

521 3049

# Making Dictionaries

*Preserving Indigenous Languages of the Americas*

EDITED BY

William Frawley  
Kenneth C. Hill  
Pamela Munro

UNIVERSITY OF CALIFORNIA PRESS  
Berkeley Los Angeles London

## CONTENTS

*This book is dedicated to Ken Hale (1934–2001),  
mentor, colleague, and friend.*

University of California Press  
Berkeley and Los Angeles, California

University of California Press, Ltd.  
London, England

© 2002 by the Regents of the University of California

Library of Congress Cataloging-in-Publication Data

Making dictionaries : preserving indigenous languages of the Americas / edited  
by William Frawley, Kenneth C. Hill, Pamela Munro.

p. cm.

Includes bibliographical references and index.

ISBN 0-520-22995-9 (cloth : alk. paper) — ISBN 0-520-22996-7 (pbk. : alk.  
paper)

1. Indians—Languages—Dictionaries. 2. Indians—Languages—Lexicography.

I. Frawley, William, 1953- II. Hill, Kenneth C. III. Munro, Pamela.

PM136 .M35 2002

497—dc21

2001053046

Manufactured in the United States of America

10 09 08 07 06 05 04 03 02

10 9 8 7 6 5 4 3 2 1

The paper used in this publication meets the minimum requirements of  
ANSI/NISO Z39.48-1992(R 1997) (*Permanence of Paper*).

Making a Dictionary: Ten Issues

*William Frawley, Kenneth C. Hill, and Pamela Munro / 1*

### I. FORM AND MEANING IN THE DICTIONARY

1. Theoretical and Universal Implications of Certain Verbal Entries in  
Dictionaries of the Misumalpan Languages  
*Ken Hale and Danilo Salamanca / 25*
2. Morphology in Cherokee Lexicography: The *Cherokee-English Dictionary*  
*William Pulte and Durbin Feeling / 60*
3. Lexical Functions as a Heuristic for Huichol  
*Joseph E. Grimes / 70*
4. Entries for Verbs in American Indian Language Dictionaries  
*Pamela Munro / 86*
5. Multiple Assertions, Grammatical Constructions, Lexical Pragmatics, and  
the *Eastern Ojibwa-Chippewa-Ottawa Dictionary*  
*Richard A. Rhodes / 108*

### II. ROLE OF THE DICTIONARY IN INDIGENOUS COMMUNITIES

6. Issues of Standardization and Community in Aboriginal Language  
Lexicography  
*Keren Rice and Leslie Saxon / 125*
7. A Dictionary for Whom? Tensions between Academic and Nonacademic  
Functions of Bilingual Dictionaries  
*Leanne Hinton and William F. Weigel / 155*

addition of more words and new texts created no significant problems. A natural next step was consequently to approach a task we had discussed over the years, namely, entering all of Molina's dictionary, and that has been the last goal this far.

Our ambitions have been a thoroughly worked-out system, deliberate decisions, and a high degree of precision. The system works very smoothly, and the results are highly satisfactory. Immediate plans for the future involve the entering of more texts and, above all, making CoNDiP available on the Internet.

The abundance of Nahuatl texts and the great interest among scholars in a better understanding of these texts were the incentives for starting CoNDiP, but it has become increasingly clear to us that the model created for the Nahuatl language and these texts from the sixteenth and seventeenth centuries would be useful and instructive if applied to all other languages.

## CHAPTER TEN

### What's in a Word?

#### *The Whys and What Fors of a Nahuatl Dictionary*

*Jonathan D. Amith*

1. LEXICOGRAPHY AND THE DOCUMENTATION OF THE SPOKEN WORD. Harold Bloom, in beginning one of his essays, states that "the word *meaning* goes back to a root that signifies 'opinion' or 'intention,' and is closely related to the word *moaning*." And he later mentions that "freedom, in a poem, must mean freedom of meaning, the freedom to have a meaning of one's own" (Bloom 1985: 1, 3). This rebellious drive for new meaning is not, of course, the exclusive domain of the poet, nor is new meaning a simple outgrowth of personal will and private design. Particularly in the languages of societies with an extended tradition of literacy, meaning is also part of a social system that generates popular texts and seductive jingles while preserving antiquated texts and a classic tradition. Undoubtedly related to this emphasis on (and, of course, availability of) written texts is the fact that the Western lexicographic tradition, as exemplified by studies in philology and etymology, considers variations in meaning more a temporal than a spatial or social phenomenon.<sup>1</sup> In societies dominated by print media, the modern lexicographer thus becomes a rather hapless figure, entrusted with the Sisyphean

It has been a pleasure to work with Bill Frawley, Ken Hill, and Pam Munro, who were kind enough to invite me to participate in this volume even though my work on Nahuatl lexicography and linguistics is still mostly unpublished. Their attention to details and insightful suggestions have helped this essay considerably and saved me from numerous embarrassments. I would also like to thank my friends and colleagues José Antonio Flores Farfán, Donna Perry, and Roberto Zavala for their critical readings of earlier versions. All the preceding, however, are absolved of any responsibility for the errors or inconsistencies that remain. The fault for these lies with the usual suspect. The latest version of the Nahuatl Learning Environment can be found at <http://www ldc.upenn.edu/nahuatl>.

1. The temporal variations, however, are often those of very long-term linguistic shifts. For an innovative approach to semantic change over short time periods that uses a cognitive and prototype approach to meaning, see Geeraerts 1997.

task of documenting the never-ending and now seldom forgotten linguistic creativity of literate others. Yet alongside this exacerbating and increasingly futile quest for completeness there is another tendency—that for purity of form and orthodoxy of meaning. Lexicography, therefore, often becomes a sort of institutionalized antipoetics. New meanings (in Bloom's term, "moanings") are sanctioned by a self-styled board of *parole*<sup>2</sup> (in certain countries a nationally certified "academy"), an increasingly besieged gatekeeper determining which individual utterances may enter "the prison-house of language," a collective abstraction of both structure and meaning.<sup>3</sup>

For lexicography, therefore, literacy—or perhaps better put, a tradition of the printed word—clearly matters.<sup>4</sup> An early focus on the "great divide" between orality and literacy, and on the autonomy of literacy as an independent variable affecting a wide range of social and cognitive activities, has been deservedly critiqued by researchers in a variety of fields, including social anthropology, sociolinguistics, cultural psychology, rhetoric, folklore, and history.<sup>5</sup> The point is not that societies with a written tradition present a more homogeneous linguistic community but that the printed text often represents a secularization and standardization of language that filters *parole* before it even reaches the lexicographer's desk.<sup>6</sup> Dictionary makers, however, have perhaps been less concerned than other linguists

2. An editor at Merriam-Webster once told a friend of mine, who had offered as a neologism a new and previously undocumented inflected form of a word, that a word usage found only once was idiosyncratic but found twice was idiomatic. Though perhaps somewhat overstated, the editor's assertion, a rather free adaptation of the legal adage that "twice makes a custom," certainly brings to the fore one key problem in lexicography: the number of independently documented occurrences needed to substantiate a valid dictionary entry. In the same vein, for a short sketch on neologisms in English-language dictionaries, see Schoen Brockman (1999), who gives *fuggedaboudit* as a word "[on] the waiting list, pending more evidence," of the *Oxford American Dictionary*.

3. The term is taken from Jameson (1972), who notes, in reference to Saussure's dichotomization of *langue* and *parole*, that "thus, at one stroke, all purely articulatory matters, all questions of local accent, mispronunciation, personal style, are eliminated from the new object under consideration, becoming themselves problems for a different science, that of the *parole*" (p. 26).

4. Linnel (1982, 1988) explores the impact of literacy and writing on linguistics. According to Linnel (1988: 47), it was Vološinov who in the early part of this century (before 1930) had suggested it was the work of linguists and philologists on written texts in foreign languages that had necessitated the construction of dictionaries, with their standard definition presupposing that lexical meanings can be given in terms of a fixed configuration of semantic features. See also the collection of articles in Frawley 1981.

5. This list of fields is that presented by Besnier (1995: 3), who, in his introduction, cites the relevant literature, particularly that which constituted a critical response to the works of Goody, Havelock, and Ong. For a similar critical response, see also Street 1984, 1988, and Street's other works cited in Street 1988.

6. For a notable and impressive exception to common lexicographic practice in English, see Cassidy and Hall 1985--.

with the different ways in which written and nonwritten languages relate to the particular problems that are the topic of their research. Recognizing this, one linguist, Ragnar Rommetveit (1988: 15), has referred to the "written language biases within various branches of semantics," a bias that becomes clear not only in any quick glance at textbooks and theoretical treatises on semantics but in much of the literature on lexicography as well.<sup>7</sup> In general, then, little work has been done on what may be loosely considered *comparative lexicography*, the ways in which the structure and content of dictionaries (and the lexicon/grammar interface) can and should be adapted to the particularities of specific cases—speech communities, lexical corpora, and target audiences.

These particularities, clearly, go far beyond the simple fact of whether the documented language is written or spoken. Much more basic and problematic is the task of identifying or determining the language community that serves as the source of lexical data and the reading or listening community that is the target of the lexicographic project. Secret languages, the specialized lexicons and grammars of rituals and initiation, the esoteric language of religion, verbal dueling, and an innovative poetics of expression as well as a rote poetics of tradition<sup>8</sup> all belie any facile claims of a unified language source group, even in communities that at first glance appear closed and cohesive.<sup>9</sup> Returning somewhat to Bloom's metaphor for meaning, what a field lexicographer hears when he or she enters a community that speaks an undocumented or little-documented language is often a harrowing cacophony, a perplexing admixture of sound and meaning that in the best of circumstances can be associated with particular registers, contexts, or social groups but in the worst seems distressingly idiosyncratic, located at the edge of what some would call "correct" and "incorrect" speech. Eventually these puzzles are "resolved"—in my experience first according to form (phonology and orthography), then according to function (morphosyntax and grammar), and finally (and most fastidiously) according to meaning (semantics and lexicon), the three essential elements of a lexicographic representation.

7. This is apparent in even the most cursory review of the proceedings of the various EuroLEX conferences, as well as in "state-of-the-art" collections on formal semantics (Lappin 1996).

8. Poetics, from this perspective, involves a highly formalized structure of discourse that is oriented not to breaking new semantic ground but to encapsulating a static, at times highly ritualized, form. Thus Berman (1983: 59) notes in regard to early legal traditions that "the dramatic and poetic qualities of Germanic law were associated with the plasticity of its substance. . . . The expression of legal rules in poetic images helped to stamp them on the memory," adding that "the earliest Irish law was expressed in the form of poetry."

9. See the bibliographies in the review articles by Cicourel (1985) and Chafe and Tannen (1987).

If the fragmented nature of the "linguistic community" is one source of uncertainty, the diversity of the potential readership of a lexicon is another. Both factors must be considered in dictionary design, the first more in regard to input, the latter in regard to output. In the case of Nahuatl—considering its function in the colonization and conversion of New Spain and its present role as the principal linguistic icon of Mexican national identity promoted by the nation-state and popularly endorsed in much of central Mexico and among the Chicano community in the United States—the groups that might use a dictionary (linguists, historians, anthropologists, theater and dance troupes, heritage language speakers, and simply the curious, as well as native speakers) would perhaps be more diverse than those groups that would wish to access similar material for other native languages of Latin America.<sup>10</sup> A particularly challenging aspect of Nahuatl lexicography, therefore, is to create a basic corpus that can be used to generate material (through electronic manipulation of the original input) that would be of use to readers as diverse as comparative linguists, on the one hand, and native speakers, on the other.

With such a variety of language sources and pedagogical ends, and trying to meet many needs in a single text, lexicography at times threatens to become a bottleneck, a constricted space of lexical representation that has resulted from standardized materials and narrowed goals. One reaction to this situation is exemplified by the increasing tendency to specialization within the lexicography of major languages: native-speaker dictionaries, collocational dictionaries, monolingual learners' dictionaries, bilingual learners' dictionaries, translators' dictionaries, visual dictionaries, dictionaries (mono- and bilingual) of particular terminologies, explanatory combinational dictionaries, among many others.<sup>11</sup> Although such a fragmented approach provides an inappropriate model for the lexicography of less documented languages—if simply because, given the limited resources (human and financial) for such endeavors, there is little chance of funding for the independent development of such specialized tools—it does suggest that there is an underlying centrifugal force that constantly threatens to pull apart any effort to construct a single "jack-of-all-trades" lexicon.

Another solution (and that undertaken for the Nahuatl material that I have studied) involves a methodology that takes advantage of electronic media to store vast amounts of material in an open-ended database constructed in a flexible and multilevel format. In essence this approach

10. Recently, in fact, I have been contacted by prisoners who wish to learn Nahuatl (as well as by department of corrections officials who want to know what prisoners have been writing) and by an individual who wanted the Nahuatl word for "eternal life" added to a tattoo on his back.

