

septembre 22, 2005

COLUMN

ON DATA*

Claire Lefebvre
Université du Québec à Montréal

Introduction

In the introduction of his book on Fula, De Guiraudon (1894, pp. V–VIII) reviews the literature on the language. The following passages are representative of his style of presentation.

ANONYME. — *Vocabulaire foule*, publié en 1845, dans le tome II des *Mémoires de la Société ethnologique*, d'après un manuscrit de la Bibliothèque nationale. Ce vocabulaire remonte au siècle dernier, et semble avoir été recueilli par un missionnaire français. Outre de nombreuses erreurs, l'écriture du manuscrit est très mauvaise, de sorte que l'éditeur, ne connaissant pas la langue, a mal lu presque partout et a ajouté de nouvelles erreurs (1894, p. V). [ANONYMOUS. — *Vocabulaire foule*, published in 1845, in volume II of the *Mémoires de la Société ethnologique*, based on a manuscript of the Bibliothèque nationale (National Library). This vocabulary dates back to the last century and seems to have been gathered by a French missionary. The manuscript comprises numerous errors; moreover, the quality of the writing is very poor. Thus, the editor, who did not master the language, misread almost everywhere and added errors of his own.]

* The research underlying this column was financed by SSHRCC, FCAR, and UQAM's institutional research funds. I would like to thank the following people for their very useful questions and comments on an earlier version of this column: Anne-Marie Brousseau, Viviane Déprez, Robert Papen, Ingo Plag, Jean-François Prunet, Armin Schwegler, Jeff Siegel, and Donald Winford. The usual disclaimer applies. I would also like to thank Renée Lambert for helping me with the documentation, Catherine Léger for copy-editing the manuscript, and Andrée Bélanger for formatting it.

REV. C. A. L. REICHARDT. — *Grammar of the Fulde language* (London, 1876). L'auteur de ce long et ennuyeux livre, mercenaire allemand au service des missions anglaises, ne semble pas avoir cherché à apprendre la langue de la bouche des indigènes. Cette chose était sans doute au-dessous de sa dignité, et il a une manière bien plus originale de procéder. Au lieu de déduire les principes de la grammaire de l'étude des textes recueillis par son prédécesseur et homonyme, il a préféré inventer d'emblée des théories grammaticales fausses et absurdes, après quoi il n'a pas craint de falsifier les textes pour les mettre d'accord avec les règles écrites dans son cerveau malade (1894, p. VI). [REV. C. A. L. REICHARDT. — *Grammar of the Fulde language* (London, 1876). The author of this long and boring book, a German mercenary at the service of the English missions, does not seem to have tried to learn the language through native speakers: this task was probably beneath his dignity and his way of proceeding was much more original. Instead of deducing the grammar principles from the study of the texts gathered by his predecessor and equivalent, he preferred to invent straightaway false and absurd linguistic theories; afterwards, he did not mind altering the texts so as to make sure that they would be in agreement with the rules hatched in his sick mind.]

In the linguistic literature of the late nineteenth century, it is not rare to find this kind of remarks on data. While the tone of the debate surrounding this topic may have improved in some cases over the last century, there are still a lot of ongoing discussions on data on the basis of languages of various families. In the recent literature in creole studies, several facets of the problem surrounding linguistic data have been at the heart of controversies. For example, some authors have raised the problem of what constitutes a linguistic fact in creole studies (e.g., Bickerton, 1996; DeGraff, 2000). Others have raised the question of whether historical texts constitute a valid source of data for linguistic analysis (e.g., Muysken, 1995). The problem of inconsistencies in some types of linguistic data has also been addressed (e.g., Labov, 1975). This column is a discussion on data, based on some thirty years of experience in data collection and analysis, mainly carried out on Quechua, French, Martinican Creole, Haitian Creole, and Fongbe. My thoughts are organized around three major themes: the non-neutral character of linguistic data, how to overcome the limits of particular databases, and the problem of “inconsistencies” in elicited data.

