap those of) other members of the category."

ti^{1}*mi*^{3}*i*^{4} (derived from *ti*^{1}- ‘prefix for animals’ and *mi*^{3}*i*^{4} ‘solitary’) was frequently characterized as a highly aggressive solitary bee. Yet the two species collected with this name (*Eufrieseamexicana* and *Eulaema polychroma*, both Euglossini [orchid] bees) are not particularly aggressive. This suggests that another bee still to be collected might be the prototype, combining morphological and behavioral characteristics that match the consultant descriptions.

As suggested by the *ti*^{1}*mi*^{3}*i*^{4}case, mismatches between morphology and behavior often indicate definitional complexity and problematic field identification by consultants. Such mismatches should motivate additional fieldwork. In the Nahuatl-speaking community of Oapan, a revealing mismatch occurred with an insect called *yēlōtlapṓhwikātsīn*, literally ‘the one that incenses green ears of maize’. For some time I had collected several *yēlōtlapṓhwikātsīn*, all various nondescript small Diptera, which didn’t make much sense in terms of expected behavior. The *yēlōtlapṓhwikātsīn* was described as a small insect that during the month of September is found hovering around green ears of maize (*yēlōtl*), as if incensing the plant. (Indeed, at the same time of year, *campesinos* go to their fields to leave an offering, incensing the field as they do.) One time, however, I collected a hoverfly (Syrphidae) also said to be a *yēlōtlapṓhwikātsīn.* It was *Toxomerus politus*, a monophagous insect that feeds only on the pollen of *Zea mays* and thus is often found hovering around immature maize as it develops. This was clearly the prototypical referent of the Nahuatl term. The previously collected small Diptera were the result of confusions, confusions that were perhaps induced by a pestering ethnobiologist asking questions in the wrong context.

The preceding discussion was presented not only to demonstrate the complexity of determining Indigenous biosemantics but to problematize the taxonomic treatment of native terms for local flora and fauna. It may be, of course, that the complexities of category inclusion, structure, and delimitation are of relatively minor interest for a given study. Translations to a single term are simplifications, and caveat expressions to cover both polyphyletic and paraphyletic exceptions to one-to-one correspondence between Indigenous and Western scientific taxa beg certain questions. Likewise, intensional (necessary and sufficient conditions) definitions are misrepresentative while extensional (list of category members) definitions are interminable and uninformative. Greater promise is presented by ostensive definitions, a scattering of illustrative examples, though these would best be accompanied by varied (but not “necessary and sufficient”) criteria by which category membership is evaluated (such as the long antennae and tree girdling behavior of Cerambycidae) and an encyclopedic discussion of the relevant denotata.[48] In all these cases, however, an extensive understanding of local flora and fauna is necessary, an understanding that can only be achieved through the collaborative support of many biologists to which must be added the sensitivity of the ethnobiologist to the details of lexicosemantic description.

48 An excellent example of such a discussion is in the Kalam-English dictionary (not consulted for this article), an entry from which is given in Pawley (2001:237–8). Cf. also Ellen (1993:154–7) for the Nuaulu term kauke. Si (2011:178) mentions the value of ostensive definitions for natural history terms, including biotaxa.

2.2 Speaker-specific and intracommunity variation in nomenclature and classification For many reasons, knowledge of the nomenclature, classification, and use of local flora and fauna is unequally distributed within native communities. Obviously there is a learning curve to traditional ecological knowledge: as with all types of information, children learn as they grow. Natural history knowledge may also be unequally distributed by sex, by the activities in which different social groups engage (many of which are tied to gender), and by speaker-specific aptitudes for natural history. Variations along these criteria are common and are occasionally reflected in lexicographic treatments, such as when specific head words are marked by register or as specialized terminology.

Another source of speaker-specific variation relates to the mechanisms by which consultants apply biosemantic nomenclature beyond a prototypical referent. Such extensions often lead to divergent classificatory schemes within a community. A clear example from my fieldwork involves two spiny *Solanum* plants classified differently by three consultants from San Miguel Tzinacapan (altitude 865 meters). The first *Solanum* is *S. rudepannum* Dunal, a medicinal plant called *itsk^{w}inpahwits* (lit., 'dog- medicine thorn(y)') that is common and extremely well known in mid- and low-altitude villages (under 1000 meters). One day, however, I and three consultants all went to a highland village where we came across, at 1485 meters, *Solanum chrysotrichum* Schltdl., a species superficially similar to *S. rudepannum*. Each consultant responded differently when asked the name of this second *Solanum*, which is found mostly at altitudes greater than 1000 meters and which is thus absent from the lands of their home village.

One stated that he didn't recognize the plant; another gave its name as *itsk^{w}inpahwits*, while recognizing that it was morphologically distinct from the *itsk^{w}inpahwits* found near Tzinacapan. A third consultant named it *itsk^{w}inpahwitsitahtāy* (*'itsk^{w}inpahwits* 'its look-alike').[49] Although all speakers probably had never seen *S. chrysotrichum*previously, they recognized it as a plant close but not identical to *S. rudepannum*, which they all knew well. The first consultant had a general tendency to limit referents to the prototype (he demonstrated the same naming strategy for *tālāmat*, prototypically the medicinal *Desmodium caripense* Kunth; he was reluctant to call other short, low-lying *Desmodium*, which most speakers do consider *tālāmat*, by this same name). The third consultant generally tried to extend categories as widely as possible, a classificatory strategy that was frequently commented on by other, more conservative consultants, some of whom nicknamed him *'itahtāy'* ('the look-alike'). Finally, the classificatory scheme of the second consultant seems to indicate a distinct categorical structure: for him the nature of the *itsk^{w}inpahwits*name/category was potentially polytypic. This flexibility in categorization was distinct from the categorical structure of the other consultants who excluded the highland plant from the *itsk^{w}inpahwits*category either definitively (first case) or lexically, establishing a boundary by using the term*itahtāy*.

Functional considerations of flora and fauna may also vary not only by speaker but also by context of discussion or elicitation. Nomenclature is affected by a continual shift in many endangered language communities from local sources of material culture and medicine to commercial substitutes. As this occurs, many of the utilitarian benefits of plants become remembrances of things past, although functionally descriptive names may

49 The utilization of a term effectively meaning 'look-alike' is not uncommon in Mesoamerican biosemantic terminology, e.g., Yoloxóchitl Mixtec has the word ta1ni1 and Balsas Nahuatl itlahtlāk (and cognate forms).

survive frozen in nomenclature even though the utility is now absent from daily activity. Thus in Yoloxóchitl only one elderly consultant knew the Mixtec name for *Bletiacoccinea* Llave et Lex.: *i³ta² nda¹ka¹*, which he translated as 'flower glue', a reference to a use of this plant that dates to prehispanic times. The three other speakers present not only did not know this name, but they had never heard the word *nda¹ka¹*, an archaic term for 'glue', which is now referenced by a Spanish loan.

Speakers may also deny a certain functionality in an interview situation (e.g., the desirability of a given wood for fenceposts), a denial that is belied by actual use. The discrepancy between discourse and actual use manifests the problematic results that normative elicitation statements may create. Thus, when asked, Oapan Nahuatl speakers (Balsas Valley) did not name *chikomolin* (or *chikimolin*; *Leucaena matudae* (S. Zarate) C. Hughes) as particularly good for fence or house posts. Nevertheless, a non-quantitative assessment suggests that this species is indeed commonly so used. In this case, the disjunction between normative descriptions of utility and actual use reflects the geographical distribution and relative scarcity of preferred species (e.g., *Comocladia* spp.): *Leucaena matudae* is abundant, particularly in the lower altitudes near the river valley village.

The preceding examples briefly illustrate some patterns of local variation in nomenclature and classification: speaker-specific interpretations of category membership, normative statements of utility, and the implications of antiquated uses frozen in nomenclature and, at times, the imagination. Such variations are facilitated by the nature of floristic and faunal categories in Indigenous communities: referents are classified together even though they do not share one or more necessary and sufficient conditions for category membership. Rather it is a series of overlapping attributes, family resemblances in Wittgenstein's original formulation, that establish links among the referents of a category term. And it is precisely this feature of categorization that has motivated the approach taken in my research: the most interesting lexicosemantic and cognitive problems are found in those categories of the local floristic and faunal environment that have the most potential named members. Moreover, the clearest way to understand the characteristics at the foundation of the family resemblances is natural discourse, particularly recorded and transcribed conversations that can be electronically searched and analyzed (a basic goal of language documentation), accompanied by an exhaustive inventory of species present (a basic goal of floristic studies).

2.3 The lexicosemantics of biotaxa A final consideration in the lexicographic treatment of biosemantic categorization is the challenge of polysemy and the hierarchical nature of meaning, specifically in regards to hypernymy and hyponymy, within the semantic domain of flora and fauna. The consideration of these types of semantic relations is particularly important given that Paul Taylor, one of the few ethnobiologists who has stressed a linguistic approach to biosemantics, has suggested (1990:46–47) that co-hyponomy is one strategy for discovering covert categories:

> The method of co-hyponymy consists essentially of identifying a set of terms that can be shown to directly contrast in at least one of their senses, but which have no superordinate term to label the entire set. Having posited a FLORAL FORM domain by this method, we still have not resolved

> the problem of the boundaries of the domain, although it must minimally include the full range of the three subordinate terms [Note: in this case 'tree', 'vine' and 'herbaceous weed'] on whose basis the FLORAL FORM class was posited. To more directly establish the boundary of the FLORAL FORM domain, we may turn to the method of "definitional implication."

In practice, however, the discovery of co-hyponomy is often problematic. One example illustrates this point. Balsas Valley Nahuatl has a series of terms that refer to distinct species of Formicidae. A list of prototypical referents of the basic terms is given in Table 1.[50] Many names are extremely well known, even to children.[51] The hierarchical organization of the Formicidae domain is, however, neither obvious nor consistent across speakers.

TABLE 1. Formicidae in Oapan Nahuatl

Oapan Nahuatl term	Meaning	Prototypical referent
āskatl chīchīltik	āskatl '?small ant' *chīchīltik* 'red' 'red small ant' *yo:n* k^{w}*extikeh* 'the very small ones' *yo:n we:imeh* 'the large ones'[52]	*Solenopsis xyloni* *Solenopsis geminata*
k^{w}*itlayā́k* *molōnki*	k^{w}*itla-* 'excrement' *(i)yāk* 'fetid' *molōnki* 'smelly' 'foul-smelling ant that smells bad'	*Forelius damiani*
k^{w}*itlayā́k* *xmolōnki*	*x-* negation 'foul-smelling ant that doesn't smell bad'	*Paratrechina longicornis* (and probably *Forelius pruinosus*)
Tsontetl	(cf. *kowtsontetl* *kow* 'tree' + *tsontetl* '? immobile' *kowtsontetl* 'stump (of a tree)' *yo:n* k^{w}*itlayoh* 'the one with visible fungus dumps' (literally, 'the excrementy one') *yo:n xk*^{w}*itayoh* 'the one without visible fungus dumps'	*Atta mexicana*

50 A more detailed discussion of other possible references of the Nahuatl terms is beyond the scope of this article.

51 There are two primary schools in Oapan and the uniforms children wear to each are distinct: red-checkered shirts for one and black-checkered shirts for the other. From the first grade, children learn the nicknames for students from each school: tēkwāntsīkatl ('red harvester ant') for the red-shirted students and kwitlayā´k for the black-shirted students.

