

LAGS WORKING PAPERS, FIRST SERIES (1981)

WORKING PAPER NUMBER SIXTEEN

LAGS Typescripting: A Preliminary Program

Marvin Bassett

LAGS Typescripting: A Preliminary Program

INTRODUCTION

A useful contribution of the Linguistic Atlas of the Gulf States (LAGS) Project to descriptive linguistics would be the design and implementation of a suitable method for typescripting its corpus of field records. This would give students of American English access to a maximum amount of general information about this survey of Southern speech. This objective is in keeping with the LAGS practice of making available in useful formats as many analogues of the systematically gathered data as need dictates. A running typescript of a field record, done in conventional orthography, would represent the most comprehensive and least abstract analogue of an interview produced by LAGS to date. The other analogues include 1) the idiolect synopses, brief summaries in narrow phonetics of phonological, grammatical, and lexical characteristics of the speech of each informant; 2) the protocols, more extensive descriptions in narrow phonetics of both work-sheet and conversational items occurring in the speech of each informant; and 3) the concordance, a comprehensive alphabetical list in conventional orthography of the information recorded in the protocols. This paper explains the process necessary in formulating a workable system of typescripting that is both useful to students and responsive to the problems in the data it describes. First, the need for and uses of typescripts are discussed, ranging from their utilitarian function in providing the contexts in which atlas data was elicited to their historical significance in representing a body of spoken English. The background of this

program is then explained, indicating both the problems involved and the gradual development of the present system. Finally, the proposed method of typescripting is discussed and illustrated by representing a portion of a LAGS field record.

USES OF TYPESCRIPTS

Typescripts of LAGS field records not only would readily enhance interpretation and clarification of both the data contained in the interviews and the methods of elicitation but also would facilitate a variety of studies of regional and/or social usage. More particularly, these analogues would be useful in 1) establishing contexts for linguistic forms, 2) conducting exhaustive studies of subregional speech and providing a basis for a global description of spoken English, 3) acting as a pedagogical device by allowing for the intimate observance of colloquial usage, 4) serving as an evaluative instrument in gauging the effectiveness of fieldworkers and scribes, and 5) providing historical documentation by preserving a corpus of regional speech.

When composing the protocols, scribes tried to record all linguistic forms clearly, especially through the use of marginal glosses. Nevertheless, the abbreviated nature of these analogues masks the full context from which these forms were extracted. The statement appearing on the title page of every LAGS protocol makes explicit the function of the protocol in relation to the field record:

These phonetic transcriptions of a tape-recorded interview serve as a guide to that primary text, and all of these observations are to be confirmed

by the reader through auditions of the field record (the tape-recorded interview).

Rather than resorting to the field record for purposes of clarification and/or elaboration, the reader of the protocol would have the more convenient alternative of consulting a typescript to determine the precise context of the item in question. If assembled in a coherent and explicit way, typescripts can be useful not only in verifying ambiguous or otherwise questionable data in the protocols but also in providing full information about such matters as sense, style, incidence, familiarity, and meaning. In addition to expanding the base for the contextual study of linguistic forms, then, typescripts would also reinforce an underlying principle in LAGS of making the data completely verifiable.

The above-mentioned distinction between protocol and field record also has implications both for idiolectal, community, or subregional studies and global descriptions of speech based on typescripted data. That is, such investigations, whether confined to a single feature of an idiolect or ranging over the phonological, grammatical, and lexical/semantic patterns of speakers in an entire region, could utilize all of the information preserved in a field record, not simply that evidence recorded in the protocols or idiolect synopses. Use of typescripts, for example, would make possible a systematic examination of all singular/plural sets of nouns occurring in the speech of informants, to reinforce the limited data of those 13 forms that were regularly transcribed in the protocols. The feasibility of making global descriptions of spoken English from typescripted data has been illustrated in the discussion of the FIELD RECORDS in Chapter 6 of the LAGS Handbook, by an inventory of a short

conversational passage from LAGS field record CG 283.01. The useful application emerging from such description is the identification of forms needing further investigation, specifically, phonological contexts for structural description, grammatical forms for morphological and syntactic patterns, and lexical and semantic forms for studies of synonymy and meaning. Also mentioned in that chapter was the suitability of this format for non-inventorial analysis, e.g., the study of contextual phonology, semantics, syntax, paralinguistics, oral composition, and folklore.