11. Snell-Hornby (1990: 232), for example, suggests that bilingual dictionaries should, like their monolingual counterparts, become "increasingly user-specific."

accepts and indeed embraces the inherent centrifugality of lexicography; for this reason it concentrates not simply on defining and delimiting but on building structures that will link together the very many facets of meaning, particularly those that involve the articulation of syntax and semantics.<sup>12</sup> Though dealing with an entirely different range of materials (those from dominant Western languages), several lexicographers have already suggested the direction such an approach might take. Beryl T. Atkins, for example, calls for theoretical linguists "to devise a typology of vocabulary items and a parallel typology of defining strategies suited to each" (1992–93: 26), and others, noting the promise of electronic reference works, have claimed the need for "an almost total restructuring of the way in which [words] are treated in lexicography" (Atkins, Kegl, and Levin 1988: 110).<sup>13</sup> Still others have stressed the study of syntagmatic over paradigmatic relations and have questioned the appropriateness of the word as the primary unit of analysis and presentation, calling for a greater attention to collocation (e.g., Cop 1990; Meyer 1990).<sup>14</sup> Additional shifts in focus should also be considered. It seems clear that in the best of circumstances studies of indigenous, nonwritten languages should be constructed more along the lines of an encyclopedia or "cultural dictionary" than according to the model of a bilingual translators' dictionary with its focus on substitutability of target language for source language terms. Care might also be taken to situate words within semantic and morphosyntactic categories that would prove useful to determining both meaning and

12. The importance of using electronic databases to relate lexical semantics to syntactic behavior is suggested by Pustejovsky, who states, for example, that "it will soon be difficult to carry out serious computational research in the fields of linguistics and N[atural] L[anguage] P[rocessing] without the help of electronic dictionaries and computational lexicographic resources" (1995: 5). See the reference to Dixon 1984 in note 42 below.

13. The original citation has "verbs" instead of "words." With this change, the full quote presents a position similar to that elaborated in the pages below: "We would argue that if dictionaries are to take a quantum leap rather than simply a series of tottering steps into the future—particularly into a future which holds out the enticing prospect of electronic reference works—then what is necessary is not merely a matter of elaborating or modifying existing entries, but rather an almost total restructuring of the way in which verbs are treated in lexicography. Any given verb participates in only a subset of the possible alternation patterns. However, if a dictionary is to provide comprehensive and consistent information about alternations, these must at least be available to the dictionary designer and to the lexicographer during the process of compilation" (p. 110).

14. Cop, for example, stresses a semasiological (receiving rather than producing texts) approach to collocations that focuses on the collocator rather than the base. In discussing the design of monolingual learners' dictionaries, Rundell (1988: 134) argues for a dictionary design that "rather than treating meaning as central, . . . give[s] equal weight to all relevant features, including grammar, style and register, collocational properties, pragmatic and connotative features, relationships of synonymy and hyponymy, contextual and syntagmatic preferences, and so on." Mackin (1978) briefly discusses work on collocations in the *Oxford Dictionary of Current Idiomatic English*.

use. Previously such solutions were unrealizable, given the problem of manuscript length, a prime consideration in printed versions, though a virtually nonexistent factor in electronic publishing (CD-ROMs/DVDs and Web sites). The most practical solution to the tension between a drive toward more complete lexicographic studies and cutbacks in the publishing industry would be to produce the complete lexicographic study in electronic format and issue a reduced, practical synthesis in print that would interface with the former.

This section began with a discussion of meaning and of the persistent tension between lexicography and poetics, between the documentation and the fabrication of meaning. It was suggested that a parallel tension exists between written and spoken discourse, not necessarily between these genres of expression per se, but in regard to lexicographic practice—which tends to the prescriptive when the source is the former, descriptive when the source is the latter. A starting point for a new lexicographic approach to the study of indigenous languages was mentioned: the need to design a dictionary project according to the characteristics of the language and language community studied, the nature of the documentation to be processed or produced with the aid of the dictionary, and the needs of the users who will seek access to the final product. Word DEFINITIONS, in this presentational strategy, provide only one facet (albeit perhaps the most basic) of a multidimensional approach to lexicography and, more generally, to the exploration and learning of a less common language for which oral texts provide the basic data.

The goal of this chapter, then, is not to outline the specific problems that Nahuatl poses for lexicography (although some of these will briefly be mentioned) but to explore the manner in which electronic media can be particularly helpful in dealing with issues of lexicography and semantics that have often been mentioned but, in printed lexicons, are difficult to address. The incipient project that I briefly describe below, the Nahuatl Learning Environment, makes use of Internet technology that not only provides a unique tool for exploring the intricacies of the Nahuatl language, but facilitates the cross-linking of corpus, lexicon, and grammar in a single learning and research environment.<sup>15</sup> One goal

15. This learning environment (now at <http://www ldc.upenn.edu/nahuatl>) was made possible by the extraordinary support I received from the Linguistic Data Consortium at the University of Pennsylvania. Mark Liberman, the director, has been supportive of the Nahuatl project from the beginning and has given generously of his time and resources. Steven Bird has been most directly responsible for the current search engine for the on-line dictionary of Ameyaltepec Nahuatl. He spent days adapting his *Hyperlex* program to my material and has given me invaluable advice on computational linguistics and lexicography in general. Brian Robinson of Yale University placed the early grammar lessons and exercises on the Yale server (<http://www.yale.edu/nahuatl>), wrote the script now being used, and established the CGI links between the lessons and exercises and the lexicon. My sincere thanks to all these individuals.

of this project is to develop a model that joins linguistic research and language pedagogy by providing the grammatical context for using dictionary material while furnishing the appropriate lexical base to operationalize a learned grammar. A more comprehensive and long-term goal is to use electronic and Internet technology to link corpus, lexicon, and grammar so that both specialists and nonspecialists can access primary linguistic data (corpus), semantic interpretation (lexicon), and structural analysis (grammar) and in this way more effectively study and learn a language (see fig. 10.1). At the same time, it will then be possible for others to analyze and critique the interpretations offered by the original researcher. The first two sections below explore how these goals—heuristically approached through a discussion of the complementary goals of research and pedagogy—can be met. The third section briefly discusses the relationship between indigenous-language dictionaries and the indigenous community.

2. RESEARCH: ENTRY DESIGN AND LEXICON STRUCTURE. For just over four and a half years I lived in two neighboring Nahuatl-speaking communities located near the Balsas River in central Guerrero—Ameyaltepec (3 years) and San Agustín Oapan (1½ years). During this period, I conducted ethnographic fieldwork on intervillage relations and at the same time studied Nahuatl—continually writing down on file cards phrases heard in everyday speech, asking friends about the meaning and use of specific words, and taking notes for a reference grammar. By the end of my fieldwork period I had amassed approximately 20,000 such cards (organized by “headword”), the vast majority documenting phrases I had heard in Ameyaltepec, although approximately 2,000 to 3,000 have words and phrases from San Agustín Oapan. Approximately one hundred hours of recordings from various villages will eventually yield interlinearized textual material that will provide additional lexical and syntactic information and eventually be incorporated into the electronic database as primary source material.

Over the past decade, I gradually began to input the Ameyaltepec phrases into an open-ended database format, a simple text file with field delimiters and separators. Later this data was imported into Shoebox, a program written by JAARS, the software developer associated with the Summer Institute of Linguistics. Shoebox permits both flexible field formats and fonts and rapid searches and filters; it also simplifies export according to style sheets (for publication) or in delimited format (for use with other database search engines). Preliminary fields (see partial list in table 10.1) were established for the following purposes: (1) to input primary lexical material along with English and Spanish translations; (2) to facilitate cross-referencing among entries

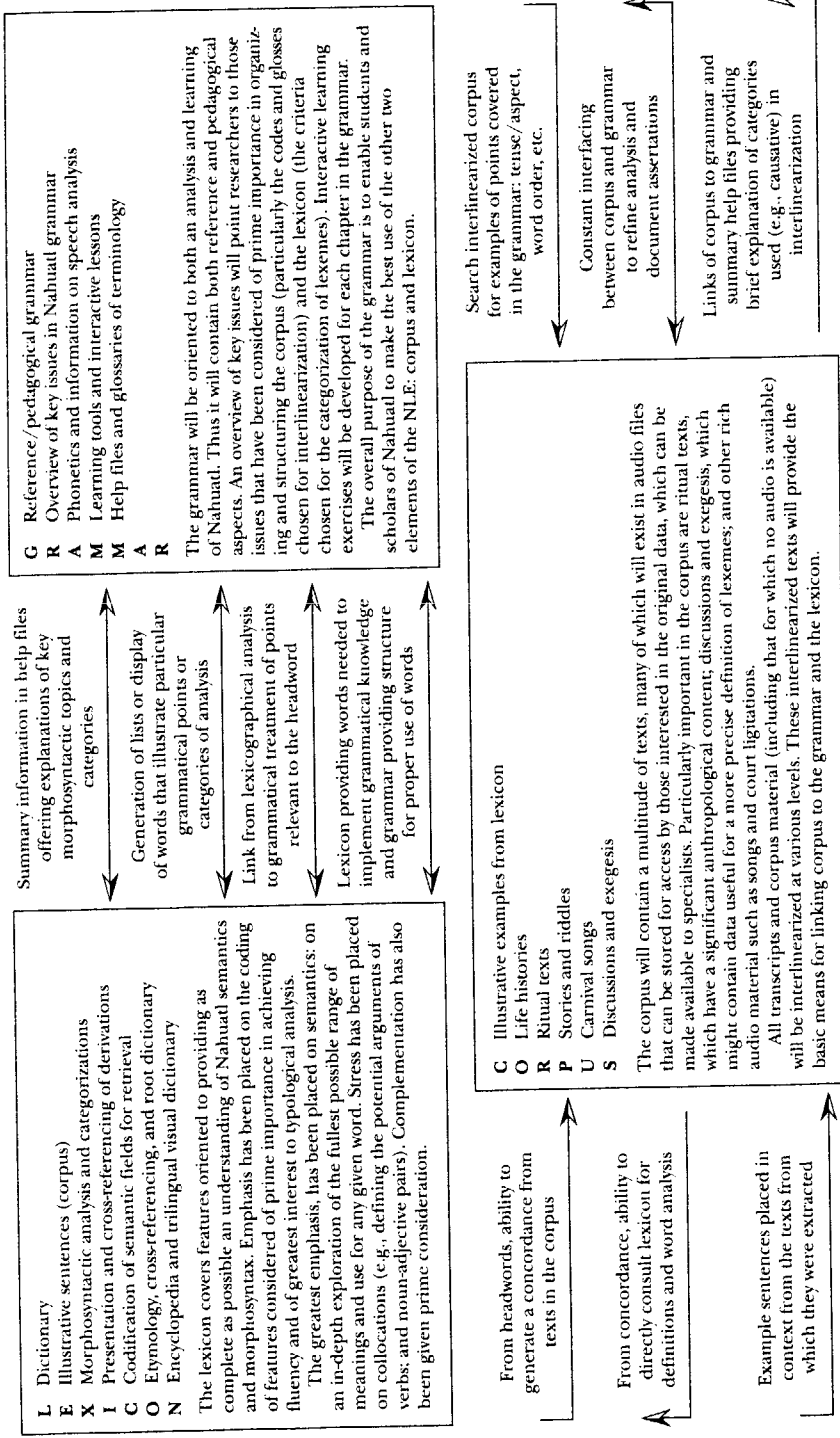


Figure 10.1. The Nahuatl Learning Environment: Schematic Representation of Links between Elements.

TABLE 10.1. Fields for Ameyaltepec Nahuatl Database Entry

| Field Delimiter | Data Input   |
|-----------------|--|
| \w              | Headword   |
| \ew and \sw     | English/Spanish definition   |
| \ee and \ss     | English/Spanish glosses for searching, dictionary reversal, and interlinearization   |
| \cat            | Grammatical category and derivational morphology (verb, noun, demonstrative, quantifier; or applicative, causative, denominal, deverbal; e.g., <i>V-3-d-ca</i> indicates a derived ditransitive causative; <i>V-1-b</i> indicates a basic, underived intransitive) |
| \com            | Schematic part of speech representation for compounds (e.g., <i>N-V1</i> indicates noun stem incorporated into an intransitive verb; <i>V-ka-V2</i> indicates a participial verb, ending in <i>-ka</i> ; incorporated into a transitive verb)                      |
| \mor            | Morphological analysis (e.g., <i>katlatla</i> would be <i>kat-tlatla</i> house-burn; <i>cho:ktia</i> would be <i>cho:k(a)-tia</i> cry-CAUS)  |
| \inf            | Verb class for inflectional paradigms, or patterns of noun pluralization and possession  |
| \sem            | Semantic field (e.g., <i>Nat-pl-med</i> for medicinal plants; <i>bd</i> for body parts, <i>son</i> for sounds, <i>tex</i> for textures)  |
| \p__            | Illustrative phrases   |
| \ep__ and \sp__ | English/Spanish translations of illustrative phrases   |
| \x__            | Various cross-references for roots and the basic stems of derivations  |
| \enc and \snc   | English/Spanish cultural and semantic discussions and notes  |
| \nt             | Notes and observations for future study or reference   |

(for instance, by linking base verbs to their causative and applicative forms or by giving stems that can be linked to produce a root dictionary); and (3) to provide fields for the coding of semantic and morphosyntactic categories (such as agentives, colors, and medicinal plants) that clarify meanings while providing data for linguistic and cultural analysis.

The decision as to how to list headword entries is at times fairly straightforward, with several caveats. Intransitive verbs are entered in the least marked form, the third-person singular present indicative. Transitive verbs are recorded by stem, without the object prefix that would be necessary for a well-formed word in Nahuatl, a head-marking language. For various reasons, including their somewhat irregular formations, causatives are given separate entries (e.g., *wetska* 'to laugh' and *wetski:tia* 'to make laugh'), whereas anticausatives, which use the reflexive marker

to detransitivize, are not separately listed.<sup>16</sup> In most of these cases relating to transitive and intransitive verbs, there is little difficulty in selecting the proper entry form. The major decision has been to follow tradition and list transitives without any object prefix, resulting in headword entries that are not naturally occurring words.<sup>17</sup>

At other times, however, the determination of entry format and of what words to include is problematic. This is particularly true in regard to incorporation, compounded forms indicating several types of motion, and reduplication.<sup>18</sup> Several criteria are relevant to a decision as to whether to accord a specific form of an incorporated, compounded, or reduplicated word a separate entry. Such words that have been lexicalized or whose meaning is not predictable are listed as headwords; forms whose meaning is covered by highly productive processes described in the grammar are not separately listed. In general an effort has been made to avoid a proliferation of entries and to keep together under one headword variants (such as short- and long-vowel reduplication of verb stems, different possessive markers on a nominal stem) that might provide interesting data for comparison (e.g., the differences between two types of reduplication or possession). An electronic format solves the most basic problem posed by more consolidated entries, since users will be able to employ a search function to find forms (such as *wa:lnemi*, discussed below) that have been listed as senses under a more basic headword (in this case, *nemi*).

