

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

8. Language Renewal and the Technologies of Literacy and Postliteracy:
Reflections from Western Mono
Paul V. Kroskrity / 171

III TECHNOLOGY AND DICTIONARY DESIGN

9. An Interactive Dictionary and Text Corpus for Sixteenth- and Seventeenth-
Century Nahuatl
Una Canger / 195
10. What's in a Word? The *Whys* and *What Fors* of a Nahuatl Dictionary
Jonathan D. Amith / 219
11. The Comparative Siouan Dictionary Project
David S. Rood and John E. Koontz / 259

IV SPECIFIC PROJECTS AND PERSONAL ACCOUNTS

12. Writing a Nez Perce Dictionary
Haruo Aoki / 285
13. On Publishing the *Hopi Dictionary*
Kenneth C. Hill / 299
14. Writing a User-Friendly Dictionary
Catherine A. Callaghan / 312
15. The NAPUS (Native American Placenames of the United States) Project:
Principles and Problems
William Bright / 322
16. Alonso de Molina as Lexicographer
Mary L. Clayton and R. Joe Campbell / 336

BIBLIOGRAPHY / 391

LIST OF CONTRIBUTORS / 423

INDEX / 429

An Interactive Dictionary and Text Corpus for Sixteenth- and Seventeenth-Century Nahuatl

Una Canger

1. THE UNIQUE STATUS OF NAHUATL AMONG LANGUAGES OF THE AMERICAS. Among the world's languages, only a few were documented and recorded as comprehensively in the sixteenth and seventeenth centuries as Nahuatl. By 1645 Nahuatl had been described in at least five grammars, and in 1571 Fray Alonso de Molina published an expanded version of a Spanish-Nahuatl, Nahuatl-Spanish dictionary, containing more than forty thousand entries. The corpus of texts in Nahuatl from that same period is, both in kind and in quantity, equally impressive and gives it a unique position among the indigenous languages of the Americas. The Copenhagen Nahuatl Dictionary Project (CoNDiP) brings together the lexicographic comprehensiveness of Molina with the contextual usages of the words as found in the texts. The organization of CoNDiP and our application of the dramatic new possibilities in lexicography provided by computers are the focus of this chapter. But first a few words about the abundant corpus of Nahuatl documents and their significance are called for.

The Nahuatl texts are abundant. They represent a great variety of types and genres, comprising historical accounts and chronicles, myths, and poetry pertaining to pre-Columbian culture; and colonial-period administrative documents and Christian devotional texts, such as catechisms, moral dialogues, confessionaries, and stories about the saints. Some of the texts were written by indigenous people who had learned writing and Latin in schools established by the Spaniards shortly after the conquest. Others were collected by Spanish friars. Most of the historical accounts

The project described here has been supported by the Carlsberg Foundation.

and chronicles have no known authors or compilers, but among the important texts we also find the work of an early indigenous historian, Chimalpahin Quauhtlehuanitzin (1958, 1997).

The work of Fray Bernardino de Sahagún, who strove to bring together material for his *Historia general de las cosas de Nueva España* (General History of the Things of New Spain, 1570), provides a major source of information. With the help of young bi- or trilingual Aztecs he visited wise and knowledgeable elders in various towns and collected information in Nahuatl on all aspects of the culture, including detailed descriptions of animals, plants, and other natural phenomena. His organized and edited version of this material is known as the *Florentine Codex*. It consists of twelve volumes, beautifully composed in two columns, Nahuatl and Spanish, and with numerous colored drawings (Sahagún, 1579 [1950–82]). It is undoubtedly the most frequently cited source on all aspects of Aztec culture. Some of the original source material for the *Florentine Codex* is also preserved (Sahagún 1997).

In their desire to convert the indigenous people to Christianity, the friars felt a great need to speak and to be understood. They found that it was easier for them to learn the indigenous languages than it was for the Indians to learn Spanish. Consequently, in part because of the friars' insistence that Nahuatl be used also in official matters and in part because Nahuatl had functioned as a lingua franca already before the arrival of the Spaniards, it continued to be widely used, both practically and officially, all through the sixteenth century and well into the seventeenth (see Heath 1972). As a consequence, there is, in addition to literary texts and historical records of preconquest events, an abundance of documents in or from local archives: land documentation, testaments, municipal documentation, petitions, and correspondence. The literature written in the Nahuatl language from the sixteenth and seventeenth centuries thus represents a rich source for the study of Aztec culture—the political and social system, religion, ethics, and so on—as well as for the study of the Nahuatl language itself.

However, Nahuatl from the sixteenth century is not an easily understood language, and the documents, many of which are written in a highly elaborated style with metaphors and other figures of speech, were hidden away in archives for centuries. Scholars rediscovered the texts and the tools with which to gain access to them only toward the end of the nineteenth century. The first half of this century saw German publications of many of the important texts with translations and discussions of their content (Eduard Selser, Walter Lehmann, Gerdt Kutscher, Günter Zimmermann). During the second half of the twentieth century, scholars from the United States, Mexico, and France (e.g., Arthur J. O. Anderson, John Bierhorst, Charles E. Dibble, Jacqueline Durand-Forest, Angel

María Garibay K., James Lockhart, Miguel León-Portilla, Alfredo López Austin, Susan Schroeder) have added tremendously to the wealth of bilingual publications of Nahuatl texts from the sixteenth and seventeenth centuries.

A profound understanding of the significance of Nahuatl texts and the desire to produce a lexicographic tool that, in terms of word entries and information, would go beyond that of Molina and make its users better equipped for reading and interpreting the many texts has led us in Copenhagen to initiate the Copenhagen Nahuatl Dictionary Project.

This presentation of CoNDiP is organized in the following way: after a brief, informative description of the basic aim and structure of CoNDiP, I expound the lexicographic merits and subsequent achievements made possible by advances in electronic technology; then I discuss the problems we have been faced with in producing CoNDiP, some of which are common to all types of lexicographic work and some specific to the new kind of situations created by the computer; and finally, I give a short account of the history of CoNDiP.

2. BRIEF DESCRIPTION OF CoNDiP. The primary and basic aim has been to make the abundant Nahuatl text material immediately available with as little interpretation as possible, to avoid the inevitable personal bias that befalls any translator. In preparing CoNDiP, we have had in mind both scholars working with Aztec culture and those who are interested more specifically in the Nahuatl language.

CoNDiP is above all a computer-based compilation, organization, and analysis of words from sixteenth- and seventeenth-century Nahuatl texts; it exploits the advantages of the computer and is at variance with the traditional dictionary form to the extent that it can in no way be converted into one or more printed books. Thanks to the electronic medium and the organization of CoNDiP, there is no limit to the amount of material (texts, words, and comments) that can be entered continuously into the system, with no end point, only updated versions.