52 Probably the size refers to different castes within a species, not different species.

māwēweyak	*mā-* 'arm' or 'hand' *wēweyak* 'long (pl)' 'long-armed'	*Aphaenogaster ensifera*
Panochēroh	< *panocha* (Spanish) 'hardened brown sugar cake' + Spanish agentive (-ero)	*Camponotusatriceps* (winged female)
Yewaltsīkatl	*yewal* 'night' *tsīkatl* 'ant'	*Camponotusatriceps* (worker)
kōlōtsīkatl	*kōlō-* 'scorpion' *tsīkatl* 'ant' 'scorpion ant' (probably for its painful sting)	*Pseudomyrmexmajor* (and perhaps *Pseudomyr-mexgracilis*)
Kowtsīkatl	*kow-* 'tree' or 'wood' *tsīkatl'* 'ant' 'tree ant'	*Camponotus rubroniger*
*tēk*ʷāntsīkatl	*tēkwān(i)* 'one that bites' *tsīkatl* 'ant' 'biting ant'	*Pogonomyrmexbarbatus*

In English, relations of hypernymy and hyponymy among series of biosemantic terms are often widely shared among speakers. Thus *maple*, *oak*, *pine* are co-hyponyms of the hypernym *tree*; *dog, cat, seal* are co-hyponyms of the hypernym *mammal*; *fire ant, carpenter ant, red harvester ant, crazy ant* are co-hyponyms of the hypernym *ant*. In many languages (and even in English to some degree), however,these relations of hyponymy/hypernymy are more nuanced.

For example, in many Indigenous languages an ostensible hypernym is often a generic co-hyponym (residual term) of more specific terms, a relationship that is discussed in more detail in §3. Thus n*tsīkatl* (loosely translatable as 'ant') is not inarguably the hypernym of the Oapan Nahuatl terms listed in Table 1. A more accurate translation might be as a residual category of unnamed and relatively uncommon ants, though even this translation must be contextualized.[53] Speakers seem to find a question such as "Is a *k*ʷítlayāk a type of *tsīkatl*?" somewhat perplexing. The speakers that were asked such questions, however, seemed most inclined to categorize as *tsīkatl*ants with compound names that include the lemma *tsīkatl* (e.g., *kowtsīkatl*) and less inclined to include those ants (e.g., *māwēweyak*, *tsontetl*) whose compound names do not include this lemma. No speaker, however, was willing to consider an āskatl (*Solenopsis* spp.) a type of*tsīkatl*. Thus there is a potential contrast set, āskatl ~ *tsīkatl*, with the former limited to *Solenopsis* spp. and the latter term representing a category, with a complex internal structure and unclear boundaries, of progressively more inclusive relations in different discourse contexts. An ethnoentomologist could certainly create a situation in which a native speaker would group *Solenopsis* spp. and *Camponotus* spp. together (e.g., triad card-sorting) but the key criteria and justification for asserting the cultural relevance and cognitive saliency of such as association should be taken from a less artificial environment.

53 Residual categories are mentioned frequently in the literature (Hays 1979: 257; Hunn 1976:511 ff., 1977: 281 ff. and passim., 1982; Taylor 1990:64–65). Hunn (1976) cites Fowler & Leland (1967) who also look at residual categories (though they do not use this term) in Northern Paiute ethnobotanical classification. A further discussion of residual categories is in §3 below.

It is also possible that in learning the nomenclature and classification of Formicidae, children begin by using the residual term of knowledgeable adults (*tsīkatl*) and gradually carve out portions of the domain in learning the terms listed in the first column of Table 1, though my impression is that the term āskatl is learned at a very young age. Unfortunately I am not aware of any research in native speaker communities on the development of nomenclature and classification of flora and fauna in early childhood, though different biosemantic domains are undoubtedly learned in different orders.[54] An American child would probably become familiar with 'bird' (equivalent to a class term, *Aves*, or subclass term Neornithes) before the lower level term 'robin'; with 'ant' (equivalent to a family term, Formicidae) before 'leaf-cutter ant', but with 'dog' (equivalent to a subspecies term *Canis lupus familiaris*) *before* any of the higher taxa. That is, the learning of classificatory relations can begin from the higher or lower taxa and the order of learning might well affect the nature of categorical structures as well as reflecting the degree of human involvement and "cultural saliency."

The second linguistic tool for determining cognitive categories is what Taylor (1990:47) calls "definitional implication, through which terms applying to animals or plants are used to derive covert classes such as the implied class of subjects of verbs such as '"tweet' or 'chirp'or those animals that possess a 'beak'." Taylor suggests that covert classes can be posited when a term relating to a plant or animal cannot be defined without reference to this covert category and "alternative definitions cannot suffice to define the term in question. It is insufficient to argue that, because terms like 'leaf' or 'wing' apply only to plants or animals, they presume the existence of a PLANT or ANIMAL class" (1990: 49). Taylor was not the first to utilize a linguistic approach to category definition. Berlin in his early work noted the value of numeral classifiers in setting the limits of semantic domain. He was referring to what Grinevald (2000) has called measural classifiers (such as 'herd') in English. Berlin & Kimball (1964) give examples of such classifiers and at one point present a list of nouns that occur with either of two such terms that communicate 'aggregations of globular-shaped objects' (*b'uhs* indicating horizontal extension 'spread out' and *t'ol* ' indicating vertical extension, 'piled up'). Objects that can be so arranged include corn kernels, coffee beans, peanuts, chili peppers, stones, pieces of corn dough, and eggs, among other items. Again, the list is not properly considered taxonomic (although the class may have an internal structure of core and periphery) but is rather an ad hoc collection of referents that do not correspond to a class that exists apart from the linguistic criteria that delimit it.

Definitional implication, however, is hard to operationalize. The first difficulty is conceptualizing the formal structure of a definition that references a covert (unnamed) category. For example, the online OED gives the first definition of 'beak' as: "The horny termination of the jaws of a bird, consisting of two pointed mandibles adapted for piercing and for taking firm hold: a bird's bill." If we pretend that English were to lack a lexical item (i.e., 'bird') covering the avefauna, it is not easy to imagine a definition that would go beyond an ostensive definition employing a set of illustrative examples: "The horny termination of the jaws of robins, sparrows, wrens, hawks, penguins, and similar animals, consisting of two pointed mandibles adapted for piercing and for taking firm hold." Evenif

54 In Balsas Nahuatl the first term children learn is the "baby-speak" wīwih (from wiyōni, intrans. v., 'to wiggle') loosely translatable as 'bug' or 'creepy crawler'.

an extensional definition is used, the list of members of the category of "beak-possessing" animals would include tortoises and turtles as well as birds.

An important caveat of definitional implication and subject/possessor associa tion with a given term is that the semantics of most terms results from the context of utterance, both associated words and social situation, a dependency of meaning precisely expressed by Firth's famous admonition: "You shall know a word by the company it keeps." (1957:11; cf. also Fillmore's perspective in his frame semantics). Thus 'bird' in collocation with 'chirp' has a different set of referents as compared to 'bird' in collocation with 'feathers'. The latter covers all avefauna while the former covers only a subset, a subset that is not lexically expressed in any commonly used term except, perhaps, songbird. Thus an inductive approach to categorization from a corpus of non-generic feathered subjects of 'sing' would probably yield something close to 'songbird' (suborder Passeri) although the boundaries and internal structure of this category (feathered agents of the activity 'sing') would undoubtedly vary across speakers. Such a category definition could be approximated with an extensive corpus theoretically leading to, if not an inductive definition, then at least an ostensive one.

This contextual limitation to Taylor's example—"bird" is "the implied class of subjects of verbs like tweet or chirp (compare hoot and its implied subject owl; it is also likely to be found in a definition of beak, (to) perch, or feather" (1990: 47)—should thus be regarded with caution. With the exception of 'feather', the possessors or subjects of 'tweet', 'chirp', 'hoot', '(to) perch', and 'beak' do not constitute a class coterminous with 'bird' or, in the case of 'hoot', 'owl'. 'Hoot' is commonly associated with 'owl' but not all owls hoot. Thus the voice of a crested owl (*Lophostrix cristata stricklandi*) is described as: "A deep, throaty, slightly frog-like, emphatic growl, *ohrrrr* or *gurrrr*, at close range a rapid, stuttering introduction may be audible *g'g'g'g'grrrr*, repeated every 5–10 s" (Howell &Webb 1995:359). Not all birds (e.g., buzzards, eagles) 'tweet' or 'chirp'. Indeed, the Oxford English Dictionary (online) defines the intransitive verb 'chirp' as follows: "To utter the short sharp thin sound proper to *some small birds and certain insects*" (emphasis added). The phrase "some small birds" clearly indicates that chirping in an activity limited to a subset of birds.[55] The "certain insects" would seem to be limited to those in the Orthoptera order and Cicadidae family. Indeed three terms (including two mentioned by Taylor)—'feather', 'perch', 'sing'—intersect the natural worlds at three different nodes: aves (class); Passeriformes (order); Passeri (suborder), respectively.[56]

It should be clear that in most of the above cases the definiendum does indeed point to a covert category but usually not to a category that is in any way coterminous with well-known biotaxa or named objects and categories of daily import. In some cases ('feather' [avifauna], 'body hair' or 'milk' [mammals]), a single term might indeed be a necessary and sufficient condition for category delimitation. But in most cases the categories established by the means presented in the previous paragraphs (e.g., the potential subjects of 'chirp', the possessors of 'beaks', the objects that can be arranged in a manner indicated by *b'uhs*, the owls that 'hoot') reflect cognitive categories that, at least from

55 In a similar manner the set of referents of Spanish pájarois not equivalent to that of 'bird', a frequent translation, but rather a set that is some what fuzzy but that may be closer to 'songbird'.

56 For a similar type of analysis in regards to verbs, see Levin (1993: 2) who notes that "native speakers can make extremely subtle judgments concerning the occurrence of verbs with a range of possible combinations of arguments and adjuncts in various syntactic expressions."

a biotaxonomic perspective, are if not arbitrary then are at least not those commonly discussed in the ethnobiological literature.

2.4 Summary This section has focused on the complexity of ethnobiological classification from a wide range of perspectives: (1) the ways in which Indigenous nomenclature and classification intersect with Western taxonomic hierarchies and the problems this creates for a precise lexicosemantic treatment of native terms; (2) the family resemblance nature of Indigenous biosemantic categories and the possibility this affords for significant intra- and intercommunity variation of the boundaries and internal structure of native categories; and (3) the complexity of lexicosemantic relations among biotaxa nomenclature and classification including the manner in which purely linguistic criteria can create cognitive categories that are not coincident with the discontinuities in nature that are often cited as the basis for the boundaries mentioned in ethnobiological studies of nomenclature and classification.

The argument presented below is that the preceding cognitive issues can best be addressed through close collaboration between documentary linguists concerned with how speakers talk about the natural environment (§3) and biologists intent on describing its taxonomic complexity (§4).

3. The contributions of endangered language documentation and corpus linguistics to the lexicosemantic treatment of the nomenclature and classification of biotaxa[57]

Linguistics and, particularly, documentary linguistics, has significant potential to facilitate understanding of biosemantics: the meaning and use of terminology relating to flora and fauna. In addition to the identification of basic nomenclature and variation, discovery methods for taxonomic hierarchies and covert categories, the latter a term that Berlin has used to indicate cognitive categories that are not labeled lexically, has been continually debated in the ethnobiological literature. Ellen, who has explored the problems of classification in most detail and with the greatest insight, notes the problems of formal elicitation: "While controlled elicitation has the considerable advantage of generating large amounts of data quickly for the purpose of quantitative and comparative analysis, it is often the case that the ethnographer is demanding tasks which might otherwise never be performed" (1993: 25).[58]

Elicitation removes important contextual elements from discourse about the

57 Language documentation efforts to record native natural historians' narratives on the environment—including flora and fauna, geography and landscape, material cultural, hunting and fishing—have multiple benefits. They produce material of great use to communities and document an endangered realm of knowledge often incompletely, if at all, documented in endangered language research projects. This point is most ele gantly expressed by Si (2011, 2016). While fully in accord with his perspective, and the necessity of encyclopedic entries for natural history ref erents, the pages below offer an additional value to documenting natural history: the creation of large sets of material that can be analyzed following the methodology of corpus linguistics.