Incorporation of certain nominal stems, such as body parts, is highly productive in Nahuatl. The decision of whether to establish a separate entry is

16. Thus one finds both *tlapowi* 'to become open' and *tlapowa* 'to open' as headwords but only *tsakwa* 'to close' and not the anticausative *no-tsakwa* REFL-close 'to become closed'.

17. This format follows that used in many Nahuatl dictionaries, particularly those from the colonial period, and has the weight of tradition behind it. An exception is Brewer and Brewer 1971, which lists transitives with the third-person object prefix *k(i)-*. Note that with intransitives, the third-person subject marker is the zero morpheme; the decision here has been to translate entries with the infinitive, which does not exist in Nahuatl. Below I discuss the utility of changing this format for a dictionary produced especially for native speakers. These points, as well as questions regarding the proper morphophonemic representation for certain stems, can be intensely debated and I have deliberately refrained from touching on all these points. Instead, a brief exposition of problems of entry format and selection associated with incorporation, compounding, and reduplication is offered to acquaint readers with some problems of headword entry design that are particularly prevalent in Nahuatl.

18. Another problem not dealt with here relates to criteria for separate listings of "extrinsically" and "intrinsically" possessed forms (e.g., in regard to *nakatl* 'meat' → *inakaw* 'his meat' and *inakayo* 'his flesh (of the body)'; or in regard to *kahli* 'house' → *ikal* 'his house' and *ikahlo* 'its roof (of a house)'). In general both types of possession are listed under the same headword, though accorded different senses. Only when a *-yo* form exists as a freestanding noun (with an absolutive ending) is it separately entered (e.g., *a:tl* 'water' → *ia:w* 'his water'; *a:yo:tl* 'juice' → *ia:yo* 'its juice'); in these cases each absolutive form is cross-referenced.

usually based on the natural occurrence of a lexeme in the corpus; with few exceptions no effort has been made to produce words that are perfectly well formed but unlikely to occur. Thus *ma:posteki* 'to break one's arm' (*ma:posteki* 'arm-break'), which is found throughout the corpus, is listed as a headword, whereas the perfectly well formed *mapiiposteki* 'to break one's finger' (which did not occur in the corpus, but which any speaker would accept, and which might well occur in a much larger corpus) is not. With incorporation, then, the criteria of inclusion is one of lexicalization, natural occurrence in the corpus, and lack of predictability of meaning.

Nahuatl verb morphology includes a series of elements that tend to indicate some sort of spatial, and occasionally temporal, motion: two prefixes (*on-* and *wa:l-*) and three pairs of suffixes (*-ti/-ki*, *-to/-ko*, and *-to:ya/-ko:ya*); the first of each pair indicates movement away from a deictic point of reference and the second toward such a point. In general if a verb takes one set (e.g., intraverse motion) it will take the other. In certain cases, however, only one form occurs. Thus *nemi* 'to live', or 'to be present (at a specific location)' occurs as *wa:lnemi* 'to be born'; the past tense is *o:nemiko* 'to have been born'. Here simply a separate sense under the headword *nemi* is given. In another case separate entries are given for *one:wa* 'to depart (away from a deictic reference point)' and *wa:le:wa* 'to come (toward a deictic reference point)'. These forms have been lexicalized: the intransitive verb *e:wa* no longer exists without a directional, and with *one:wa* and *wa:le:wa* the prefixes no longer alter with suffixes in certain tenses and aspects (thus *one:was* 'he will leave' and *o:one:w* 'he left', but compare the alternation in *onitlakwa:s* 'he will go to eat' and *o:tlakwa:to* 'he went to eat').<sup>19</sup>

There is also a series of verbs that appear as suffixes and indicate movement associated with a primary event. For example, *-tasi* refers to an action carried out upon arrival at a particular location away from a deictic reference point: *tlakwa* 'to eat' forms *tlakwatasi* 'to eat upon arriving there'. In general an effort has been made to avoid separately listing these forms of associated motion unless the meaning is not predictable from the component elements. Under this criterion *tlakwatasi* is not accorded a separate entry, but *komo:ntasi* 'to hit the ground with a thud' (derived from the verb *komo:ni* 'to make a booming sound or to thunder') is separately entered; the related form *komo:nateko* 'to hit the ground with a thud (toward a deictic reference point)' is also listed, but users are here pointed to the more common *komo:ntasi*.<sup>20</sup> In other cases in which the meaning of such compounds

19. Nahuatl does not distinguish gender. However, following certain conventions that perhaps should be avoided, I have translated the third-person singular pronoun as 'he'.

20. The endings *-tasi* and *-teko* differ only in terms of deixis; *-teko* indicates an action or event that occurs when the subject arrives at a point toward a deictic reference point (usually the speaker). Thus *tlakwatateko* 'he (will) eat upon arriving here'.



might seem idiosyncratic, an example of the relevant form is included under the entry for the principal verb. Thus under *isa* 'to wake up' one finds *isate:wa* 'to suddenly wake up'. With the set of elements described in this paragraph, an effort has been made to avoid separate listings. When meaning is not predictable, however, either a separate entry is given or a separate sense (appropriately marked) is given under the headword.

Finally, there is the problem of reduplication. Reduplicated nouns such as *kakahli* 'canopy' ← *kahli* 'house' and *susuwa:tl* 'effeminate man' ← *suwa:tl* 'woman' are accorded their own entries. Such forms are lexicalized, and their meaning (which involves a metaphoric extension from the nominal base) is not completely predictable. The same pattern of reduplication can be used to indicate toy or play items: *metlatl* 'metate' → *memetlatl* 'toy metate'. This is a productive, though rarely used, construction. Yet to avoid a proliferation of entries, such reduplicated words are not separately listed.<sup>21</sup> Reduplication of adjectives is used mostly to indicate plurality; no separate entries are assigned to these forms.

Unlike the case with nouns, the reduplication of verbal stems is highly productive, a situation that according to the criteria I have established argues against separate listings. In Ameyaltepec, reduplication occurs in two basic shapes: (C)V- and (C)V:-.<sup>22</sup> Often there is a clear semantic difference between each form:

|           |                               |                             |  |
|-----------|-------------------------------|-----------------------------|--|
| no:tsa    | 'to address or call to'       | kino:tsa itah <sup>23</sup> | 'his father calls to him'                      |
| nono:tsa  | 'to converse with'            | kinono:tsa itah             | 'his father talks to him'                      |
| no:no:tsa | 'to give advice'              | kino:no:tsa itah            | 'his father counsels him'                      |
| teki      | 'to cut (in a single action)' | kiteki                      | 'he cuts it (i.e., once, all the way through)' |
| teteki    | 'to cut (repeatedly)'         | kiteteki                    | 'he cuts it (e.g., cloth with scissors)'       |

21. For convenience, irregular reduplicated plural forms, which are now rare, are separately listed, although the user is simply remitted to the singular. Thus under *ko:koneh* one finds "pl., see *kone:tl*."

22. In many other dialects the first pattern of reduplication is (C)Vh- (e.g., *kinohno:tsa*), which in Classical Nahuatl would be (C)V?-(*kino?no:tsa*). In Ameyaltepec, word-internal *h* is deleted at the surface level, which leads to problems for interdialectal comparison, some of which are discussed below.

23. Ø-ki-no:tsa-Ø      Ø-i:tah-Ø  
3sgS-3sgO-call-SG.IND      3SGS-3SGPOSS-father-SG.POSSED

|         |            |           |   |
|---------|------------|-----------|---|
| te:teki | 'to slice' | kite:teki | 'he slices it (e.g., bread with a knife)' |
|---------|------------|-----------|---|

One goal in organizing data input into the lexicon is to identify areas of interest to potential users and to structure data entry to facilitate easy retrieval and analysis. Given that the most salient problems will vary among different languages, dictionary design should vary accordingly. In Nahuatl, one particular problem needing research is the range of semantic concomitants to verbal reduplication. Entries into the Ameyaltepec lexicon have therefore been designed to maintain reduplicated forms of a single verb within one entry, providing for easy and rapid comparison between base verbs and both (C)V- and (C)V:- reduplications.<sup>24</sup> A provisional English-only entry for *no:tsa* exemplifies this organization of data:<sup>25</sup>

**no:tsa** (*V trans.*; REG. INFLEC. see **ki:sa**; ROOT: *no:tsa*) *applic.* → **no:chilia 1:** (often with prefix *on-* or *wa:l-*) to address or call to *Mitsno:tsa*, *itlah mits-ihli:sneki*. He's calling you, there is sth he wants to say to you. **2:** to visit or pay a call on *Nikno:stiki:sa mo:stla*, *a:man xnukaxilia*. I'll pass by and pay a call on him tomorrow, today I don't have time. **3:** to speak to (in the sense of being on good terms with them) *Xne:chno:tsa*, *ne:xtlawe:lita*. He doesn't speak to me, he hates me. **4:** (*rdp-short*) to carry on a conversation with *Timonono:tsan pa:mpa kwahli timowi:kan*. We talk to each other (a lot) because we get along well. **5:** (*rdp-long*) to give advice to or counsel *Ne:chno:no:tsa nona:n para ma:ka itlah nikchi:was*. My mother gives me advice so that I won't do anything (bad).

At present the underlying field markers of illustrative sentences (not visible in the output but marking the fields in the database) encode certain morphological features of the headword as it occurs in the example phrase. One of these features is reduplication (illustrated below). Here the use of distinct field markers facilitates searches for any illustrative phrase that contains a reduplicated form of the headword:<sup>26</sup>

24. When there is significant and unpredictable semantic content to reduplication, the definitions of reduplicated forms are separately numbered (as exemplified with *no:tsa* below), each with a short gloss marking the morphological form being defined (e.g., *rdp-short* for (C)V- reduplication and *rdp-long* for (C)V:- reduplication).

25. Here, as elsewhere, an English format and translations are given, although Spanish versions also exist. Current plans are for two separate electronic forms of the Nahuatl Learning Environment (including the lexicon), one in Spanish and one in English. Printed versions would also be separate, unless a publisher prefers a fully trilingual (Nahuatl to English and Spanish) lexicon.

26. Thus \ph indicates a simple illustrative phrase; \prt indicates a phrase that manifests (C)V- short-vowel reduplication of the verbal headword; and \prt indicates a phrase that

|                 |  |
|-----------------|--|
| Field delimiter | Illustrative phrase and translation                          |
| \ph             | <i>Xne:chno:t̥sa, ne:xtlawe:lita.</i>                        |
| \eph            | He doesn't speak to me, he hates me.                         |
| \prs            | <i>Timonono:tsan pa:mpa kwahli timowi:kan.</i>               |
| \eprs           | We talk to each other a lot because we get along well.       |
| \prl            | <i>Ne:chno:no:t̥sa nona:n para ma:ka itlah nikchi:was.</i>   |
| \eprl           | My mother gives me advice so that I won't do anything (bad). |

The employment of distinct delimiters is but a temporary (and unattractive) solution to the problem of accessing particularly salient morphological elements and processes (such as reduplication, directional and purposive motion affixes, certain tense and aspect markers). It is not designed to mark phrases in which a word other than the headword is reduplicated; and it is completely ineffectual in providing a fully searchable text (which could, for example, be used to locate stems or to isolate inflectional morphemes). Only interlinearization would permit such searches (e.g., a search for *rdps* or *rdpl* on the *\morph* line below would isolate (C)V- and (C)V- reduplication, respectively):

|        |   |
|--------|---|
| \ph    | <i>Xne:chno:t̥sa, ne:xtlawe:lita.</i>   |
| \morph | x-Ø-ne:ch-no:t̥sa-Ø Ø-ne:ch-tlawe:l+ita-Ø   |
| \gloss | NEG-3SGS-1SGO-call-SG.IND 3SGS-1SGO-hate+see-SG.IND   |
| \trans | He doesn't speak to me, he hates me.  |
| \prs   | <i>Timonono:tsan pa:mpa kwahli timowi:kan.</i>  |
| \morph | ti-mo-rdps-no:t̥sa-n pa:mpa Ø-kwal-li ti-mo-wi:ka-n   |
| \gloss | 1PLS-REFL-RDPS-call-PL.IND because 3SGS-good-ABS<br>1PLS-REFL-carry-PL.IND                                      |
| \trans | We talk to each other a lot because we get along well.  |
| \prl   | <i>Ne:chno:no:t̥sa nona:n para ma:ka itlah nikchi:was.</i>  |
| \morph | Ø-ne:ch-rdpl-no:t̥sa-Ø Ø-no-na:n-Ø para ma:ka itlah<br>ni-k-chi:wa-s  |
| \gloss | 3SGS-1SGO-RDPL-call-SG.IND 3SGS-1SGPOSS-mother-<br>SG.AL.POSS so.that NEG-IMP something 1SGS-3SGO-<br>DO-SG.FUT |
| \trans | My mother gives me advice so that I won't do anything (bad).  |

manifests (C)V- long-vowel reduplication of the verbal headword. Subsequent field delimiters indicate English and Spanish translations of each type of phrase. Other phrase delimiters include *\pd* to mark sentences that manifest the use of a directional and *\pti*, which is used to mark phrases with an aspectual auxiliary fused to the verb.

