LAGS WORKING PAPERS, FIRST SERIES (1981)

WORKING PAPER NUMBER NINE

A Plan for the LAGS Concordance

Lee Pederson

Susan E. Leas

A Plan for the LAGS Concordance

This is an outline of the background and a description of the method for the composition of an exhaustive finders' list of all phonetically transcribed data in the corpus of the Linguistic Atlas of the Gulf States.[1] Because the atlas has, since its inception, been recognized as a research tool for various kinds of investigation, the presentation of basic materials is a crucial operation.[2] Unlike three atlases of English dialects completed in the past 40 years, as well as three others that remain in progress,[3] LAGS includes all of the phonetically transcribed data in its inventory of basic materials. Furthermore, this project aims to present all of that information in a coherent and uncomplicated format that will accommodate the widest possible variety of readers.[4] The plan offered here identifies the goals of the concordance program, the complexity of the corpus, and the method of composing this reference instrument that will stand at the very center of the atlas and its data.

During the past four years (1976-1980), several attempts to organize the LAGS material manually have led to the present plan. This includes five considerations: 1) a description of the LAGS corpus, 2) a summary of preliminary efforts to organize the data, 3) a rationale for computer assistance, 4) an outline of the aims and methods of composing the proposed concordance, and 5) the expected results.

1. THE LAGS CORPUS. The LAGS corpus includes more than 5,000 hours of tape-recorded speech (the field records), 1,118 protocols (the phonetic notation of data preserved in the field records), a microfilmed reproduction of the protocols, idiolect synopses, A Manual for Dialect Research in the Southern

States (3rd ed.), and A Compositional Guide to the LAGS Project (2nd ed.), all of which will be published by University Microfilms International.[5] All of these basic materials will be described in the handbook to be published by the University of Georgia Press. All of these materials will be complete--organized, edited, and ready to be indexed--when the proposed plan is to be implemented in 1981.

The average duration of a LAGS field record is five hours and thirty minutes, with the entire collection including more than thirty million spoken words. To control a corpus of this size, several indexes are needed. Of these, the protocol serves as the primary reference instrument. Constructed by American linguistic geographers as a questionnaire, atlas work sheets serve a dual function in the LAGS project. In addition to directing the course of the field investigation, the work sheets--when converted to protocol form--become a guide to the contents of the tape-recorded interview, giving the reader of the protocol both the phonetic notation of the target items and direction for audial verification. Finally, the protocol also identifies positions in the tape of interesting passages of conversation for the study of contextual phonology, semantics, and syntax, as well as paralinguistics, oral composition, and folklore.[6]

Even this skeletal outline of the field records (x 1,118) comprises a collection of more than one million entries, exclusive of cross references, which, as described below, will approximately double the size of the entry list. Furthermore, however useful the phonetic notation may be for the study of phonology and grammar, that writing system does not lend itself to easy interpretation, especially by those unfamiliar with the idiosyncrasies of any scribal hand.[7] To put this corpus into a readily accessible form requires the transliteration of all narrow phonetic writing in the protocols into a

conventional orthography. In this form, the data can be indexed and described for all readers of the English language, with the phonetic notation remaining intact in the protocols for atlas composition and general use of that research tool.

2. PRELIMINARY EFFORTS TO ORGANIZE THE LAGS CORPUS. In the summer of 1976, the contents of the LAGS protocols were identified with the possibility of organizing the descriptive text of a linguistic atlas in the format of a historical dictionary.[8] In September, 1976, a prospectus was circulated to 19 LAGS consultants, outlining the principal characteristics of the atlas in its presently proposed form. With the recommendations of consultants included, a methodology was developed through several experiments, all of which are summarized in the LAGS GUIDE (1977/78), which was prepared originally for the LAGS staff and its consultants. These efforts included the following work and recognition:

A. The first editorial experiment (Inventory One) included the transliteration of 16 representative LAGS protocols (one from each of the regional sectors).[9] The phonetics were converted into conventional orthography described in Webster's Third New International Dictionary of the English Language (1961), hereafter W3. These orthographic strings[10] were recorded on 4 x 6 slips during the winter of 1976-1977, and these were sorted and alphabetized during the spring and summer of 1977. These included more than 20,000 slips that were organized as a word list and evaluated during the summer of 1977.[11]

B. Immediately thereafter, the LAGS GUIDE was prepared with a plan to enlarge the alphabetized list to include 96 protocols (Inventory Two). This work extended through the summer of 1978.[12] With that work completed, several facts became clear:

1) to cross reference and hand sort the entire collection would require 20,000 hours of clerical work for the composition of slips and the organization of the list;[13]

2) to alphabetize and cross reference a corpus of more than one million basic entries and their permutations (more than two million orthographic strings) would introduce an extremely large element of error and would further burden the work schedule;

3) a simple concordance program could be prepared and implemented at a small fraction of the costs required to complete the manually-produced list;[14]

4) the elaborate scheme for the systematic reduction of phonetic notation to conventional orthography created as many problems as it solved, often grossly violating the conventions of English orthography.[15] It has, therefore, been decided to follow orthographic practices of W3 and to provide special spellings only when absolutely necessary, all of which is summarized below under rules for concordance composition;[16]

5) the schedule outlined below reflects the experience of composing the two inventories, the organization of a rationale, the discussion with computer consultants Ronald Wood of Emory University and John Nitti of the University of Wisconsin, and lexicography consultant Audrey Duckert of the University of Massachusetts,[17] and a recent transliteration of a LAGS protocol.

3. THE RATIONALE FOR THE COMPUTER-ASSISTED CONCORDANCE PROGRAM. In addition to those matters of economy and practicality summarized above, the contents of the protocol corpus prohibit the use of simpler instruments, e.g., "smart typewriters" and word processers. No one has publicly demonstrated the efficacy of either of those machines in the manipulation of more than three million units of linguistic data. A description of the protocol corpus, a summary of priorities, and an outline of procedures are included here to demonstrate

the viability of the proposed method.

A. The Protocol Corpus. The 1,118 protocols are considerably varied in their relative completeness and complexity. Approximately 300 protocols report the findings of incomplete field records (the tape-recorded interviews), and these vary considerably in their relative thoroughness of coverage. Loosely defined, an incomplete protocol includes entries for less than 80% of the items in the work sheets, with most of these including at least 50% of the items.[18] Most of these incomplete protocols report records that have been designated as secondary units in the LAGS collection, i.e., although the evidence will be recorded in the concordance and in the legendry, data from these protocols will not be reported in the maps and computations, which will be restricted to the speech of primary informants.[19]

A fair estimate of the number of phonological strings to be converted into conventional orthography is made in the following way:

1) the average complete basic protocol (i.e., with at least 80% of the items recorded and without the urban supplement) includes 1,200 entries;

2) the average complete protocol with the urban supplement includes 1,500 entries;

3) the average incomplete protocol (either basic or supplemented forms) includes 700 entries, yielding:

700 x 1,200 = 860,000 entries from basic protocols

100 x 1,500 = 150,000 entries from supplemented protocols

300 x 900 = 270,000 entries from incomplete protocols

a total of 1,280,000 entries from all protocols.

As will be demonstrated below, the number of cross-referenced items in each protocol will probably approximate the number of basic entries. Thus, the

total number of all entries to be recorded in the concordance will probably exceed 2,500,000, and may well exceed 3,000,000 converted strings.[20]

B. Problems of Conversion in the Context of Priorities. To investigate the range and complexity of the problems in converting phonetics to orthographic strings, an exemplary LAGS protocol was put to the test.[21] This required several decisions in establishing priorities that match those for the entire project, first articulated in 1968.[22] All of these reflect the aim to provide a useful finders' list, not an analytical or descriptive record of the protocol corpus.

1) Completeness. Because of its broad generalization of phonological information, conventional (W3) orthography provides the most effective writing system to identify the data for all readers. With few exceptions,[23] notably the extensive assimilation and simplification of postvocalic consonant clusters, as in wasp /wɑs/, desk /dɛs/, and nest /nɛs/, W3 is remarkably thorough in its treatment of phonological variants, with or without special spellings.

2) Consistency. Although in several instances a good case might be made for special spellings, e.g., the sequence /swɪvəld/ for shriveled, with the distinct possibility there of popular etymology, the impulse must be resisted, for the sake of consistency.[24] It is important for both the editors and the users of the concordance to understand that all such possibilities can be explored in the analysis of the LAGS data indexed therein. To introduce semantically sensitive orthographic strings would violate the principle of consistency.

3) Simplicity. Although this principle is clearly subordinate to the other two, it is an important consideration for both editors and readers. Every reader of the concordance will no doubt offer suggestions for different kinds of presentation, reflecting their special interests. The approach se-

lected here seems most effective for all readers because it identifies each datum in the protocol corpus in a consistent way.

These three criteria determine the method illustrated below. Although the protocol is certainly richer in data than the average LAGS record, its contents were useful because they offered a closer approximation of the full range of problems, exclusive of foreign-language and loan words. The total number of primary entries in the protocol is 1,821 with 1,753 cross-reference entries. All of the rules outlined below were developed, revised, and rewritten as the conversion was done. All of these reflect the rationale of W3, which for better and worse provides a basis for all decisions of spelling, transliteration, and composition.

4. THE AIMS AND METHODS OF CONCORDANCE COMPOSITION. As a descriptive finders' list, the concordance will not offer data for immediate analysis; it will identify the information in the protocols, explaining the contents of all phonetic entries on all pages of the protocols. When completed, the concordance will be published in microfiche with a summary list that will be published in hard copy by the University of Georgia Press as the LAGS Index, the second volume of the atlas. The Index will then provide the word list for the legendry, the descriptive volumes that will comprise the analyzed data of the survey.

A. Procedures. The procedures outlined here include preliminary, interim, and final phases of concordance composition. Although much of the preliminary editing has been completed in the preparation of the protocols for photography and many of the interim and final operations will be defined as the project develops, a general plan can be summarized here.

1) Preliminary Work. The main task here is the preparation of the protocols, including a proofreading of all of them and an auditing of as many

questionable entries as time permits.[25] With this completed, the remaining schedule will include:

a) the roster equitably divides the protocols among the three editors, Bassett, Leas, and Pederson;

b) with the weekly schedule available, editors will read through several of their assigned protocols to isolate problems, all of which will be discussed prior to entering the information into the computer;

c) a list of all place-names, foreign words, and other complicated phonetic strings will be prepared in an effort to find authoritative spellings. When these cannot be established, the strings will be filed as withheld forms.[26]

d) staff meetings will be held at least once a week to discuss these problems and to seek ways to solve them.

2) Interim Work: In addition to the proposed input of one protocol daily by Bassett, Leas, and Pederson, these procedures will be followed:

a) daily proofreading of the printout copy. This may be done best with one person reading the printout aloud and another reading the protocol to be sure nothing has been misrepresented or omitted.

b) periodic corrections within the computerized corpus will be made according to the recommended schedule, i.e., corrections will be entered as frequently as possible to minimize the amount of correction to be made at each time;

c) as the work proceeds, we will accelerate the preliminary reading schedule to isolate all conversion problems well in advance of the actual moment of input;

d) the final work of the interim phase will concern the verification of withheld forms.