58 Ellen (1993:25) cites in this regard Hays (1976) and Healey (1978–9). Brown (1974) and Taylor (1984, 1990) make the same point and note that formal elicitation tasks are activities quite divorced from everyday communicative activity. Taylor (1990:44) notes that besides the use of a nu merical classifier in Tzeltal for plants, the other techniques mentioned by Berlin, Breedlove, & Raven (1968) for determining a "unique beginner" are all "based on tests for perceived similarities among organisms." Hays's (1976) methodology for determining covert categories is also not convincing. He suggests that distinct plants that different consultants label with a single name constitute a covert category. There is no evidence, however, that such a grouping indicates anything more than perceived similarity and probable confusion, not a culturally salient grouping.

natural environment in regards to naming, classification, and use. Simple questions such as “Is X a type of Y?”, “How many types of X are there?”, or “What is X used for?”, often used in elicitation sequences, are themselves contextualized and situated within a series of expectations that the interviewer and consultant bring to the exchange.[59] Elicitation involves a reseaercher-consultant dynamic that affects responses. One example was given above, the adequacy of *chikomolin* (*Leucaena matudae*)for posts and construction. A normative response (and elicitation frameworks tend to favor normative responses) is “no”; but this “no” is belied by actual practice. Card sorting and triads tests are equally problematic. As Brown (1974:327) notes, they “often present informants with culturally irrelevant options coercing them to sort items together which they rarely, if ever, group together on an ordinary *day to day basis*” (emphasis in original). That is, the stimulus, in most cases a visual image or dried or otherwise preserved voucher specimens, implies an organizing principle, in these cases morphology, that channels responses. Recorded bird calls, a quite efficient tool for identification of native referents, would undoubtedly lead to classification patterns distinct from those prompted by visual imagery, and might be less prone to generate hierarchical patterns.[60]

Given the caveats associated with formal elicitation through structured questioning and decision-making tasks, other methods need to be developed to discover nomenclature variation and category membership, boundaries, and internal structure. This is particularly true for covert higher level categories such as life-forms and unique beginners. Taylor (1990:46) mentions “natural conversation” in his account of polysemy of certain Tobelo terms and, indeed, clearly all ethnobiologists, particularly those who have taken a more literary, conversational, and encyclopedic approach to Indigenous natural history (e.g., the works of Nabhan, Rea, Breedlove and Laughlin, Bulmer, Majnep, and Pawley) rely on overheard natural exchanges among consultants as well as unstructured conversations and, in the best of cases, extensive recorded discourse on relevant topics.[61] Language documentation, which creates the resources for searches through large corpora of transcribed texts, offers a unique opportunity for both qualitative and quantitative analysis of how members of a given community designate flora and fauna and converse about their natural environment. Several examples from both unrecorded natural discourse and corpus material illustrate the importance of conversational data for the study of ethnobiological nomenclature and classification.

Careful attention to unelicited natural conversation can provide important clues to the structure of native categories. In the previous pages brief mention was made of Oapan Nahuatl nomenclature and classification for a group of insects designated ‘Formicidae’ in Western taxonomy. One of the most salient ants throughout the Mesoamerican cultural area is the leaf-cutter *Atta mexicana*. In Yoloxóchitl Mixtec distinct terms are used for the (edible) winged queen, the winged male, and wingless female soldiers. In the Sierra Nororiental de Puebla *tsīkatl* is a monotypic term reserved exclusively for this species. And in the Balsas region, speakers distinguish between colonies of *A. mexicana* (*tsontetl*)

59 This same point is made by Ellen (1993:225): “Informants, unprompted, rarely in the course of their ordinary lives will uses expressions such as ‘is x a kind of y?’, or ‘how many kinds of y are there?’.”

60 Hunn (1992) suggests using bird calls as a stimulus for identification and in lieu of physical specimens as vouchers.

61 A pioneering effort in this regard is the work of Majnep & Bulmer (1990). The Taller de Tradición Oral de CEPEC and Pierre Beaucage (1988) produced exemplary material, though the corpus was not subjected to direct analysis.

with large, visible surface fungal dumps (the contents of which are used to fertilize plants, such as cilantro, cultivated as condiments) and those without. In the Balsas and Sierra Nororiental de Puebla regions it is, therefore, rather unusual to ask whether a given ant is a "type of *tsīkatl*". Whether the answer is affirmative or not, the categorization of *tsontetl* (Balsas) or other non-leaf cutters (Puebla) as a type of *tsīkatl*is not a common perspective.

Nevertheless, one day I was entering my house (in San Agustín Oapan, Balsas valley) with an elderly woman who had consistently helped me in the study of arthropods. At the entrance was a large *A. mexicana* nest. The woman looked at me and joked, "*Nō tihpia motsīkaw*" (lit., 'You also have your ants'). Though it was obvious that we both knew that this was a leaf-cutter mound, she not only used a generic term, *tsīkatl*, instead of the specific *tsontetl*, but she used a possessive construction.

This brief exchange demonstrates the inclusion of *tsontetl* within the *tsīkatl* category, at least by this speaker in a given context: a joking reference to the house-site infestation of leaf-cutter ants. Even though red harvester ants (*Pogonomyrmex barbatus*; locally *tēk^{w}*āntsīkatl) are selected to be appeased through an offering of maize kernels in exchange for a person's soul,[62] it is the *A. mexicana* that is considered particularly prone to take revenge on those who disturb its nest, often very visible and marked by large surface mounds. Thus unlike other ants, particularly fire and harvester ants, colonies of *tsontetl* are tolerated and not exterminated with pesticides. This stability in location,[63] a combination of the effect of large underground nests and traditional beliefs in the danger of provoking the ire of these ants, undoubtedly also contributes to the possibility of "possessing" these ants (or "having the insect lodge in one's residence"), a relation generally not found with insects.[64]

62 The selection of P. barbatus for these offerings is undoubtedly related to their seed-gathering activities: they rapidly find the offered maize kernels and take them into their nests. Harvester ants also are used ritually, though in a different manner, in southwestern United States (Groark 1996, 2001). In obtaining the Western scientific species identification for the ritually used ants, Groark (2001:141) comments on the value of voucher specimens: "That such an identification can be confirmed more than a century after the species' last known use is eloquent testimony to the importance of voucher specimens in anthropological research, as well as to the importance of the collections that preserve such materials." The argument presented in this article is that best practice would preserve both the voucher specimen and native discourse on the taxa.

63 The term tsontetl is a compound noun from tson 'hair' + tetl 'stone' or 'rock'. The compound has no other meaning but in regards to an implication of immobility note that kohtsontetl (with the added element koh 'tree' or 'wood') means 'tree stump'.

64 One notable exception is Polybia occidentalis nigratella du Buysson, swarming-founding wasps that often build their nests under house eaves and are jokingly referred to as the house owner's 'little chicks'.

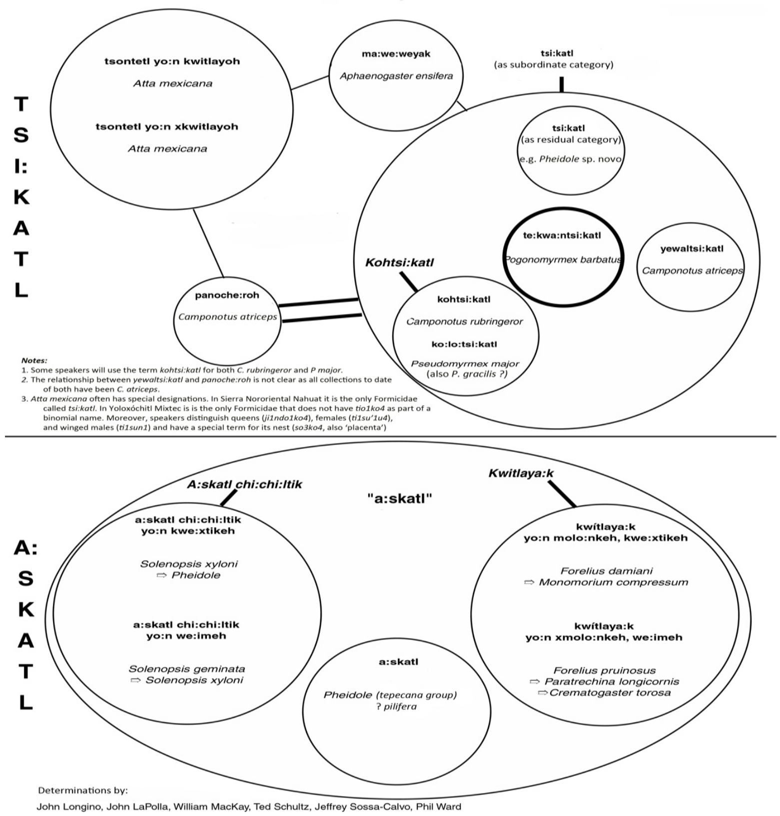

Figure 2. Classification of Formicidae in San Agustín Oapan Nahuatl

As is the case with the off-hand comment *Nō tihpia motsīkaw*, natural conversations produce some of the most revealing insights into Indigenous categorization. Several have already been noted: (1) the comment that *Psorthaspis formosa* (family Pompilidae) was an unusual winged ītskwin tiōpixki indicated that winglessness was a central characteristic of this category; (2) the discussion of the status of *Solanum chrysotrichum* revealed speaker-specific judgments of classificatory criteria and boundaries. In Oapan, Guerrero, the statement of one consultant in regard to a stick insect, *Nō chapolin, nō nok^{w}a* ('It is also a *chapolin*, it is also edible'), demonstrated clearly that edibility was one criteria for class inclusion into one (perhaps the most inclusive) of several senses of *chapolin* (see Appendix for an encyclopedic discussion of the *chapolin* category). Indeed, the criteria by which category membership is judged can best be obtained by careful attention to dialogues, discussions, and disagreements among native speakers, particularly in regard to the categorization of peripheral items.

Insight into Indigenous nomenclature and classification of biotaxa can be also gleaned from semantic and statistical analysis of digitally recorded discussions among native natural historians particularly on topics such as: local flora and fauna; material culture including house construction, fencing, and animal traps; food preparation; and medicine. Large (> 1 million words) topically relevant corpora that focus on the semantic domains just mentioned can be explored to address many issues in the lexicosemantics of biotaxa. The following discussion, based on such an analysis,addresses two such issues: residual categories and patterns of nomenclature and classification.