In addition to those descriptive uses, these analogues might help fulfill a fundamental obligation of LAGS to students. As suggested in the Manual, "no operation in linguistic geography, or any other application of general linguistics, is more important than the training of students." Used pedagogically, typescripts meet this requirement by bringing an observer into close contact with an extended passage of colloquial usage, allowing him to perceive a variety of structural alternations that might escape him in ordinary conversation. It becomes possible, then, to distinguish between important and trivial differences in habits that may, for example, reflect the same morphophonemic process (e.g., the plural of fist and the preterit of ask). With the expectation that students will actually have a hand in the composition of a significant number of texts, this use of typescripting should be realized immediately.

Typescripts might also help to improve field and scribal procedures by showing investigators precisely what they did and what they failed to do. By recording the complete text of the interview, including the utterances of the fieldworker, the typescript affords a valuable means of establishing a profile of a successful field record. An examination of a typescripted field record

would indicate the degree to which the fieldworker was successful in such crucial matters as perceptive questioning and establishing rapport with the informant. The former would involve checking the text for 1) the interviewer's diligence in investigating regular work-sheet items, 2) his willingness to pursue related matters that emerge from routine questioning, 3) his ability to discern and pursue work-sheet items occurring in free conversation, 4) his effectiveness in stimulating wide-ranging conversation by asking general or "shotgun" questions, and 5) his skill in having the informant provide concrete illustrations and details about artifacts and clear definitions of terms. The typescript should also reveal whether the fieldworker has successfully made the interview situation relaxed and uninhibited. This could be judged from the interplay between the two individuals and by the relative garrulousness of the informant. By noting the amount of conversation originating with the fieldworker, one should be able to determine whether the interviewer has heeded McDavid's precept maintaining that "the less conspicuous the fieldworker, the better." With regard to scribal evaluations, a check of the protocol against the typescript will quickly reveal any phonological, grammatical, or lexical information that has been missed in transcription. Furthermore, judgment of the scribe's accuracy and diligence in making all types of glosses as well as in the identification of conversational passages will be facilitated by use of the typescript.

Finally, by preserving a body of regional speech, LAGS typescripts would have real value as historical documents. When published, these texts, representing approximately thirty million words of spoken English, will complement existing records such as the Corpus of Present-Day Edited American English at

Brown University. This collection consists of about one million words of connected discourse drawn from various categories of printed English. Collectively, these records would permit valuable studies of current spoken and written usage of the language.

DEVELOPING THE SYSTEM OF TYPESCRIPTING

Although the application of typescripting LAGS field records has been recognized for several years, the practical and technical problems of converting speech to conventional orthography have been deferred as the basic research was completed. The progress toward developing a system for typescripting can be outlined by a discussion of 1) an initial investigation into the practicability of having the work done by court reporters or secretaries, 2) a format for composing inventory slips, proposed in Working Paper #5, A Compositional Guide to the LAGS Project (Guide), 3) experimental typescripting of a field record during Pederson's seminar in general dialectology at Emory in 1977, and 4) the original concordance system of conversion.

The conventional notion of having court reporters or secretaries compose the typescripts was rejected on both technical and practical grounds. Pederson identified these problems in Working Paper #4, "Toward the Publication of the Linguistic Atlas of the Gulf States," as follows:

During the past year [1976], we have learned that the Urban Language Series approach to typescripting, i.e., turning it over to a secretary, is wholly unacceptable. Only trained phoneticians can do that work effectively, and much additional thought and discussion are also needed in the organization of an automatic (or mechanical or systematic) method of dialect writing with conventional orthography...The implication in 'Tape/Text and Analogues' that all field records will be typescripted

must be corrected at once. We find that court reporters trained to do this work would need \$250,000 to do the entire job.