3) Final Work. The aim of this phase will be to complete the input operation in 52 weeks or less, with these additional operations:

a) a close study of the entire alphabetized list to determine the form of the index to be extracted from it, specifically:

(1) how many entries will the index include?

(2) what will be excluded?

(3) how much statistical information will be admitted?[27]

b) with those questions answered, the concordance and the index will be produced by the computer. The results expected from this work are summarized below, following a description of the methodology that will produce the concordance.

B. Conversion Rules. To find the range of problems in the composition of the concordance, the exercise described above was undertaken. A LAGS protocol was converted into the conventional orthography of W3 following these 9 rules. In the composition of the concordance, these rules are applied to the narrow phonetic notation of the informant's speech, the subject of the protocol. All of the utterances are interpreted as either Uninflected Phonetic Strings (UPS) or Inflected Phonetic Strings (IPS). UPS include all forms free of inflection; IPS include all forms marked by inflection or characterized by deleted inflection or other grammatical forms.

RULE 1: Convert all UPS according to the first (preferred) W3 spellings. With the exception of the specific instances identified below in RULE 4, all variation of phonemes and sequences of phonemes below the level of the syllable in UPS are converted according to the first (preferred) W3 spelling.

RULE 2: Disregard all deleted or excrescent internal syllables in UPS, with the exception of the specific instances identified below in RULE 4.

RULE 3: Record all deleted initial and final syllables of UPS in parentheses, with these two exceptions:

1. Accept deleted derivational suffixes as pronounced;
2. Record partial syllable deletion involving vowel loss in parens.

RULE 4: In the conversion of UPS, use special spellings only in these instances:

1. Animal calls and animal noises without substantiated reference;
2. Nonce forms without substantiated reference;
3. Corrupted forms, both folk etymologies and malapropisms, that involve the composition of unique words through the combination of nonce or well-established morphemes as substitutes for conventional forms;
4. Dialect spellings of the words recurrently used by American linguistic geographers.

RULE 5: Convert all IPS according to the first (preferred) W3 spellings. With the exception of WITHHELD forms and the special inflections identified below in RULE 6, all variation of phonemes and phonemic sequences below the level of the syllable in IPS is converted according to the first (preferred) spelling of W3. Until they are verified, all WITHHELD forms are regarded as IPS.

RULE 6: Mark all deleted or special inflections according to the conventions of LAGS CONCORDANCE abbreviations. Special inflections include all nonstandard markers at the level of bound morphemes.

RULE 7: Mark all deleted morphological and syntactic grammatical units according to the conventions of LAGS CONCORDANCE abbreviations. These deleted units include all nonstandard markers at the level of morphological free forms.

RULE 8: Observe LAGS conventions of textual abbreviation, remarks, and marginalia.

RULE 9: Observe LAGS conventions of Phantom Space.[28]

C. Phantom Space Rules.

RULE A: Use Phantom Space (PS) to join all forms to be bound in the concordance count and index. In addition to the specific instances listed below in RULES B, C, and D, use PS to bind all multiword proper nouns, all greetings, all animal calls and animal noises, and all brief epigrammatical sayings under 50 characters;

RULE B: Use PS to join all compounds and interdependent words recorded in the W3 word list. In addition to these, use PS in compound nouns, complex verbs (e.g., verbs with essential postpositionals), multiword adjectives, adverbs, prepositions, verb auxiliaries, function words, and idioms.

RULE C: PS all compounds and phrases that provide explicit synonyms for worksheet items that seek such information. For all marginally related structures of modification, decisions will be made at staff meetings.

RULE D: PS all structures that comprise noun + noun, whether one of the members functions as adjective, adjunct, or possessive qualifier.

In addition to these rules, several other conventions are observed in these conversions. These include corrupted forms, whether the products of misunderstanding, folk etymology, or word play. When, e.g., a folk informant offers [wɔʃəmətɪɘɨə] for washeteria, the form is converted washemateria[=washing machine + cafeteria?], because washomatic is also possible. Familiar folk etymologies, such as bodock for bois d'arc, juice harp for Jew's harp, and wheelbarrel for wheelbarrow, are recorded as pronounced without special notation. Others, such as wholecake for hoecake, when the informant offers the etymological explanation, are also given with bracketed information, wholecake [=hoecake], following the style of the usual form. In the case of word play, e.g., [kæstrəsaɨz], the form is converted castracize[=castrate + ostracize?] because the etymology is not explained.

D. Glosses and Textual Codes.

Glosses include the three columns to the right of the CRT page. These are ordered according to their importance in the first column and according to grammatical category in the second column.

SCRIBAL AND CONVERSION GLOSSES

F French

G German

S Spanish

a probable scribal error of text or marginal gloss

b variant (or inappropriate) line use

c disambiguation of entry

d false start and hesitation

e vocal qualifier, including contrastive/emphatic stress

f usage and illustration

g citation

GRAMMAR GLOSSES

h singular number, pronoun and noun

i plural number, pronoun and noun

j nominative case, pronoun and noun

k possessive case, pronoun and noun

l objective case, pronoun and noun

m pronoun

n noun

o infinitive tense, verb

p 3rd singular, present (tense) indicative (mood), verb

q present tense, verb (excluding 3rd singular)

r preterit tense, verb

s present participial tense, verb

t past participial tense, verb

u imperative mood, verb

v verb

w comparison

x adjective

y adverb

z function word

! sic in either SCRIBAL or GRAMMAR column

? query in either SCRIBAL or GRAMMAR column

RUN-ON TEXT (more than 50 characters and spaces)

y

TEXTUAL CODES

[ ] frame deleted grammatical forms and signals, syntactic and inflectional units

( ) frame deleted initial and final syllables, derivational units

< > frame remarks, i.e., contexts and glosses in conventional orthography provided by the scribe

These are the symbols appearing within brackets:

Form-Class Words

N- for noun

V- for verb

M- for pronoun

C- for copula (linking verb)

A- for adjective

B- for adverb

Function Words

P- for preposition

X- for verb auxiliary

D- for determiner (article)

J- for conjunction

R- for relative pronoun

T- for dummy subject, e.g., there

QUALIFYING SYMBOLS

-Ø for any deleted form-class or function word

These are the symbols for deleted inflections of form-class words:

-i for deleted pronoun or noun plural marker

-k for deleted pronoun or noun possessive marker

-p for deleted verb or copula 3rd singular, present indicative marker

-q for deleted verb or copula present tense marker (excluding 3rd singular)

-r for deleted verb or copula preterit marker

-s for deleted verb or copula present participial marker

-t for deleted verb or copula past participial marker

-w for deleted adjective or adverb comparison marker

E. Sample Conversion. The following eight pages from protocol CG 283.01, #444, (1A, 12, 16, 38, 52, 53, 70, and 78) are converted here to the form proposed for computer input.[29] In the following sample, each string to be entered at the CRT terminal is preceded by the protocol page and line number; each cross-reference entry is indented. Textual codes, glosses, and remarks follow the entries. Conversion rules are identified here only for readers of this paper. These rules will not appear in the concordance, nor do examples of every rule appear in the sample below.

	ENTRY	GLOSS(ES)	RULE(S)
1A.1	eighteen-to-forty-five*draft <conscription in WWI> to-forty-five*draft, eighteen- forty-five*draft, eighteen-to- five*draft, eighteen-to-forty- draft, eighteen-to-forty-five*	f	1, D
1A.1	<the> thirties <1930s>		
1A.1	forty-five five, forty-		
1A.1	forty dollars dollars, forty		
1A.2	nineteen hundred hundred, nineteen		
1A.2	seventy		
1A.2	thousand		
1A.2	millionaire		
1A.2	thousand		
1A.2	million		
1A.3	twentieth		
1A.3	eighth grade grade, eighth		
1A.3	ninth grade grade, ninth		
1A.3	(e)leventh of September		

	ENTRY	GLOSS(ES)	RULE(S)
	of September, (e)leventh September, (e)leventh of		
1A.3	first		1
1A.3	third [J-Ø] fourth grade [J-Ø] fourth grade, third fourth grade, third [J-Ø] grade, third [J-Ø] fourth		7
1A.3	fifth		
1A.3	fourth-grade*education grade*education, fourth- education, fourth-grade*		D
1A.3	second		
1A.4	sixth	b	
1A.4	seventh	b	
1A.4	all*at*once at*once, all* once, all*at*		C
1A.5	twice as much as last year as much as last year, twice much as last year, twice as as last year, twice as much last year, twice as much as year, twice as much as last		1
1A.5	twice as big as big, twice big, twice as		1
1A.6	August the twentieth the twentieth, August twentieth, August the		1
1A.6	September		
1A.6	November		
1A.6	January		
1A.6	February		
1A.6	March		
1A.6	April		

	ENTRY	GLOSS(ES)	RULE(S)
1A.7	April		1
1A.7	May		
1A.7	May		
1A.7	October		
1A.8	June or July or July, June July, June or		
1A.8	November		
1A.8	December		
1A.9	August		1
1A.9	September		
12.1	bathroom <off the back porch>		
12.2	I got two living wives got, two living wives, I two living wives, I got living wives, I got two wives, I got two living		
12.2	I got mine too got mine too, I mine too, I got too, I got mine		
12.3	never did hear did hear, never hear, never did	o	
12.3	I've heard it heard it, I've it, I've heard	t	
12.3	hearing	s	1
12.3	hear	o	
12.3	heard	r	
12.4	<have> heard*of him		B, C
12.5	no, I haven't I haven't, no, haven't, no, I		

	ENTRY	GLOSS(ES)	RULE(S)
12.5	I haven't seen him yet haven't seen him yet, I seen him yet, I haven't him yet, I haven't seen yet, I haven't seen him		
12.5	he hasn't been by hasn't been by, he been by, he hasn't by, he hasn't been		1
12.6	[M-Ø] [C-Ø] [D-Ø] ewe, ain't it? <it's a ewe...> [C-Ø] [D-Ø] ewe, ain't it?, [M-Ø] [D-Ø] ewe, ain't it?, [M-Ø] [C-Ø] ewe, ain't it?, [M-Ø] [C-Ø] [D-Ø] ain't it?,[M-Ø] [C-Ø] [D-Ø] ewe, it?, [M-Ø] [C-Ø] [D-Ø] ewe, ain't	f	7
12.7	some of them has hot tea <instead of coffee> of them has hot tea, some them has hot tea, some of has hot tea, some of them hot tea, some of them has tea, some of them has hot	b	
12.8	it does make it nice does make it nice, it make it nice, it does it nice, it does make nice, it does make it		
12.9	I [X-Ø] be dogged [X-Ø] be dogged, I be dogged, I [X-Ø] dogged, I [X-Ø] be	?	7
16.1	field		1
16.1	field		
16.1	pea*patch <larger than a lot; a small field> patch, pea*		C, D
16.1	corn*patch <larger than a lot; a small field> patch, corn*		C, D
16.1	lot <=an acre>		
16.2	fence		
16.2	palings		