CoNDiP consists of three permanently and concurrently interactive components: a word dictionary, a root dictionary, and a text component; the links between the elements of these three components form the backbone of the system (see fig. 9.1). The typical entry in the CoNDiP word dictionary contains what would be expected in a traditional dictionary: the entry word, grammatical information, and a gloss. In addition, it displays the root(s) contained in the word in the order in which they occur as well as example sentences from texts. A window for comments written by the compilers as well as by the users also appears. The root(s) given in a word entry are linked to entries in the root dictionary. In this the typical root entry includes the entry root, a gloss, and a list of all the

WORD ENTRY

WORD	class	Eng. GLOSS
gram.inf.		Sp. GLOSS
ROOT(S)		
EXAMPLE SENTENCES/(COMMENT)		
.....		
.....		

ROOT ENTRY

ROOT	Eng. GLOSS
	Sp. GLOSS
(COMMENT)	
...	
WORDS containing the root	
...	
...	

TEXT

Standardized form
...
...
Paleography
...
...
Translation
...
...

Figure 9.1. Interactive Components of CoNDiP.

words that contain the given root; the root dictionary also provides the possibility for comments on the individual root. The interactiveness of the program is realized in the two-way links between words and roots. The example sentences in the word entries come from the text corpus. Every word in the text corpus is, with its full sentence, linked to its word entry in the word dictionary; and in the text component, texts are divided into

numbered sentences, given in paleography,¹ in a standardized version, and with a translation.

The user may look up a word, a root, or a text, including page or folio number, and may naturally profit from the links to move around between the components. In the present version of CoNDiP, the user may not add comments or entries; it is possible, however, to lift out of the system a word entry, a root entry, or the entire list of roots.

Eventually the text component will comprise the relevant sixteenth- and seventeenth-century texts in Nahuatl. We have included Molina's 1571 dictionary as a text; this means that every entry has, in the form of "sentences," all the entries from Molina's Spanish-Nahuatl and Nahuatl-Spanish sections that contain the word in question. Since it is entered as a text, it is possible to jump directly to the text component in order to see the preceding and following entries in Molina's dictionary by scrolling up and down in it. The Molina text is given only in paleography.

The basic structure of CoNDiP was developed in the late 1980s, and the interactive components and the windows for comments were all created in a DOS framework. CoNDiP has not yet been rewritten for a Windows format. Additional features and options, some more closely related to traditional lexicographic aspects and some beyond the normal possibilities of a dictionary, are discussed in the following sections.

3. LEXICOGRAPHIC MERITS OF CoNDiP.

3.1. ALPHABETICAL ORDER OBLITERATED. Words and roots are nowhere in CoNDiP organized in any alphabetical order. They are numbered chronologically according to when they were entered into the program. This number is covert, however, and is of no significance to the user, who accesses a word entry simply by typing the word or by going from the root dictionary or the text component, and accesses a root entry through the word dictionary link or by typing the root. The great advantages of this complete obliteration of alphabetical ordering—radical as it may seem in a dictionary—may not be immediately obvious.

The first advantage is that the dictionary has been representative from the start, in that the first version was not limited to words beginning with certain letters of the alphabet, and when expanded it is expanded evenly over the alphabet. An obvious problem with the obliteration of alphabetical ordering is that one cannot check the words just above or below in

1. Among Nahuatl scholars the term "paleography" refers to the exact rendering of a text from a manuscript, as opposed to any standardization or transcription of the text.

the alphabet, but the inclusion of Molina's dictionary and the easy access to word lists in the root dictionary should remedy this loss.

The second advantage is of an entirely different nature. It has to do with Nahuatl orthographies in general, since alphabetical ordering is in no way a straightforward matter in Nahuatl lexicography. In the sixteenth century, orthography was generally not as consistent as it is today. We find a great deal of variation in choice of letters in Spanish texts from the period. Similar orthographic inconsistencies are found in the Nahuatl texts, and they correspond partly to variation in Spanish writing but stem also from phonological differences in the two languages, differences that caused the Spaniards problems. It is a question both of underdifferentiation (Nahuatl distinguishes long and short vowels and has a postvocalic glottal stop, traditionally called *saltillo*, both unknown features in Spanish) and of overdifferentiation (Spanish has five vowels versus four in Nahuatl).

As a result of differing traditions in the writing of Nahuatl and of recent attempts at an orthographic systematization, the alphabetical order in extant dictionaries varies frustratingly. Because all these dictionaries contribute to our understanding of the language, scholars who work with sixteenth-century texts will constantly consult all of them. But because of their different orthographies and above all because of their different alphabetical order, it takes a tremendous amount of practice, and one really has to mind one's p's and q's to consult them successfully.

The choice of problematic letters and their critical order as found in the major dictionaries and in J. Richard Andrews's (1975) grammar are charted below. Table 9.1 gives the critical letters (keyed to the phonemes in the leftmost column) that are used in the dictionaries. In general I have included only those found in word-initial position. Certain intricacies, such as the writing of /w/ and /k^w/ in syllable-final position, are thus not included in the tables. ([], / / and < > are used according to the traditional conventions for phonetic, phonemic, and orthographic symbols, respectively.) No significance should be attributed to the vertical order of the phonemes/letters in table 9.1. Table 9.2 displays the ORDERING of words according to initial letters in the selected sources (not all the letters of the alphabet are included; <a>, <e>, and other letters that in all the sources are identical and appear in the same place in the alphabet have been left out). Here the vertical order of the letters represents the alphabetical order found in the dictionaries; the phonemic column is the rightmost because it has only an auxiliary function. Thus in Molina's dictionary, for example, we find words beginning with <ca> before those that begin with <ça>, those beginning with <ça> before those that begin with <ce>, and so on. The notations <ca, ça> and <co, ço> in the column under "Siméon" are explained below under (1). A comparison of the two