Nevertheless, given the immense amount of labor involved in parsing the illustrative phrases in a dictionary of about ten thousand entries (let alone more than one hundred hours of additional texts), a temporary solution to data access is needed. One such solution (expedient though limited) is the use of specialized delimiters for particular categories that are problematic for grammatical or semantic analysis or that might prove most interesting for cross-linguistic comparison. Such a design also facilitates the isolation and export of illustrative sentences so marked.

Another structural mechanism that has been used to facilitate data access is the employment of complex coding within particular fields. As with delimiters, I have adopted an initial approach that might be described as "practical overcoding," in that a single code is used to mark several categorical features that together isolate terms deemed in need of further study. For any printed or screen version, an interface can convert these overspecific codes to shorter "human readable" forms (e.g., *V-1-nondir-wi* for "intransitive verb manifesting nondirected alternation," where the number 1 indicates the number of core arguments, becomes *V intrans.*). Yet the existence of the underlying code will facilitate searches and subsequent determination of the boundaries and internal structure of selected categories and paradigms.

For example, consider that many dialects of Nahuatl manifest a sequence of apparently denominal forms—adjectivals, inchoatives, and causatives—that manifest the endings *-k(i)*, *-ya*, and *-lia*, respectively, on what seems to be a single nominal stem (see table 10.2).<sup>27</sup> This sequence is one of the most complete and regular paradigms in Nahuatl. By coding each member according to grammatical category (*Adj-d*, *V-1-d*, and *V-2-d*) while identifying the paradigm by a special code (*-k/ya/lia*), it becomes a simple matter to create a system whereby users of the on-line version of the dictionary can "click" on any member of the paradigm and generate either a list of all related forms (e.g., all adjectivals with a morphology similar to that, say, of *yenkwik*), or a complete set of adjectivals, inchoatives, and causatives (displayed in a format similar to that of table 10.2).

Another feature of the data entry structure chosen is that it permits comparison across categories. This tool is especially useful for examining

27. See Dixon 1977: 27 ff. for a brief discussion of the relationship between "adjectives" and inchoative and causative verbs. I mention that the stem is apparently nominal because in most present-day dialects of Nahuatl only a few of the stems occur in nominal forms. One such occurrence is *istail* 'salt', which in denominal forms signifies 'white' (see table 10.2). A few other cases occur: *xokotl* 'plum' and *xokók* 'sour or tart'; *yell* 'bean' and *yeti:k* 'heavy'. Nevertheless, the majority of these adjectival, inchoative, and causative sequences are built on stems that no longer exist in nominal form. Launey (1992: 110) also refers to these adjectivals as denominal.

TABLE 10.2. Examples of a Word-Formation Paradigm in Nahuatl: Adjectivals, Inchoatives, Causatives

| (Nominal) Stem      | Adjectival                    | Inchoative                  | Causative                   |
|---------------------|-------------------------------|-----------------------------|-----------------------------|
| *yenkwi             | yenkwi:k ('new')              | yenkwi:ya ('to become new') | yenkwi:lia ('to make new')  |
| *yema:ni            | yema:nki ('soft')             | yema:nia ('to become soft') | yema:nilia ('to make soft') |
| istatl ('salt')     | istá:k ('white') <sup>1</sup> | ista:ya ('to become white') | ista:lia ('to whiten')      |
| Database code       | Adj-d-k/ya/lia                | V-1-d-k/ya/lia              | V-2-d-k/ya/lia              |
| Human readable code | Adj. (denominal)              | V. intrans.                 | V. trans.                   |

NOTE: <sup>1</sup>The Nahuatl orthography I have chosen indicates accent only when it does not fall on the penultimate syllable, as is the case with *istá:k*.

morphological processes such as noun incorporation, which both creates new lexical items and affects the argument structure of the clause. Noun incorporation—which has been the topic of much comparative and typological work, particularly in regard to the semantic roles of incorporated nouns—can thus be documented at the lexicosemantic level (by according the incorporated compound an entry as a headword) while being analyzed at both the morphological (as a compound) and syntactic (according to the valency of the newly formed verb) levels.<sup>28</sup> Therefore, it would seem desirable to structure lexical entries so as to provide for the most complete access to the semantic, morphological, and syntactic concomitants of this process. Incorporation has been and is highly productive in Nahuatl, which for this reason offers much data on the process. Making primary data available for use and analysis by linguists is one goal of the present project.<sup>29</sup> A detailed discussion on how incorporation may be represented and analyzed in a Nahuatl dictionary is beyond the scope of this chapter. Nevertheless, a few points may be made regarding how the database design of a lexicon can facilitate exploration of this topic as well as increase our understanding of the semantic range of certain nominal stems. The sample

28. The best overview and overall study of noun incorporation remains Mithun 1984. Both Mithun (1984: 77 ff.) and Baker (1995: passim) take various examples from Nahuatl. For a specific study of Nahuatl, see Merlan 1976. For recent studies of incorporation, see various works in Chappell and McGregor 1996. For a recent review of the subject, see Gerdtz 1998.

29. Croft (1990: 4), for example, notes that the removal of primary data from published typological studies is a serious problem for linguistic research. On-line presentation of this material should help to solve this problem.

entries below and the brief discussion that follows illustrate the benefits of a system devised to permit multilevel comparison.<sup>30</sup>

**a:te:mi** (*V intrans.*; *N+VI*; REG. INFLEC. see **ki:sa**; ROOTS: *a:*; *te:m*) **1:** for a fruit (plum, melon) to become almost ripe (as it 'fills with water') *Kimich a:te:mis un xokotl*. That plum is just about to become ripe. **2:** to become filled with water (a ditch, a hole in the ground) *Yo:a:te:n kwentli*. The furrows have gotten filled with water (e.g., after a heavy summer rain). **3:** to have a swollen belly (a pregnant female) *Un ne:nkah suwa:tl, yo:pe:w a:te:mi, ne:si ye o:stli*. That woman over there, she has a swollen belly, it appears that she is pregnant.

**chi:lkwa** (*V intrans.*; *N+V2*; REG. INFLEC. see **kwa**; ROOTS: *chi:l*; *kwa*) *caus.*→ **chi:lkwaltia** to eat chile *Ma:ski pitentsi:n, wel chi:lkwa*. Even though he is small, he can eat chile.

**koma:tlapa:na** (*V trans.*; *N+V2*; REG. INFLEC. see **ki:sa**; ROOTS: *koma:l*; *tlapa:*) to break the clay griddle (*comal*) of *O:ne:chkoma:tlapa:n*; *o:kalakiko tlawa:nke:tl, san o:wets*. He broke my clay griddle; a drunk came in, he just fell.

**kone:miki** (*V intrans.*; *N+VI*; REG. INFLEC. see **ki:sa**; ROOTS: *kone:*; *miki*) to lose a child in death *Ye ye:xpa o:kone:mik, xwel tlanemi:tia*. This is already the third time that she has lost a child (i.e., that a child has died on her), she cannot keep children alive.

**kopaxokonono:tsa** (*V trans.*; *N+V2*; REG. INFLEC. see **ki:sa**; ROOTS: *kopa*; *xoko*; *no:tsa*) to brag or boast to, particularly about doing sth that one cannot do, or about having sth that one does not have; to talk big to, to promise something to and not deliver *Te:kopaxokonono:tsa, xtlah kipia*. He brags to people, he doesn't have anything (i.e., property, money, etc.).

**kwilxi:ni** (*V intrans.*; *N+VI*; REG. INFLEC. see **ki:sa**; ROOTS: *kwil*; *xi*) to get covered with maggots or worms *Yo:kwilxi:n nonakaw, a:man xok wel nihkwa:s*. My meat got covered with maggots, now I won't be able to eat it anymore.

**panwetsi** (*V intrans.*; *N+VI*; REG. INFLEC. see **ki:sa**; ROOTS: *pan*; *wetsi*) **1:** to reach the summit (of a hill or mountain, of a high building) *Wekapan, xwel tipanwetsis*. It's high, you won't be able to reach the top. **2:** (*fig.*) to become wealthy; to attain power *Panwetsis, momo:stla tekiti wan xkaman tlai*. He'll do very well, he works every day and never gets drunk.

**suwa:kochi** (*V intrans.*; *N+VI*; REG. INFLEC. see **ki:sa**; ROOTS: *suwa:*; *kochi*) to sleep with a woman *Mo:stla nisuwu:kochis, a:man nikochis san nose:lti*. Tomorrow I'll sleep with a woman, today I'll just sleep by myself.

**tlá:kamiki** (*V intrans.*; *N+VI*; REG. INFLEC. see **ki:sa**; ROOTS: *tlá:ka*; *miki*) to die like a man (i.e., bravely) *O:tlá:kamik, xcho:kaya, san nochi o:kiyo:wih*. He died like a man, he didn't cry, he bore up under everything.

30. Note that this "multilevel comparison" is not simply a search on two fields. Rather, it involves the initial coding of a single feature (in this case, valency) in two fields that represent different levels of analysis: morphological and syntactic.

TABLE 10.3. Noun Incorporation in Ameyaltepec Nahuatl

| <i>search \cat for</i> | <i>search \com for</i> | Type of Incorporation Selected For                 | Example  |
|------------------------|------------------------|--|--|
| V-1                    | N-V2                   | Type I (Mithun 1984)<br>"saturating" (Launey 1998) | <i>nichi:lkwa</i><br>'I eat chile'                       |
| V-2                    | N-V2                   | Type II (Mithun 1984)<br>"modifying" (Launey 1998) | <i>ne:chkoma:tlapa:na</i><br>'he breaks my clay griddle' |

Incorporation may occur on intransitive ( $V_1$ ), transitive ( $V_2$ ), or ditransitive ( $V_3$ ) verbs. With transitive and ditransitive verbs, the nominal stem may be in a patient relationship to the main verb, reducing its valency in forming the new lexical item (cf. the intransitive *chi:lkwa* above, derived from the noun stem *chi:l-* 'chile' incorporated into the transitive verb *kwa* 'to eat').<sup>31</sup> With verbs of any valency, however, "the case role vacated by the incorporated nouns may be occupied by another argument, leaving the valency of the verb unchanged" (Mithun 1984: 859). In such cases the "modifying" relationship of the incorporated noun to the verb (either intransitive or transitive) may be quite varied—adverbial, locative, and comitative, among others.<sup>32</sup>

One key aspect of noun incorporation, therefore, is its effect on argument structure. For headwords, this can be determined by comparing the valency of the verb with no incorporated noun to that of the incorporated form. In the Nahuatl lexicon of Ameyaltepec, this is accomplished by simultaneously searching in the fields *\cat* (which gives the valency of the new lexical item) and *\com* (which gives the valency of the verb to which the noun has been incorporated). In table 10.3, the compound *koma:tlapa:na*, like *tlapa:na* from which it is derived, is a transitive verb (i.e., the incorporation of *koma:l-* 'clay griddle' does not affect valency), whereas *chi:lkwa* is an intransitive derived through compounding from the transitive *kwa* 'to eat'.

The ability to search for roots along with shifts in valency will enable users of the Nahuatl dictionary to determine how the incorporation of any particular noun stem affects the valency of a particular verb. The

31. This is Mithun's (1984) Type I incorporation, which Launey (1998: 5), along with others, refers to as "saturating incorporation."

32. For a brief discussion of these relationships in Nahuatl, see Launey 1998. A quite common relationship, particularly when the incorporated noun is a body part, is one that has been called "possessor raising": the incorporated noun is a possessed subject (intransitive) or object (transitive), and the possessor is raised to subject (of an intransitive) or object (of a transitive). For a discussion of the term "possessor raising," see Chappell and McGregor 1996.

TABLE 10.4. Incorporation of *koma:l-* with Ameyaltepec Nahuatl Verbs

| Noun Stem      | Transitive Verb               | Intransitive Compound                           | Transitive Compound  |
|----------------|-------------------------------|---|--|
| <i>koma:l-</i> | <i>chi:wa</i><br>'to make'    | <i>koma:lchi:wa</i><br>'he makes clay griddles' | <i>ne:chkoma:lchi:wilia</i><br>'he makes clay griddles for me' |
| <i>koma:l-</i> | <i>tlapa:na</i><br>'to break' | —   | <i>ne:chkoma:tlapa:na</i><br>'he breaks my clay griddle'       |

noun stem *koma:l-* provides a case in point as to how the nature of incorporation may vary according to verb (see table 10.4).

At one level the dictionary reveals that, unsurprisingly, making clay griddles (*koma:lchi:wa*) is a culturally salient activity (and thus glossed by an intransitive verb that is a distinct lexical item), whereas breaking clay griddles is not (given the absence of an intransitive lexeme, \**koma:tlapa:na*, *koma:tlapa:na* exists only as a transitive verb, the result of "possessor raising," in which a possessor is encoded as primary object: *kilapa:na nokoma:l* 'he breaks my griddle' → *ne:chkoma:tlapa:na* 'he griddle-breaks me'). At another level the contrast between *ne:chkoma:lchi:wilia* and *ne:chkoma:tlapa:na* reveals that the affectedness of someone who has his or her clay griddle broken is greater than that of one who has a griddle made. In both cases, *ne:chkoma:lchi:wilia* and *ne:chkoma:tlapa:na*, the noun stem is incorporated. However, *koma:lchi:wilia* is an applicative of *koma:lchi:wa* "an intransitive predicate denoting a unitary concept," whereas with *koma:tlapa:na* incorporation (like that which often occurs with body parts) vacates a case role, here filled by the adversely affected possessor.<sup>33</sup> The structure of the database (which includes a field for the stems that make up any given word) facilitates the extraction of this and similar information on incorporation and compounding.