	ENTRY	GLOSS(ES)		RULE(S)
	would climb like a dog, them old wild*hogs climb like a dog, them old wild*hogs would like a dog, them old wild*hogs would climb a dog, them old wild*hogs would climb like dog, them old wild*hogs would climb like a			
52.2	up yonder yonder, up			
52.2	up there there, up			
52.4	what?			1
53.1	poor farmers			
53.1	poor			
53.2	orchard			
53.4	[T-Ø] wasn't nobody [R-Ø] slept in it <of a bed> wasn't nobody [R-Ø] slept in it, [T-Ø] nobody [R-Ø] slept in it, [T-Ø] wasn't [R-Ø] slept in it, [T-Ø] wasn't nobody slept in it, [T-Ø] wasn't nobody [R-Ø] in it, [T-Ø] wasn't nobody [R-Ø] slept it, [T-Ø] wasn't nobody [R-Ø] slept in			7
53.4	I said I don't want [P-Ø] go up there said I don't want [P-Ø] go up there, I I don't want [P-Ø] go up there, I said don't want [P-Ø] go up there, I said I want [P-Ø] go up there, I said I don't [P-Ø] go up there, I said I don't want go up there, I said I don't want [P-Ø] up there, I said I don't want [P-Ø] go there, I said I don't want [P-Ø] go up		y	7
53.4	and ask him for no credit ask him for no credit, and him for no credit, and ask for no credit, and ask him no credit, and ask him for credit, and ask him for no	f	y	1
53.5	they got now to where dairies is got now to where dairies is, they now to where dairies is, they got to where dairies is, they got now where dairies is, they got now to dairies is, they got now to where is, they got now to where dairies		y	

	ENTRY	GLOSS(ES)		RULE(S)
53.5	[D-Ø] prettiest dairy I've ever seen prettiest dairy I've ever seen, [D-Ø] dairy I've ever seen, [D-Ø] prettiest I've ever seen, [D-Ø] prettiest dairy ever seen, [D-Ø] prettiest dairy I've seen, [D-Ø] prettiest dairy I've ever		y	7
53.6	I don't think [T-Ø] [C-Ø] any fox*squirrels don't think [T-Ø] [C-Ø] any fox*squirrels, I think [T-Ø] [C-Ø] any fox*squirrels, I don't [T-Ø] [C-Ø] any fox*squirrels, I don't think [C-Ø] any fox*squirrels, I don't think [T-Ø] any fox*squirrels, I don't think [T-Ø] [C-Ø] fox*squirrels, I don't think [T-Ø] [C-Ø] any squirrels, I don't think [T-Ø] [C-Ø] any fox*	f	? y	7, C, D
53.6	in here here, in		y	
53.6	he's just a wild animal that will aggravate you just a wild animal that will aggravate you, he's a wild animal that will aggravate you, he's just wild animal that will aggravate you, he's just a animal that will aggravate you, he's just a wild that will aggravate you, he's just a wild animal will aggravate you, he's just a wild animal that aggravate you, he's just a wild animal that will you, he's just a wild animal that will aggravate	d	y	1
53.6	around your house, getting stuff like that your house, getting stuff like that, around house, getting stuff like that, around your getting stuff like that, around your house, stuff like that, around your house, getting like that, around your house, getting stuff that, around your house, getting stuff like		y	
53.6	like a weasel [J-Ø] (o)possum a weasel [J-Ø] (o)possum, like weasel [J-Ø] (o)possum, like a [J-Ø] (o)possum, like a weasel (o)possum, like a weasel [J-Ø]		y	3, 7
53.8	you'd never get to where you didn't hear one never get to where you didn't hear one, you'd get to where you didn't hear one, you'd never to where you didn't hear one, you'd never get where you didn't hear one, you'd never get to you didn't hear one, you'd never get to where didn't hear one, you'd never get to where you hear one, you'd never get to where you didn't one, you'd never get to where you didn't hear	f	y	

	ENTRY	GLOSS(ES)	RULE(S)
53.8	all day long <=you always hear helicopters> day long, all long, all day	f y	
70.1	pretty close <to midnight> close, pretty		
70.2	almost give it to you give it to you, almost it to you, almost give to you, almost give it you, almost give it to	o	1
70.2	like*to have slipped to have slipped, like* have slipped, like*to slipped, like*to have		B, C
70.3	a few minutes few minutes, a minutes, a few		
70.3	just full of them full of them, just of them, just full them, just full of		
70.4	<I> don't remember that far back remember that far back, <I> don't that far back, <I> don't remember far back, <I> don't remember that back, <I> don't remember that far		
70.4	how far is it to [N-Ø]? far is it to [N-Ø]?, how is it to [N-Ø]?, how far it to [N-Ø]?, how far is to [N-Ø]?, how far is it [N-Ø]?, how far is it to		7
70.5	there it is it is, there is, there it		
70.6	brain*tumor tumor, brain*	f	D
70.6	ruptured gallbladder gallbladder, ruptured	f	
70.6	brain-scanning <device>	f	1

	ENTRY	GLOSS(ES)	RULE(S)
70.6	<brain-> skinning[=scanning] <device>	f	4
70.7	appointment		
70.7	(a)cetylene*lights lights, (a)cetylene*	f	3, D
70.7	(a)cetylene*gas gas, (a)cetylene*	f	3, D
70.7	concrete*floor floor, concrete*	f	D
70.7	scramble some eggs some eggs, scramble eggs, scramble some	f	1
70.8	gather <=collect, harvest>	f	1
70.8	rear*wheels wheels, rear*	f	D
70.8	chilled	f	
70.8	boll*weevil weevil, boll*	f	D
70.9	plantation	f	1
70.9	deer	f	1
70.9	bicycle	f	
78.1	wound		
78.2	proud*flesh flesh, proud*		C
78.3	iodine		
78.3	Merthiolate or Mercurochrome or Mercurochrome, Merthiolate Mercurochrome, Merthiolate or	d	
78.4	quinine		
78.4	capsules of quinine of quinine, capsules quinine, capsules of		
78.5	passed*away there away there, passed* there, passed*away		B, C

	ENTRY	GLOSS(ES)	RULE(S)
78.5	when they first backed this pond up they first backed this pond up, when first backed this pond up, when they backed this pond up, when they first this pond up, when they first backed pond up, when they first backed this up, when they first backed this pond	y	1
78.5	people died (a)round here like flies <before 1926> died (a)round here like flies, people (a)round here like flies, people died here like flies, people died (a)round like flies, people died (a)round here flies, people died (a)round here like	y	1, 3
78.6	kicked*the*bucket the*bucket, kicked* bucket, kicked*the*		1, B, C
78.7	what caused his death caused his death, what his death, what caused death, what caused his		
78.7	<what he died> from		
78.8	graveyard		
78.8	cemetery		

5. THE EXPECTED RESULTS OF THE LAGS CONCORDANCE PROGRAM. After the orthographic strings are entered into the computer and after the computer has automatically produced all of the necessary permutations--the indented forms in the preceding examples--the entire list of strings will be alphabetized. The expected results of this work can be summarized in terms of substance and form.

A. The Substance of the Concordance. With each entry and all of its permutations presented in alphabetical order and identified with the informant who provided the form, as well as the page and line of the LAGS protocol where the string was elicited,[30] the user of the concordance will be able to find this information:

1) a representation in conventional orthography of each phonetically transcribed word, phrase, and clause, all immediately retrievable in an alphabetized organization;

2) the incidence of linguistic features, exclusive of phonetic and phonemic forms;

3) the regional distribution of each recorded form;

4) the social distribution of each recorded form;

5) glosses of meaning, content, usage, style, and context;

6) all relevant recorded remarks in all protocols.

In addition to this, the methodology of the concordance program provides a system of conventional orthographic writing for the conversion of other linguistic texts (tape recordings) for consistent typescripting, e.g., the LAGS field records. Finally, the concordance will provide LAGS editors with the material for the composition of the hard-copy index and the word list for the legendry.

B. The Form of the Concordance. Two formats are proposed for the presentation of the contents produced by the computer, both of which will be camera-ready copies. The microfiche production, the concordance, will be produced in a double-column format of 25,000 pages or a single-column format of 50,000 pages.[31] These estimates are based on a projected corpus of 3,000,000 strings (including permutations) printed either with 120 (60 x 2) or 60 (60 x 1) strings to a page. The computer will also produce an index to be published in hard copy by the University of Georgia Press. If the descriptive glosses are preserved and summarized from the concordance, this form will also be produced in a double-column format in a text of fewer than 400 pages. This estimate includes the possibility of run-on lines needed for summary description. The physical dimensions of the index cannot be properly estimated, however, until the concordance reveals the

number of different words in the LAGS corpus. With this information available, it will be possible to organize the index and establish the word list for the legendry.

NOTES

1. This essay was originally composed for members of the NEH Computer Panel and for consultants to the LAGS project. The plan is subject to all constructive modification suggested by any of these readers. In its present form, the plan reflects suggestions of Guy Bailey and Marvin Bassett, as well as the recommendations of Frederic G. Cassidy, Audrey R. Duckert, and James B. McMillan.

2. The meanings and implications of the term basic materials in linguistic geography is reviewed in the "Introduction to the Basic Materials" of this text (fiche #1). For a descriptive analysis of these basic materials, see Bassett's chapter, "The Basic Materials," in the forthcoming LAGS handbook.

3. H. Kurath, et al., Linguistic Atlas of New England (1939-1943); H. Orton et al., Survey of English Dialects: Basic Materials (1962-1971); H. Allen, Linguistic Atlas of the Upper Midwest (1973-1976). The first two fascicles of the Linguistic Atlas of the Middle and South Atlantic States, edited by R. McDavid and R. O'Cain, have been published, but much work remains to be done. Two other atlas projects directed by McDavid are also underway: the Linguistic Atlas of the North Central States and the Linguistic Atlas of Oklahoma, works originally undertaken by A. Marckwardt and W. R. Van Riper, respectively.

4. This goal reflects the influence of H. Allen, who, to a greater extent than any other linguistic geographer, tried to make his data useful to a large audience, including teachers and their students. In their second edition of Kurath et al., Handbook of the Linguistic Georgraphy of New England (1973), A. Duckert and R. McDavid composed valuable indexes, including a reverse index of maps to work sheets, an inventory of maps and commentary, and a word-index to

the inventory for readers of LANE. The referential plan of LAGS is an elaboration of the innovations made by those three linguistic geographers.

5. An idiolect synopsis (figure 1) includes abstracted phonological, grammatical, and lexical data to suggest the composition of each idiolect in the LAGS collection. Selected synopses will be published in hard copy by either UMI or the University of Georgia Press.