TABLE 9.1. Choice of Letters in Major Nahuatl Dictionaries

Phon	Molina 1571	Siméon 1885	Campbell 1985	Andrews 1975	Karttunen 1983	Bierhorst 1985
/k/	c/_a,o,C,# qu/_i,e	c/_a,o,C,# qu/_i,e	c/_a,o,C,# qu/_i,e	c/_a,o,C,# qu/_i,e	c/_a,o,C,# qu/_i,e	c/_a,o,C,# qu/_i,e
/s/	c/_e,i ç/_a,o z/_C,#	c/_e,i ç/_a,o z/_C,#	c/_e,i z/_a,o z/_C,#	c/_e,i z/_a,o z/_C,#	c/_e,i z/_a,o z/_C,#	c/_e,i z/_a,o z/_C,#
/č/	ch	ch	ch	ch	ch	ch
/k ^w /	cu/_e,i qu/_a	cu/_e,i qu/_a	cu/_e,i qu/_a	cu/_e,i qu/_a	cu/_e,i qu/_a	cu/_e,i qu/_a
/ç/	tz	tz	tz	tz	tz	tz
/w/	v	u	u	hu	hu	hu
/y/	y	y	y	y	y	y
/i/	i	i	i	i	i	i
/o/	o,v	o,u	o	o	o	o
/ʔ/	-	-	-	h	h	h

TABLE 9.2. Order of Letters in Word-Initial Position in Major Nahuatl Dictionaries

Molina 1571	Siméon 1885	Campbell 1985	Andrews 1975	Karttunen 1983	Bierhorst 1985	Phon
ca	ca,ça	ca	ca	ca	ca	/ka,(sa)/
ça						/sa/
ce	ce	ce	ce	ce	ce	/se/
ch	ch	ch				/č/
ci	ci	ci	ci	ci	ci	/si/
co	co,ço	co	co	co	co	/ko,(so)/
ço						/so/
			ch	ch	ch	/č/
			cua	cua	cua	/k ^w a/
cue	cue	cue	cue	cue	cue	/k ^w e/
cui	cui	cui	cui	cui	cui	/k ^w i/
			hu	hu	hu	/w/
i,y	i,y	i	i	i	i	/i,(y)/
qua	qua	qua				/k ^w a/
que	que	que	que	que	que	/ke/
qui	qui	qui	qui	qui	qui	/ki/
t	t	t	t	t	t	/t/
tl	tl	tl	tl	tl	tl	/ʎ/
to	to	to				/to/
tz	tz	tz	tz	tz	tz	/ç/

(continued)

TABLE 9.2. (continued)

Molina 1571	Siméon 1885	Campbell 1985	Andrews 1975	Karttunen 1983	Bierhorst 1985	Phon
va,ve,vi	ua,ue,ui	ua,ue,ui				/wa,we,wi/
vC	uC					/oC/
		y	y	y	y	/y/
		z	z	z	z	/s/

tables reveals that there is broad agreement concerning choice of letters in the dictionaries and that it is in the ordering of the letters that they differ, creating difficulties for those who continually consult them.

We can see the scope of the difficulties by looking at the seven crucial problems below.

1. Molina (1571) and Siméon (1885) agree on rendering the phoneme /s/ before /a/ or /o/ as <ç>, but whereas Molina gives all the words that begin with <ca> (/ka/) before those beginning with <ça> (/sa/), Siméon alphabetizes according to all the letters in the words, thereby giving at first a few that begin with <ca>, then some with <ça>, then more with <ca>, and so on. The two lexicographers differ in the same way in the case of words beginning with <ço>. The more recent sources avoided this problem by choosing to write <z> for /s/ before /a/ and /o/.
2. There is some disagreement as to where to place words beginning with <ch> (/č/). Molina, Siméon, Campbell (1985), and Bierhorst (1985) intermingle them with words beginning with <c>; Andrews (1975) and Karttunen (1983) place them in a separate section after all the words that begin with <c> for /k/ and /s/.
3. Before /a/, the phoneme /k^w/ is written <qu> by Molina, Siméon, and Campbell; Andrews, Karttunen, and Bierhorst write <cu>. As a consequence, words beginning with /k^wa/ are found in different places in the two groups of sources. They all agree on writing <cu> for /k^w/ before /e/ and /i/. /k^w/ does not occur before /o/.
4. In agreement with Spanish tradition in the sixteenth century, Molina and Siméon use both <i> and <y> for the consonant phoneme /y/, placing words beginning with /y/ in the section with words that begin with /i/. All later sources reserve the letter <i> for the vowel phoneme /i/ and the letter <y> for the consonant phoneme /y/.
5. The phoneme /w/ divides the sources differently: Molina writes it with the letter <v>, Siméon and Campbell prefer <u>; Andrews, Karttunen, and Bierhorst use <hu>. One thus has to look for words beginning with /w/ in totally different places in the sources.

6. The sources disagree on how to handle words that begin with the letter <ɔ>, which may represent just /t/ or, followed by <l> or <z>, two other phonemes, namely, /λ/ and /ç/. Only Andrews and Karttunen establish three distinct sections, one for each of the three phonemes. The others simply organize words beginning with the letter <ɔ> alphabetically.
7. Saltillo was generally not indicated in the texts from the sixteenth and seventeenth centuries. It is not found in Molina, Siméon, or Campbell—the latter two being derived from the first. Andrews, Karttunen, and Bierhorst, however, write it with the letter <h> and include it in the alphabetic ordering. The saltillo does not occur word initially, so it is not included in table 9.2, and the problems this additional letter creates are limited.

3.2. SEVERAL ORTHOGRAPHIES. Eliminating the alphabetical order in CoNDiP does not solve all the problems created by the diversity of orthographic standards. Scholars who work with Nahuatl texts from the sixteenth and seventeenth centuries have a variety of motives and backgrounds and therefore also different preferences as to orthography; some are more linguistically inclined than others, who generally prefer the words to be written the way they find them in the texts and who see no need for the linguist's exotic symbols or for an indication of saltillo and vowel length. This situation can be easily and elegantly taken care of on a computer.

CoNDiP offers the user the option of three different alphabet standards that we call (1) sixteenth-century orthography (ORT); (2) the normalized version (NOR); and (3) transcription (TRA). The selected alphabet standard is given in root entries, in word entries, and in the standardized version of the texts; it does not operate within comments, where we normally use sixteenth-century orthography. This selection of one of the three alphabet standards is in no way definitive; one can shift between them at will.

SIXTEENTH-CENTURY ORTHOGRAPHY most closely resembles the orthography we find in the documents, but it is consistent. A given form is always and everywhere written in the same way.

THE NORMALIZED VERSION differentiates short and long vowels: /a/ versus /ā/, /calli/ 'house' versus /cālli/ 'raven'; and an apostrophe indicates saltillo, for example, /ta'tli/ 'father'. A capital vowel indicates that we have no information about vowel length or a possible saltillo: for example, /kAntli/ 'cheek' stands for /kantli/ or /kāntli/, and /çIwAktli/ 'a plant' stands for /çiwaktli/, /çiwāktli/, /çi'waktli/, or /çi'wāktli/.