The non-neutral character of linguistic data

Data collection is not a neutral activity. It is always designed within the framework of a particular methodology. Different research paradigms and questions call for different methodologies, hence for different types of databases. Data are thus not independent from research paradigms and questions. For example, the content of lexicons (or dictionaries) varies according to the research tradition of the authors. The methodology used to establish official terminologies (e.g., Chansou, 1997; Gacic, 1994; Nagao, 1994; Thoiron *et al.*, 1997) differs from that used to record the lexical entries of everyday speech. For one thing, whereas the latter may provide all the forms used to refer to a given object, including dialectal or regional variants, the former will provide the one form that has been selected as the “official” one. Likewise, the methodology established by Weinreich (1984) for lexicographers specifies that the definitions of lexical entries should be valid for the entire linguistic community rather than for particular idiolects. In contrast, lexicons built in the tradition of lexical semantics are based on the mental lexicons of individuals. Thus, in this approach, the content of lexical entries corresponds to particular idiolects. In writing lexicons (or dictionaries), lexicographers and lexical semanticists have different goals. Lexicographers seek to make dictionaries of particular languages. Lexical semanticists, however, seek to describe the knowledge speakers have that enables them to use particular lexical entries. The methodology that guides the content of lexical entries is thus not exactly the same in the two traditions. For example, in Valdman’s (1996) English-Haitian Creole lexicon, we find that the English verb ‘to escape’ may be rendered by two Haitian words: *chape* and *sove*. In Valdman (1996), the content of these two lexical entries is as in (1).

- (1) a. *chape* escape, make it, pull through, rescue
 b. *sove* break out, escape, fly the coop, get out
sove ak/avè(k) make off with
sove kite flee

The English lexical entry ‘to escape’ described by Levin (1989, p. 120), a lexical semanticist, is as in (2).

- (2) a. The convict escaped.
 b. Preposition Drop:
 a) The convict escaped from the police.
 b) The convict escaped the police.
 c. *Causative/Inchoative Alternation:
 a) The convict escaped.
 b) *The collaborators escaped the convict.
 (on the reading of “cause to escape”)
 d. Depictive but not resultative phrases:
 The convict escaped exhausted.
 e. Adjectival Passive/Perfect Participles:
 an escaped convict, the recently departed guests
 (a convict that has escaped, not a convict that someone has escaped from)
 *an escaped jail
 f. *The convict escaped the soles off his shoes.
 g. *The convict escaped his way to freedom.

The content of the lexical entries in (1) and (2) reflects the different goals of lexicographers and lexical semanticists, respectively. This shows that the type of data researchers collect and the type of information recorded in lexical entries are not independent from research paradigms and questions.

Likewise, we can contrast the types of grammars that can be written on a given language. For the present purpose, I will contrast prescriptive grammars with generative grammars. Prescriptive grammars seek to provide the official or accepted way of saying things in a given language. They thus provide a list of licit structures and sometimes a list of structures identified as “do not say”. The structures identified as “do not say” are often possible ones in the language, in the sense that some people use them. They are, however, identified as illicit on the basis of normative judgments. In contrast, generative grammars seek to provide an account of the knowledge speakers have that enables them to use their language. The databases of generative linguists mainly consist of grammaticality judgments. Generative grammars thus provide the structures that are judged to be grammatical by a given speaker; they mark as ungrammatical those that have been identified as such by the same speaker. Structures considered illicit by normative grammarians are not necessarily illicit for generative linguists. “Truly” ungrammatical sentences (in the generative grammar sense) are usually not provided by normative grammarians; they are, however, crucial to the analyses produced by generative linguists. The two types of grammars thus call for databases that overlap in some points, but that are far from being identical.

Another contrast between two research paradigms further shows that data collection is not a neutral activity. The methodology guiding the study of ongoing linguistic change within a given community, as established by Labov and his associates, calls for large corpora of natural speech data drawn from a representative sample of members of this community (e.g., Labov, 1972a; Sankoff (Ed.), 1980; etc.). Similarly, longitudinal studies of child language acquisition require corpora of recorded data drawn from the speech of children (e.g., McDaniel *et al.* (Eds.), 1996, and the references therein). In contrast, the methodology of what is being referred to in the literature as “field linguistics” does not call for the same type of data. The aim of field linguistics is to provide descriptions of the languages of the world in order to reveal the range of structures available to human languages. These descriptions can be used to build a database useful for comparing the languages of the world. In field linguistics, the