As noted above (n. 22) residual categories are apparent hypernyms that have an additional sense as a co-hyponym of more specific terminology at a lower level. Such categories were first, I believe, extensively discussed by Hunn (1977), who suggested that certain terms may be both the hypernym of specifically named taxa and a co-hyponym (residual category) of these same taxa. For example, *pehpen*$_2$(residual) is a co-hyponym of named butterfly taxa (*species*$_1$, *species*$_2$, *species*$_3$): it covers and categorizes, in Hunn's analysis, butterflies not specifically named. *Pehpen*$_1$, as a hypernym, covers the named species (1, 2, 3, ...) as well as those of the unnamed residual category, *pehpen*$_2$. Taylor (1990:64–65) offers a nuanced critique of always considering "residue" (i.e., unnamed taxa at a given taxonomic level) a residual *category*, although he does accept that such categories exist (e.g., *o iuru* as a residual term for the wingless forms of all ants except the specifically named weaver ants).[65]

Several researchers have suggested that residual categories are often marked by a term meaning 'just' (Hunn 1976, 1977; Berlin 1992:114). Corpora large enough to permit study of the use of a native term meaning 'just' with different named biotaxa, therefore, can suggest the relative degree to which different taxa may mark a residual category, although "just" may modify biosemantic terms for other reasons. The table below presents the four most common words that follow the Sierra Nororiental de Puebla Nahuat terms *xiwit*, *kowit*, *komekat*, *sakat*, and *mōsōt*.

Table 2. Association of *sah* with various biotaxa terms[66]

Term	Following word	Occurrences	Percent
Xiwit applied to certain herbaceous plants (but not grasses, *sakat*) 1152 total occurrences[67]	sah ('just')	150	13.0 %
	wān ('and')	122	10.6 %
	tein (relativizer)	50	4.3%
	mochīwa ('become', 'grow')	19	1.7%
Kowit 'tree' or 'wood' 1577 total occurrences	wān ('and')	171	10.8%
	mochīwa ('become', 'grow')	80	5.1%
	tein (relativizer)	80	5.1%
	sah ('just')	74	4.7%
Komekat 'vine' or 'liana' 442 total occurrences	wān ('and')	55	12.4%
	tein (relativizer)	27	6.1%
	sah ('just')	21	4.8%
	mochīwa ('become', 'grow')	13	3.0%

65 Berlin, despite early acceptance of residual categories, later adopted the more critical perspective of Taylor (Berlin 1994:114–16).

66 The entire corpus comprises approximately 1.5 million words; topics are skewed towards flora and fauna, material culture, and trapping and fishing.

67 This figure does not include the 233 times that xiwit (a homophone) is clearly used in the sense of 'year'.

Sakat 'grass' and 'sedge'[68] 98 total occurrences	wān ('and')	9	9.2%
	tein (relativizer)	4	8.2%
	sah ('just')	8	4.1%
	mochīwa ('become', 'grow')	1	0.0%
Mōsōt *Bidens alba* (specific name) or a category term including *B. alba* and *B. reptans* 164 total occurrences[69]	wān ('and')	33	20.1%
	tein (relativizer)	6	3.7%
	mochīwa ('become', 'grow')	3	1.8%
	sah ('just')	0	0.0%

The preceding table demonstrates that *sah* occurs almost three times as frequently after *xiwit* as it does alongside other "life forms." Analysis of unrecorded natural conversation and recorded dialogues suggests that *xiwit* is indeed a residual category while the other three life-form terms (*kowit*, *komekat*, *sakat*) are less so. There seems to be an expectation that *xiwit* is not further named. Thus a question about the name of a herbaceous plant could elicit the response "Āmo kipia ītōkāy, xiwit sah" ('It doesn't have a name, it is just a (n unnamed) herbaceous plant') whereas a question about the name of a tree could elicit the response "Āmo nikmati" ('I don't know') or "Āmo kipia ītōkāy" ('It doesn't have a name') but rarely if ever ? "Āmo kipia ītōkāy, kowit sah" ('It doesn't have a name, it is just a tree'). *Kowit* also signifies 'wood' and many of the occurrences of *kowit sah* refer to the fact that a material item is made of 'just wood' (i.e., not some stronger materials). Other occurrences of *kowit sah* refer to the fact that some trees grow *just from wood* (a branch) stuck in the ground and not only from seed.

Xiwit may also be used to contrast cultivated from uncultivated (wild) herbaceous plants and thus the collocation *xiwitsah* may also mean that the plant in question is not planted or tolerated. In such cases the *xiwit* might well be named, but the context of utterance makes clear that *xiwit* is being used to contrast the referent from other cultivated or useful plants. That is, *xiwit* is part of three sometimes but not necessarily overlapping contrast sets: *named ~ unnamed*, *cultivated ~ wild*, and, often, *useful ~ not useful* (though some *xiwit* may be used, predominantly as fodder)[70]

Corpus analysis may also shed light on nomenclature patterns. Table 3 reveals the occurrences of all lemmas that include the unanalyzable stem *mōsōt*, a very common term for two basic plants: *Bidens alba* (also the very similar *B. odorata*) and *B. reptans*. In an unmarked sense, mōsōt_1 refers to the higher level node; in a marked sense, $\textit{mōsōt}_2$, it refers to *Bidens alba* (and *B. odorata*). To clarify which taxonomic level is meant, speakers may signal the marked use through a compound *mīlahmōsōt* ('cornfield *mōsōt*') based on the fact that this plant is frequently found in highly disturbed areas including abandoned or fallowed fields as well as along clear roadside sites. *K^{w}amōsōt*

68 Actually only a subset of sedges (family Cyperaceae), as about a half-dozen sedges are given specific names.

69 Only those occurrences not followed by an end-of-phrase marker (period, comma).

70An interesting meaning of xiw- is found in the incorporated-noun verbal compound xiwnekwisti 'to smell like a weed' used by a protagonist in Silvestre Pantaleón (Olivares &Amith 2011) in reference to Chenopodiumnuttalliae that had been fertilized by a chemical and not a natural, organic fertilizer. The plants that had been fertilized by commercial products had a "weedy" taste, not the taste of the plants fertilized with fungus from an Atta mexicana nest or from guano.

is a much larger plant, a climbing vine, also found along roadsides though frequently climbing on other plants. The two *mōsōt* are quite different and the common feature, the double barbed awn of the achene, has never been mentioned by any consultant as a defining characteristic of the *mōsōt* category.[71] This suggests that the classification of the two *Bidens* together is not based on the feature that botanists would use to establish the generic category but rather simply learned as a nomenclatural unit.

Table 3. *Mōsōt* in the Sierra Nororiental Corpus

Occurrences	Lexeme
123	mōsōt (once as mōsōtsīn)
110	k^{w}omōsōt, k^{w}amōsōt, komōsōt, kowmōsōt, k^{w}omōsōt
57	Mīlahmōsōt
15	k^{w}omekamōsōt, k^{w}amekamōsōt, komekamōsōt (an alternative name for *k^{w}omōsōt*)
7	Tālmōsōt
6	istāk mōsōt
5	kowtah mōsōt
4	yēkmōsōt (once as yēkmōsōtsīn)
1	xōkihyākmōsōt
1	Mōsōtakōt
1	māweweyak mōsōt (speaker gives this as another name for *mīlahmōsōt*)
1	īknīw mōsōt (lit. ‹mōsōt›s brother›, a descriptive relationship applied to other Asteraceae, particularly *Melampodium divaricatum*)
333	**Total occurrences**

Table 3 reveals patterns in the term *mōsōt*, including *yēkmōsōt* and *yēkmōsōtsīn* (lit. 'true *mōsōt*') for *B. alba*/ *B. odorata*, indicating that this species is the prototypical referent of *mōsōt*. Particularly interesting is the emergence of the term *tālmōsōt* in 2014 for *Erigeron karvinskianus*, a daisy-like Asteraceae that frequently grows in rocky crevices and walls. The plant was known to several consultants, who also were aware of, and had commented on, its favored habitat. However, it was only in 2014 that one highly knowledgeable consultant designated the plant a *tālmōsōt* (lit., 'earth *mōsōt'*), classifying it as a *mōsōt* along with the very common and well-known *mīlahmōsōt* (*'milpa* [cornfield] *mōsōt'*, *Bidens alba*) and *kwamōsōt* ('tree or woody *mōsōt*', *B. reptans*). By that time this consultant, and others, had understood that the project involved the "classification of discontinuities in nature," to borrow the subtitle of an influential book by Eugene Hunn, and had begun to recategorize plants, creating groups naturally, independent of formal elicitation. As in other cases (e.g., the extension of the *ndi^{3}xi^{4}tu^3* / *kohtekine* "woodcutter" category previously discussed), the extension of the *mōsōt* category to *Erigeron*

71 Interestingly, in Costa Rica the loan into Spanish, mozote, is applied (extended) to Triumfettalappula L. (Malvaceae, formerly Tiliaceae), another plant with prickly haired fruits that likewise stick to one's clothing and body (Chízmar F. 2009:319), an extension that suggests that "burr-like" fruits are a key element of the category definition for mōsōt, at least historically. Note that Bidens derives from Latin: bis 'twice' and dens ('a tooth') (Hyam & Pankhurst 1995).

karvinskianus, though an innovation, is not uninteresting. It demonstrates both the impact of project participation on the cognitive categories of native natural historians and the morphological characteristics that may orient patterns of extension over time. Starting after one consultant identified *E. karvinskianus* as *tālmōsōt*, he and another consultant present at the time have consistently included this term as one "type of" *mōsōt*.[72]

The examples in tables 2 and 3, the first on the utilization of *sah* ('only') in modifying terms for biotaxa (particularly life-forms) and the second on the relative occurrence of different terms including the word-stem *mōsōt*, suggest the value of corpus-based research on the nomenclature and classification of biotaxa. One final point should be added, again relevantto a discursive and linguistic basis for cognitive research in ethnobiology. In discussing Brent Berlin's theory of taxonomic hierarchy, Eugene Hunn (1977:44) gives six examples of "life-form" categories: 'bird', 'mammal', and 'fish' in the animal domain and 'tree', 'vine', and 'grass' in the plant domain. Taxonomically all terms occupy nodes at the same level of their respective taxonomic trees (e.g., bird:robin :: tree:maple), just above intermediate or folk generic forms in Berlin's (e.g., 1992, esp. Chap.4) analytical scheme.

The analogical "equality" of life-forms at the taxonomic level disappears, however, when one examines patterns of usage incommunicative events. That is, use of terminology at any level, from terminal taxon to life-form, is part of a process of "situationally adapted" (Ellen 1993:3) information exchange that is greatly influenced by sets of complex issues. Thus, if we take Hunn's three animal life-forms (bird, mammal, fish) and count usage in two large corpora of American English (see Table 4), we see that 'mammal' is rarely used.[73] The reason for this is not, I think, related to the relative rarity of mammals in daily life but rather to the general lack of informational relevance of this term for most situations of discourse. For example, if I were to tell a friend about a dead duck, rattlesnake, or raccoon that I had passed on the side of a highway, I would be most likely to comment in the following ways:

1. I saw a dead duck/bird/animal on the side of the road.
2. I saw a dead rattlesnake/snake/reptile/animal on the side of the road.
3. I saw a dead raccoon/? mammal/animal on the side of the road.

A strictly formal approach would suggest that duck:bird:animal::rattlesnake:snake:animal::raccoon:mammal:animal. The gap in felicitous expressions? "I saw a dead mammal on the side of the road" is not explained by any formal taxonomic analysis. It reflects, instead, the low relevance of the category in quotidian exchanges. Two basic questions in beginning a cognitively oriented ethnobiological study should be, then, the context in which nomenclature and classification are to be studied and the different registers in which terms at the same taxonomic level are used. The argument presented

72 Breedlove & Laughlin (1993, vol. 1: 2) note another level of impact on consultant responses: group dynamics:

> The impact of group dynamics upon the choice of plant names was terrible to behold. Of paramount importance was the social criteria of the collector within the team, such as his age or bonds of friendship. Social position and temperament often determined who would align with whom in assigning which names. Efforts to prevent collectors from influencing their colleagues' decisions frequently were unsuccessful.

73 Given the absence of syntactic parsing, 'fish' would show up regardless of its use as a noun or verb.

in this essay is that linguistic and conversational data is particularly interesting in this respect and that corpus-building as part of a language documentation project is especially valuable in creating this resource.