Some letters in NOR differ in a few contexts from ORT:

ORT	NOR	
quiqua	quicua	'eats it'
açaca	ázaca	'carry water'

TRANSCRIPTION is the standard that least resembles what we see in the documents. In TRA every phoneme is represented by only one symbol, the same one in all contexts.

The following list exemplifies the differences between standardized sixteenth-century orthography and our transcription.

ORT	TRA	
chantli	čānli	'home'
cactli	kakli	'sandal'
quenin	kēnin	'how'
quiqua	kik ^w a	'eats it'
quicuepa	kik ^w epa	'turns it'
neuctli	nek ^w li	'honey'
čan	san	'only'
cecec	sĒsĕk	'cold'
nez	nĕs	'appeared'
nezque	nĕske'	'appeared (pl.)'
tlacatl	lākaλ	'man'
tzatzi	ča'çi	'shouts'
huehuetl	wĕwĕλ	'drum'
huauhtli	wawli	'amaranth'

TRA is the master orthography in which roots, words, and texts are entered into the program and from which the two other orthographies are mechanically derived. The system is sensitive to the three alphabet standards and responds differently depending on the standard in operation. TRA is prim and recognizes only the correct spelling with letters that belong in TRA; it will not recognize, for example, <tl> and <ch>. This strict requirement does not include vowel length and saltillo unless these features are the only indication of a distinction between two or more words. Thus if, for example, one wants to look up /kisa/ and types it with a short *i*, the correct entry will appear. However, in the case of word pairs that differ only as to vowel length or saltillo, for example, /toka/ 'follow' and /tōka/ 'bury', only the form typed with correct vowel length will appear. In ORT both words will appear in a suggestion list. And here the tolerance is limited only by potential ambiguities. The word <çacatl> 'grass', for example, can successfully be looked up in any one of the following ways: *sakatl*, *sacatl*, *zacatl*, *zakatl*, *çacatl*, *çakatl*, *sakaλ*, *sacaλ*, *zacaλ*, *zakaλ*, *çacaλ*, *çakaλ*. Thus ORT gives much more leeway than TRA for looking up roots and words, and TRA offers more information, which, however, is of interest to fewer users.

Occasional Spanish loanwords appear in the Nahuatl texts. These are dealt with according to other rules.

3.3. INSIGNIFICANCE OF GLOSSES. Since the aim of CoNDiP is to give immediate and comprehensive access to occurrences of words in texts and thereby to encourage new and more well-founded interpretations, the English and Spanish glosses in the word entries lose their importance. This appraisal of the supplied glosses is supported by the immediate access to complete entries from Molina's dictionary. As a consequence, little effort has been spent on finding the absolutely "right" glosses. In keeping with this view, only little space is allocated for glosses. In fact, only after sufficient text material has been entered would it make sense to work out detailed glosses on the basis of those texts.

The same is true of the text translations. We allow ourselves to stay close to the original, both syntactically and lexically. We also introduce a kind of methodical consistency that leads to somewhat awkward and rigid English. However, we adhere to this consistency because it makes the translating process more appropriate for our demands and because it offers information to the user. Furthermore, since the text in all its versions—paleography, standardized, and translation—is divided into sentences, each sentence should, if at all possible, have a meaningful translation that is sufficiently independent of that of the preceding and following sentences. We avoid the traditional dilemma for any translator, namely, striking a balance between being true and close to the original and producing a text acceptable in the target language.

No function for looking up words through the English or Spanish glosses is implemented. However, a more general search function allows the user to search for an English or Spanish gloss.

4. ACHIEVEMENTS. Some of the possibilities permitted by the electronic medium go well beyond the limits of the traditional book format, as discussed below.

4.1. FULL TEXTS. The first overwhelming advance is the possibility of having an almost endless number of full texts next to the word entries. That the meanings of words are best understood through their use is the obvious reasoning behind this central feature of CoNDiP. In good traditional dictionaries, words are exemplified by illustrative phrases or sentences—in the best cases, by genuine text sentences. In CoNDiP the use of words is illustrated not only by a few examples but also, ideally and eventually, by the sentences in a comprehensive corpus of sixteenth- and seventeenth-century Nahuatl texts in their full context. The inclusion of full texts and, above all, the linking of these texts to the word entries in the dictionary are cardinal features of CoNDiP. For a quick assessment the reader may look at the list of occurrences of a given word in the

narrow context of the sentence, but for every occurrence, there is direct access to the full text in the text component.

4.2. MORPHOLOGICAL ANALYSIS. During the initial phase of CoNDiP, the first ten to thirty pages of text were subjected to quite a bit of experimentation. Our experiments addressed two distinct areas. The first had to do with the presentation of the texts and how they should appear in the text component and how the text sentences should be displayed in the word entries. This involved both the precise configuration of the screen in the two components and decisions about alphabet standards. The second dealt with the adaptation of texts for their entry into the program and with the crucial linking of what we call a WORD FORM to a word as found in the entry of a dictionary. In this second area, our discussions and decisions were dictated not by pedagogical considerations but by the structure of the Nahuatl language and the inconsistent orthography found in the documents. To establish a morphological analysis that will support the mechanical linking of inflected words to dictionary words is an overwhelmingly complex task for any language, and that Nahuatl is a polysynthetic language with a rich inflectional system employing prefixes, suffixes, compounding, and noun incorporation only adds to the complexity of that task.

A mechanical linking of word forms to dictionary entries was not intended in the first phase of CoNDiP. But even for manual linking, the preparation of texts still involved two connected sets of critical tasks: (1) interpreting the inconsistent orthography found in the texts and (2) producing a morphological analysis. The two are interconnected in that the saltillo (generally not indicated in the texts) and other crucial phonological features play a significant role in the morphology. This has required much debate and work. I discuss some of the details of the two connected sets of tasks.

Simple inconsistencies in the orthography involve, for example, the variable use of several letters for a vowel, /i/ or /o/, and a consonant, /y/ or /w/. Thus <i, y, j> can represent /i/ and /y/; <o, u, v> can represent /o/ and /w/. This situation is further complicated by the fact that there is no phonetic distinction between /ia/ and /iya/ or between /oa/ and /owa/. Wherever the syncretism can be resolved, the resolution is based on morphological evidence (see Canger 1980: chap. 1). In the *Florentine Codex*, the word /yōli/ 'he lives' is found in the following variants: <ioli, iuli, yoli>; /yālwa/ 'yesterday' has four orthographic variants: <ialoa, ialhoa, yaloa, yalhua>; and /iwiyan/ 'calmly' has twelve: <ihuiian, iuiian, iuijan, iujian, iviian, ivijan, juijan, jviiian, yuiyan, yujian, yviian, yviiian>.