informant is the source of information and the evaluator of the utterances submitted by the investigator (e.g., Bouquiaux & Thomas, 1992; Burling, 1984; Kibrik, 1977; Longacre, 1964; Payne, 1997; Samarin, 1967; Vaux & Cooper, 1999; etc). Samarin stresses the fact that “(...) without an informant one cannot test hypotheses (...) and cannot make statements concerning the productivity of morphemic relations — one cannot predict” (1967, p.123). Bloomfield was a leader in this tradition of research, mainly through his work on Algonquian languages (e.g., *A Leonard Bloomfield anthology*, edited by Hockett, 1970, and “Leonard Bloomfield’s descriptive and comparative studies of Algonquian” by Goddard, 1987). Most of the books which are part of grammar collections of several major publishing houses are based on this tradition of research. For example, in March 2000, Mouton de Gruyter had in one of its collections twenty grammars of languages as diverse as Pipil, spoken in El Salvador (Campbell, 1985), Breton (Press, 1986), Kilivila, spoken in the Trobriand Islands (Senft, 1986), Slave, an Athapaskan language (Rice, 1989), Afrikaans, spoken in South Africa (Donaldson, 1993), Berbice Dutch Creole (Kouwenberg, 1994), Wardaman, spoken in the Northern Territory of Australia (Merlan, 1994), Tukang Besi, spoken in Indonesia (Donohue, 1999), etc. The grammars of a significant number of languages (e.g., African, Amerindian, Asian, Australian, European, Indonesian languages, etc.) have been described within the research tradition of field linguistics. It goes without saying that the methodology and the type of data that define this paradigm of research cannot be used to address the questions of researchers working in sociolinguistics and language acquisition. The reverse is also true. The large corpora of data required in sociolinguistics and language acquisition could not be used as such in the research program set forth by field linguists. This is another example showing that the type of data researchers collect is determined by their research paradigms and questions.

All the examples above show the non-neutral character of linguistic data. Several authors have even pointed out that linguistic theories constitute instruments that are helpful in searching for data relevant to solving problems (e.g., Aoun, 1992; Botha, 1976; Mulder,

1996). As Aoun puts it: “(...) the uncovering of the data is theory driven: it is the theoretical or analytical apparatus that explicitly or implicitly guides linguists in their search for the data” (1992, p. 77). Furthermore, as was illustrated above, what should be recorded in the course of data collection for a given project depends on the nature of individual research programs. It should be pointed out that the non-neutral character of data has become a central notion in epistemology. As advocated by Popper (1972), researchers need to know what to observe and why in order that their observations be meaningful. This is true for all sciences, including linguistics. In this respect, we may wonder why the types of data produced by the methodologies of field linguistics and generative linguistics have been so unpopular with many colleagues in creole studies, given the fact that these methodologies are being used successfully for a wide range of languages.

Since scholars who work on creole languages come from different research traditions, it should not come as a surprise that the data they report on are of different types. In my view, data coming from different research traditions on a given language are complementary. They reflect the different angles from which the linguistic facts of a given language have been addressed. Each type of data contributes something to our overall understanding. Each type presents advantages and disadvantages. There are ways of overcoming the limits of individual types of databases, however. These topics are discussed in the next section.

How to overcome the limits of particular types of databases

There are different types of data that researchers can use as sources for their analyses. The following paragraphs discuss the advantages and disadvantages of the various types of databases and present ways of overcoming their respective limits.

Older texts in a given language constitute a possible source of data for linguistic analysis. This source of data may provide information on earlier stages of a particular language. As has been pointed out by Muysken, however, “part of the problem in dealing with historical

materials is determining exactly what type of speech is reflected by them. This problem is even greater in creole societies, where often dramatic linguistic differences occur within one speech community” (1995, p. 335). This observation by Muysken, based on Negerhollands historical materials, carries over to Haitian historical materials. As has been pointed out in Lefebvre (1998a, pp. 68–69), the Haitian written sources covering the period between 1776 and 1936, listed in Baker and Corne (1982, pp. 273–274), have been questioned with respect to whether they reflect the speech of the majority of the African slave population in Haiti at the time they were written. Chaudenson (1977, p. 259), Dejean (cited in DeGraff, 1993, p. 90, note 56), and Lefebvre and Lumsden (1994, p. 136) are of the opinion that they do not. Hence, for these authors, Haitian historical materials are not considered as a valid source of data (but see Carden & Stewart, 1988, pp. 26–27, for a more moderate position on this matter). Whether historical materials constitute a valid source of data for linguistic analysis may vary from language to language, depending on the adequacy of the written sources. Hence, while they may be considered as a valid source of data in some cases (e.g., Muysken, 1995, on Negerhollands), they may not in other cases (e.g., Lefebvre, 1998a, pp. 68–69, on Haitian).