The third edition of the Manual will include the revised LAGS grid (expanded in 1975 to include all of Georgia and Arkansas, as well as East Texas as far west as Fort Worth) and the urban supplement (201 lexical items added for use in the cities of the Gulf states). This edition will be edited by L. Pederson, C. Billiard, S. Leas, and M. Bassett.

6. Such identification is scored to match the tape from the original field record: number of reel (1-n), track (A or B), and position on the track (1-10). With all LAGS interviews recorded at 3 3/4 ips, on 600-1800 ft. reels, each datum on the tape is readily retrieved with this system, as the past 12 years of editorial work with the field records have proven. This method of scoring is used instead of the meter counts available on the Uher and Tandberg machines used in LAGS because too many discrepancies emerged through the use of leaders and the individual differences of the 10 machines used in the project.

7. Although some scribes in the LAGS project have limited calligraphic skills, close study of any hand makes it quite readable without much effort. That commonplace among medievalists applies here as well, but the casual reader requires an easier reference.

8. The manuscript, "Toward the Publication of the Linguistic Atlas of the Gulf States" (Working Paper #4), was sent on 2 September 1976, to these consultants:

John Algeo, Harold Allen, Walter Avis, Charles Billiard, Frederic Cassidy, A. L. Davis, Audrey Duckert, C. W. Foster, Hans Kurath, Raven McDavid, James McMillan, Raymond O'Cain, Hood Roberts, Willis Russell, Gerald Udell, W. R. Van Riper, Juanita Williamson, Rex Wilson, and Gordon Wood. Many of the revisions in the current plan reflect their valuable assistance.

9. The eight-state territory covered in the LAGS project is divided for convenience into four zones: the Eastern Zone (East Tennessee, Upper and Lower Georgia, and East Florida), the East Central Zone (Middle Tennessee, Upper and Lower Alabama, and West Florida and Gulf Alabama), the West Central Zone (West Tennessee, Upper and Lower Mississippi, and Gulf Mississippi and East Louisiana), and the Western Zone (Arkansas, West Louisiana, and Upper and Lower Texas). The 16 sectors include the aforementioned subregions within each zone.

10. In the description of the concordance, the term string is used in two constructions: an orthographic string is a succession of letters, spaces, and abbreviations in the style of W3 representing a phonetic string, a succession of phonetic symbols recorded in a LAGS protocol. See also the application of the terms uninflected and inflected phonetic strings in the rules. Given the restrictions of the proposed concordance format, orthographic strings are limited to 50 positions, including spaces and abbreviations.

11. The number of hours expended in this, as well as the following, experimental effort cannot be accurately calculated. In addition to the work of eight scribes and four clerks, the latter of whom worked full-time throughout the summer of 1978, Leas independently converted more than 400 protocols to slips. The problems summarized below indicate the really prohibitive nature of a manually-composed concordance. A conservative estimate of the time expended in that effort certainly exceeded 2,000 hours of work.

12. The LAGS Guide was revised to include these developments in the fall of 1978.

13. This estimate is quite conservative because it is impossible to project the number of errors that would inevitably occur at the hands of clerks with limited linguistic training. If the work were restricted to trained phoneticians, recruitment and funding problems would certainly be serious. These points are belabored here because a general skepticism toward machine processing of data has prevailed in this project since its earliest days. This is to say, even now, we would much prefer to make the concordance by hand if it were possible.

14. The total computer component of the current LAGS proposal is budgeted at $60,000, excluding salaries of personnel who would be needed to supervise a manually-produced set of slips. Included in this projected work will be the camera-ready text and the abstracted index (for the University of Georgia Press publication), both of which would require an additional $50,000 minimum to type 3,000,000 entries, to proofread the copy, and to compose and type the index.

15. Among LAGS consultants, this point was made most effectively by John Algeo.

16. The DARE project is developing an authoritative list of dialect spellings, and this responsibility, among other problems that relate to the general composition of American English, is yielded by LAGS to F. Cassidy and his associates. For a good sample of the representation of dialect spelling in DARE, see Joan Hall, "DARE: The View from the Letter F," Dictionaries 1 (1979): 25-46.

17. In addition to her general consultation with LAGS, A. Duckert spent the entire winter quarter of 1978 at Emory as a Visiting Professor in the Department of English. During those months, she gained a full understanding of the project and provided valuable assistance throughout her stay.

18. This estimate is based on the conversion of 400 protocols to 4 x 6 forms by Leas, as well as the contents of inventories one and two.

19. LAGS secondary informants, or records, are not to be confused with American atlas and DARE project designations of auxiliary informants. No LAGS record includes the speech of more than one subject. See Pederson, "Anglo-American Linguistic Geography," Zeitschrift für Dialektologie und Linguistik 46 (1979): 231-46.

20. See the previous reference to the limitations of 50 positions in the concordance for each orthographic string (including words, spaces, and abbreviations) and 11 additional positions for the code described below in note 30. This conditioning factor could be instrumental in producing a large number of unanticipated slips.

21. As a fourth-generation native of the community named after his maternal great-grandfather and residing in a small satellite settlement named for his paternal great-grandfather, the informant is a model folk speaker in the context of American linguistic geography. Despite his travels, this retired mechanic is classified insular, rather than worldly, because his experience has been limited to a rather narrow range of social interaction. He is characterized in the Personal Data Sheet in this way:

> Having traveled all over the Southeast, he has observed much of his working-class world as a road mechanic, paver, and driver. He is now nearly blind, but quite alert and good-natured, despite recent surgery. A good example of South Alabama folk speech--labialized /r/ and some lexical forms suggest Gulf coast influence, but he is clearly an inlander of the Coastal Plains. Thoroughly honest and open about himself, he reflects this in his linguistic usage: when in doubt concerning correctness, he offers two forms without abashment or real concern.

This interview was conducted by Bailey (9/8/76) and transcribed by Pederson (10/16/76). The field record is of six hours and forty-five minutes duration.

22. Pederson, "The Linguistic Atlas of the Gulf States: Preliminary Considerations," read at SAMLA-ADS, Jacksonville, FL, 1968, published as "Structural Description of Linguistic Geography," Papers in Language Variation, D. L. Shores and C. P. Hines, eds. (1977), 19-24.

23. Most of these are characteristics of Southern American English dialects and their derivatives preserved in the speech of migrants who left the South for the urban North and West, e.g., /ɪ ~ ɛ/ homophony before nasals, /t ~ d/ for /θ ~ ð/, especially under weak stress, and the assimilation of /r/ and /l/ in consonant clusters, as in April /epəl/, aggravate /ægəveɪt/, bulb /bʌb/, and gulf /gʌf/.

24. See the discussion of corrupted forms after Rule D.

25. A systematic proofreading of all protocols and an auditing of selected field records began in 1979 and was concluded in January, 1981.

26. Forms that cannot be verified prior to the composition of the legendry will be included there as withheld forms, following the precedent established by Joseph Wright in the English Dialect Dictionary.

27. E.g., it will be desirable to identify incidence of all forms with as much additional information about regional and social distribution as possible. A minimal entry will include the form, the protocol page and line, and the number of occurrences: mosquito hawk 60A.4 (250), but it might also indicate the subregions where the form recurs, as well as the age, educational, and social groups who share the usage. Such glosses will be as elaborate as space permits.

28. Phantom Space is the designation to indicate multiple free forms that will be bound in the word list, but parsed individually in the concordance. An asterisk

(*) between the forms indicates the occurrence of phantom space.

29. The eight protocol pages are reproduced in the appendix as figures 2-9. The protocol code at the upper margin of each page identifies the social characteristics of the informant and the community he represents:

MMY: M(ale), M(iddle class), Y (Caucasian), 76 (Age), 1 (Elementary-school education), A (insular type). See note 21 above for a summary of his social characteristics.

LA (Lower Alabama) GANTT (the locality)

CG (the grid unit, comprising Coffee and Covington counties)

283 (Covington County)

.01 (the accession number, i.e., the first LAGS record made in 283).

30. Informants will be identified in the concordance by numbers 1-911, consecutively according to the counties and parishes they represent and their age, education, social class, etc. E.g., the oldest folk speaker from Johnson County, Tennessee, will be #1, and the youngest cultivated speaker from Cameron County, Texas, will be #911. Five characters will be required in each entry line for this identification, and five more characters with an intervening space or mark will be required to identify the protocol page and line number, e.g., 104.1 Thus, each line in the concordance will include this code: 104.1 0000, and the entire code and string will require 61 spaces.

31. See figure 10, Concordance Sample. This alphabetized list includes all orthographic strings from LAGS protocol CG 283.01, pp. 1A, 12, 16, 38, 52, 53, 70, and 78. A single-column format is used therein because a computer printout sheet is not available.

FIGURE 1

LINGUISTIC ATLAS OF THE GULF STATES
IDIOLECT SYNOPSIS

444

MMY 76 1A
GB/76:LP/76

LA GANTT
CG 283.01

/ɪ/	* hwɪ˃ᵊp	* kr̥ɪ˃bz	* t̪'ɪ˄ᵊn	hɪ˄.ᵊl̥	* ɪːə
/ɛ/	nɛ˄k	* leˇ.ɪg	t'ɛ˃n	nɛ́˄.ʃɪ̆	mɛ̂˄r̆ɪ kr̥ɪ́˄smə̆s
/æ/	* glæᵉs	* bæ.ᵉgz	* hǽᵋmmə̆	vǽ.ʃɪ̆˄	* mɛ́˄.r̆ɪ̆d
/ʊ/	* p'ʊ˂.ʃ	* wʊ˂ᵊd	wúː.mn̥̆	* p'ʊ˂l̥	ʃʊ˂ᵊ
/ʌ/	* bʌ́ːkɪ̆t	* hʌ́˂ᵊzbn̥̆	s̪ʌ́˂nʌ˂.p	* bʌ̂ᵊɯb	fʌ́˂.ʳə̆
/ɑ/	* kr̥a˃.p	fá.ɣə̆	p'a˃ᵊm	k̪'á˂ʃɪ̆dʒ	* k'a˃ˠ
/i/	ji˃.st	* ˢθr̥i˃	bi˃.nz	* fi˃ᵊl̥d	bɪʲəd
/e/	eˇɪt	me˃ɪ	* stre˃ɪn	re˃ɪl	mé.r̆ɪ̆
/u/	t'ʊ˂ʉθ	bǽ.tʔn̥̆ rəʉdʒ	wʊ͡ʊ˂ː nᵈ	* mjʉ.lz	* p'ʊ˂ə̣
/o/	k'o˂ʊt	əğó˂ʊ	* ho˂ʊm	k'o˂ʊtd	hɔ˄.əs
/ɔ/	dɔˇ.ɔ˄t̆ə	dɔˇ.o˄g	gɔˇon	sɑᵊtt	* hɔˇ.ʔs
/ɜ/	* tʃɜ̆.ɪtʃ	θə̣d	* wɜ̆.mz	* gɜ̆.t	stɜ̆́.ə˄ps
/aɪ/	ra˃.ɪt	ra:ɪd	na:ᵋn	* maˑ.ɪl̥z	* wa̋.ɪə̆
/au/	* ha˄ʊ˂s	* k'a˄o˂z	* da˄o˂n	* a˃.əl̥z	flá˃.wə̆z
/ɔɪ/	ɔ́ɪˇ.stə̆z	p'ɔ́˄ɪˇzn̥̆	dʒɔ˄ə˄nts	ɔˇ.ᵊl̥	—
PL	p'o˂ʊst	* pa˄onᵈ	ˢʃrɪ̃ːmp	wɒ˄ᵊsts	—
FW	*kwɔ̂.t̆ə̆ dɬ	k'ʌ̂˂mn̥̆ tʊʉ mìː	rǽᵋn ʌ˂p ɔ̀ːn	ɪ̆t ɣə̆ stʌ́˂.mɪ̆k	—