This rich orthographic variation is crucial when looked at from the perspective of how to correlate an orthographic variant with an analyzed

word form. The readings of a word like <quioalioalioaia> /Ø-ki-wāl-yawaloā-ya-/ (3.S-3SG.O-DIR.towards-surround-IMPF-PL) 'they surrounded him' can be used as an example of the complexity; it gives 406 possible readings.

The interplay between problems of phonology and morphology can be illustrated by a few examples:

- | | | |
|------------------|---------------------------------|-----------------------------|
| (1) <tlaqua> | (a) /la-k ^w a/ | 'he eats' |
| | (b) /la-k ^w a-'/ | 'they eat' |
| | (c) /la-k ^w a'/ | 'he ate' |
| (2) <tlaquatiuh> | (a) /la-k ^w a'-ti-w/ | 'he goes away while eating' |
| | NONREF.O-eat-LIGATURE-go | |
| | (b) /la-k ^w ā-tiw/ | 'he goes to eat' |
| | NONREF.O-eat-DIR.away | |

The three analyses of ambiguous <tlaqua> in (1) are explained by the fact that saltillo functions in some tenses as a marker of plural of subject, and some verbs, for example, /k^wa/ 'eat', have two stems, one ending in a vowel and one in saltillo. The two analyses of <tlaquatiuh> involve (2a), a composite construction with the stem found also in (1c), and (2b), a directional suffix /tiw/ added to a third stem form of the verb 'eat', /k^wā/. In both cases, arguments for the correct analysis will come from the context.

The precise analysis needed for the linking of word forms in texts with words as they appear in the dictionary thus yields a valuable by-product, namely, the comprehensive morphological analysis of word forms in the texts. In preparing the texts, word forms must be analyzed into morphemes and the morphemes are marked as STEMS, DERIVATIONAL AFFIXES, and INFLECTIONAL AFFIXES. Stems (in many cases identical with roots, see §5.1) are marked by a preposed hyphen, derivational affixes by a preposed equal sign, and inflectional affixes by a plus sign.

- | | |
|-----------------------------|-----------------------|
| (3) +ti+k-itti=ti'+ke' | 'we showed it to him' |
| 1PL.S-3O-see-CAUS.PERF-PL.S | |

Reduplication serves a number of functions in Nahuatl (see Canger 1981), for example: 'plural' of nouns, /koyō-λ/ 'coyote', /kō-koyo-'/ 'coyotes'; 'distributive', /in-ča'-čan ō Ø-ya'-ya'-ke'/ (3PL.P-DISTR-home past 3S-DISTR-GO.PERF-PL.S) 'they went, each to his own home'; 'repetitive', /Ø-šamā-ni/ 'it breaks', /Ø-ša-šama-ka/ 'it breaks in many places'. Some of these functions form part of the inflectional system, and in other cases they belong either in the productive derivational system or among fossilized derivational affixes (see §5.5). In addition to being marked as inflectional or derivational, these prefixes also receive a special mark, ≈, for reduplication:

- (4) +=kō-koyo+ 'coyotes'
PL-coyote-PL
- (5) ==ka-kala=ka 'clank'
REP-sound-INTR

In the present version of CoNDiP this morphological analysis is displayed overtly in TRA; in NOR, only stems are marked as such:

- (3) TRA NOR
+ti+k-itti=ti'+ke' tik-ittiti'ke' 'we showed it to him'
1PL.S-3O-see-CAUS.PERF-PL.S

But the combination of alphabet standard and morphological analysis can obviously be separated completely, allowing the user to have any possible combination of the two options.

The declared intention of forcing as little interpretation of the texts as possible on the user may seem to be disregarded in the comprehensive morphological analysis manifested in the transcription. However, the interpretation exhibited in the transcription never conceals the unadulterated ambiguous original, because whatever alphabet standard the user selects, a paleographic version of the text will always appear alongside the analyzed one. The only analysis applied in the paleography is a division of the text into words and sentences. While the division into words is fairly uncontroversial, the identification of clauses is the result of a syntactic analysis that may be questioned, but it is overt and has no concealed consequences, and some division of the text is necessary for purely practical purposes.

The consistent morphological analysis has in fact made the linking partly mechanical. The analyzed text is typed in manually, but subsequently the system, beginning from the first word of the text, suggests the word entry to which that word form is to be linked. I accept it or—in the cases when the suggestion is not correct—I find the correct one, link, and go on to the next word.

For entering Molina's dictionary, we went a little further. A multifaceted program was written that interprets the orthography, provides morphological analyses of a given word, and identifies the dictionary word to which it should be linked. Inflected word forms in Molina's dictionary are limited to a few standardized ones, so as a "text" it is naturally not as complex and demanding as any regular text would be. It is the plan, however, to expand this first pilot parsing program with the goal of creating a tool to aid the user in analyzing inflected word forms. It will handle the same two separate problems: giving a reading to the orthography and identifying and marking the morphemes. Given the profusion of word forms that present multiple readings, the parsing program will in many cases provide suggestions rather

than give final analyses. A final and fully developed parsing program will also serve in entering new texts.

4.3. COMMENTS AND OPEN DISCUSSIONS. Another advantage of the electronic medium is the possibility of comments and open discussions about every word and root. The almost unlimited capacity of the computer today is being exploited in CoNDiP in the concept of comments on entries of individual words and roots. Furthermore, the comments are thought of as a forum for discussions. Wherever called for, my arguments for the phonological interpretation, for a gloss, for root identification, and so on, appear in the form of comments. Others may provide added information or disagreement with my analyses or may contribute with comments on other entries. Such reactions will be entered; if there is more than one comment on a word or root, these are organized chronologically.

4.4. LIST OF WORDS WITH A SHARED ROOT. Our understanding of the semantics is not only improved through the many text examples. The root entries, which include a list of words containing the given root, provide a better grasp of the basic meaning, history, or etymology of a word.

The word /çak^wa/, for example, is traditionally translated by 'close', but /ā-çak^wa/ 'stop water', /i'iyō-çak^wa/ 'suffocate', and /e'ēka-çak^wa/ 'protect against the wind' suggest that the meaning is rather 'block (off)'. Precisely this grouping of words around their roots is also found in Campbell's dictionary (1985). He reorganized the material from Molina's Nahuatl-Spanish dictionary in entries according to what he calls "main morphemes," a concept that roughly corresponds to our concept of root.

5. PROBLEMS HANDLED BY CoNDiP. I have already commented on a number of problems related to the way Nahuatl was written in the sixteenth century. In our work with CoNDiP we have naturally encountered many other problems of different types. I shall touch on a few of these, some well known in traditional lexicography and some tied to the medium and organization of CoNDiP. They deal with (1) the delimitation of certain words; (2) the inclusion of productively derived words; (3) sources for saltillo and vowel quantity; (4) the concept of the root and homonymy in roots; (5) the distinction between derivation and inflection; and (6) the organization of the word lists and texts.