For these reasons, the efforts of the staff were redirected to more pressing concerns until the following year.

A tentative system for converting phonetic transcription to conventional orthography appeared with the publication of the Guide in 1977. This system was formulated by Pederson and his staff for the purpose of conducting a preliminary inventory of protocol data that would provide an estimate of the size and complexity of an exhaustive index of protocol material. The aims of the inventory were enumerated in this way:

1. To identify all the linguistic information in all protocols;
2. To transliterate all phonetic notation into conventional orthography;
3. To establish a system of retrieval to accommodate all data not covered by the systematic investigation of the WORK SHEET/PROTOCOL items/entries;
4. To establish a WORD LIST for the composition of the DESCRIPTIVE MATERIALS (p. 124).

A system of conversion that met these objectives also had the potential for application in the production of typescripts, a fact recognized in the Guide:

TYPESCRIPT FORMAT will follow the style identified above for the inventory...Specifically, no phonetic notation will appear in any typescript, all assimilated, deleted, excrescent, or otherwise transformed surface structure syntax will be written out in full and underscored, and special orthography will be used only in the composition of words that have no written precedents (p. 141).

In addition to the principles mentioned above, the system of conversion, described in the Guide under GUIDELINES FOR COMPOSING SLIPS (pp. 124-36), specified 1) usage of first spellings in Webster's Third New International Dictionary of the English Language (W3), 2) a means of showing significant phonemic variation from these spellings, 3) the use of parentheses to indicate added forms, 4) the use of special slips for forms not conveniently represented in conventional orthography, e.g., grunts of affirmation, negation, and hesitation, and 5) symbols indicating the presence of marginal glosses in the protocol. Appropriate parts of this system were then utilized in the experimental typescripting of a field record.

In the fall of 1977, students in Pederson's seminar in general dialectology were assigned one-hour portions of LAGS field record CG 283.01 for typescripting according to the principles for conversion set forth in the Guide. That experiment revealed not only the dimensions of a text representing a full interview but also some procedural difficulties. Of particular concern were 1) the inordinate amount of time necessary in identifying phonemic variations from W3 pronunciations and 2) the inability of some students, especially those with little training in phonetics, to detect some types of articulatory phenomena, e.g., deletion of final consonant clusters. In addition to indicating a need for further modification in adapting the conversion rules to typescripting, this experiment verified the earlier-stated principle of using only trained phoneticians to compose the typescripts.

Further refinements in this system of conversion followed the decision in 1980 to produce an exhaustive finders' list of all phonetically transcribed data in the protocols by means of a computer-assisted concordance. Pederson's

description of the methodology to be utilized in the composition of the concordance identified principles of conversion for 1) dealing with phonemic, rather than phonetic, variation, 2) using, without special spelling, all pronunciations described in W3, 3) marking phonemic variation not described in W3, 4) using special spellings in W3, 5) spelling and marking words not included in W3, 6) capitalizing appropriate words, 7) representing calls to animals, 8) using parentheses for expanding forms that have been reduced and for showing deleted grammatical signals, 9) using punctuation, and 10) dealing with false starts and stammering. In addition to these rules, abbreviations were specified for identifying grammatical forms and scribal glosses. A second conversion of protocol CG 283.01 according to the revised rules showed that the process took far too much time, and, again, the culprit proved to be the requirement of identifying and marking phonemic variations from the pronunciations recorded in W3. Principally for this reason, the next several months were almost exclusively devoted to revision and experimental application of the conversion rules with the aim of composing, in an optimum amount of time, a concordance that was rich in information.