Large corpora of recorded and transcribed material present advantages and limits (e.g., Isquierdo, 1998; Issoufi, 1998; Kuo, 1997). As for the advantages, they allow us to make word and structure frequency and probabilistic analyses, discourse analyses, and language variation analyses along various dimensions. As for the drawbacks of recorded texts, we find the following: recorded texts do not contain readily available information on the meaning of words and structures anymore than written texts do. For example, the Haitian lexical entries *sa* and *sila* were shown to manifest slightly different interpretive patterns across speakers. The three patterns of grammar reported on so far are listed in (3) (from Lefebvre, 2000, forthcoming).

- (3) G1 *sa* [+ proximate] *sila* [– proximate]
 G2 *sa* [□ proximate] *sila* [– proximate]
 G3 *sa* [□ proximate] *sila* [□ proximate] (= (15) in Lefebvre, forthcoming)

As is pointed out by Valdman (1970, p. 204), when the subject has been questioned, the morpheme *ki* must appear in the vicinity of the verb.¹ Obligatoriness of *ki* in sentences of the type of that in (5) is shown by the ungrammaticality of the sentence in (6), where *ki* is missing.

- (6) *ki-moun _ te vini an
 which-person _ ANT come DEF (=16) in Koopman, 1982, p. 211)

To my knowledge, Koopman (1982) is the first author to have provided examples of long-distance Wh-movement involving the subject. Such a case is illustrated in (7).

- (7) Ki-moun Jan kwè Mariz te di ki te vini an?
 which-person John believe Mariz ANT say COMP ANT come DEF
 ‘Who is it that John believed that Mariz said that came?’

(=34) in Koopman, 1982, p. 216)

Koopman (1982, pp. 219–220) points out that *ki* must occur in the embedded clause from which the subject has been extracted by long-distance Wh-movement. Without the embedded *ki*, the clause in (7) would be ungrammatical. Furthermore, Koopman shows that *ki* only occurs in cases of Wh-subject extraction, whether it involves short- or long-distance movement. It cannot occur in other cases; namely, it is excluded in cases of object or adjunct extraction. Déprez (1992a, 1992b, 1994), Law (1992, 1994), Lumsden (1990), and Sterlin (1988) provide data that show the same distribution for *ki* as the one in Koopman.² DeGraff reports on a slightly different distribution of *ki*, whereby *ki* “(...) is not obligatory in COMP of

¹ There are competing analyses for *ki* in the literature. Some authors analyze it as a complementizer governing the empty subject position; others, as a resumptive pronoun occurring in subject position. The glosses in the examples respect the authors’ analyses.

² Most of the cited authors work in different cities — some even in different countries. The fact that their informants came up with the same distribution for *ki* is thus significant.

Fournier), and the transcriptions constitute a source of information against which it is possible to check the grammaticality judgments of the informants in some cases. For example, texts give direct access to surface word order(s) in the subsystems of a creole grammar (e.g., the nominal structure subsystem; the tense, mood and aspect subsystem; etc.). Interpretive data of the type of those in (3) and (4) cannot be checked against texts, for, as was pointed out earlier, texts do not provide this type of data. Texts do not provide researchers with negative evidence of the type of that given in (6) nor with data that require syntactic tests (e.g., see (7)). It is worth pointing out that particular structures may not be instantiated in recorded texts. For example, in the Lefebvre and Fournier corpus, there is not a single occurrence of subject extraction involving long-distance Wh-movement. Hence, in this case, it is not possible to confront the different grammaticality judgments in (7) and (8) with data produced spontaneously. Nonetheless, whenever possible, checking the grammaticality judgments of speakers against recorded material is a good way of making sure that these two types of data are not in disagreement. Based on my experience, I would say that the two types of data are generally not in disagreement. Apparent problematic cases are discussed in the next section.

The problem of “inconsistencies” in elicited data

The validity of data drawn from elicitation sessions with informants has been challenged on some occasions. For example, Labov (1975) has raised the question of (in)consistencies between informants and within a given informant. I will address these two aspects of the problem in turn on the basis of my own fieldwork experience on Haitian Creole and its source languages.