— / r̥o˂ʉ̂ˇz / r̥ɪ́˃.ᵊzn̥̆	dra.ɪv ~ dra.ᵋv / dro˂ʊv / —
dræ.ᵉg / — / —	i˃.t / eːt / i˃.t ~ iːˑtʔn̥̆
drɪ̃ːŋk / dræːᵋŋk / drǽᵋŋk	hɛ˃ᵊp / hɛ˃ᵊpt / hɛ˃ᵊp
da˃.ɪv / do˂ʊv / dɪ́˃ᵊvn̥̆	kla.ɪm / — / kl̥ʌ˂m ~ kla.ɪmd

fáːə dɔˇ.o˄gz	mǽ.nɪ̆l̥ boːˑə˃d	k'a˄ʊ˂ p'ɛ̃ːn	—
*p'eːɪpə̆ bæᵉgz	* kró˂ʊkə̆ sæᵋk	* ha˃.ˠp	* ra̋.ɪd̆ʌ̂ hɔ˄.ɤs
ə̆ flǽᵋmbò˂ dɪ̂˃.ᵊl	* r̥ó˂ʊ bo˂ʊt	p'ʊ́˂ʃɪ̆ bõ˂.ʊ˂n	* p'ǽᵋn kè.ɪks
* sa˃.ʊ˂s	k'á˃.tᶴɪ̆dʒ tʃɪ̀˃.z	* mʌ˂.ᵊʃ	ə̆ prɛ́˃ᵊs (art.+n.)
kl̥ɪːə si˃.dz	* p'ɪːndə̆z	* t'a˃mɪ̆tò˂	strɪ̃ːŋ bi˃.nz
p'ɛ́˃kə̆wʊ˂ᵊdz	*ɪ̂˄ŋlɪ̆ʃ wɜ̆.mz	* go˂.ʊfə̆	* kró˄ɔ˄fɪ˄ʃ
* mə̆skí˃.t̆ə̆ hɒ˄ᵊks	* rɛ˃.d bʌ̀ˠgz	—	—

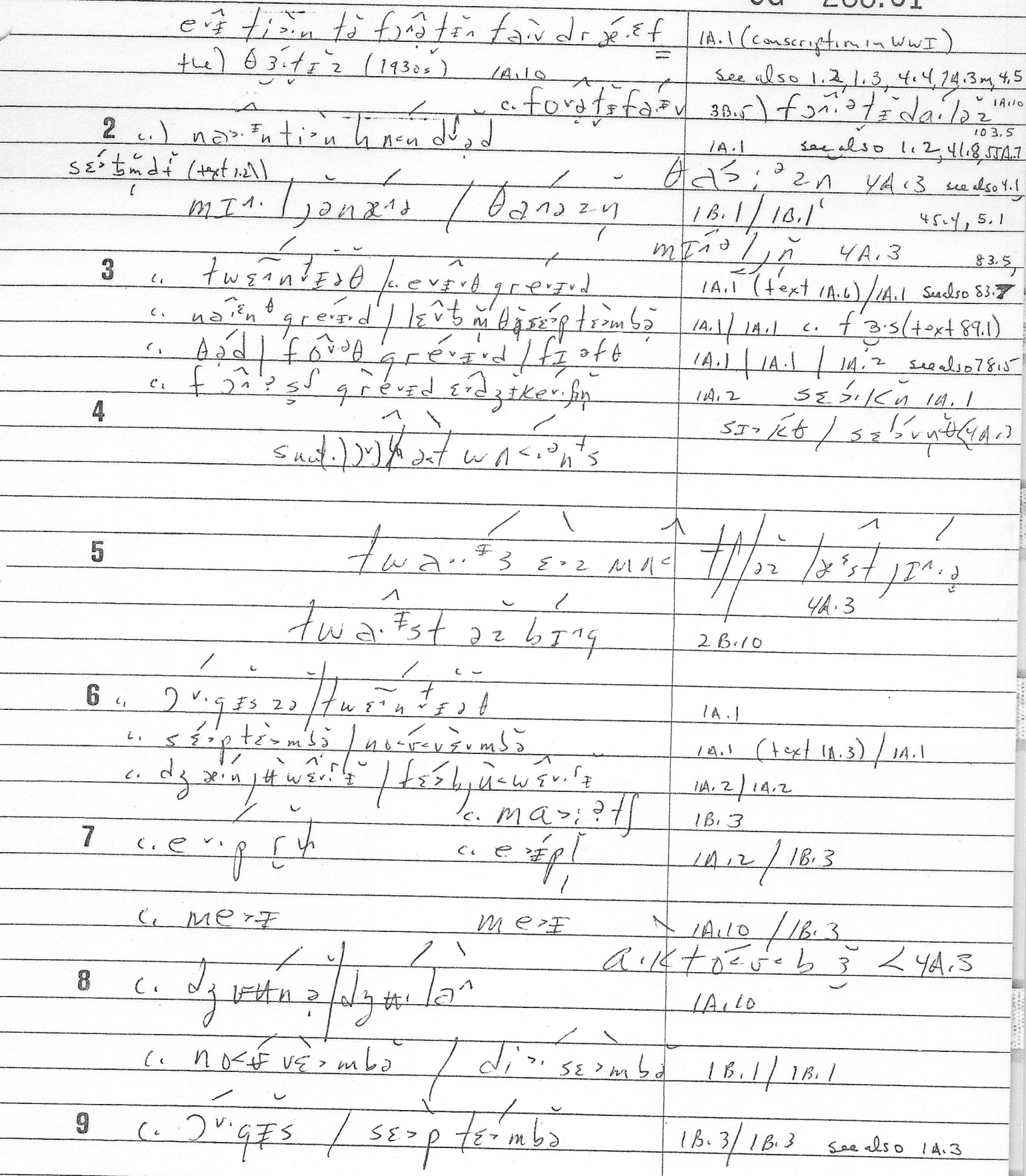
MMY 76 1A

1A

LA GANTT

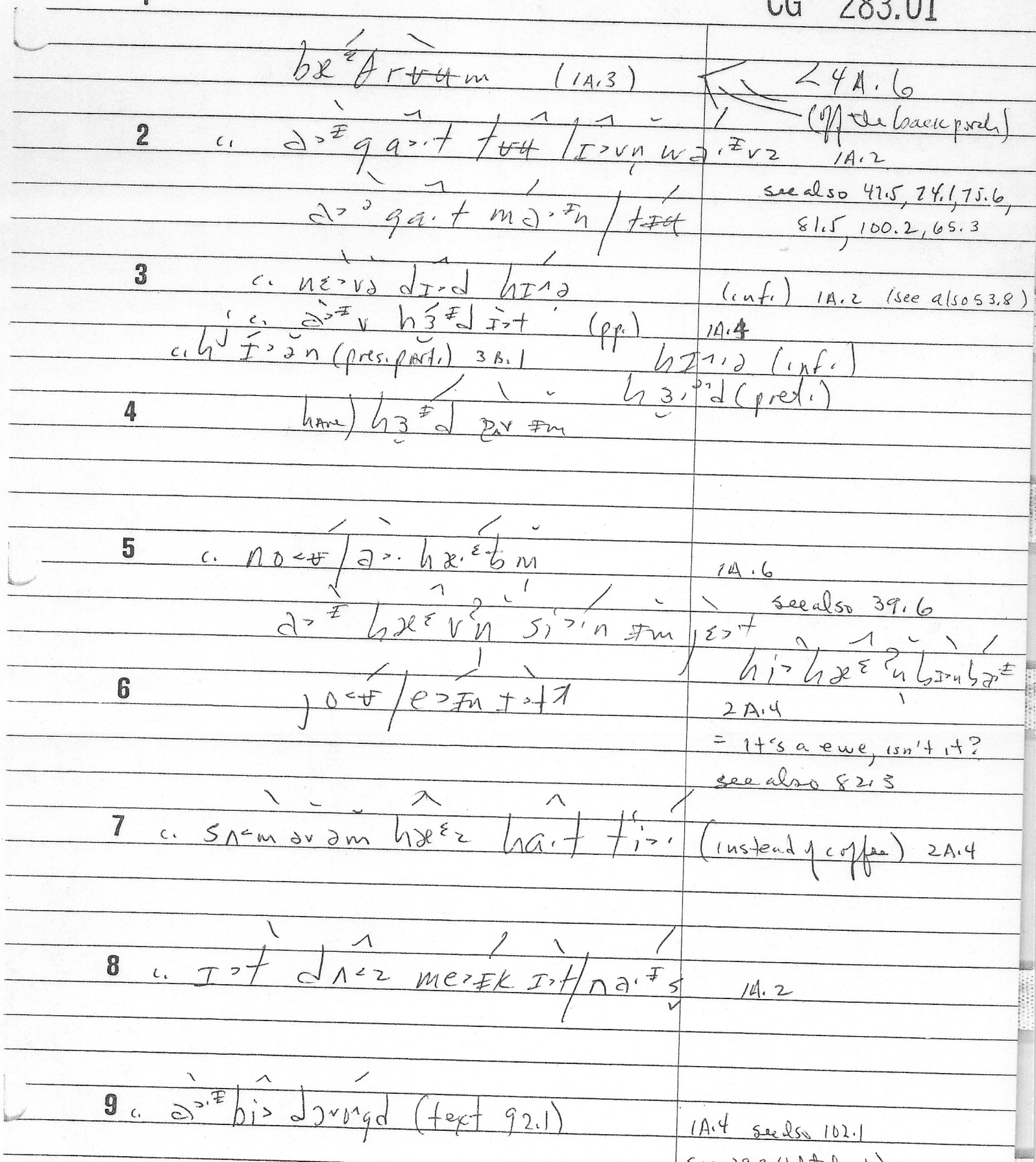
CG 283.01

FIGURE 2



MMY 76 1A 12 LA GANTT

FIGURE 3 CG 283.01



MMY 76 1A 16 LA GANTT

CG 283.01

FIGURE 4

1

2

3

4

5

6

7

8

9

1B.3 > a patch is larger than 1B.2 a lot; a small field 1A.3/1B.3/1D.3 ↑ see also 98.6

s.) p'e>ɪ/n 2 < l

1" spruce pickets, nailed to 2 rails

1A.4 p'O<ʊst (pl.)