Table 4. Occurrence of Hunn's six cited life-forms in two corpora of American English

	Corpus of Contemporary American English 450 million words			Corpus of Historical American English 400 million words		
	Singular + Plural	**Total**	**Percentage**	**Singular + Plural**	**Total**	**Percentage**
bird	20,986 + 21,532	42,518	43%	23,736 + 23,043	46,779	58%
mammal	1,102 + 2,938	4,040	4%	396 + 1,657	2,053	3%
fish	50,179 + 1,805	51,984	53%	28,972 + 2,736	31,708	39%
tree	36,205 + 40,563	76,768	75%	42,092 + 49,534	91,626	72%
vine	2,221 + 2,805	5,026	5%	3,798 + 4,627	8,425	7%
grass	18,363 + 2,232	20,595	20%	24,599 + 1,869	26,468	21%
platypus	68 + 0	68		36 + 19	55	
rattlesnake	590 + 363	953		925 + 424	1,349	
marigold	208 + 299	507		186 + 235	421	
tulip	802 + 814	1,616		585 + 659	1,244	

In regard to Table 4, certainly polysemy ('smoke some grass' for marijuana) and multiple part-of-speech functions ('fish [verb] for compliments') might account for some variation in frequency, although a corpus with part-of-speech tagging and more rigorous syntactic analysis might obviate some of these problems. But the rarity of 'mammal' and the fact that the plural is much more commonly used than the singular (a distributional fact not shared by the other terms) suggest that other factors play a role. One is the saliency of subordinate categories: most mammals, at least mammals that are frequently the topic of conversation, are named and easily recognized. Speakers probably would tend to use the lower-level term when appropriate. This does not explain, however, my native speaker intuition that in phrase 3 above I would be more likely to use 'animal' than 'mammal' even though the former is inclusive of snakes and birds. This suggests that 'mammal' is a term of restricted register.

Considering discourse pragmatics, prototype theory of categorical relations, and situational caveats to lexical relations of (co-)hyponymy and hypernymy, it is clear that taxonomic relations represent but one expression of classificatory criteria among many cross-cutting patterns of contextualized lexicosemantic relations. A taxonomic hierarchy represents the expression of one context: elicitation formalized by either triadic selection or set questions of the type mentioned earlier: "Is X a type of Y?". Functional grouping represents another categorization pattern in which prototypical status may reflect a combination of normative (the best X for Y function) and geographic (the best X for Y function given locational constraints) criteria. Another type of pattern is brought out by linguistic analysis of natural speech, three types of which have been mentioned above: (1) discourse revelations of hypernymy, hyponymy, and co-hyponymy relations; (2) co-occurrence patterns (e.g., terms that share the feature of potential subjects of a given verb, possessors of a given noun); and (3) statistical analysis of medium-sized corpora, particularly those topically oriented to natural history. All this material invariably reveals patterns distinct from those obtained in formal elicitation settings.

The final section addresses the issue of interdisciplinary collaboration with biologists in projects that address the empirical data and research objectives of all participants. From the discussion in the preceding sections it is clear that a cognitive focus on ethnobiology—exploring the nomenclature and classification of local flora and fauna—is enhanced by large data sets of natural conversation and comprehensive inventories of local taxa, both those named and unnamed, as well as those known and unknown to native natural historians. Certain taxonomic groups are more likely than others to pose both methodological and theoretical questions for cognitively oriented ethnobiological research, inevitably requiring more resources—from fieldwork through determinations to analysis. Finally, the advantages of multisited research to address issues of cultural history and diachronic shifts in the biosemantic lexicon can only be realized by a decided shift to a new form of interdisciplinary collaboration, one that involves the use of DNA barcoding and molecular analysis to shift the burden of determinations to species away from taxonomists and toward technicians.

4. The role of biology in ethnobiological research

4.1 Introduction At a very basic level, primary ethnobiological data comprises: (a) a native speaker account, (b) the context or field situation in which communication between researcher and native natural historian takes place, (c) a physical specimen including all relevant collection data, and (d) a determination to scientific species by a taxonomist. Each collection event constitutes a data point (single in terms of the botanical specimen and determination, often multiple in terms of ethnographic information) that, when multiplied, provides an increasingly detailed sketch of native interpretation of the local floristic and faunal environment. Multiple data points, of course, increase the definition of the cognitive sketch of the environment (e.g., unequal distribution of knowledge in a community, peripheral or central status of a reference in a classificatory scheme), a significant goal of linguistically based ethnobiological research. Yet proliferation of collections/data points has a cost—particularly in fieldwork resources, in administration and processing of vouchers, and in imposition on collaborating taxonomists. Moreover, an accurate and extensive portrait of ethnobiological knowledge is complicated by the fact that *both* local natural historical knowledge and Western taxonomic expertise is endangered.[74] And while

74 From the Western scientific perspective, there is a "dwindling pool of taxonomists" (Hebert et al. 2003:313) able and willing to determine voucher specimens to species, particularly specimens having no direct relation to their own research agenda. Indeed, NSF had recognized the problem of diminishing taxonomic expertise by creating the PEET (Partnerships for Enhancing Expertise in Taxonomy) program (now discon tinued), which sought "to enhance taxonomic research and help prepare future generations of experts." The dearth of taxonomists has not, how ever, abated and from a practical perspective documenting the nomenclature, classification, and use of local flora (particularly in diverse neo tropical environments) is fraught with difficulties because of this.

native experts, familiar with local flora can invariably identify sterile (without flower or fruit) specimens, Western taxonomists are reluctant to receive sterile material for determination. Over the past half-dozen years of my research, however, several projects have emerged that have greatly increased the level of mutually beneficial multidisciplinary collaboration by addressing issues and goals of significance to the different collaborating disciplines and by more fully engaging native speaker collaborators in botanical research.

The first shift toward a more collaborative, multidisciplinary, and interdisciplinary effort targets the most complex and extensive classificatory sets for extensive ethnobiological and biological documentation. Essentially this involves intensification of research in particular taxonomic groups, a triage approach to research that focuses on those taxa (i.e., families and genera) that are highly diverse in the local environment and particularly problematic for native naming and classification.

The second shift involves using DNA barcoding to facilitate ethnobotanical research while creating a lasting resource—preserved voucher specimens and DNA extractions—that will advance not only phylogenetic and systematic biological research in the coming generations, but also biodiversity studies and ethnobiological research. Each of these is discussed below as "taxon-targeted" and "inventory-based" research, respectively.

4.2 Targeted taxon–based research Each culture that I have studied has certain groups of flora and fauna that, from a cognitive perspective, are less interesting than others. For example, *Cuscuta* spp. (dodder), Loranthaceae (mistletoe), Dermaptera (earwigs), and Myrmeleontidae larvae (doodlebugs) are rather stable and unchallenging in terms of native systems of categorization. Speakers may distinguish among the parasitic *Cuscuta* and Loranthaceae by host plant, color, or size but these seem to be ad hoc distinctions rather than lexicalized subgroupings of set categories. Indeed, the above mentioned groups manifest little internal structure (i.e., differentiation between prototype and peripheral referents) and relatively unproblematic boundaries.

Other groups, such as orchids in the Sierra Nororiental de Puebla, are inevitably recognized by native speakers. In this case the "native" category is apparently the result of Western influence (including trafficking of these plants) and thus the Indigenous and scientific categories are essentially co-terminous: the scientific family Orchidaceae.[75] From the native perspective, the internal structure of this category is rather uninteresting. Only approximately 10 percent of the over 100 orchid species found in the municipality of Cuetzalan are locally named in Nahuat and in most cases there is a simple one-to-one correspondence between the Indigenous name and Western scientific species or, in a few cases, genus. In the case of orchids, extensive speaker-led collection of the taxon reveals only minor complications to intercultural category correspondences but the facility with which native consultants recognize this Western taxon can be advantageous to botanists working on orchids. Although unnamed in Nahuat, speakers can easily recognize an orchid and through collections can contribute to the research agenda of a collaborating

75 Orchids are heavily, and illegally, trafficked in the Sierra and this may well contribute to the highly salient boundaries to this category, which is named only by the Spanish loan orquídea. Interestingly ferns and fern allies are not so easily categorized.

expert with little cost to a project that focuses on "positive" Indigenous knowledge: those taxa named, classified, and used by local residents.

An exclusive focus on such positive knowledge, however, may diminish the capacity for research into questions of category extensions and boundaries and may complicate any effort to determine the unequal distribution of nomenclature, classification, and use across different Western taxonomic groups.Those species that are not on the native speaker cognitive map are also likely to be more interesting from a purely botanical and floristic perspective. This introduces a methodology that I suggest be called "negative cognitive mapping," ethnobotanical research that is focused not on plants that are locally named, classified, and used but rather on those that are not known to local experts. The plants so selected are not simply those for which no native name exists but, more significantly, those that have gone unnoticed.[76]

Table 5. Taxonomic complexity from an Indigenous nomenclatural and classificatory perspective

Plants:	*Piper* spp. (but not *Peperomia* spp., the other major Piperaceae genus; in the Sierra Nororiental de Puebla),
	Asteraceae
	Burseraceae (*Bursera* spp., in the Balsas River valley of central Guerrero)
	Commelinaceae (in the Sierra Nororiental de Puebla)
	Lauraceae (in the Sierra Nororiental de Puebla)
	Leguminosae
	Melastomataceae (in the Sierra Nororiental de Puebla)
	Solanaceae (particularly *Cestrum*, *Physalis*, and *Solanum*)
	Rubiaceae
	Senna spp. (in the Pacific Coast of Guerrero)
Insects:	Apidae
	Cerambycidae
	Formicidae
	Mutillidae (along the Pacific Coast and Balsas Valley, Guerrero)
	Orthoptera (in the Balsas Valley, central Guerrero)
	Vespidae

In the three areas I have studied, the preceding taxa have proven to be particularly challenging in regard to native language nomenclature and classification. Certain groups (*Bursera* spp., *Piper* spp., Melastomataceae, Formicidae, Orthoptera, Vespidae) enjoy high levels of consistent recognition. Lauraceae manifests what appears to be extensive over determination, intraspecific distinctions made based on fruit color and formation.

76 A expert native natural historian will have wide knowledge of plants, including many for which he or she has no name. Breedlove & Laughlin (1993, vol. 1:6) regretfully note that they "kept no tally of the number of plants that [their]consultants were unable to identify." I think this should, however, be a part of ethnobotanical projects when resources permit particularly since knowledge gaps may be significant to understand ing traditional ecological knowledge.