The problem of inconsistencies between speakers

It could be the case that not all speakers of a given language share the same judgments on a given structure. The determiners occurring in the Haitian nominal structure will be used to illustrate this point. It is a well-known fact that count nouns in Haitian Creole may be followed

As can be seen in (10c), the surface sequence *yo a* is possible. All authors agree, however, that, in this case, *yo* is interpreted as the third person plural possessor (e.g., DeGraff, 1992, p. 108). All the authors who have discussed the properties of these determiners agree on (9b) and (9c). The property on which authors, and hence Haitian speakers, disagree is the one in (9a), exemplified in (10a). The remainder of the discussion will concentrate on this property.

When I did the research underlying Lefebvre (1982) on the Haitian determiners, I was not surprised to find that *la* and *yo* could co-occur in the grammar of the speakers I worked with, for other authors had already acknowledged this possibility (e.g., d'Ans, 1968, p. 105; Faine, 1937, p. 83; Fournier, 1977, p. 43; Goodman, 1964, p. 45; Sylvain, 1936, p. 55; Valdman, 1978, pp. 194–195). Furthermore, there are instantiations of co-occurring *la* and *yo* in the Lefebvre and Fournier (1976) corpus. In the literature published after Lefebvre (1982), we find the following reports. Lefebvre and Massam (1988), as well as Ritter (1992, pp. 207–209), also note the possibility of co-occurrence of *la* and *yo*. DeGraff (1992, p. 107) and Lumsden (1989, p. 65) report on some Haitian speakers for whom *la* and *yo* cannot co-occur.⁵ Joseph (1988, p. 201) also reports on some speakers for whom *la* and *yo* cannot co-occur, but he specifically mentions that *la* and *yo* can co-occur for some speakers; he provides examples of these co-occurrences, one of which is reproduced in (11).

(11) Sīriyis kraze tab la yo.

Sīriyis break table DEF PL

‘Sīriyis broke the tables.’

(=(6.6.b) in Joseph, 1988, p. 201)

Some may consider the above data as inconsistencies between speakers. Others, however, may consider these facts as a case of variation between speakers; according to this

⁵ DeGraff (1992, p. 107) and Lumsden (1989, p. 65) independently propose an analysis in which both *la* and *yo* occupy the same position. This analysis cannot hold for speakers who accept the co-occurrence of both morphemes.

view, there would be two slightly different grammars in Haitian. These are schematized in (12).⁶

- (12) a. G₁, where *la* and *yo* can co-occur HAITIAN
 b. G₂, where *la* and *yo* are mutually exclusive

The assumption that the “variation” view is the correct way of interpreting the data raises the following question: how can we be sure that the facts in (12) represent significant patterns of variation in the language?

In my opinion, there are several ways one can make sure to have discovered significant patterns of variation in the language. First, the fact that different authors independently report on the same subsets of data argues that they may represent significant patterns of variation. This is the case for the two grammars in (12), found independently by more than one author.

A second methodological procedure consists in enlarging one’s sample. While doing fieldwork on Haitian and Fongbe, I had recourse to this methodology every time I found variation between a small number of speakers on a given structure. This procedure allowed me to find a limited number (generally two or three) of systematic patterns of variation for a given structure or construction. Examples of such systematic patterns of variation obtained after enlarging my sample of speakers can be found in Lefebvre (1998a, pp. 219–248, for the distribution of the determiner in the clause; pp. 369–370, for the distribution of the determiner in verb doubling constructions) and Lefebvre (forthcoming, for the interpretation of deictic terms).

A third way of finding out whether the variation encountered between speakers of a given language is significant or not is to determine if the same patterns exist in a different (but related) language. For example, assuming that Haitian Creole is historically related to Fongbe — (and other West African languages that were present at the time Haitian Creole was formed), along the lines of Lefebvre (1998a, and the references therein) —, we may ask

⁶ See also d’Ans (1968, p. 105), for further discussion of this point.

worked with, \mathcal{D} and $l\mathcal{E}$ may co-occur. In the literature, however, there is one author who points out that, for some speakers, the determiner \mathcal{D} cannot occur in the environment of $l\mathcal{E}$ (Agbidinoukoun, 1991, p. 149).⁷ This shows that, as is the case in Haitian, there are two slightly different grammars of Fongbe with respect to the possibility of co-occurrence of the definite determiner and the plural marker. These two grammars are represented in (15).