1B.1

fences/walls not made of rock here

see salmons txt 55A.9

MMY 76 1A   38   LA GANTT
CG 283.01

FIGURE 5

gɛˀt ʌˀp (or touch him with a switch) | < 2A.10 single click "just click to him" to get him started.

w ɒ.ɦ

ʰwʌɦ.pɔ / ʰwʌɦ.pɔ / ʰwʌɦ.pɔ / ʰwʌɦ.pɔ (to call them) falsetto

ʃ ʊˀ. ɪʌ (to scatter them)

ɦɪˀk / ɦɪˀk / ɦɪˀk / ɦɪˀk falsetto

ʃɪɦ (to scatter them)

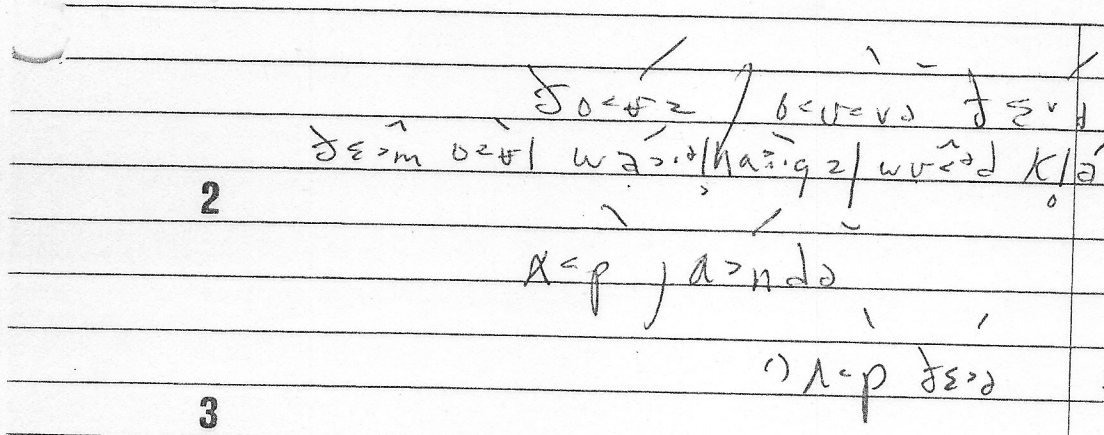
hɪˀt / ʌˀp ɔɔ hɔˀ ɔɪsɪɹ

h ɒʌˀɔnɪs (n.)

MMY 76 1A 52 LA GANTT

CG 283.01

FIGURE 6



2B.6

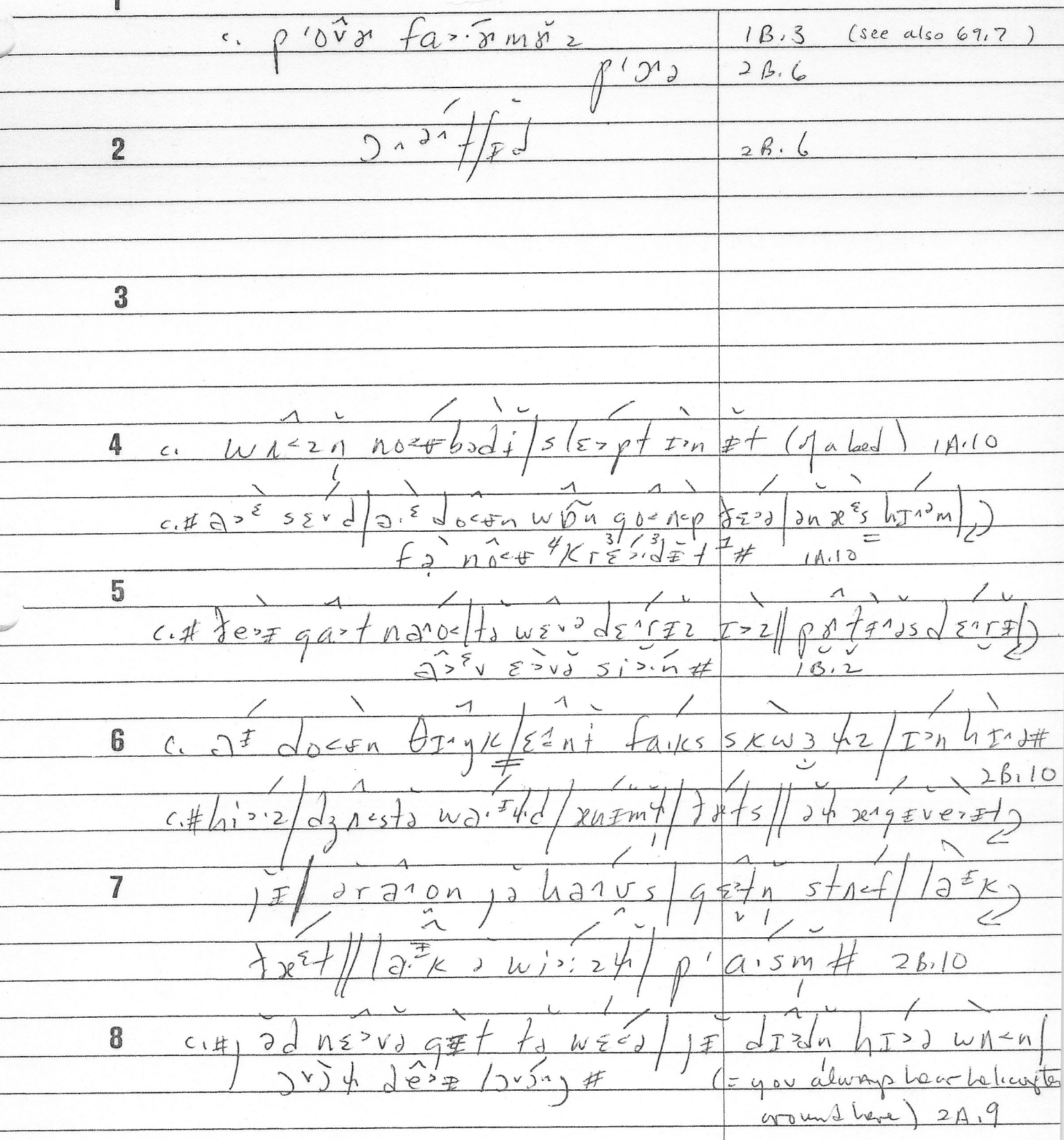
2A.8

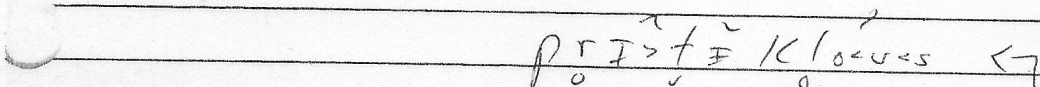
2B.10

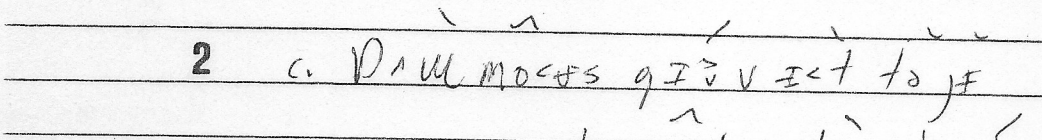
text 53.4

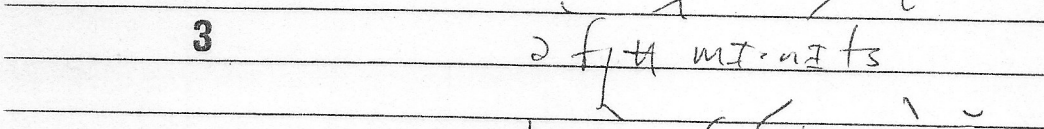
See 85.4

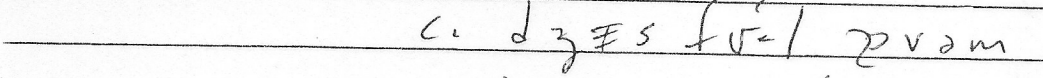
FIGURE 7

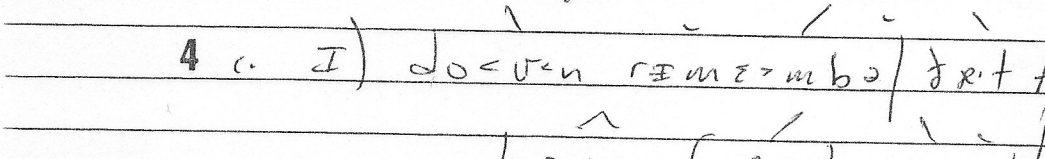

1
1B.3 (see also 69.7)
2B.6
2
2B.6
3
4
1A.10
1A.10
5
1B.2
6
2B.10
7
2B.10
8
(= you always hear helicopters around here) 2A.9
9

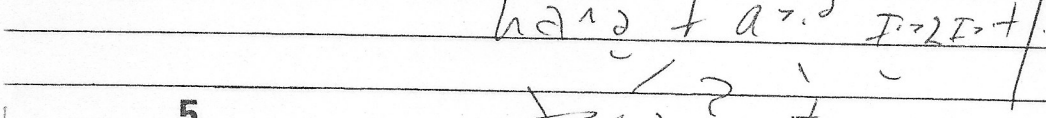
1  <3A.6 (to mid-night)

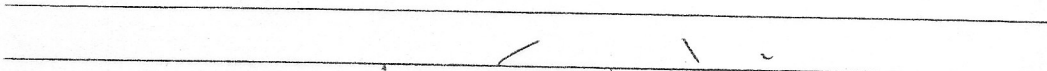
2 c.  1A.8 (inf.)