As noted, in many instances incorporated elements are in some modifying or adverbial relationship with the verb. As the examples given above illustrate, in Nahuatl these relationships are highly variable and include some that are interesting in comparative perspective (e.g., *nikwilxi:ni*, lit. 'I [am] worm-scattered', where the subject is the location at which an event involving the incorporated noun occurs; and *panwetsi*, lit. 'on-fall', in which the incorporated element is a relational noun). An initial understanding of the semantics of incorporation is facilitated by an electronic version of a lexicon that has been structured

33. The citation, and this perspective in general, is based on Mithun 1984.

so as to permit rapid retrieval of the pertinent cases. However, researchers may also be interested in determining the connotations of particular stems that enter into a modifying relationship with a verb. To illustrate, one example given above is *kopaxokonono:tsa*, literally, 'to converse with someone like a *kopaxokottl* tree', a tree known to flower but not come to fruit. People who brag, who talk big, are like the *kopaxokottl* in that they promise "fruit" that is never delivered. Although here the metaphor involves a very specific cultural understanding that would need to be explained in the lexicon, in many cases this would not be true. Thus, for example, the metaphoric extension that establishes the meaning of *tla:kamiki* 'to die bravely' is not as obscure, although it is definitely a culturally specific interpretation of the implications of "man." Moreover, a user who so wishes could generate a list of all words that contain the root *tla:ka* 'man' and thus determine the semantic range of the connotations of this root. Indeed, a search function could easily be developed that would automatically list all words containing the roots (or any single root) of any given headword.

This first section has explored the research implications of a lexicographic project. I have suggested several ways in which project design can help attain certain scholarly goals. The inclusion of reduplicated CV- and CV'- verb forms (with both definitions and example sentences; see example of *no:tsa* above) under an unreduplicated headword was proposed as a means to facilitate comparison between such forms and thus promote an understanding of the semantics of reduplication. Likewise, by setting up separate field delimiters for illustrative sentences containing a reduplicated version of any given headword, users can easily retrieve all entries that have examples of short- or long-vowel reduplication (a system that would become obsolete after interlinearization). A positive side effect of this organization is the reduction in dictionary size, as measured by the number of headwords.<sup>34</sup> A second aspect of dictionary design oriented to contributing to linguistic research was the coding of morphological and syntactic information, thus facilitating retrieval of certain morphosyntactic phenomena. The case of noun incorporation was mentioned. Other codes

34. The positive aspects of this effect, I think, are well worth contemplating. A dictionary of a polysynthetic language runs a great risk of "overdocumenting" lexical items. This can easily occur with body-part incorporation and it would be easy to elicit hundreds of entries documenting such words. Likewise, Nahuatl has many endings indicating associated and purposive motion that can form "new words" from almost all verbs (e.g., *cho:ka* 'he cries' and *cho:katiw* 'he goes along crying'; *wetska* 'he laughs' and *wetskatinemi* 'he goes around laughing'; and *tlakwa* 'he eats' and *tlakwatiwetsi* 'he eats in a hurry'). With only a few semantic limitations, *-tiw*, *-tinemi*, and *-tiwetsi*, as well as some half dozen other verbal suffixes of associated motion, can combine with almost any verb. Few of these combinations have been accorded a separate entry in the Amejaltepec Nahuatl lexicon.

were either only briefly alluded to (e.g., some semantic codes such as *bd* for "body parts") or not mentioned (e.g., *N-ag* for "agentive nouns").

The basic premise offered above is that filtered lexical information (i.e., a lexicon that can be manipulated and searched according to multiple criteria) and the corpus of texts from which a lexicon is constructed constitute primary linguistic data that should be made available in electronic format to scholars and others. Clearly, the targeting of those elements that might prove most useful for comparative or typological analysis or that are most salient for an understanding of the language being documented is not a trivial problem. It must be based on an awareness of general research interests in the field and how a particular language might provide information relevant to these interests. Just as clearly, establishing a format for accessing certain morphosyntactic features (e.g., statistical documentation of word order in example sentences or text corpora) is much more time-consuming and difficult than others (e.g., the generation of a list of causative verbs). Yet the possibilities of significantly advancing linguistic understanding through the widespread availability of electronic versions of lexicographic databases and their associated text corpora should serve to encourage their development and diffusion.

3. PEDAGOGY: TRANSLATION AND LEARNING THROUGH LEXICON AND GRAMMAR. Dictionaries of indigenous languages of the Americas are, almost perforce, "bilingual": the source language (headword) is the indigenous language (the  $L_2$  of most users) and the target language is the Western (often colonial) language (the  $L_1$  of most users). Given this format, these dictionaries are best suited to passive use, oriented to the translation of received indigenous-language texts by nonnative speakers. The bilingual format is ill suited to the production of source (indigenous) language texts and of limited use for either the reception and translation of texts in the Western language or for the production of such texts by speakers of the indigenous language.<sup>35</sup> Given that the bilingual language format of these dictionaries (with the target language being  $L_1$  for most users) is clearly more by default than by design, it might be appropriate to ask how the traditional structure of such dictionaries can be modified so as to provide the most benefit to the greatest number of users.<sup>36</sup> In this section two uses

35. This use is problematic, although the remarkable *Hopi Dictionary* (Hopi Dictionary Project 1998), with its exhaustive English-Hopi Finder List (pp. 801-60) of perhaps some seven thousand English words, goes a long way toward facilitating such an application.

36. The relationship among source and target language, on the one hand, and  $L_1$  and  $L_2$ , on the other, has considerable effect on the strategies employed for defining source language terms. Thus Meyer (1990: 178) notes that "for  $L_2$ - $L_1$  use, explanatory equivalents may be used when there is no precise equivalent in the TL, since the user's native-speaker

are explored: translation of received texts in the indigenous language and learning of the indigenous language.

3.1. TRANSLATION AND STRATEGIES OF DEFINITION. The nature of the documentation to be processed with the aid of a Nahuatl dictionary is a key variable affecting project design. Nahuatl presents an almost unique case among languages native to the Americas in that a tradition of written documentation in this language—including grammars, lexicons, colonial texts, and chronicles—began almost five hundred years ago: the first Nahuatl grammar (Olmos 1547) was contemporary with the earliest grammars of French and English, and a mid-sixteenth-century dictionary (Molina's, in 1571) remains one of the great achievements in the lexicography of indigenous languages of the Americas. Nahuatl was used in the administration and proselytization of indigenous peoples, resulting in a profusion of materials—documents from imperial archives, native-language chronicles, and religious texts—that have been an important source for the study of the colonization and conversion of indigenous peoples in New Spain. Therefore, among the goals of Nahuatl lexicography should be that of providing a reference work capable of aiding in  $L_2(\text{Nahuatl}) \rightarrow L_1(\text{Spanish/English})$  translation and interpretation of a wide range of textual material. This orientation to the decoding rather than the encoding of language, to the reception rather than the production of texts, is a primary factor that should influence the project design and presentation of Nahuatl lexicographic materials.

That decoding occurs in the native language of the translator offers considerable protection against infelicitous phrasing of translated items and permits some concession, as is the case with passive dictionaries, to the standard criteria of substitutability for the construction of bilingual translation dictionaries. Many words in a Nahuatl dictionary may, of course, simply be “translated” and not defined. Such a situation would occur, for example, with *pa:mpa* ‘because’ and *me:stli* ‘moon’. Undoubtedly, translation is

---

intuition can help him come up with a translational equivalent which best fits into the TL context; for  $L_1$ - $L_2$  use, however, translational equivalents are more desirable, which means that the lexicographer must anticipate all possible TL contexts and their corresponding TL equivalents.” Meyer also notes, in general, that a major weakness of bilingual dictionaries is “an attempt to serve  $L_1 \rightarrow L_2$  and  $L_2 \rightarrow L_1$  users simultaneously” (p. 177). These weaknesses are divided into two broad categories: “weaknesses which impede the user's selection of a TL equivalent for a source language (SL) items . . . [and] weaknesses which impede the user's combination of a TL item, once correctly selected, with other TL items in context” (p. 175). There has been much work on bilingual dictionaries, particularly by Mary Snell-Hornby and Ladislav Zgusta, but little that has been specifically oriented to indigenous languages. For an exception, see Bartholomew and Schoenhals 1983.

also highly appropriate for communicating the referents of certain highly specific lexical terms, such as plant and animal names (here one may think of plant guides in which, particularly for “speakers” of the restricted discourse of Linnaean classification, the genus and species names provide the best unambiguous code for recognizing the meaning of vernacular terms). Yet, despite the functional goal of providing a translators' dictionary, with almost all words of a language such as Nahuatl for which no “backup” monolingual dictionary exists, in general it would be best to offer explanatory constructs (as opposed to substitutable “synonyms”) of headword entries and to rely on the native-speaker proficiency of the dictionary user to accommodate the most felicitous word or short phrase into the target language translation.

The lack of “backup” monolingual dictionaries, or sometimes even any other adequate bilingual tool for the understanding of meaning in an indigenous language, places a heavy burden on definitions and the semantics, broadly conceived, of a lexicographic project. Even when a primary focus is on the translation of received texts, there is a need for careful definition of many culturally specific terms or of words for which the “equivalents” of source and target language do not overlap. For example, even though ‘bird’ and ‘leg’ would be correct glosses for *to:to:tl* and *ikxi:tl*, respectively, the English translations should include the clarifications that *to:to:tl* refers only to small birds (as is the case with Spanish *pájaro* as opposed to *ave*) and that when referring to animals, *ikxi:tl* signifies only the back legs (the front legs of animals are its “arms,” *matli*). Care must always be exercised in providing the caveats and distinctions that result from incomplete overlap or equivalence between the terms of each language. Target language synonyms and substitutable lexemes may be offered, but the general strategy should be to present a more elaborate definition and place greater responsibility on the user or translator to supply the most adequate target language word. The ideal structure of the indigenous-language dictionary, therefore, should be a hybrid between a bilingual dictionary (with the target language being the native language of the user) and a translated monolingual dictionary (in which extreme care is given to the precise meaning and use of specific terms and proper formation of collocations is of prime concern).<sup>37</sup>

Considering the wide-ranging and serious responsibilities of an indigenous language dictionary to fully document meaning and the general inadequacy of a translational model, the best definitional strategy is one that embraces three primary facets of word use and meaning: the

37. On the utility of combining elements of a monolingual learners' dictionary and a bilingual dictionary in one text oriented to learning a second language, see Kharmā 1985. See also Magay 1988 on bilingual learners' dictionaries.

semasiological, the onomasiological, and the syntagmatic.<sup>38</sup> The first starts from the sign (the word) and tends to present meaning as word dependent. The second starts from a shared semantic content and groups together words that share meaning. The third approach, syntagmatic, “goes beyond the individual sign and focuses on the combinational aspects of the words” (Svensén 1993: 18). These three approaches do not represent separate aspects of meaning but instead are divergent paths to achieve an understanding of the complexity of meaning as it applies to specific words and groups of words. The necessity of combined use of all three strategies can be illustrated through a discussion of the appropriate definitions and descriptions of *koto:ni* and its transitive form *koto:na*<sup>39</sup> (see the entries below as well as fig. 10.2).<sup>40</sup>

**koto:na** (*V trans.*; REG. INFLEC. see **kisa**; ROOTS: *koto*.; SEM accepts intensifier *te*.; APPLIC. *koto:nilia*) **1**: to snap (sth long that can be stretched or pulled, such as a strap or rubber band) *Xkoto:na un i:loh!* Break off that thread! (e.g., by pulling on it hard or biting it after having finished sewing sth) **2**: to pull apart or pull off into pieces, to shred (e.g., chile, or an onion, etc.) *Xko:koto:na chi:hli para kwaltias tli:n titlakwa:s, ma koko:ya!* Shred some chile (pulling it apart with your fingers and, implicitly, adding it to the

38. Approaches to these aspects of meaning and lexicographic strategies are present in much of the literature on the subject. For a concise presentation of these three approaches, see Svensén (1993: 17 ff. and passim).

39. In the Ameyaltepec Nahuatl dictionary, entries have been marked according to various classes. For example, verbs are distinguished among those that form a morphologically marked causative (e.g., *wetska* ‘to laugh’ and *wetski:tia* ‘to make laugh’), those that form a morphologically marked anticausative (e.g., *tsakwa* ‘to close’ *trans.* and *notsakwa* ‘to close’ *intrans.*), and those that show an unmarked (or nondirected) alternation between intransitive and transitive (e.g., *koto:ni* and *koto:na*). These categories should help users to better anticipate meaning and will prove useful for typological research. The above distinctions are based on Haspelmath 1993.

Note, as indicated in the entry, that the verb *koto:ni/a* accepts the use of the intensifier *te* (e.g., *te:koto:ni* ‘to suddenly snap’). The verbs that accept this prefix have been marked in the lexicon and seem to belong to a certain semantic type. The Nahuatl dictionary being elaborated will mark these morphosyntactic properties and then discuss word groups (including semantic types) in a separate section.