5.1. DELIMITATION OF WORDS. In describing how texts are handled in CoNDiP, I have said that the division into words may be less controversial than that into sentences; in many cases the wealth of easily identifiable prefixes and suffixes provides an easy solution. However, there are cases of disagreement among Nahuatl scholars as to what a word is, and the

handwritten texts provide little guidance in word division because what we recognize as spaces seem to appear randomly. An example of such disagreement is /ō/ indicating 'past'. Traditionally it is considered a prefix; however, according to a general phonological rule, it does not belong to the verb form. The prefix for third-person singular object has two variants, /ki/ and /k/; the distribution of the two is related to a simple rule that allows only the following syllable types: V, CV, VC, CVC. In other words, Nahuatl has no word-initial or word-final consonant clusters. When the prefix for third-person singular object is the first prefix in a verb and precedes a consonant, it has the form /ki/: /Ø-ki-k^wa/ 3SG.S-3SG.O-eat 'he eats it'. If the verb begins with a vowel or the prefix in question is preceded by another prefix that ends in a vowel, the other variant, /k/, appears: /Ø-k-a'si/ 3SG.S-3SG.O-reach 'he reaches it', /ni-k-k^wa/ 1SG.S-3SG.O-eat 'I eat it'. Now, if the element /ō/ 'past' precedes the prefix in question, it does not influence the choice of variant: /ki/ occurs, hence /ki/ is in word-initial position and /ō/ is a separate word: /ō Ø-ki-k^wa/ PAST 3SG.S-3SG.O-eat.PERF 'he ate it', not */ō-Ø-k-k^wa/. Another argument supporting this analysis is that the particle /iw/ 'thus' and a few other particles occasionally appear between /ō/ 'past' and the verb.

A case in which other considerations come into play is /a', a morpheme indicating negation. It occurs most frequently in the word /a'-mo/ 'no, not' and with pronominals and various particles, /a-yāk/ 'no one', /a'-λε/ 'nothing', and so on. However, /a' also occurs attached to a noun or a verb, and here we distinguish between cases in which /a' functions as a separate word from those in which it clearly forms part of a word, for example:

- (6) (a) /ni-n-a'-lāsa/ 'I am fighting death'
1SG.S-1SG.REFL-neg-let.go
(b) /a' ō ni-k-nek/ 'I did not want it'
neg past 1SG.S-3SG.O-want

In (6a) /a' is preceded by the personal prefix for subject, whereas in (6b) it precedes the prefixes as well as the particle /ō/ 'past'. Whether /a' in /ni-n-a'-lāsa/ should be analyzed as an inflectional element or whether the word should be considered a compound is an entirely different question.

Our general principle is that wherever there is no clear evidence in favor of considering /a' as forming part of a word, it is analyzed as a freestanding particle. This kind of decision has certain consequences that have played a role in our debates. It is a question of recognizing either the free variant or the attached variant of /a' as the unmarked one. If the attached variant is considered the unmarked one, we would have to establish a great many words beginning with /a', the list of words in the entry for /a' in the root dictionary would be very long, and every word

would be exemplified by few sentences in the word dictionary. But if the free variant is chosen as the unmarked one, fewer words would have to be established, and many uses of /a' would be exemplified in its entry in the word dictionary.

5.2. INCLUSION OF PRODUCTIVELY DERIVED WORDS. Decisions about whether to include words that are productively derived (a strategy that may lead to unfortunate results; cf. Malkiel's (1962) "ghost words") cause no problems for CoNDiP since all words that occur in a text, including in Molina's dictionary, must be given separate entries. If they have not been found in texts, they have no entry.

5.3. SOURCES FOR SALTILLO AND VOWEL QUANTITY. Problems involving decisions about saltillo and vowel quantity are not new. They have been dealt with by those of us who have tried to establish a systematic phonemic way of writing sixteenth-century Nahuatl. This has been done most comprehensively by Andrews (1975), Launey (1979, 1986), Karttunen (1983), and Bierhorst (1985). The material available for establishing vowel quantity and the occurrence of saltillo are above all from Carochi (1645), who not only indicates saltillo and vowel length but also occasionally discusses these features. In addition to this cardinal work, we have a few other sources from the seventeenth century and some vocabularies of several quite distinct contemporary dialects from the second half of the twentieth century. The scholars mentioned, who have established their own word corpus and rules for assigning length to the vowels, are not in perfect agreement. The disagreements may be a result of varying access to the sources and also of differing interpretations of the sources. I do not expect uniformity in the phonology of the dialectally and socially diversified Nahuatl spoken by millions of people in the sixteenth and seventeenth centuries. For CoNDiP, we have a practical attitude toward these details of phonology: saltillo and vowel length are included wherever we have information about these features. If our information about them is not straight out of Carochi's grammar, we list our sources of information in a comment on the root. As already mentioned, if we have no information from these sources for some word, we indicate this explicitly by writing the vowels in question with a capital letter. As for vowel quantity that forms part of the inflectional and derivational system, we have established a consistent set of rules, based on phonological analysis and on Carochi's explicit comments.

5.4. THE CONCEPT OF THE ROOT. In traditional dictionaries, an entry generally subsumes several or many inflected word forms; and one form for each word class is selected to represent all the inflected forms. For modern European languages, the infinitive is the form of the verb that

appears in dictionaries, and according to a firmly cemented tradition, this form subsumes all the possible inflected forms of the verb. In this sense the infinitive is an abstraction. The number of inflected forms subsumed under the dictionary form obviously depends on the morphology of the language. This tradition relates to the abstract concept of the "word" that is well established in European culture.

The stem or root is a form that plays a role in traditional lexicography. It is a form that does not appear in isolation in spoken or written texts and is in fact the linguist's construct. Roots in general represent a greater degree of abstraction than dictionary words and are mostly further removed from word forms. It has been mentioned that the purpose of the root dictionary is to compile all the words that contain a given root, assuming that such a compilation will be useful in studies of the derivational system of the language and will give insight into Nahuatl semantics. In the establishment of roots in CoNDiP, we have tried to reconcile three factors or requirements: (1) a certain consistency in the morphological analysis; (2) a view to what may be in keeping with the expectations of the various users of CoNDiP; and (3) some considerations for our knowledge about the history of Nahuatl.