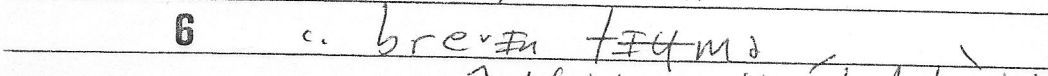
3  see also 46.5m, 53.6, 65.3m, 73.6, 79.4, 82.7, 89.4

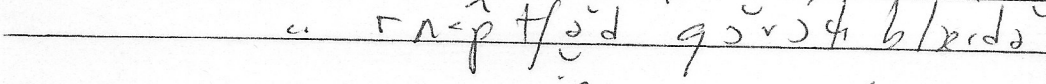
c.  2B.10

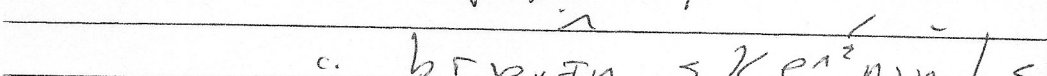
4 c.  1A.1

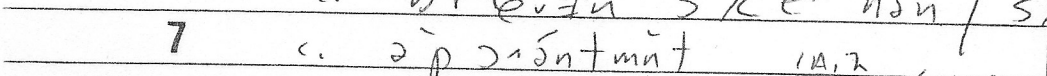
5 

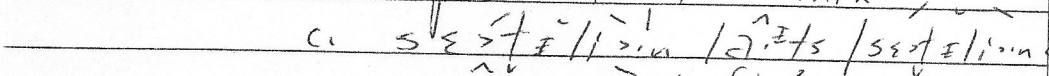
6 c.  (brain tumor) 1A.2

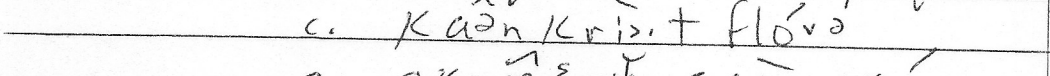
c.  1A.2 (ruptured gall bladder)

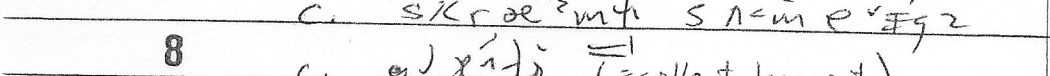
c.  (brain-skinny[device])

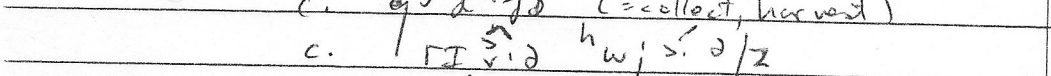
7 c.  1A.2 (brain scanning) 1A.2

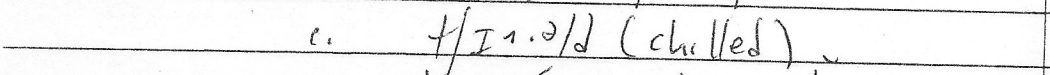
c.  (acetylene) 1A.3

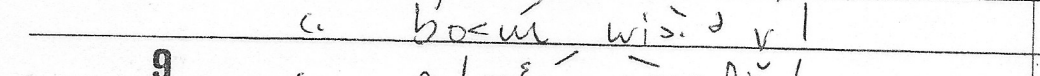
c.  1A.3 (concrete floor)

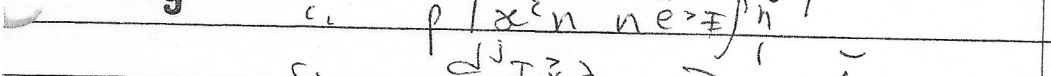
c.  1A.9 'scramble'

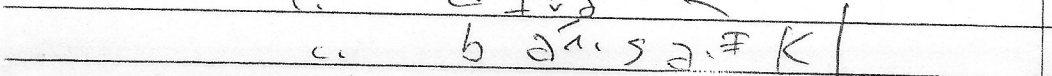
8 c.  1B.1 ('gather')

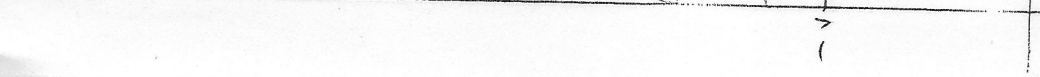
c.  1B.1 ('rear wheels')

c.  1B.2

c.  1B.3 (boll weevil)

9 c.  1B.3 (plantation)

c.  1B.3 (door)

c.  1B.10 (bicycle)

MMY 76 1A 78 LA GANTT

CG 283.01

FIGURE 9

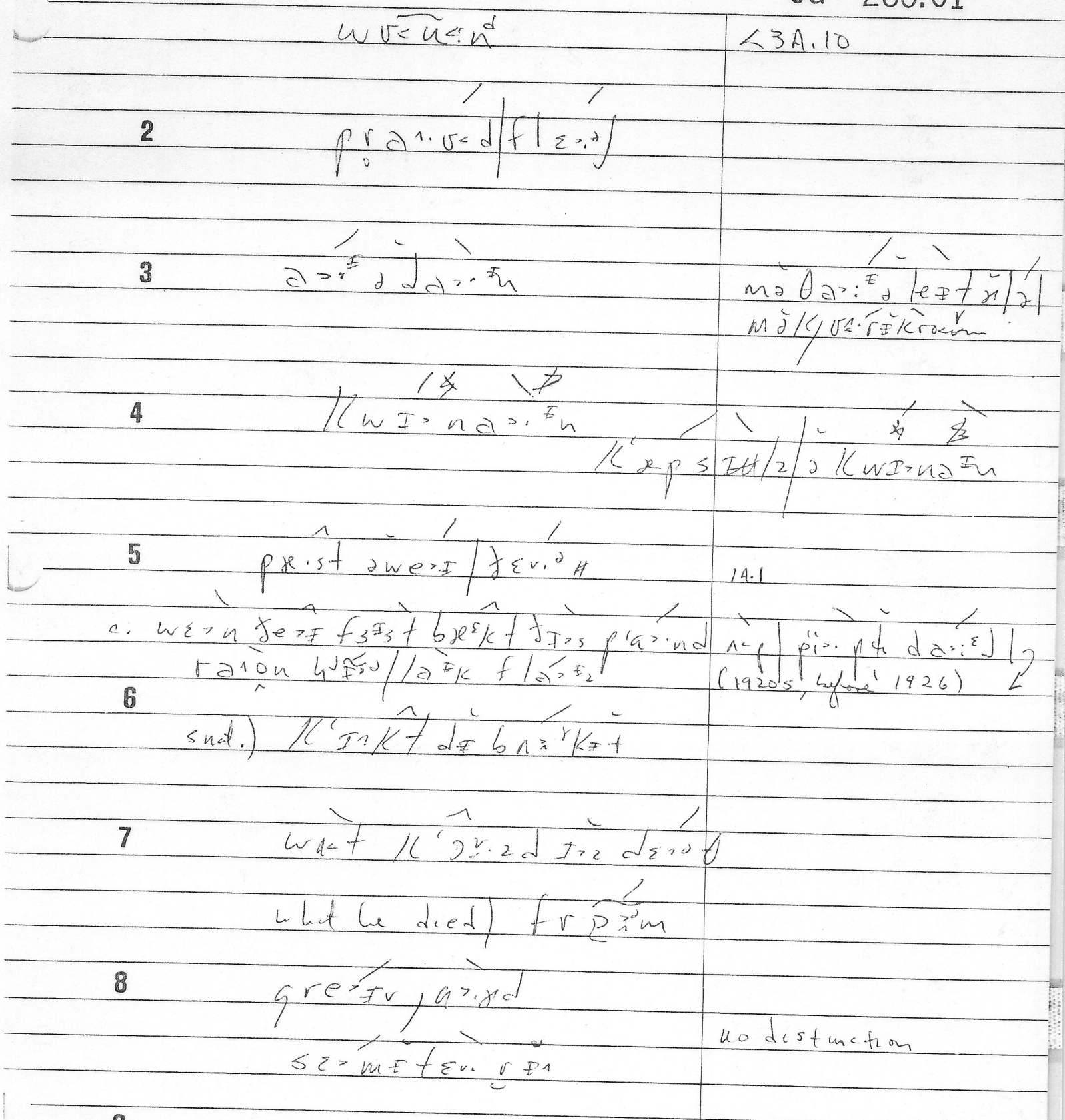


FIGURE 10: CONCORDANCE SAMPLE

The following sample is based on the conversion of the contents of LAGS protocol CG 283.01, informant #444, pp. 1A, 12, 16, 38, 52, 53, 70, and 78. Entries are arranged in word-by-word alphabetical order, followed by scribal and grammatical glosses, remarks (< >) that will not be parsed or included in the alphabetizing, protocol page and line number, and informant number. Although in the full concordance all deleted (Ø) forms will be listed separately, in this sample they are listed within each entry where they occur, and are alphabetized according to form-class or function word symbol. E.g., all deleted determiners are alphabetized here under [D-Ø], at the beginning of the entries for the letter D. Phantom space (*), included in the conversion in the text, is omitted here because it will not appear in the printed concordance.

a dog, them old wild hogs would climb like				52.1	444
a few minutes				70.3	444
a weasel [J-Ø] (o)possum, like			y	53.6	444
a wild animal that will aggravate you, he's just	d		y	53.6	444
(a)cetylene gas	f			70.7	444
(a)cetylene lights	f			70.7	444
aggravate you, he's just a wild animal that will	d		y	53.6	444
ain't it?, [M-Ø] [C-Ø] [D-Ø] ewe <it's a ewe...>	f			12.6	444
all at once				1A.4	444
all day long <=you always hear helicopters...>	f		y	53.8	444
almost give it to you		o		70.2	444
and ask him for no credit	f		y	53.4	444
animal that will aggravate you, he's just a wild	d		y	53.6	444
any fox squirrels, I don't think [T-Ø] [C-Ø]	f	?	y	53.6	444
appointment				70.7	444
April				1A.7	444
April				1A.7	444
(a)round here like flies, people died <before 1926>			y	78.5	444
around your house, getting stuff like that			y	53.6	444
as big, twice				1A.5	444
as last year, twice as much				1A.5	444
as much as last year, twice				1A.5	444
ask him for no credit, and	f		y	53.4	444
at once, all				1A.4	444
August				1A.9	444
August the twentieth				1A.6	444
away there, passed				78.5	444
back, <I> don't remember that far				70.4	444
backed this pond up, when they first			y	78.5	444
barb wire	f			16.3	444
bathroom <off the back porch>				12.1	444
be dogged, I [X-Ø]	?			12.9	444
been by, he hasn't				12.5	444
bicycle	f			70.9	444
big, twice as				1A.5	444
boll weevil	f			70.8	444
brain-scanning <device>	f			70.6	444
brain tumor	f			70.6	444
bucket, kicked the				78.6	444
by, he hasn't been				12.5	444