40. The schematic representation in figure 10.2 illustrates some, but not nearly all, of the annotation conventions and search functions associated with word entries. The verb *koto:ni* has only one stem, and for this reason the entry has no morphological information. However, a form such as *kechkoto:na*, ‘to snap or break the neck of’, would be coded *N-V2* (noun incorporated into a transitive verb) and would have both *kech* and *koto*: listed as roots. And the entry for the transitive form *koto:na* would reference the ditransitive applicative *koto:nilia* as part of a cross-referencing system that points to morphologically marked derivations such as applicatives and causatives. Not represented in the diagram, though an important part of the Learning Environment, are links, particularly for words such as nouns, to extended cultural information (e.g., medicinal plants would link to illustrations and explanations of their use).

food) so that what you’re eating tastes better, so that it becomes spicy hot! **3**: to divide up, particularly in order to distribute or apportion to various persons; to take off a section or portion of *Ma tihkoto:natin notla:l, nikte:makas ilakotipan!* Let’s go to fraction off (by measuring, marking, and titling) a portion of my land, I’m going to give about half away! **4**: to end or break off (sth such as a custom or practice, or litigation, particularly acts or processes that take place and continue over long periods of time) *O:kikoto:n plei:toh, xok o:kinek kinenemi:li:s.* He broke off the suit, he didn’t want to pursue it any longer. NOTE: For both the transitive and intransitive forms, the intensifier *te* is used only when referring to the snapping of an object such as a band, rope, or strap but not to the dividing of land, or ending of litigation, etc.

**koto:ni** (*V intrans.*; REG. INFLEC. see **kisa**; ROOTS: *koto*.; SEM accepts intensifier *te*.) **1**: to snap (sth long that often can be stretched or pulled, such as a strap or rubber band) *Xok kixi:ko:s, ye kokoto:nis. Xtla:lili mejó:r se: yewan ye:nkwik!* It won’t bear up anymore, it is about to snap apart in pieces. Better put on one that is new! **2**: to fragment, to break apart in isolated portions (e.g., a brook that during the dry season dries up, leaving strings of poorly connected ponds; or clouds that become scattered and broken up by the wind *O:tsi:tsikiliw moxtli, yo:kokoto:n, kas yeye:kakawis.* The cloud cover has ripped open here and there, it’s come apart, there is a chance a driving hard rain will start to fall. **3**: (for sth, such as a pattern of behavior, a tradition or custom, or litigation) to cease to occur *Yo:koto:n, xok kitta:lian bake:ros.* It’s stopped, they don’t stage the dance called *vaqueros* anymore. **4**: (for a lineage or family) to end *O:koto:n i:nel-wayo, nika:n xok kipia a:kin ke:n kita.* His family line has come to an end, he doesn’t have any relatives here anymore. See **posteki**.

The Nahuatl verb *koto:ni* accepts a series of subjects that seem to have several characteristics in common: they can be snapped (straps) or pulled apart (clouds, onions) and include events or situations that “stretch” through time (litigation, customs, kinship relations). The transitive form of the verb, *koto:na*, typically takes a volitional animate agent acting on an object, most of which can also function as the subject of the intransitive *koto:ni*. Yet the overlap between subjects of the intransitive and objects of the transitive is by no means complete. The water of a brook can be the subject of the intransitive yet not the object of the transitive (given that the fragmenting of the water into ponds is the result of a natural process: the drying up of the landscape during the winter months); in a reverse manner, land can be the object of the transitive (*Kikoto:nas i:tla:l* ‘He will break up his land’) but not the subject of the intransitive. Clouds can be either subject or object, although speakers tend to prefer the oblique expression of agentivity (*O:kokoto:n moxtli ika o:yeye:kak* ‘The clouds became scattered as a result of the wind’ rather than a transitive form *O:kikoto:n moxtli yeye:katl* ‘The wind broke up the clouds’).

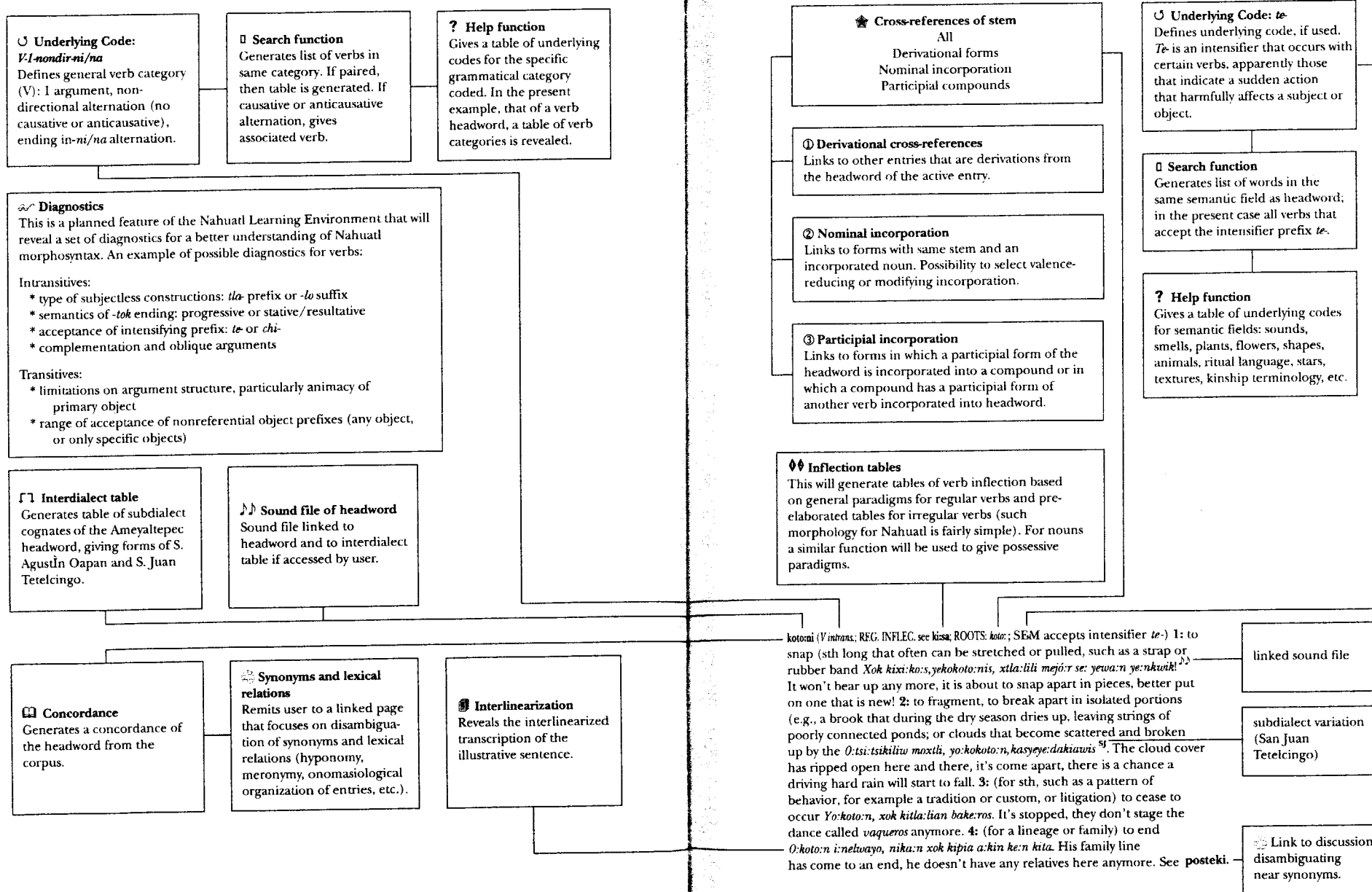


Figure 10.2. The Nahuatl Learning Environment (Lexicon): Schematic Representation of Verb Entry.



The contrast between possible subjects of intransitive *koto:ni* and objects of transitive *koto:na* demonstrates that whereas the latter include all potential objects that can be affected ("snapped" or "fragmented") by the volitional action of an agent, the former includes only those that can (but do not necessarily) suffer this change of state without the intervention of an agent.<sup>41</sup> Thus with an argument such as 'land', the verb *koto:na* cannot "detransitivize," since 'land' cannot become fragmented without human volitional action. For the fragmentation of 'land', when no agent is specified or known the transitive is used with a reflexive marker (in this case *no-* '3sg') indicating a type of impersonal passive: *O:nokoto:n i:ila:l* 'His land has been split up'; here there is an implication of agentivity that is lacking in the underived intransitive. This use of the reflexive marker is completely predictable from the grammar. Note, however, that with *koto:na* a reflexive marker can also be used with an animate subject. In this case (since, in effect, an animate cannot be the patient of a verb meaning 'to snap', as might be suggested by a reflexive form) the verb acquires a new sense: 'to break loose (from a tether or other object that can snap)'. This meaning requires that the action of the subject be carried out with little control while at the same time affecting an implied object (the tether) in a way that benefits the subject: *O:nokoto:n noburroh, o:cholah* 'My donkey broke free (from its tether), it ran away'. This use of the reflexive (as opposed to its function with a passive meaning) must be lexically marked and, therefore, accounted for in an entry that relates the possible argument structures of the transitive verb *koto:na* to its meaning.<sup>42</sup>

As the present case demonstrates, the "agent-oriented meaning component" of a verb, a component that affects semantics and the set of possible subjects and objects of intransitives and transitives, is not inherent in the verb itself. Rather, it is an aspect of meaning that emerges from word

41. The verb *koto:ni* manifests many characteristics of what have been called "unaccusative verbs"; see the early article by Perlmutter (1978) and more recently Levin and Rappaport Hovav (1995), who cite much of the literature. Haspelmath (1993) discusses the implications of "agent-oriented meaning components"; Levin and Rappaport Hovav (1995: esp. chap. 3) explore this issue further.

42. It is possible that a recognition and encoding in the lexicon of certain "semantic types" might facilitate the representation of types of argument structures. In this regard note Dixon (1984: 594), who suggests the implications of semantic types for understanding "general patterns of correlation between semantic and syntactic properties." A similar perspective on the relation of semantics to syntax is set forth in the analysis of verb entries (through a case study of *bake*) by Atkins, Kegl, and Levin (1988). They stress the significance of lexical semantic theory for lexicographic practice in noting that "a theory of lexical organization is needed in order to provide the context for building the entry. And it is the lack of theory in the semantic-syntactic interdependency area that makes . . . dictionary entries less than adequate" (p. 100).

combinations and, more generally, syntactic behavior.<sup>43</sup> Thus even within a basically semasiological approach, dictionary entries should take care to represent potential combinatory (syntagmatic) sequences and note the semantic implications of such collocations. With verbs this will often mean a discussion of potential subjects of intransitives and potential objects of transitives, although the implications for verb meaning of complements and adjuncts should be kept in mind. With attributive adjectives a combinational approach would involve potential collocations with nouns. However, in all cases, to aid users in extrapolating from the few collocators given in the definition to other potential combinations, care must be given to discuss the common denominators that link these bases.<sup>44</sup>

In addition to semasiological and syntagmatic considerations, a third approach to meaning that should be incorporated into a dictionary is onomasiological. The nature of this approach can be illustrated by its treatment of collocations. Whereas the semasiological approach departs from the collocator (e.g., asking for the possible objects of *koto:na*), the onomasiological approach departs from the base (e.g., asking what one can do to land, *ila:hli*).<sup>45</sup> An electronic database format can be of immense help in articulating a semasiological and onomasiological approach to collocations. With verbs, for instance, a field of potential subjects and objects can be created and a search function developed that would generate possible collocators for any given base. Thus under words like *koto:na* 'to divide up (a field)', *pupwa* 'to clear (a field)', *to:ka* 'to plant (a field)', *o:melia* 'to plow (a field) a second time before planting', and *ye:kpuwa* 'to replant (a field) where the first seeds did not sprout', the word *ila:hli* would be listed in a field dedicated exclusively to collocation bases. An entry under *ila:hli*

43. A similar aspect of meaning occurs with a typical unaccusative verb such as *break*, which is considered unaccusative in its intransitive form. Thus, although one may have "He broke the vase" and "The vase broke," with other arguments only a transitive or an intransitive use, but not both, is acceptable: "He broke his promise" but ?"His promise broke"; and ?"He broke his voice"; but "His voice broke." Levin and Rappaport Hovav (1995: 105) give the former example (with *promise*) and suggest that intransitive subjects are a subset of transitive patients. However, as the second example (with *voice*; cf. the Nahuatl example with *brook*) demonstrates, there are also cases of intransitive subjects that cannot be encoded as the objects of transitive verbs even though they are patients.

44. This factor is related to what Mel'čuk (e.g., 1996) refers to as lexical inheritance. See the article by Grimes in this volume. Also pertinent in this regard is Apresjan's (1992-93: 80) concept of lexicographic types, which he defines as "a group of lexemes having a number of properties in common that are sensitive to the same or similar sets of linguistic rules—morphological, syntactic, prosodic, semantic, etc. . . . Every lexicographic type should be treated in the dictionary in a unified way." Both Mel'čuk and Apresjan have dealt with verbs of emotion; see Mel'čuk and Wanner 1996 and Apresjan 1992. See also Dixon 1984.

45. See the discussion in Cop 1990.

in an onomasiological section of the dictionary could then present and analyze information on combinational patterns.