PROCEDURES FOR TYPESCRIPTING

The current plan is to apply the latest form of the concordance rules for conversion, with slight modification, to the typescripting of LAGS field records. These rules are reproduced in Working Paper #9, "A Plan for the LAGS Concordance." In reading them, one should remember that they were originally designed for the purpose of converting phonetic transcriptions to conventional orthography but, for the most part, are readily transferrable

to the task of representing connected discourse. The following conversion glosses and textual code have no application to typescripting:

- a - probable scribal error of text or marginal gloss
- b - variant (or inappropriate) line use
- c - disambiguation of entry
- f - usage and illustration
- g - citation
- < > - frame remarks, i.e., contexts in conventional orthography provided by the scribe

The grammar glosses would probably never be used either, since the form would be apparent from its context in the typescript. The rules for phantom space (*) may provisionally be considered a part of the typescripting system on the assumption that computers might eventually be used in the analysis of the typescripts. One should keep in mind, however, that the use of phantom space is dependent upon the kinds of things to be counted (e.g., recognized by the computer as a unit) and that the rules for using phantom space will probably require modification to facilitate the desired analysis.

This apparatus for conversion is applied to the typescripting of a short portion (about 3 minutes in duration) of LAGS field record CG 283.01, work sheet item 17.5, below. Here the original symbols for the glosses are retained although some modification will, no doubt, be necessary to allow for the superfluous glosses (see above). The glosses are placed in brackets within the text. The reader is referred to Working Paper #10, "LAGS Field Records: Form and Content," for information about aural problems that may

be encountered when working with the taped interviews.

F[fieldworker]: Uh-huh. And what were some of the names of some of the things that you used? For example, what would you fry eggs in?

I[informant]: Just a frying*pan.

F: Uh-huh. Ever hear that called anything else? People have any other names for frying pan?

I: No, no, GRUNT (N).

F: What about, did you ever call it a skillet?

I: Skillet, yes, you can get a skillet or Dutch*oven or . . .

F: What's a Dutch oven?

I: A Dutch*oven's an outfit you can get it under [d], over a [!] open fire, and, and you can put your biscuits [d], make up your biscuits and put them in that, bottom of it, and then put your top on it to fit down in it, kind of sealed it up, and just set it out in the coals and rake [d], cover it up with coals, and, and leave it so long, and you just pull it out, and you had some of the prettiest flour*bread you [X-φ] ever seen.

F: Aw.

I: [M-φ] used to go down on the river, fellow [d], this German family over here, they cooked on, on the fireplace a long time, you know. The old man had, had two sons. They [C-φ] all dead now. And there was Conrad and Fritz, and the old man, I forget what was his name [!], but he had something, he used to come (a)round every morning with something to sell; he had what he called his bee*town, he had about 75 hives of bees.

F: Huh.

I: And he, he had honey in the comb and honey that was been strained*up [!]

and slung*out in little containers, and he got (al)most any kind of vegetable you'd think of, year around; I don't know how he'd do it, but he'd have turnips and, and mustard and stuff like that.

F: Huh.

I: Always something to sell. And he had him an old horse, was [d] Hans was the horse's name, I never [X-Ø] forget it. And he had him a kind of buckboard*deal, a wagon*deal on springs, springs on each side of it. [M-Ø] always thought the world of old man [friend's name].

F: Uh-huh.

I; When I was a little fellow, I'd go over there and, if you wanted to go in his bee*town with him now don't ever make no fight, if one (a)lights on you, don't knock him off.

F: Uh-huh.

I: You'd make him mad. And he'd soon have that whole bunch on you. And I've been in there with him a time or two, but I, I didn't never feel comfortable, he'd just, just pick up a handful of them, stuff like that, you know.

F: Huh. That's amazing.

I: Yes, it is, I . . .

F: You mentioned that Dutch oven. Did people ever call that a spider? Or was that spider something different?

I: [D-Ø] spider was different*to [!] that.

F: Could you explain what a spider is?

I: Well, I just, just good for [d], I reckon that Dutch*oven without a lid would be called a spider.

F: Uh-huh. Did it have a frying pan with legs on it?

I: Well, no, a spider was a [d], I'd call that a trivet, the little-old outfit they'd make and put legs on to set your frying*pan on, that's [d], that was a spider or a trivet.