flesh, proud				78.2	444
flies, people died (a)round here like <before 1926>			y	78.5	444
floor, concrete	f			70.7	444
for no credit, and ask him	f		y	53.4	444
forty dollars				1A.1	444
forty-five				1A.1	444
forty-five draft, eighteen-to- <conscription in WWI>	f			1A.1	444
fourth-grade education				1A.3	444
fourth grade, third [J-Ø]				1A.3	444
fox squirrels, I don't think [T-Ø] [C-Ø] any	f	?	y	53.6	444
<what he died> from				78.7	444
full of them, just				70.3	444
gallbladder, ruptured	f			70.6	444
gas, (a)cetylene	f			70.7	444
gather <=collect, harvest>				70.8	444
get to where you didn't hear one, you'd never			y	53.8	444
get up <or touch him with a switch>				38.1	444
getting stuff like that, around your house			y	53.6	444
give it to you, almost		o		70.2	444
go up there, I said I don't want [P-Ø]			y	53.4	444
got mine too, I				12.2	444
got now to where dairies is, they			y	53.5	444
got two living wives, I				12.2	444
grade education, fourth-				1A.3	444
grade, eighth				1A.3	444
grade, ninth				1A.3	444
grade, third [J-Ø] fourth				1A.3	444
graveyard				78.8	444
harness		n		38.6	444
has hot tea, some of them <instead of coffee>	b			12.7	444
hasn't been by, he				12.5	444
have slipped, like to				70.2	444
haven't, no, I				12.5	444
haven't seen him yet, I				12.5	444
he hasn't been by				12.5	444
hear		o		12.3	444
hear, never did		o		12.3	444
hear one, you'd never get to where you didn't			y	53.8	444
heard		r		12.3	444
heard it, I've		t		12.3	444
<have> heard of him				12.4	444
hearing		s		12.3	444
here, in		s	y	53.6	444
here like flies, people died (a)round <before 1926>			y	78.5	444
he's just a wild animal that will aggravate you	d		y	53.6	444
him for no credit, and ask	f		y	53.4	444
him, <have> heard of				12.4	444
him yet, I haven't seen				12.5	444
his death, what caused				78.7	444
hitch*up the horses				38.6	444
hogs would climb like a dog, them old wild				52.1	444
horses, hitch up the				38.6	444
hot tea, some of them has <instead of coffee>	b			12.7	444

house, getting stuff like that, around your			y	53.6	444
how far is it to [N-Ø]?				70.4	444
hundred, nineteen				1A.2	444
I don't want [P-Ø] go up there, I said			y	53.4	444
I don't think [T-Ø] [C-Ø] any fox squirrels	f	?	y	53.6	444
I got mine too				12.2	444
I got two living wives				12.2	444
I haven't, no				12.5	444
I haven't seen him yet				12.5	444
I said I don't want [P-Ø] go up there			y	53.4	444
I [X-Ø] be dogged	?			12.9	444
in here			y	53.6	444
in it, [T-Ø] wasn't nobody [R-Ø] slept <of a bed>				53.4	444
iodine				78.3	444
is it to [N-Ø]?, how far				70.4	444
is, there it				70.5	444
is, they got now to where dairies			y	53.5	444
it does make it nice				12.8	444
it is, there				70.5	444
it, I've heard		t		12.3	444
it?, [M-Ø] [C-Ø] [D-Ø] ewe, ain't <it's a ewe...>	f			12.6	444
it nice, it does make				12.8	444
it, [T-Ø] wasn't nobody [R-Ø] slept in <of a bed>				53.4	444
it to [N-Ø]?, how far is				70.4	444
it to you, almost give		o		70.2	444
I've ever seen, [D-Ø] prettiest dairy			y	53.5	444
I've heard it		t		12.3	444
[J-Ø] fourth grade, third				1A.3	444
[J-Ø] (o)possum, like a weasel			y	53.6	444
January				1A.6	444
July, June or				1A.8	444
June or July				1A.8	444
just a wild animal that will aggravate you, he's	d		y	53.6	444
just full of them				70.3	444
kicked the bucket				78.6	444
last year, twice as much as				1A.5	444
lights, (a)cetylene	f			70.7	444
like a dog, them old wild hogs would climb				52.1	444
like a weasel [J-Ø] (o)possum			y	53.6	444
like flies, people died (a)round here <before 1926>			y	78.5	444
like that, around your house, getting stuff			y	53.6	444
like to have slipped				70.2	444
living wives, I got two				12.2	444
long, all day <=you always hear helicopters...>	f		y	53.8	444
lot <=an acre>				16.1	444
[M-Ø] [C-Ø] [D-Ø] ewe, ain't it? <it's a ewe...>	f			12.6	444
make it nice, it does				12.8	444
March				1A.6	444
May				1A.7	444
May				1A.7	444
Mercurochrome, Merthiolate or	d			78.3	444
Merthiolate or Mercurochrome	d			78.3	444
million				1A.2	444
millionaire				1A.2	444

mine too, I got				12.2	444
minutes, a few				70.3	444
much as last year, twice as				1A.6	444
[N-Ø]?, how far is it to				70.4	444
never did hear			o	12.3	444
never get to where you didn't hear one, you'd			y	53.8	444
nice, it does make it				12.8	444
nineteen hundred				1A.2	444
ninth grade				1A.3	444
no credit, and ask him for		f	y	53.4	444
no, I haven't				12.5	444
nobody [R-Ø] slept in it, [T-Ø] wasn't <of a bed>				53.4	444
November				1A.6	444
November				1A.8	444
now to where dairies is, they got			y	53.5	444
October				1A.7	444
of him, <have> heard				12.4	444
of quinine, capsules				78.4	444
of September, (e)leventh				1A.3	444
of them has hot tea, some <instead of coffee>		b		12.7	444
of them just full				70.3	444
old wild hogs would climb like a dog				52.1	444
once, all at				1A.4	444
one, you'd never get to where you didn't hear			y	53.8	444
(o)possum, like a weasel or [J-Ø]			y	53.6	444
or July, June				1A.8	444
or Mercurochrome, Merthiolate		d		78.3	444
orchard				53.2	444
over there, those				52.1	444
[P-Ø] go up there, I said I don't want			y	53.4	444
palings				16.2	444
passed away there				78.5	444
people died (a)round here like flies <before 1926>			y	78.5	444
patch, corn <larger than a lot; a small field>				16.1	444
patch, pea <larger than a lot; a small field>				16.1	444
pea patch <larger than a lot; a small field>				16.1	444
picket fence <1" square pickets nailed to 2 rails>				16.2	444
plantation		f		70.9	444
pond up, when they first backed this			y	78.5	444
poor farmers				53.1	444
post		h		16.5	444
post[N-i]		i		16.5	444
post[N-i], two		i		16.5	444
prettiest dairy I've ever seen, [D-Ø]			y	53.5	444
pretty close <to midnight>				70.1	444
proud flesh				78.2	444
quinine				78.4	444
quinine, capsules of				78.4	444
[R-Ø] slept in it, [T-Ø] wasn't nobody <of a bed>				53.4	444
rail fence				16.4	444
rear wheels		f		70.8	444
remember that far back, <I> don't				70.4	444
rock <fences/walls not made of rock here>				16.6	444

ruptured gallbladder	f			70.6	444
said I don't want [P-Ø] go up there, I			y	53.4	444
scramble some eggs	f			70.7	444
second				1A.3	444
seen, [D-Ø] prettiest dairy I've ever			y	53.5	444
seen him yet, I haven't				12.5	444
September				1A.6	444
September				1A.9	444
September, (e)leventh of				1A.3	444
seventh	b			1A.4	444
seventy				1A.2	444
shoo <to scatter them>				38.5	444
sixth	b			1A.4	444
<brain-> skinning[=scanning] <device>	f			70.6	444
slept in it, [T-Ø] wasn't nobody [R-Ø] <of a bed>				53.4	444
slipped, like to have				70.2	444
some eggs, scramble	f			70.7	444
some of them has hot tea <instead of coffee>	b			12.7	444
sooey <to scatter them>				38.3	444
squirrels, I don't think [T-Ø] [C-Ø] any fox	f	?	y	53.6	444
stuff like that, around your house, getting			y	53.6	444
[T-Ø] [C-Ø] any fox squirrels, I don't think	f	?	y	53.6	444
[T-Ø] wasn't nobody [R-Ø] slept in it <of a bed>				53.4	444
tea, some of them has hot <instead of coffee>	b			12.7	444
that, around your house, getting stuff like			y	53.6	444
that far back, <I> don't remember				70.4	444
that will aggravate you, he's just a wild animal	d		y	53.6	444
the bucket, kicked				78.6	444
the horses, hitch up				38.6	444
the twentieth, August				1A.6	444
them has hot tea, some of <instead of coffee>	b			12.7	444
them, just full of				70.3	444
them old wild hogs would climb like a dog				52.1	444
there, I said I don't want [P-Ø] go up			y	53.4	444
there it is				70.5	444
there, passed away				78.5	444
there, those over				52.1	444
there, up				52.2	444
they first backed this pond up, when			y	78.5	444
they got now to where dairies is			y	53.5	444
think [T-Ø] [C-Ø] any fox squirrels, I don't	f	?	y	53.6	444
third [J-Ø] fourth grade				1A.3	444
<the> thirties <1930s>				1A.1	444
this pond up, when they first backed			y	78.5	444
those over there				52.1	444
thousand				1A.2	444
thousand				1A.2	444
to-forty-five draft, eighteen- <conscription in WWI>	f			1A.1	444
to have slipped, like				70.2	444
to [N-Ø]?, how far is it				70.4	444
to where dairies is, they got now			y	53.5	444
to where you didn't hear one, you'd never get			y	53.8	444
to you, almost give it		o		70.2	444

too, I got mine				12.2	444
tumor, brain	f			70.6	444
twentieth				1A.3	444
twentieth, August the				1A.6	444
twice as big				1A.5	444
twice as much as last year				1A.5	444
two living wives, I got				12.2	444
two post[N-i]		i		16.5	444
up, get <or touch him with a switch>				38.1	444
up the horses, hitch				38.6	444
up there				52.2	444
up there, I said I don't want [P-Ø] go			y	53.4	444
up, when they first backed this pond			y	78.5	444
up yonder				52.2	444
want [P-Ø] go up there, I said I don't			y	53.4	444
wasn't nobody [R-Ø] slept in it, [T-Ø] of a bed				53.4	444
weasel [J-Ø] (o)possum, like a			y	53.6	444
weevil, boll	f			70.8	444
what?				52.4	444
what caused his death				78.7	444
wheels, rear	f			70.8	444
when they first backed this pond up			y	78.5	444
where dairies is, they got now to			y	53.5	444
where you didn't hear one, you'd never get to			y	53.8	444
whoa				38.2	444
whoopa(x4) <to call them; falsetto>				38.3	444
wild animal that will aggravate you, he's just	d		y	53.6	444
wild hogs would climb like a dog, them old				52.1	444
will aggravate you, he's just a wild animal that	d		y	53.6	444
wire, barb	f			16.3	444
wire fence				16.3	444
wire, wove				16.3	444
wives, I got two living				12.2	444
would climb like a dog, them old wild hogs				52.1	444
wound				78.1	444
wove wire				16.3	444
[X-Ø] be dogged, I	?			12.9	444
year, twice as much as last				1A.5	444
yet, I haven't seen him				12.5	444
yonder, up				52.2	444
you, almost give it to		o		70.2	444
you didn't hear one, you'd never get to where			y	53.8	444
you, he's just a wild animal that will aggravate	d		y	53.6	444
you'd never get to where you didn't hear one			y	53.8	444
your house, getting stuff like that, around			y	53.6	444