More important, an onomasiological approach also pays particular attention to shades of meaning that distinguish near-synonyms; in this sense it focuses on words that are in a paradigmatic, not syntagmatic, relation. Thus in a resource such as *Merriam-Webster's Collegiate Dictionary*, one finds that entries for *procrastinate*, *lag*, *loiter*, *dawdle*, and *dally* all point to *delay*, where a concise explanation is offered of the differences in meaning among these words. Dictionaries of indigenous languages have generally neglected this facet (disambiguation) of lexicography, a failure that is unfortunate given that such distinctions, presented in a consolidated explication, are key to understanding the subtleties of meaning that would otherwise be difficult, if not impossible, to grasp. The Ameyaltepec Nahuatl dictionary currently being elaborated will include a supplementary section that explores meaning differences of closely related words, including those that can be grouped according to fairly delimited, and often overt, semantic fields. Thus a discussion of definitions might center on words that describe different textures, different ways of hardening (e.g., the differences between verbs such as *tepi:tsiwi*, *kuhpistia*, *kuhpitsiwi*, and *kuhtia*, all having to do with hardening or stiffening), or various ways of walking. Other patterns of interlexical relations, particularly hyponymy and meronymy, would also be explored.<sup>46</sup> Presentation would include sets of related words within relatively salient categories such as medicinal plants, colors, smells, and sounds.<sup>47</sup> Still other sets would focus on activities (such as house building, cooking, farming) and objects (e.g., a house, a plow, a plant) and the different verbs and nouns associated with these particular events and objects. This onomasiological facet of the lexicographic project will provide key information on collocators associated with specific bases, on shades of meaning distinctions among near-synonyms, and on words related within given semantic fields. It will be combined with a more formal approach to categories based on the codification of word morphology (e.g., of causative verbs, which could thus be easily extracted from the lexicon). In conjunction these two approaches will provide a more complete understanding of Nahuatl than would emerge from a presentation based strictly on headwords.

This subsection has explored the way in which a basic function of a Nahuatl lexicon as a passive translators' dictionary should be expanded to embrace a greater sensitivity to the semantic complexity of the language.

46. For an overview of interlexical relations, see Cruse 1986.

47. For example, one could mark verbs that refer to animal sounds (e.g., *cho:ka*, *nanalka*, *nokwi:katia*, *tlayowa*, *tlapi:tsa*) and then in an onomasiological section on animal sounds group animals into categories according to the verbs used to indicate their respective sounds.

Indeed, the primary focus of an indigenous language dictionary should be on semantics, and extreme care should be taken to develop all aspects of meaning as fully as possible. Deficient explanations impoverish the dictionary and do a great injustice to the richness of expression in indigenous languages. Users should be able to provide the appropriate translational equivalent in their native language; but without a detailed exposition of meaning they will be unable to reinterpret or account for many usages that they are bound to encounter in their philological studies. It was suggested that the problem of meaning be broached from three directions—semasiological, onomasiological, and syntagmatic—in order to provide a multidimensional semantic perspective. These approaches move a translators' dictionary into the realm of a learners' dictionary, at least in regard to the dichotomization between a dictionary's role in the reception as opposed to the production of texts. It is to this problem of learning (of producing texts or utterances) and to the indigenous dictionary's role in pedagogy that I now turn.

3.2. LEARNING AND THE STRATEGIES OF TEACHING. Over the past decade, as lexicographers have realized that "language learning imposes its own requirements on the format of a dictionary" and that "traditional dictionaries for general use are not particularly suitable for the purpose of learning, especially a new language," increasing attention has been paid to the development of what is called a "learners [*sic*] dictionary" (Svensén 1993: 24).<sup>48</sup> The driving force behind these dictionaries, which have been elaborated for most of the "major" languages of the world, is to provide students of a given language with an additional tool for language acquisition. The philosophy behind these efforts, however, is one that runs almost contrary to that which should guide lexicographers of indigenous languages. The learners' dictionaries, in effect, presuppose a teaching environment and a fairly elaborate infrastructure of pedagogical tools. Moreover, in molding monolingual native-speaker dictionaries so that they meet both the needs and the abilities of nonnative speakers, learners' dictionaries implement changes (an abridged format in terms of number of headwords and information included, as well as a carefully delimited definitional terminology) that would be unacceptable in a serious lexicographic project on an indigenous language, which would stress completeness of coverage and exhaustive analysis (e.g., etymological, word class, etc.). The problem facing lexicographers of indigenous languages is, therefore, somewhat the mirror image of a problem that has challenged lexicographers of major languages. Simplified dictionaries of

48. For a critical review of the literature, see Dolezal and McCreary 1999.

major languages are the result of an abundance of economic and human resources, whereas simplified dictionaries of indigenous languages (many of which are better described as vocabularies) are most often the result of a scarcity of economic and human resources.<sup>49</sup> And whereas the former comprise dictionaries adapted to an extant learning environment, the latter are often elaborated in a pedagogical vacuum.

The Nahuatl Learning Environment has been designed in part to respond to the problems mentioned in the preceding paragraph. This project aims to join together a lexicon (a brief sketch of which was given earlier), a user's guide to the lexicon, a reference grammar with a pedagogical orientation, and interactive lessons for language learning.<sup>50</sup> By combining research and learning tools, it will provide a corpus of material useful to beginning students as well as expert scholars; it makes available primary linguistic data for research while organizing it in such a manner as to achieve didactic goals; it links a dictionary with a user's guide and grammar so that learning a language and vocabulary go hand in hand; and it establishes an interactive language course that makes full use of a lexicon and reference grammar that otherwise would be used mostly by specialists. The Nahuatl Learning Environment, therefore, seeks to solve a major problem in the study and learning of lesser-taught languages by providing the grammatical and pedagogical context for a dictionary while furnishing the appropriate lexical base for students to practice and implement the language skills they learn through a grammar. Key to this approach is the conviction that printed and electronic media are compatible, not conflicting, forms of representation and diffusion. When possible, even a published work such as a reference grammar should be available electronically so that it may be linked to a lexicon and to annotated primary linguistic data (such as interlinearized texts), thus providing a more comprehensive approach to the study of a given language.<sup>51</sup>

A case was already made for presenting the results of lexicographic projects in indigenous languages in electronic format, although this should by no means preclude the publication of printed material. Several benefits of the former have already been mentioned: the possibility of including large amounts of material that, because of the costs of

49. Atkins (1992-93: 9), for example, estimates that "a one-volume collegiate dictionary compiled from scratch will take over 100 person/years of work, and cost up to four million dollars."

50. Target-language material of the Nahuatl Learning Environment is bilingual (English and Spanish), although below only English examples are given (see note 25).

51. The importance of linking primary data to typological studies is mentioned by Croft (1990): see note 29.

publication, would otherwise be unavailable; and the ability of electronic formats with detailed coding to accommodate the research interests and priorities of a wide range of scholars. Another advantage is the facility of linking to the appropriate headword or text much additional material: illustrative material, ethnographic descriptions and data, and sound files (pronunciation guides and minimal pairs, example sentences, types of music described by certain entries, ritual texts). Equally significant is the possibility of using an electronic format to provide an easily accessible user's guide, with help files and a glossary as well as a tutorial for incipient users on how to best extract and exploit the multiple levels of information contained in the lexicon.<sup>52</sup> Lexicographic and semantic data on Nahuatl is of interest to a wide range of people and care should be taken to ensure that they will be able to make full use of the material available. That most similar lexicons have been developed with little pedagogical concern (both in the sense of language teaching and in regard to making the indigenous language dictionary more user-friendly) has undoubtedly adversely limited the potential audience for many lexicographic resources. Two examples should suffice to demonstrate the value of linking learning to the lexicon through electronic media.

One characteristic of Nahuatl that lends itself to a pedagogical use of an electronic lexicographic database is the morphological complexity of words (in terms of both derivational processes and compounding). To teach this aspect of Nahuatl, users can be taken through a short tutorial that would include both a discussion of Nahuatl word formation (such as the manner in which nominal stems can be joined together or incorporated into verbs) and a guide to using Steven Bird's *Hyperlex* search engine to explore this aspect of Nahuatl morphology.<sup>53</sup> By embedding hypertext links with the correct queries into the HTML tutorial text, students who encounter difficulties in defining the parameters of a given search they are requested to perform can easily activate the search engine. This will generate the appropriate list or table of entries, followed by the *Hyperlex* query box with the appropriate data. Table 10.5 is a small sample of a

52. For a user's guide to specific learners' and students' dictionaries in English, see Underhill 1980. Such materials seems to be lacking for many indigenous-language dictionaries, although it is perhaps here that they are most needed. Terms that linguists and other academics might take for granted (*incorporated noun, causative, applicative, stative*) are quite opaque to the majority of people.

53. The processes described below, as indeed the entire electronic format of the Nahuatl lexicon, is made possible by Steven Bird's *Hyperlex* search engine, which he generously adapted to the Nahuatl data I have been working with. Again, through their expertise and generosity, both he and Mark Liberman of the Linguistic Data Consortium have made this project possible. Bird's *Hyperlex2* should be working by mid-2002.

TABLE 10.5. Sample Result of Morpheme Search on *a-* 'water'

| Morphemes        | Words  |
|------------------|--|
| a:               | <b>a:tl</b> ( <i>N basic</i> ) 1: water 2: ( <i>rel.</i> ) fontanel, soft spot on a young child's head 3: (- <i>yo</i> ) juice, nectar (of a fruit or vegetable)   |
| a:; i:           | <b>4:</b> (- <i>yo</i> ) broth (of a cooked dish such as beans, chicken)<br><b>a:tl</b> ( <i>V intrans.</i> ) to drink water and, by extension, other (nonalcoholic) beverages   |
| a:; xi:x         | <b>a:tl</b> : <b>tia</b> ( <i>V trans.</i> ) to give (sb) water or, by extension, other (nonalcoholic) beverages to drink<br><b>a:xi:xa</b> ( <i>V intrans.</i> ) 1: to urinate on 2: ( <i>refl.</i> ) to urinate 3: ( <i>fig.</i> ) to excrete a liquid (e.g., as a tree does sap)<br><b>a:xi:xaltia</b> ( <i>V trans.</i> ) to cause to urinate<br><b>a:xi:xltli</b> ( <i>N deverbal</i> ) urine<br><b>i:a:xi:x burroh</b> ( <i>N complex</i> ) a type of mushroom that often grows on donkey dung |
| a:; xi:x; -eh    | <b>a:xi:xeh</b> ( <i>N derived</i> ) person who always has to urinate, who urinates a lot (same as <i>a:xi:xpal</i> )  |
| a:; -yo:         | <b>a:yo:tia</b> ( <i>V trans.</i> ) to add water to (sth, particularly foods being boiled such as beans or sauces)   |
| a:; -yo:; ki:sa  | <b>a:yo:kisa</b> ( <i>V intrans.</i> ) to ooze or secrete a liquid (particularly in reference to the pus or liquid of an infection)  |
| a:; -yo:; wa:tsa | <b>a:yo:wa:tsa</b> ( <i>V trans.</i> ) ( <i>refl.</i> ) to lose one's bodily fluids  |

table generated onscreen by searching for all words that include the stem *a-* 'water'. The actual query is illustrated in a reconstructed screen shot (fig. 10.3).

The facility with which complex queries to *Hyperlex* can be embedded in HTML texts on-line is not only useful for tutorials; in addition, it offers an extremely powerful tool for linking lexicon and grammar into a single learning environment. Queries can be embedded in the electronic text of a reference grammar to permit the extraction of relevant information and paradigms from the lexicon (which can be continually updated) or other primary linguistic data (such as interlinearized texts). For example, in a section of the grammar on causative verbs, students can be offered three types of supplementary learning tools: a short, 100- to 200-word, summary explanation of causativity; a table generated by *Hyperlex* of all causative verbs, their meanings, and derivations; and an interactive lesson teaching the formation and analysis of causative constructions. The same material could be easily accessed from the other (lexicographic) direction. Thus a student working with the lexicon who encounters a causative verb could automatically activate the same short summary explanation of causativity, move to the relevant chapter in the reference grammar, generate a list or

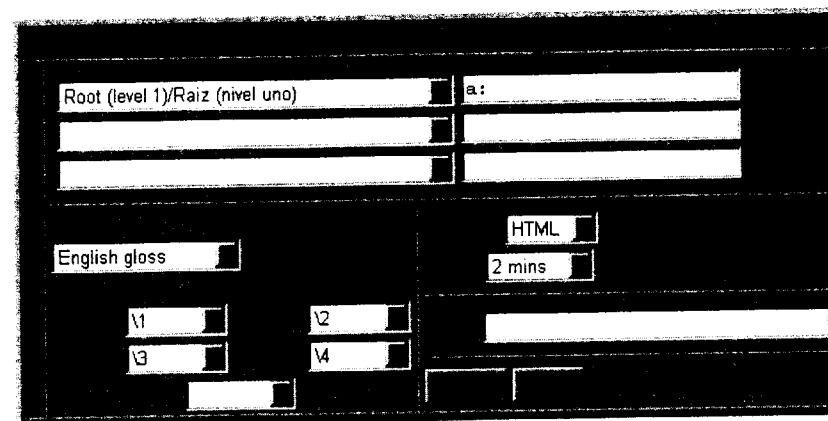


Figure 10.3. Query Used to Generate Table 10.5.

table of verbs similar to the causative verb on the screen, and practice the formation and analysis of these types of verbs through an interactive lesson on the topic.