The desire for consistency is a requirement in morphological analysis in general, but we have not adhered strongly to any specific model of morphological analysis and have given priority sometimes to processes and sometimes to items. There are cases in which our consistency results in words that include a stem form different from that of the root. For the words /k-itki/ 'he carries it', / λ a-tki- λ / 'belongings', and /itko/ 'people carry it', we have a root that, according to the word, appears in different shapes (*itki*, *tki*, *itk*). Here we choose the longest form (*itki*) to represent the root.

Simple nouns provide another example. Most simple nouns occur in three morphological contexts (1) as unpossessed nouns with a so-called absolutive suffix (- λ i/C₋, -li/L₋, - λ /V₋); (2) as possessed nouns, either with a possessive suffix (-wi/C₋, -w/V₋) or with no suffix; and (3) with no suffix, as the first element in a compound.

UNPOSSESSED	POSSESSED	COMPOUND	
okič- λ i	'man'	i-okič-wi	okič-ēwa 'attack bravely' (man-rise)
siwā- λ	'woman'	i-siwā-w	siwā-tōtol-in 'turkey hen' (woman-turkey)
kaši- λ	'bowl'	i-kaš	kaš-peč- λ i 'plate' (bowl-bed)
naka- λ	'meat'	i-nak	naka-ok ^w il-in 'maggot' (meat-worm)
šōči- λ	'flower'	i-šōči-w	šōči-k ^w āl-li 'fruit' (flower-good)

The roots we have established for these examples are /okič/ 'man', /siwā/ 'woman', /kaš/ 'bowl', /naka/ 'meat', and /šōči/ 'flower'. The determining criteria have been the form of the root found in compounds and the desire to recognize only three variants of the absolutive suffix for the 'singular' (- λ i, -li, - λ). The consequence of this decision is that not all stem forms are identical with the root; thus the word /kaši λ / consists of a variant of the absolutive suffix, / λ /, and a stem, /kaši/, that differs from the form that we establish as the root, /kaš/.

A more intricate case involves regular sets of derived verbs: an intransitive verb in /-wi/, a transitive one in /-oa/, and an applicative one in /-lwia/.

INTRANSITIVE	TRANSITIVE	APPLICATIVE	
i' λ aka-wi	i' λ ak-oa	i' λ aka-lwia	'damage'
peçi-wi	peç-oa	peçi-lwia	'shine'
poli-wi	pol-oa	pol-wia	'disappear'

The rules that govern the derivation of the transitive and the applicative forms in these sets are fairly simple (see Canger 1980: chaps. 3, 4), but they were not productive in the sixteenth century, so in this case the identification of the root involves the history of the language. And the user who knows only the transitive verb of one of these sets, for example, /poloa/, cannot with certainty predict the root, /poli/. Here considerations for a historically plausible analysis outweigh consistency and transparency of the morphological analysis. As a consequence, it is in general simpler to enter the root dictionary through the word dictionary precisely because a root represents a hard-to-describe abstraction that has its origin in previous stages of the language.

Although it can be a problem to find the phonological shape of a root that ties together a great family of disparate words and satisfies the stipulated ambitions, there are cases in which one is required to decide, often on shaky semantic grounds, whether to establish one or two roots. The question of how to deal with true homonymy is probably not very different for our kind of dictionary than for traditional dictionaries. Should we enter /toma- λ / 'tomato' and /tomā-wa-k/ 'thick, fat' under one root (according to one analysis, the stem-final vowel is lengthened before the suffix /wa/), or is this a case of two homonymous roots? Campbell (1985) and others believe the two words share one root. However, the only argument I can find in support of this interpretation is the semantic imagination of those scholars, and that seems insufficient. Consequently, in CoNDiP they are registered under two different roots.

An even trickier problem involves near-homonyms resulting from lack of information about vowel quantity and saltillo. For this, Carochi provides help. He has a list of near-homonyms, differing only in vowel length

or saltillo. In this list he mentions "Tēma nic, 1. larga, echar maiz, trigo, reales, &c. téma nic 1. breue, bañar en temazcal, assar en hornillo, ò bar-bacoa" (Carochi 1645: 128).² But, in fact, the situation is not completely clarified by this information; a third root is involved, /tēmi/ 'swell, be filled', that is homonymous with an intransitive verb, /tēmi/ 'lie together (cats, dogs, corn, gourds)', which is based on the same root as Carochi's first verb. We thus have the following sets:

		BASIC ROOT MEANING	
1. intr.	/tēmi/	'lie together (cats, corn, gourds)'	(involves discrete entities)
tr.	/tēma/	'pour (corn . . .)'	
2. tr	/tema/	'heat in oven, bathe in temazcal'	(involves heat)
3. intr.	/tēmi/	'swell, be filled'	(involves liquid or fill)
tr	/tēmi-tia/	'fill'	

These sets display partial phonological overlap and reveal causes for semantic problems. Because the basic meanings of the roots have not been recognized, 'pour' and 'be filled', as well as 'bathe' and 'fill with liquid', have been paired and have thereby created confusion.

5.5. DERIVATION AND INFLECTION. The distinction between derivation and inflection plays the same role in CoNDiP as in traditional dictionaries: a derived word has its own dictionary entry, whereas inflected words are subsumed under one unifying entry in some inflected form based on a conventional choice. The only difference is that the dictionary word in CoNDiP is without inflectional affixes; this decision is based on the widespread use of inflectional prefixes in Nahuatl. One exception to this principle is the singular, nominal, absolutive suffix, which is the only inflectional affix that will appear in dictionary words. Thus the transitive verb 'eat' is given as /k^wa/, a nontext form, whereas the word for 'house' appears with the absolutive suffix, /kal-li/, a perfectly common word form.

Now, the distinction between inflection and derivation is in general not clear-cut (see Anderson 1985), but our notation requires that we categorize affixes as inflectional or derivational precisely because the distinction is decisive for whether a word gets its own entry or is identified as an inflected word form to be registered under a dictionary word. For example, the diminutive suffix, also used as an honorific suffix, in Nahuatl is

2. téma nic, first [vowel] long, throw corn, wheat, coins, etc. téma.nic. first [vowel] short, bathe in sweat bath, roast in small oven, or [in] earth oven.

categorized as inflectional; as an appropriate consequence, all sentences containing *Moteucçoma* and *Moteucçomatzin* appear in one list in the entry for /mo-tēk^w-sōma/ as /mo-tēk^w-sōma+çin/ represents an inflected form of /mo-tēk^w-sōma/. Likewise, the directional prefixes /wāl/ 'toward' and /on/ 'away from' are considered inflectional.