Asteraceae are extremely common (in the Sierra Nororiental de Puebla they comprise just over 200 species, approximately10 percent of the angiosperm flora) and morphologically diverse, offering excellent grounds for studying native nomenclature and classification in robust sets of stimuli. *Piper* spp. and Solanaceae both manifest interesting patterns of nomenclature correspondences: some Indigenous terms are monotypic while others cover sets of somewhat varied taxa. Among insects Formicidae, Mutillidae, and Orthoptera are consistently recognized, and knowledge of Vespidae is often extremely detailed.[77]

Taxon-based research has to date produced rewarding cross-disciplinary collaboration by establishing synergistic research relationships with biologists specializing in generic and family taxa that are interesting from a cognitive perspective and abundant enough regionally to reward closer interdisciplinary collaboration, often including joint field ventures.[78] Statistics from the Sierra Nororiental de Puebla reveal that twelve families and one group (the latter in reference to ferns and allies) account for 60 percent of the regional flora.[79] Though not a full floristic study, targeting specific taxa does require the collection of material that is not only not locally named but often not even known to the most knowledgeable native natural historian. As mentioned above, it is more than likely that these unnamed or unknown taxa would be rarer than their named and known counterparts and thus of potentially more interest to botanists.[80]

Targeted taxon-based research does not preclude complete coverage of the nomenclature, classification, and use of local flora and fauna but rather forges greater collaboration among social scientists, biologists, and native experts in taxa of greatest local diversity and native recognition. Joint fieldwork—among native natural historians, ethnobiologists, and taxonomists—is particularly productive as each group offers expertise in distinct areas. Yet it is important to note that such collaboration is a learning experience for all participants and that native speaker collaborators are as eager to acquire new knowledge as are the Western natural and social scientists. Thus, for example, at one point a group of native speakers gathered many specimens of *mātalin* for five vouchers. I was careful to separate out *Commelina erecta* from *C. diffusa* and show the two speakers who were helping in pressing the material how to distinguish the two species. Before the collection was complete, all five consultants had internalized the classificatory distinction and lexicalized it in an emergent nomenclatural distinction: *mātalin wehwei* ('large *mātalin'* for *C. erecta*) and *mātalin tsikitsīn* ('small *mātalin'*for *C. diffusa*).

Thus the highly integrated research teams necessary for both targeted taxon-based research and inventory-based research (§4.3) must, of course, be completely open to full synergy among participants. Much as it would be unimaginable to expect native speakers who collaborate on grammars or lexicons not to acquire the relevant analytical and theoretical tools, so too is it naïve to expect collaborators in (ethno)biological

77 For Cerambycidae, see Amith & Lingafelter(2016).

78 To date this triage focus on cognitive ethnobiology has resulted in over a dozen expert taxonomists (botanists and entomologists) either having joined in fieldwork or made commitments to do so in the near future.

79 The statistics are as follows: Ferns and allies (12.8%), Asteraceae 10.9%, Leguminosae 7.5%, Poaceae 6.2%, Solanaceae 3.4%, Rubiaceae 2.9%, Labiatae 2.9%, Orchidaceae 2.4%, Euphorbiaceae 2.3%, Malvaceaesensulato 2.3%, Fagaceae 2.1%, Piperaceae 2.1%, and Melastomataceae 1.9%.

80 This is, however, not always the case. After several years of having a reference to a plant called yōlpoliwkāxiwit, it was finally collected. It turned out to be Peperomia mexicana (Miq.) Miq. (Piperaceae), a species that had last been collected in the state of Puebla in 1945 (Guido Matheiu, personal communication). Other Peperomia not named by Indigenous consultants are much more common. Thus named plants may indeed be rare.

research not to learn some of the perspectives of Western science in establishing classificatory boundaries for flora and fauna. This will, of course, influence their own knowledge base (as collaboration with native communities influences Western understanding of natural history) but I have not found it difficult to work with native collaborators in differentiating sources of their expanding field of expertise. Indeed, this growth of knowledge is a fundamental part of developing native natural historians who can interpret their environment, particularly local flora and fauna, from multiple perspectives.

4.3 Inventory-based research and DNA barcoding

Comparative study of the nomenclature and classification of biological species (flora and fauna) has been an important tool for studying the cultural history of language groups. In addition to a general concern with reconstruction and proto-forms, two more specific research topics have emerged.[81] The first, exemplified by scholars such as Catherine Fowler (1972a, 1972b, 1983), Paul Friedrich (1970), Frank Siebert (1967), and K. W. Whistler (1977) has reconstructed the lexicosemantics of proto-language terms for biotaxa and taken the reconstructed meanings as reflecting ancestral homeland ecosystems. In the depth of his study of Proto-Indo-European, Friedrich broke new ground in rigor by linking protosemantic reconstruction of flora to prehistoric ecosystems. Siebert, in turn, was the first to apply this methodology to American languages. Fowler continued this effort to use lexical evidence to document ecological clues to homelands but at the same time noted the limitations of this approach "given the *quality* and *quantity* of data presently available" (Fowler 1983:224; emphasis added).

A second direction of research regarding the lexicosemantics of biological nomenclature relates to contact phenomena and what has been called linguistic stratigraphy: "the systematic investigation of the layering of grammatical and lexical material in a language or dialect which reflects its historical development and past contacts between its speakers and bearers of other linguistic and cultural traditions" (Andersen 2003b:1). Within this area of research a small subset of studies has either focused exclusively on biological nomenclature (Bowern 2007; Bowern & McConvell 2011; Bostoen 2007; Meroz 2013) or relied heavily on terms from this semantic domain (Dakin 2003). The strategy of these studies differs, as Yoram Meroz notes, from most comparative lexical surveys in that to elucidate genetic relationships among languages a set of basic words most resistant to change is preferred (see Haspelmath &Tadmor (2009), particularly Chap. 1–3). The nomenclature of flora and fauna, however, is probably more sensitive than basic vocabulary to change through contact and thus is a particularly propitious semantic domain in which to study migration and contact.[82]

The loans that are relevant to such stratigraphic studies, in turn, may be either of form (i.e., loan words) or meaning (calques or loan translations). The former is common, for example, among Nahuat speakers of the Sierra Nororiental de Puebla who

81 Undoubtedly the most complete attempt for deep historical reconstruction of terms for flora and fauna is the Oceanic Lexicon Project (https://sites.google.com/site/theoceaniclexiconproject/) in which volumes three and four, of a proposed seven-volume effort, deal with plants and animals, respectively, and volume one with material culture (see Ross, Pawley, & Osmond 1998–2011).

82 Balée & Moore (1991) and Berlin et al. (1969, 1973) have looked at the factors that shape the rate of retention and loss in closely related languages among different types of ethnobiological nomenclature.

have borrowed many, some even quite basic, terms from Totonac (*xopepe* ‘cockroach’, āltsimit ‘wasp’, and *chokoy* ‘puss caterpillar’ [Megalopygidae]), the borrowing probably the result of Nahuat migration into the area. The second type of loans are calques, loans in which meaning is translated from one language to another but the term itself is not borrowed. Calques have been used to support the definition of Mesoamerica as a cultural area, though these loan translations are only one of several features, many morphosyntactic, that are regionally shared (Campbell, Kaufman, &Smith-Stark, 1986; Smith-Stark 1982, 1994; about 18 percent of the calques these authors reference denote flora or fauna, such as ‘mother of the leaf-cutter ant’ for ‘coral snake’). In my own ethnobiological research I have discovered a few more calques (e.g., camel spiders and whip scorpions are both called “shame [animal]” among the Aztecs, modern Nahuatl speakers from central Guerrero, and the Coastal Mixtecs) and examples of cultural beliefs associated with animals and plants (e.g, Mutillidae are considered omens in several Totonac, Mixtec, Triqui, and Mazatec languages). The viability of using loan words, calques, uses, and shared cultural beliefs of biological nomenclature for studying historical contact is, however, also hindered by the same poor quality and quantity of data noted by Fowler. The project described in this section represents an effort to provide such necessary data.

Despite the importance of comparative research on the nomenclature and classification, as well as use, of local flora and fauna, unified efforts to document this knowledge across communities face problems that only begin with the complexity and fluidity of local knowledge.[83] A comparative project that targets a collection of 1,000 specimens from half a dozen communities embedded at different points in a regional ecosystem would produce 6,000 plants. If botanical best practice is followed and only fertile material collected, the task of assembling the inventory becomes even more daunting. Employing multiple teams of field botanists is one way to address this issue, but this often occurs at the expense of ethnographic detail and understanding. Moreover, with such a large collection, taxonomists would be overwhelmed by material, much of which would be common and of routine identification.

Inventory-based research and DNA barcoding, the focus of a project underway in the Sierra Nororiental de Puebla, attempts to solve the difficulties of multi-sited research by facilitating the determination to species of vegetative material (e.g., leaf tissue) through the use of genetic markers.[84] The botanical aspects of this project comprise four steps: (1) collection of flowering specimens that represent the floristic inventory of the Sierra Nororiental de Puebla; (2) identification to species of these voucher specimens; (3) DNA sequencing of up to four regions of the genome of each species to create a DNA barcode reference library; and (4) use of this reference library to facilitate the identification of a small sample of vegetative plant material that will be used to document the nomenclature, classification, and use of local flora in Sierra Nororiental Indigenous villages.

This methodology occasions two major shifts, one at the field end and one at the herbarium end of the project. At the level of field collection, speakers can collect a single sterile voucher specimen, along with vegetative material dried in silica gel, linked to ethnobotanical knowledge (a name, a classification, a use). This simplifies and accelerates

83 Cf. Laughlin’s comments on the variability of taxonomic knowledge in Zinacantan (Breedlove & Laughlin 1993, vol. 1:8).

84 This is an NSF, Documenting Endangered Languages award (BCS 1401178), and an NEH, Preservation and Access, award (PD-50031) entitled “A Biological Approach to Documenting Traditional Ecological Knowledge in Synchronic and Diachronic Perspectives.”

collection, reducing the need for flowering or fruiting vouchers as well as the number of duplicate vouchers needed. This will empower communities to document their own ethnobiological knowledge, diminishing the need for Western trained botanists in the field and, at the herbarium end, in a majority of cases transforming identification into a laboratory process and freeing taxonomists from routine identifications of common species. In certain cases even a four-region barcode (*matK*, *rbcL*, *ITS*, *trnH-psbA*) may not distinguish among congeneric species, leaving ambiguity between two and perhaps half a dozen specimens.[85] In such cases, in which more discriminatory power is needed, supplementary methods will be developed to disambiguating the congeneric species: (1) the use of disambiguating morphological features present in sterile specimens, such as leaf form, venation, or pubescence; and (2) a clade-specific DNA locus capable of distinguishing among these congeneric species. It is likely that with these additions close to 100% accuracy in sterile species identification can be approached.

The use of DNA barcodes not only simplifies collection and shifts identification to a technical process in a laboratory, it also creates a permanent barcode reference library that can be built upby future projects. The present project in the Sierra Nororiental de Puebla is a significant start. Itshould produce a DNA barcode reference library covering approximately10 percent of Mexican angiosperms and about 25 percent of Mexican ferns and allies; for the state of Puebla the figures are closer to 38 percent of angiosperms and 60 percent of ferns and allies.

This project goes well beyond the common interests and methodologies of ethnobotanical research. It begins with a floristic inventory of a relatively large region, an inventory that probably comprises between four and five times the number of species known to Indigenous natural historians from any given community and perhaps three times the flora collectively known to Indigenous experts throughout the region. Significant effort is expended to continue to collect material even when it is clear (as it is with orchids, *Peperomia* spp. and *Piper* spp.) that all species of a given family or genus that are known to or used by Indigenous people in the region have been collected. The inventory approach of this project has generated the interest and support of the state and national herbaria as well as that of CONABIO (Comisión Nacional para el Conocimiento y Uso de la Biodiversidad), a federal agency charged with documenting and protecting Mexico's biodiversity. Moreover, during the project's first year, seven renowned taxonomists from Mexico and the United States have visited the sierra to collaborate on the project and several have indicated an interest in repeated visits to develop parallel projects integrating botanical and ethnobotanical research. Another eighty have supported the project with identifications in the families of their expertise, both of new vouchers collected in this project and of misidentified herbaria specimens previously collected in the region.