The benefits of linking grammar and lexicon can be further exemplified by the treatment accorded intransitive verbs. Most Nahuatl intransitives can be grouped into one of three major formal classes: (1) CAUSATIVE ALTERNATION: those that manifest a basic intransitive form and a morphologically marked causative derivation (e.g., *cho:ka* 'to cry' and *cho:ktia* 'to make someone cry'); (2) ANTICAUSATIVE ALTERNATION: those that have a basic transitive form and a morphologically marked intransitive (e.g., *tsakwa* 'to close', *notsakwa* 'to become closed'); and (3) NONDIRECTED ALTERNATION: those that show minor variation and little evidence for determining the direction of the derivational process (e.g., *chaya:wi* 'to become dispersed', *chaya:wa* 'to disperse or scatter').<sup>54</sup> Each type has been coded in the lexicon; phonological characteristics permit further differentiation. In the chapter on intransitives, one of the tables lists various patterns manifested by verbs showing nondirected alternation (see table 10.6). In the electronic version of the grammar, hypertext queries (see fig. 10.4) have been embedded at the intransitive form of each pair. When readers activate the link at any specific location, *Hyperlex* generates tables of similarly patterned verbs. For example, activating the query embedded at *poliwi* generates a table of all verbs

54. The classes are directly based on a typology proposed by Haspelmath (1993); see note 40. The vast majority of Nahuatl verbs manifest either causative or nondirected alternation.

TABLE 10.6. Intransitive and Transitive Verb Pairs:  
Nondirected Alternation

| Transitivity | Present Indicative | Gloss                     |
|--------------|--------------------|---------------------------|
| Intrans.     | <u>poliwi</u>      | he (it) gets lost         |
| Trans.       | ki-polowa          | he loses it               |
| Intrans.     | <u>toma:wi</u>     | he gets fat               |
| Trans.       | ki-toma:wa         | he fattens it             |
| Intrans.     | <u>koto:ni</u>     | it snaps                  |
| Trans.       | ki-koto:na         | he snaps it               |
| Intrans.     | <u>kaxa:ni</u>     | it (e.g., a knot) loosens |
| Trans.       | ki-kaxa:nia        | he loosens it (a knot)    |
| Intrans.     | <u>totomi</u>      | it gets untied            |
| Trans.       | ki-totoma          | he unties it              |

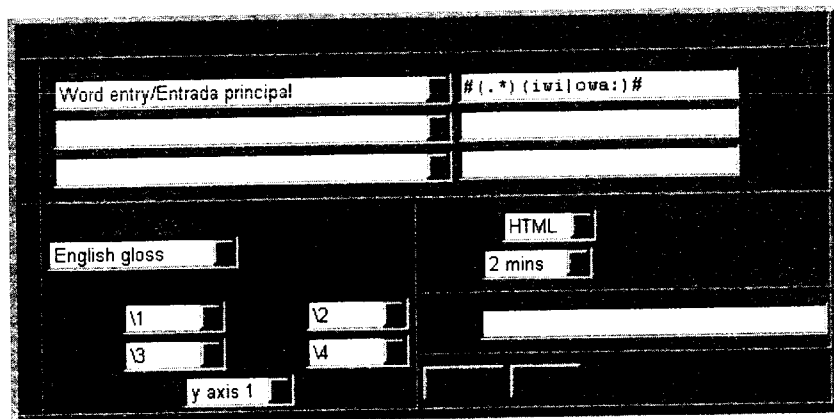


Figure 10.4. Query Used to Generate Table 10.7.

that manifest the same *-iwi/-owa* alternation (see table 10.7). This type of automatic search can be linked to any point covered in the grammar.

If dictionaries teach the meanings of words, they do not teach the meaning of language. And if grammars teach structure, they do not teach content. In isolation each tool is destined to remain a blunt instrument for learning and a dull device for communication. For many potential users of the dictionary, carefully elaborated morphosyntactic information will remain a mysterious code unless a concerted effort is made to explain and explore the significance of these data. For all the effort that might be made to present collocations, without a background in grammar, dictionary users will be unable to produce even the simplest of sentences.

TABLE 10.7. Sample of Verbs Manifesting *-iwi/-owa* Variation  
(reconstructed screen shot)

| Stem      | <i>-iwi</i>   | <i>-owa</i>  |
|-----------|---|--|
| a:pach    | <b>a:pachiwi</b> ( <i>V intrans.</i> ) to get covered with water, to become submerged in water                            | <b>a:pachowa</b> ( <i>V trans.</i> ) to submerge in water, to dunk in water                            |
| a:tots    | <b>a:totsiwi</b> ( <i>V intrans.</i> ) to become watery (a mixture, of sth to which water is commonly added)              | <b>a:totsowa</b> ( <i>V trans.</i> ) to add water to (a mixture, making it more watery)                |
| chi:mal   | <b>chi:maliwi</b> ( <i>V intrans.</i> ) to spread out in a circle (a stain, the houses at the edge of a town as it grows) | <b>chi:malowa</b> ( <i>V trans.</i> ) to spread or flatten out in a circle                             |
| chikipe:l | <b>chikipe:liwi</b> ( <i>V intrans.</i> ) to split open or burst apart  | <b>chikipe:lowa</b> ( <i>V trans.</i> ) to split open lengthwise (e.g., a pig after being slaughtered) |
| chikino:l | <b>chikino:liwi</b> ( <i>V intrans.</i> ) to become crooked or bent in places   | <b>chikino:lowa</b> ( <i>V trans.</i> ) to make crooked or bent  |

Nor, without a lexical base from which to depart, will students of a grammar be able to apply even a small portion of the knowledge they might have gained. The Nahuatl learning project attempts to make use of electronic formats to link lexicon, grammar, and corpus in a single learning environment. Rather than a learners' dictionary strictly for students interested in language acquisition, it seeks to create a pedagogical tool that can use a linguistically complex and semantically rich text to draw students and scholars into the language. Hopefully it will provide a model for future work in this direction.

4. LEXICOGRAPHY AND THE LANGUAGE COMMUNITY. In the previous pages I have noted three major categories of users, their concerns, and the textual material they would be dealing with. These are scholars interested in the comparative or lexicographic value of a Nahuatl dictionary, translators interested in the decoding and translation of received (and mostly historical) texts, and students interested in acquiring proficiency in spoken Nahuatl. The needs of these users may be met through the presentation of analytical fields for certain types of morphosyntactic and phonological material, more expanded areas oriented to textual interpretation that would include a cultural and contextually based semantics, and learning tools meant to create tools for the productive and active use of grammar and lexicon. However, another group of potential users

should be considered: native speakers of Nahuatl who wish to use the material as a bilingual learners' dictionary for Spanish or as a source of Nahuatl material that might be used in a village school. The simplest way to deal with this potential conflict of interest is to grab the bull by the horns and simply state what is an unfortunate truth: the goals of linguistics, translation, and learning Nahuatl are often not aligned with many of the goals of indigenous speakers. Indeed, if research into the lexicosemantics of indigenous languages is to benefit native speakers at all, a conscious effort must be made to work with communities of speakers.<sup>55</sup> Here I limit my comments to the ways in which an electronic database format facilitates the manipulation and export of data into a printed medium that can better serve native speakers' needs.

Such a product requires considerable effort to transform the printed output into a text that accommodates specific requirements of the language community. For example, it is important to provide a bilingual text that both departs from Nahuatl into an L<sub>2</sub> target language (oriented to the production of texts in Spanish) and one that departs from Spanish into an L<sub>1</sub> target language (oriented to the reception and interpretation of texts in Spanish). In both instances the material used in constructing the lexicon should be resynthesized in more accessible form. Some of the necessary changes, such as an orthography more in accord with Spanish conventions and the elimination of analytical fields of specialized information, could be accomplished through electronic manipulation of the primary data.<sup>56</sup> In the case of Nahuatl, work with native speakers will perhaps reveal whether certain decisions regarding headword entry should be changed. It may be decided that the obligatory object pronoun for

55. For this and other suggestions, see Jeanne 1992. Much work has been done to make linguistic research more responsive to the language goals of native speakers of indigenous languages. Particularly noteworthy is the work of the American Indian Languages Development Institute (AILDI) at the University of Arizona and Nora England's work with the Proyecto Lingüístico Francisco Marroquín in Guatemala. Robert Laughlin's work with Tzotzil speakers in Chiapas also stands out, as does that of Ofelia Zepeda with Tohono O'odham and the AILDI, and that of the Hopi Dictionary Project in elaborating the previously mentioned *Hopi Dictionary* (1998). See also the various essays in Hale 1992a.

56. I have not discussed orthographic conventions that I have chosen to use or the rationale behind the decisions taken. However, all of these conventions can be changed by fairly simple "search and replace" operations should a different perspective be adopted or, as I suggest here, should it be decided that slight variations in orthography might better suit the needs of different target audiences. Related to this issue (as well as to the question of standardization) is another one that is of particular importance but that I have not discussed here. This is the issue of "interdialectal lexicography," an extremely pressing problem for work in indigenous languages. Electronic databases can solve some of the problems that confront lexicographers who wish to work among language variants that are close but distinct (such as that of two nearby communities). For the Balsas River basin, this might involve

transitive verbs should be represented; if so, a choice must be made between one specific marker (the third person *k(i)-*) or two nonreferential markers (*te:-* for humans and *tla-* for nonhumans). Many of these changes can be made rapidly if a database format was used originally. Other changes must be more carefully contemplated and more laboriously implemented. Headwords must be carefully selected for inclusion; reduplicated forms might be accorded their own entries, and the same might be decided for irregular plurals and certain inflectional forms of verbs. If indeed a learners' dictionary is the goal, illustrative sentences should be selected or elaborated for their pedagogical value, and collocational and syntagmatic relations should be made as clear as possible. If a finders' list is utilized, care should be taken that it represents a basic vocabulary in the national language.

In this section I have briefly discussed the inadequacy of academic syntheses of indigenous lexicons, no matter how well intentioned, for addressing the communicative needs of the native speaker community in the dominant colonial language. I suggested that research material be restructured and a true bilingual learners' dictionary, in a much more traditional sense, be developed for use by native speakers of the language academics study. Yet at the same time, respecting the wishes and privacy of individual speakers, when possible the primary data of linguistic research should also be made available to the community. The utility of this material to promote the teaching and preservation of an indigenous language still as alive and vibrant as Nahuatl should not be minimized. Nor should its value in promoting pride in the indigenous language and culture, and thus indirectly helping language maintenance, be discounted.

5. CONCLUSION. I began this chapter with a brief discussion of the difference between spoken and written languages as it affects lexicography and suggested that the dictionary tools and resources that have dominated the Western tradition (particularly their focus on written texts and its general orientation to increased specialization) are not altogether appropriate for work in indigenous languages. There are various reasons for this: the implications of the use of oral data for the elaboration of a

---

work in the neighboring villages of Ameyaltepec, San Juan Tetelcingo, and San Agustín Oapan (which, for example, respectively manifest *otli*, *ótlí*, and *ohlle* for 'road'). Electronic manipulation can solve some of the problems of phonological and lexical variation, particularly in the ability of this medium to provide the necessary cross-links and to store and manipulate the vast amounts of data that would be involved in such a comparative lexicographic project.

lexicon, the specific characteristics of the language being documented, the types and needs of users who will access the dictionary (from linguists to historians to members of the speech community), the responsibility of a lexicographer of an indigenous language to provide the greatest flexibility and range of use for the final product, and the value of presenting cultural and encyclopedic information in the dictionary definitions.

The two key elements of the project presented here are a multifaceted approach to meaning (semasiological, onomasiological, and syntagmatic) and the utilization of electronic media to link corpus, lexicon, and grammar in a mutually reinforcing research and learning environment. In the realization of both goals, an electronic format is particularly appropriate. Certainly, the bottom line of a lexicon is the breadth and depth of its definitions, and particular attention to semantics is necessary to convey the full richness of the language being studied. Yet regardless of the care and precision that might go into dictionary preparation, much will be lost if corpus and grammar are not integrated into a system of lexicographic representation. An electronic structure permits the establishment of direct links between these three elements. These links reflect the way in which research involves a constant triangulation between data, definition, and structure in order to develop adequate semantic and morphosyntactic understanding. By developing tools that permit others to re-create this process, at the same time making available primary data and the tools for additional analysis (such as parsed texts and examples, cross-referencing, and so on), both further research and debate are encouraged and language learning becomes a more integrated experience.

## CHAPTER ELEVEN

# The Comparative Siouan Dictionary Project

*David S. Rood and John E. Koontz*

1. INTRODUCTION. The project we describe is the preparation of a dictionary of Comparative Siouan, that is, a reconstruction and cognate list of the vocabulary and, necessarily, much of the morphology of the language that was the ancestor of the modern Siouan languages. (See Rood 1979 for a survey of Siouan studies as they stood before the Comparative Siouan Dictionary project.) The Comparative Siouan Dictionary (CSD) is now in an advanced state of preparation, except for introductory essays and a phase of final cleanup editorial work, though it has now been about twenty years since we first conceived of the idea. This chapter describes the history of the project, including its conceptual history, the use of computers, the funding history, and the evolution of methodology as we moved from paper to computers and learned to work together as a team. Some of the specific linguistic results of the work have been published by Rankin, Carter, and Jones (1998).

The Siouan family consists of fifteen to eighteen documented languages in three major subgroups (see fig. 11.1); the exact number of languages depends on whether one classifies some of them as dialects or as separate languages. They are (1) the Missouri River Siouan group, Crow and Hidatsa; (2) the Central Siouan group, subdivided into Mandan and a large grouping called Mississippi Valley Siouan, comprising Dakotan<sup>1</sup>

1. Dakotan is a neologism intended to serve as a term for the subgroup as a whole (Dakotan) in contrast to the Santee-Sisseton and Yankton-Yanktonais version of the native name for the ethnicity (Dakhota or Dakota), which is often applied to the Santee-Sisseton dialect specifically. For a discussion of Dakotan dialectology, see Parks and DeMallie 1992. Santee-Sisseton is commonly known as Santee or Dakhota or Dakota. Yankton-Yanktonais is traditionally grouped erroneously with Assiniboine and Stoney as Nakota or Nakoda. Teton is also known as Lakhota or Lakota.