It gets more complicated when an affix functions sometimes as an inflectional affix and sometimes as a derivational one; and certain affixes in Nahuatl seem to work that way. For example, possessed nouns are marked for the person and number of the possessor: /no-kaš/ (1SG.P-bowl) 'my bowl'; and person and number of both subject and object are marked on the verb with prefixes: /ni-k-k^wa/ (1SG.S-3SG.O-eat) 'I eat it'. When the object is nonreferential, one of two prefixes, /la/ 'nonreferential, nonhuman object' or /tē/ 'nonreferential, human object', appears: /ni-λa-k^wa/ (1SG.S-NONREF.NONH.O-eat) 'I eat', /Ø-tē-miktia/ (3SG.S-NONREF.HUM.O-kill) 'he kills'. This persistent marking of a verb as transitive is retained also when verbal nouns are derived from transitive verbs. Thus 'the act of crying', from the intransitive verb /çōka/, is /çōki-lis-li/, whereas 'the act of eating', from the transitive verb /k^wa/, is /λa-k^wā-lis-li/ and no other object prefix can here be exchanged for /la/. The 'nonreferential, nonhuman object' object prefix participates in the derivation process and is therefore marked as a derivational affix: /=λa-k^wā=lis+li/. An alternative analysis would categorize the prefixes /tē/ and /la/ as derivational in all occurrences and claim that they detransitivize transitive verbs.

In other cases certain inflected forms are lexicalized; a relational noun /wān/ 'with' is constructed with possessive person prefixes, /no-wān/ 'with me' or /i-wān mo-siwā-w/ (3SG.P-with 2SG.P-woman-P) 'with your wife', where /i/ refers to /mo-siwā-w/. Occasionally this word is used with the third-person singular prefix with the meaning 'and', the reference then being at best vague, and I consider this use a lexicalization. In such cases the possessive prefix is therefore marked as a derivational affix, /=i-wān/. The same word form is thus sometimes an inflected form (in this case found in the entry /wān/) and in others a dictionary word (in this case found in its own entry /i-wān/).

5.6. ORGANIZATION OF WORD LISTS AND TEXTS. In the word list of a given root entry we typically find three kinds of words: simple words, derived words, and compounds. Because of the delight in derivation and compounding characteristic of Nahuatl speakers, most roots are found in more than twenty-five words and many in well over one hundred. These words are listed in the root entry, but I have not found any natural and consistent way of organizing these lists because of a number of obvious criteria that intercross and compete. Until now the ordering has been

based on some intuitive arrangement: first the simple words, then other words derived from them, then compounds and derivations formed from simple compounds, and at last proper names and toponyms. And within the wealth of compounds, the semantically more concrete have in general been placed before the semantically more abstract, but no absolute consistency is possible.

Text sentences in word entries give rise to similar problems. Should we order them according to texts, first all the sentences from one text, then from another, and so on? Or should we give the user the option of seeing only the first ten sentences from each text? Another possibility is to let users choose any of these options individually.

6. THE HISTORY AND PROPOSED FUTURE OF CoNDiP. The problems outlined here involve practical considerations not only for a computer-based dictionary but also for the linguistic analysis of the Nahuatl language.

6.1. PREHISTORY. A small-scale dictionary project preceded CoNDiP. When I began to teach Classical Nahuatl in the early 1970s, I soon became aware of the quality of Horacio Carochi's *Arte* (1645). His truly linguistic description, his unique notation of vowel quantity and *saltillo*, and his great fund of examples, word forms, and long sentences (see Canger 1997) are astounding. In 1975-76, with the help of three students, I compiled the Nahuatl words from his examples in the old-fashioned way: with filing cards and pencil. The material was then organized alphabetically according to roots so that a root was given in its phonemic shape, and under such a root all occurrences in Carochi's grammar of words containing the root were listed in their original form and with page references. It was typed up and mimeographed (Adrian, Canger, et al. 1976). This Carochi vocabulary gives the reader easy access to all the occurrences of a given word in their original form and to the words sharing the same root.

When computers entered the scene in the 1980s, the need to construct some kind of computer-based text dictionary for an expedient zipping around between texts and various types of well-organized information became more and more pressing. The decisive factor was, however, the appearance of Michael Thomsen, a rare student from computer science who combined an interest in the language and culture of the Aztecs with a profound appreciation of the details of philology. He became familiar with our Carochi vocabulary, and we began to discuss my dreams of a grand-scale dictionary project. Over the years CoNDiP has grown out of our discussions. Every decision has resulted from thorough debates between the two of us. Only the actual programming has been Thomsen's responsibility alone but always in agreement with the desired layout; and the linguistic aspects of the project have been my responsibility.

6.2. THE PILOT PROJECT. In summer and early fall 1988 our discussions about how such a dictionary should be organized became more specific. Funded by the Danish Research Council of the Humanities, I went to the annual meeting of the American Anthropological Association in Phoenix, Arizona, in November to present our ideas. The purpose of that presentation was to inform scholars who might have an interest in the project and to involve them in discussions about the principles. We wrote up a proposal and applied to the Carlsberg Foundation for financial support in order to secure Thomsen's participation in the pilot project. Carlsberg deemed the proposal worthy of support, and in December 1988 actual work on the pilot project was begun.

As it turned out, during the first half year, we spent much more time working on the analysis of texts than we had planned, but at all times the debates about analysis and preparation of texts involved the perspective of the whole project. By March 1989 we had prepared the first twenty to thirty pages of text; from then on, we deliberately focused on just a few pages at a time, because we had learned that the principles were constantly being revised. In April 1989 we sent the first report to fifty scholars in Mexico, Guatemala, the United States, Canada, France, Germany, and Denmark.

By the time we received reactions to the first report, we were well into revisions and into the actual programming. The paleography, the transcription, and the English translation of the texts were checked and rechecked for consistency and fortuitous errors. The roots and words were written on cards so as to facilitate the testing of the program and of the interrelationship between the components when they were finally to be entered. The entering of the roots and words from the first short texts became a long process that led to numerous improvements of the program. In late June, while our ideas of how to shape and present the results of the pilot project became clearer and more concrete, we wrote to a limited group of scholars asking them to serve as special critics of the pilot project.

Details in the program were still being revised, and we realized that whatever results the pilot project would lead to, they would not be final; but in the end we put an arbitrary stop to the revisions, tried to tidy up the most conspicuous inconsistencies and blunders, and entered the first short texts. The program worked smoothly and satisfactorily on the small test corpus. In October 1989 we sent out the second and final report on the pilot project.

6.3. SECOND PHASE. The second phase was begun in 1992. It consisted primarily of adding more material, more words and roots, and more text but also of minor adjustments of the program.

In January 1993 a full-fledged version with a great many new details and options was sent out. The basic structure was unchanged, and the

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.