The degree to which this project has been able to generate strong support in the botanical community reflects, to a great extent, the manner in which it creates a floristic inventory and enlists the collaboration of expert taxonomists to identify the fertile vouchers, the foundation of the DNA reference library, to species. Indeed one expert on DNA barcoding mentioned that the greatest value of the project rests more with the creation of a regional flora that is accurately identified by specialized taxonomists and accompanied by permanently preserved DNA extract that could be resequenced at a later date when

85 There is little doubt that in all, or virtually all, cases, the DNA barcode should be able to get the specimen to genus.

technology is more advanced and economical (Peter Hollingsworth, personal communication). Another botanist, Martin Ricker, noted that approximately 81 percent of the 11,219 collections from the *Inventario Nacional Forestal y de Suelos (2013)* are sterile. Although some 70 percent of the collection have been identified to species, the rest remain unidentified to this level (Ricker et al. 2015). One solution he is attempting is to develop models of leaf morphology that can be applied to the leaves of the collection. But the creation of a DNA barcode reference library, such as that envisioned in the present project, offers an additional methodology for identification to species. Certainly a combination of sequencing and leaf morphology along with a knowledge of the biodiversity of regional ecosystems, will greatly enhance identification of the large number of sterile specimens.

Within the endangered language community this project has also motivated interest. David Beck, an expert on Totonac, was included in the original grant to document ethnobotanical nomenclature, classification, and use in the Upper Necaxa Totonac village he studies. Recently he has suggested that three or four students replicate the process: intensive collection of vegetative material of plants named, classified, or used in other Totonac communities in the Sierra Norte de Puebla. Should this expansion take place, the project would be able to develop comparative lists of nomenclature and classification of flora in over half a dozen Indigenous villages of the Sierra Nororiental and Norte in Puebla.

A project that builds up a regional DNA barcode reference library has proven, therefore, to be highly attractive to botanists and ethnobotanists, including linguists and anthropologists. From a botanical perspective it will develop an extensive floristic inventory comprising well over 2,000 expertly determined species, each with an associated four-locus barcode and extracted DNA preserved for future study. From an ethnobotanical perspective it will develop regional sets of plant nomenclature and classification that will facilitate the study of cultural history, particularly migration and contact in a fairly wide region of Indigenous settlements.

As a final observation, it is still debatable whether the type of collaboration envisioned in the DNA-barcode project is truly interdisciplinary. In the sense advocated by the National Science Foundation, wherein interdisciplinary research should have the potential to forge new disciplinary ventures, it is probably lacking. The theoretical focus is mostly within the domain of anthropology, linguistics, and cultural history. But the core systematic questions that are increasingly addressed by molecular analysis, exemplified in such efforts as the Angiosperm Phylogeny Website (http://www.mobot.org/MOBOT/research/APweb/) are only peripherally advanced by the collection of plant tissue associated with expertly determined collections from often poorly studied regions. Full integration and the potential for a new disciplinary venture is not fully realized.

Nevertheless, the type of project mentioned above has demonstrated that a floristic and ethnobotanical project can stimulate extensive collaboration from Indigenous communities to regional and national herbaria while melding the expertise of native natural historians, anthropologists and linguists, and taxonomists and systematists into a cohesive collaborative enterprise that can have significant and long-lasting impact at many levels of activity and expertise.

Forging such varied collaboration, as deficient as it may be from a demanding definition of interdisciplinarity, is still not trivial. An ethnobiologist or linguist focusing on biosemantics will need to be sensitive to the distinct needs of communities on the one hand and herbaria and museums on the other. This may involve collections well beyond those needed for a simplified, at times simplistic, lexical entry. Field guides and exhibits for local communities should be prioritized and excellent voucher specimens with proper field data in electronic format should be made. At times the social scientist or humanist will need to go out of his or her way to acquire materials that a biologist might request, thus stimulating a desire among these natural scientists to collaborate fully with the project. For herbaria and museums, not only must the collected material be well prepared and expertly documented, but a sufficient number of duplicates should be collected so as to allow deposits in regional and national venues, as well as gifts to the specialists.

5. Summary

This essay began by presenting two perspectives on interdisciplinary research. The first focuses on the degree of intergration among disciplines to solve a research problem that requires resources beyond those of any single field. The second explored the relative degree to which participants from distinct disciplines were not simply acting as service providers to others, but were participating in an effort that met their own research agenda. Interdisciplinarity, from these perspectives, involves a high degree of integration and an equitable balance of scientific impact on the collaborating disciplines. I suggested that ethnobiological research, particularly that which focuses on cognitive issues of nomenclature and classification, tends to be poorly integrated with biological research. The challenge, then, is how to achieve greater integration and disciplinary importance between the potentially major stakeholders in ethnobiological research: (1) anthropologists or linguists, (2) biologists, particularly taxonomists, (3) native communities, and (4) herbaria and museums.

The second section explored the complexity of Indigenous patterns of nomenclature and classification in an attempt to demonstrate that the most interesting patterns were unevenly distributed across the biological spectrum and most efficiently analyzed (1) through an extensive dataset of natural speech data and (2) through a focused inventory of relevant taxa in the local environment. The first item is the domain of documentary linguistics and in §3 the argument was presented that documentary linguistics and the building of a large corpus of data topicalized on natural history and related themes (e.g., material culture, agriculture, and horticulture) offers significant insight into cognitive and communicative issues related to natural historical knowledge in Indigenous communities. Language documentation is particularly useful as it focuses on the compilation of recorded, transcribed, and translated verbal communication among native speakers. The second item, a focused inventory, is best addressed through the dedicated participation of taxonomists who specialize in those taxa that are most challenging from a nomenclatural and classificatory perspective. An initial effort to build collaborative ventures involved taxon-based research and shifting ethnobiological research to target specific families and genera. Such a shift meant assigning resources to extensive collection in the targeted groups but the result is research that addresses the interests of both social scientists and biologists. A second effort involves a more general floristic approach: creating an inventory of regional flora and developing a DNA barcode reference library from this collection. While the resource itself does not address theoretical concerns, it is highly useful across disciplines, from phylogenetic studies (given that the extracted DNA linked to expertly determined voucher specimens will be preserved for future sequencing and study), through ecology and to ethnobiological research, at a time in which both native and Western taxonomic expertise is increasingly endangered. Moreover, the resource is "cumulative" in the sense that the DNA barcode reference library can be progressively enhanced and thus can increasingly facilitate discovery and analysis in botanical and ethnobotanical research.

Acknowledgements

Research that underlies the perspectives presented in this essay was carried out with support from the National Science Foundation, Documenting Endangered Language Program (s), the National Endowment for the Humanities, Preservation and Access, and the Endangered Language Programme at the School of Oriental and African Studies, London. The following awards contributed to the development of the data and collaborations that were key to understanding the cross-disciplinary work situations discussed in this paper: NSF-Documenting Endangered Languages, Award #0756536, "Nahuatl Language Documentation Project: Sierra Norte de Puebla"; Hans Rausing Endangered Languages Project, School of Oriental and African Studies, Award MDP0201, "Corpus and lexicon development: Endangered genres of discourse and domains of cultural knowledge in Tu'un ísaví (Mixtec) of Yoloxóchitl, Guerrero"; NSF-DEL, Award ##0966462, "Corpus and lexicon development: Endangered genres of discourse and domains of cultural knowledge in Tu'un ísaví (Mixtec) of Yoloxóchitl, Guerrero"; Endangered Language Documentation Programme, SOAS, Award MDP0272, "Documentation of Nahuat Knowledge of Natural History, Material Culture, and Ecology in the Municipality of Cuetzalan, Puebla"; NSF-DEL and NEH Preservation and Access, Awards #BCS-1401178 and #PD-50031-14, respectively, "A Biological Approach to Documenting Traditional Ecological Knowledge in Synchronic and Diachronic Perspectives, National Endowment for the Humanities"; CONABIO (Comisión Nacional para el Conocimiento y Uso de la Biodiversidad), Mexico, co-PI, Gerardo A. Salazar, Instituto de Biología, Universidad Nacional Autónoma de México, Award ME010, Floristics, Biodiversity, and Traditional Ecological Knowledge in the Sierra Nororiental of Puebla, Mexico; and NEH Office of Digital Humanities, Awards HD-228866-15 and HAA-258602-18 "Comparative Ethnobiology in Mesoamerica: A Digital Portal for Collaborative Research and Public Dissemination."

At the beginning of the reference section the biologists who provided identifications to species of collections cited this paper are listed. Well over a hundred more botanists and entomologists provided identifications of other collections within the areas of their expertise. The invaluable help of all is greatly appreciated. Particularly important for specimen management and distribution has been the support and collaboration of the following individuals at collaborating institutions: Lawrence Dorr and Erika Gardner at the Department of Botany, Smithsonian Institution; Gerardo A. Salazar and David Gernandt at the Instituto de Biología, Universidad Nacional Autónoma de México; and Allen Coombes at the Jardín Botánico Universitario, Universidad Nacional Autónoma de México. Martin Ricker and Jesús Romero Nápoles generously assisted in processing the necessary collection permits. Pedro Acevedo, Frank Almeda, Allen Coombes, and

Douglas Stevens, who in addition to identifications in their own areas of taxonomic expertise, were always willing to provide initial determinations based on photos, enabling the efficient distribution of plant specimens to the proper expert taxonomists. Finally, my appreciation to Susan Penfield for inviting me to participate in the original workshop and in overseeing the editorial process of this special issue.

References

Acknowledgments

Mutillidae	Kevin Williams
Meloidae	John Pinto
Vespidae	James Carpenter, Matt Buffington
Apidae	Charles Michener, Victor González, Ismael Hinojosa, Terry Griswold, Ricardo Ayala, Robert Minckley, Sam Droege, Sandra Rehan
Cerambycidae	Steven Lingafelter
Formicidae	William Mackay, John Longino, Phil Ward, Rafael Achury, Andy Suarez, Ted Schultz, Sean Brady, Jeffrey Sossa-Calvo, Robert Johnson
Pompilidae	James Pitts
Mantidae	Gavin Svenson
Phasmatodea	Enrique Mariño
Reduviidae	Thomas Henry
Syrphidae	Christian Thomason
Passalidae	Christian Beza-Beza
Lantana	Roger Sanders
Asclepediaceae	Douglas Stevens
Sphecidae and Crabronidae	Wojciech Pulawski and Michael Ohl

Amith, Jonathan D., and Steven W. Lingafelter. 2016. Ethnoentomological and distributional notes on Cerambycidae (Coleoptera) of Guerrero and Puebla, Mexico, with notes on other insects. Submitted to [Pending].

Andersen, Henning. 2003. "Introduction," in Henning Andersen, ed., *Language Contacts in Prehistory: Studies in Stratigraphy*. Philadelphia: John Benjamins, pp. 1–10.

Andersen, Henning, ed. 2003. Language Contacts in Prehistory: Studies in Stratigraphy (Papers from the Workshop on Linguistic Stratigraphy and Prehistory at the Fifteenth International Conference on Historical Linguistics, Melbourne, 17 August 2001. Current Issues in Linguistic Theory 239. Philadelphia, Penn.: John Benjamins.

Balée, William, and Denny Moore. 1991. "Similarity and variation in plant names in five Tupi-Guarani languages (Eastern Amazonia)." *Biological Sciences* 35(4):209–62.

Berlin, Brent. 1992. *Ethnobiological classification: Principles and categorization of plants and animals in traditional societies*. Princeton, N.J.: Princeton University Press.

Berlin, Brent, and A. Kimball Romney. 1964. Descriptive semantics of Tzeltal numeral classifiers. *American Anthropologist* 66: 79–98.

Berlin, Brent, Dennis E. Breedlove, and Peter H. Raven. 1968. Covert categories and folk taxonomies. *American Anthropologist* 70:290–99.

Berlin, Brent, Dennis E. Breedlove, Robert M. Laughlin, and Peter H. Raven. 1973. "Cultural significance and lexical retention in Tzeltal-Tzotzil ethnobotany. In Munro S. Edmonson, ed., *Meaning in Mayan Languages*. The Hague: Mouton, pp. 143–64.

———. 1969. "Lexical retention and cultural significance in Tzeltal-Tzotzil comparative ethnobotany." Paper presented at the 68th Annual Meeting of the American Anthropological Association, 20–23 November, 1969.

Bianchi F. A. 1962. Notes on the biology of *Cissites auriculata* (Champion) (Coleoptera: Meloidae). *Proceedings of the Hawaiian Entomological Society* 18:111–19.

Bostoen, Koen. 2007. Bantu plant names as indicators of linguistic stratigraphy in the western province of Zambia. Doris Payne and Jaime Peña, eds., *Selected Proceedings of the 37th Annual Conference on African Linguistics*, pp. 16–29.

Bowern, Claire. 2007. "On Eels, Dolphins, and Echidnas: Nyulnyulan Prehistory through the Reconstruction of Flora and Fauna.: In Alan Nussbaum, ed., *Verba Docenti: Studies in historical and Indo-European Linguistics, Presented to Jay H. Jasanoff by Students, Colleagues, and Friends.* Ann Arbor: Beechstave Press, pp. 39–53.

———. (work in progress with Patrick McConvell). 2011. "Reconstruction in ethnobiological systems of Australian languages." PowerPoint presentation. Australian Nacional University.

Brown, Cecil. 1974. Unique beginners and covert categories in folk biological taxonomies. *American Anthropologist* 76:325–26.

Byerzychudek, P. 1981. *Asclepias*, *Lantana* and *Epidendrum* a floral mimicry complex. *Biotropica*. 13:54–58.

Campbell, Lyle, Terrence Kaufman, and Thomas Smith-Stark. 1986. "Meso-America as a linguistic area." *Language* 62:530–70.

Chízmar Fernández, Carla, et al. 2009. *Plantas comestibles de Centroamérica.* Santo Domingo de Heredia, Costa Rica: Instituto Nacional de Biodiversidad.

Dakin, Karen. 2003. "Uto-Aztecan in the linguistic stratigraphy of Mesoamerican prehistory," in Henning Andersen, ed., *Language Contacts in Prehistory: Studies in Stratigraphy*. Philadelphia: John Benjamins, pp. 259–88.

DiCanio, Christian, C. Zhang, J. D. Amith, R. Castillo García, and D. H. Whalen. forthcoming. Phonetic structure in Yoloxóchitl Mixtec consonants. *International Journal of American Linguistics.*

———, Hosung Nam, Jonathan D. Amith, Rey Castillo García, D. H. Whalen. 2015. Vowel variability in elicited versus running speech: Evidence from Mixtec. *Journal of Phonetics: Special Issue on the Impact of Stylistic Diversity on Phonetic and Phonological Evidence and Modeling* 48:45–59.

———, Jonathan D. Amith, Rey Castillo García. 2014. The phonetics of moraic alignment in Yoloxóchitl Mixtec. *Proceedings of Fourth International Symposium on Tonal Aspects of Languages*, Nijmegen, The Netherlands, 13-16 May.

———, H. Nam, D. H. Whalen, H. T. Bunnell, J. D. Amith, and R. Castillo García. 2013. Using automatic alignment to analyze endangered language data: Testing the viability of untrained alignment. *Journal of the Acoustical Society of America*, 134:2235–46.

———, H. Nam, D. H. Whalen, H. T. Bunnell, J. D. Amith, and R. Castillo García. 2012. Assessing agreement level between forced alignment models with data from endangered language documentation corpora. Proceedings of INTERSPEECH, 2012.

Portland, Oregon.

Ellen, Roy. 1993. *The cultural relations of classification: An analysis of Nuaulu animal categories from central Seram.* Cambridge, England: Cambridge University Press.

Firth, John R. 1957. A synopsis of linguistic theory 1930–1955. In *Studies in linguistic analysis*, 1–32. Oxford: Blackwell.

Fowler, Catherine S. 1972a. "Comparative Numic ethnobiology." Ph.D. dissertation, University of Pittsburgh.

————. 1972b. "Some ecological clues to Proto-Numic homelands." In D. D. Fowler, ed., *Great Basin Cultural Ecology: A Symposium*. Reno, Nev.: Desert Research Institute Publications in the Social Sciences, no. 8, pp. 105–21.

————. 1983. Lexical clues to Uto-Aztecan prehistory. *International Journal of American Linguistics* 49:224–57.

————, and Joy Leland. 1967. Some Northern Paiute native categories. *Ethnology* 6:381–404.

Fox, William J. 1895. Third report on some Mexican Hymenoptera, principally from Lower California. *Proceedings of the California Academy of Sciences*. 2a series 5:260–72.

Friedrich, Paul. 1970. *Proto-Indio-European Trees: The Arboreal System of a Prehistoric People*. Chicago: University of Chicago Press.

González, Victor, Timothy Stein, Jonathan D. Amith, Ricardo Ayala. 2014. New record and the nest of the nocturnal sweat bee *Megalopta tetewana* (Hymenoptera: Halictidae). *Pan-Pacific Entomologist* 90(1):40–43

Grinevald, Colette. 2000. A morphosyntactic typology of classifiers. In Gunter Senft, ed., Systems of nominal classification. Cambridge: Cambridge University Press, pp. 50-92.

Groark, Kevin P. 1996. Ritual and therapeutic use of 'hallucinogenic' harvester ants (Pogonomyrmex) in native south-central California. *Journal of Ethnobiology* 16:1-29.

————2001. Taxonomic identity of hallucinogenic "harvester ant" (*Pogonomyrmex californicus*) confirmed. *Journal of Ethnobiology* 21:133-44.

Haspelmath, Martin, and Uri Tadmor, eds., 2009. *Loanwords in the World's Languages: A Comparative Handbook*. The Hague: De Gruyter Mouton (particularly chap. 1: Martin Haspelmath and Uri Tadmor, "The loanword typology project and the World Loanword Database" (pp. 1–34), Martin Haspelmath, "Lexical borrowing: Concepts and issues" (pp. 35–54), Uri Tadmor, "Loanwords in the world's languages: Findings and results" (pp. 55–75).

Hays, Terence E. 1976. An empirical method for the identification of covert categories in ethnobiology. *American Ethnologist* 3:489–507.

————. 1979. Plant classification and nomenclature in Ndumba, Papua New Guinea Highlands. *Ethnology* 18:253–70.

Healey, C. 1978–79. Taxonomic rigidity in folk biological classification: Some examples from the Maring of New Guinea. *Ethnomedizin* 5(3–4):361–84.

Hebert, Paul D. N., Alina Cywinska, Shelley L. Ball, and Jeremy R. deWaard. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London, series B* 270:313–21.

Himmelmann, Nikolaus P. 1998. Documentary and descriptive linguistics. *Linguistics* 36:161–95.

Howell, Steve N. G., and Sophie Webb. 1995. A guide to the birds of Mexico and Northern Central America. New York: Oxford University Press.

Hunn, Eugene S. 1976. Toward a perceptual model of folk biological classification. *American Ethnologist* 3:508–24.

————. 1977. *Tzeltal folk zoology: The classification of discontinuities in nature*. New York: Academic Press.

————. 1982. The utilitarian factor in folk biological classification. *American Anthropologist* 84:

————. 1992. The use of sound recordings as voucher specimens and stimulus materials in ethnozoological research. *Journal of Ethnobiology* 12:187–98.

Hyam, Roger, and R. Pankhurst. 1995. *Plants and their names: A concise dict*ionary. New York: Oxford University Press.

Levin, Beth. 1993. *English verb classes and alternations: A preliminary analysis*. Chicago, Ill.: University of Chicago Press.

Majnep, Ian Saem, and Ralph Bulmer. 1990– *Aps basd skip kmn ak pak ñbelgpal = Kalam hunting traditions*. Auckland, New Zealand: University of Auckland, Department of Anthropology.

Martin, Gary John. 1996. Comparative ethnobotany of the Chinantec and Mixe of the Sierra Norte, Oaxaca, Mexico. Ph.D. dissertation, Department of Anthropology, University of California, Berkeley.

Meroz, Yoram. 2013. "Large-scale vocabulary surveys as a tool for linguistic stratigraphy: A California case study." Paper presented at the 39th Berkeley Linguistic Society Meetings, Feb. 16–17, 2013.

O'Donnell, Sean., and J. H. Hunt. 2013. Group hunting by workers of two Neotropical swarm-founding paper wasps, *Parachartergus apicalis* and *Agelaia* sp. *Insectes Sociaux* 60:369–72.

————. and Frank J. Joyce. 1999. Dual mimicry in the dimorphic eusocial wasp *Mischocyttarus mastigophorus* Richards (Hymenoptera: Vespidae). *Biological Journal of the Linnean Society*. 66:501–14.

Olivares, Roberto, and Jonathan D. Amith. 2011. *Silvestre Pantaleón*. Documentary in Nahuatl. 65 minutes.

Pawley, Andrew. 2001. Some problems of describing linguistic and ecological knowledge. In Luisa Maffi, ed., On Biocultural Diversity: Linking Language, Knowledge, and the Environment. Washington, DC: Smithsonian Institution Press, pp. 228–47.

Richards, Owain Westmacott. 1978. *The Social Wasps of the Americas*. London: British Museum, Natural History

Ricker, Martin, et al. 2015. Resultados: Determinación taxonómica de los ejemplares de herbario del re muestreo del inventario nacional forestal y de suelos 2009–2013 (año 2013). Ms.

Rosch, Eleanor, and Carolyn B. Mervis. 1975. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology* 7:573–605.

Ross, Malcolm, Andrew Pawley, and Meredith Osmond, eds. *The Lexicon of Proto Oceanic*, vol. 1: *Material culture*; vol. 2: *The physical environment*; vol. 3: *Plants*; vol. 4: *Animals*. Canberra, Australia: Australian National University.

Si, Aung. 2011. Biology in language documentation. *Language Documentation and Conservation* 5:169–86.

———— 2016. *The traditional ecological knowledge of the Solega: A linguistic perspective.* New York: Springer

Siebert, Frank T. 1967. "The original home of the proto-Algonquian people." *Contributions to Anthropology: Linguistics I. National Museum of Canada Bulletin* 214:13–47.

Taller de Tradición Oral de CEPEC and Pierre Beaucage. 1988. *Maseualxiujpajmej, Kuesalan, Puebla / Plantas medicinales indígenas, Cuetzalan, Puebla*. Puebla, state of Puebla: DIF (Desarrollo Integral de la Familia).

Taylor, Paul Michael. 1984. Covert categories´ reconsidered: Identifying unlabeled classes in Tobelo folk biological classification. *Journal of Ethnobiology* 4:105–22.}

————. 1990. *The folk biology of the Tobelo people: A study in folk classification*. Smithsonian Contributions to Anthroplogy, Num. 34. Washington, D.C.: Smithsonian Institution Press.

Turner, Nancy. 1973. Plant taxonomic systems and ethnobotany of three contemporary groups of the Pacific Northwest (Haida, Bella Coola, Lillooet). Ph.D. thesis. Dept. of Botany, University of British Columbia.

Whistler, Kenneth W. 1977. "Wintun prehistory: An interpretation based on linguistic reconstruction of plant and animal nomenclature." *Proceedings of the Berkeley Linguistic Society* 3:157–74.

Yang, Ya, et al. 2012. Molecular phylogenetics and classification of *Euphorbia* subgenus *Chamaesyce* (Euphorbiaceae). *Taxon* 61:764–89.