- (15) a. G_1 , where \mathcal{D} and $l\mathcal{E}$ can co-occur FONGBE
 b. G_2 , where \mathcal{D} and $l\mathcal{E}$ are mutually exclusive

The fact that different languages with a similar determiner system present the same type of variation in this system suggests that the variation found in each of these languages taken separately may be significant, regardless of the size of the sample of speakers who manifest what would appear to be a less frequent pattern (in this case, G_2 in both (12b) and (15b)).⁸

There are thus several strategies that one can use to find out whether the variation encountered between the speakers of a given language is significant. In my experience, what may appear at first glance to be inconsistencies between speakers generally turns out to constitute consistent patterns of variation in the language.

The problem of inconsistencies within a given speaker

When a speaker is questioned on the properties of a given structure more than once, (s)he may produce the same judgments the second or the third time. In this case, the speaker is

⁷ This suggests that, if we were to have a larger sample of speakers, we might find speakers for whom \mathcal{D} and $l\mathcal{E}$ cannot co-occur.

⁸ Some scholars might want to argue that significant patterns of variation require detailed sociolinguistic studies. While I would agree with such a statement at some level of analysis, the discussion in this section is set within the framework of another paradigm. I would say that the slightly different grammars in (12) and (15) reveal significant patterns of variation between speakers. These grammatical patterns may be considered as a “variable” (in the sense of Labov, 1972a) for community studies.

considered to have consistent judgments. It may happen, however, that a speaker will not have the same judgments the second or the third time. This situation has been used by some authors to illustrate the weaknesses of databases that rely on grammaticality judgments. For example, Labov writes: "Lack of consistency on repeated testing has been among the most serious weaknesses of the claims for idiosyncratic dialects" (1975, p. 27). As part of the methodology I have established for my fieldwork, I always test the grammaticality judgments of each speaker at least twice (sometimes more) for a given construction. Over and over again, I have found that speakers were quite consistent in their grammaticality judgments; that is, they generally had the same evaluation of the data on the repeated testing.

There are a few cases, however, where I kept getting slightly different interpretations for the same sentences from session to session. In such cases, there are two options. The first one is to conclude that the speaker is inconsistent, and sometimes, this is the right conclusion. The second one is to try to find out why there is this variation in individual speakers' judgments and to use this variation as a trigger for a new way of looking at the data. For example, while I was working on the Haitian clausal determiner (Lefebvre, 1998b), I kept registering slightly different interpretations of the data from the same speaker from one session to the next. My first evaluation of this situation was that the speaker was inconsistent in his judgments. I then started to work with other speakers. To my surprise, I found the same so-called inconsistencies with these other speakers as well. At this point, I decided to look at the data from a different point of view and to consider that there was some order in what first looked like inconsistencies. This turned out to be a quite successful methodology, for it led me to propose that the Haitian clausal determiner can occupy more than one position in the syntactic tree. Each position that can host this lexical item determines its scope; this explains the slightly different interpretations of the sentences containing the clausal determiner. Had I abandoned the topic at an early stage of my investigation, I would have missed one very

important property of the clausal determiner, that of being able to occur in more than one syntactic position in the clause structure.⁹

Native speakers as informants

It is worth noting that being a native speaker of a given language does not necessarily confer one with clear judgments on one's grammar. Some people are naturally good informants, while other people are not. Like in other areas of life, being a good informant — that is, having clear judgments on one's grammar — is a matter of talent or skill, regardless of the language under consideration. Researchers have to be able to identify speakers with such a skill, whether they work on a creole language or not.

In the case of creole communities, the situation is extremely complex for creole native speakers may speak different varieties that are quite far apart from one another. For example, the Haitian varieties referred to as basilectal, mesolectal, and acrolectal do not always present the same data (Valdman, 1978, pp. 292–295). Speakers of different varieties may have different judgments on subsets of data. On the basis of findings on varieties of English, Labov (1972a) has convincingly shown that speakers of a given variety do not have direct access to the grammar of speakers of the other varieties. Differences between grammaticality judgments among Haitian speakers may be attributable to this dimension of variation. It is possible that the variation found in the distribution of *ki*, discussed earlier (see (7) and (8)), is an example of such a type of variation. (For another interpretation of this variation, see Déprez, 1994.) Furthermore, there is variation between speakers of the same social group, as community studies have shown us over and over again (e.g., Labov, 1972a, 1972b). Community studies on Haitian Creole will most certainly reveal examples of this type of variation. Finally, there is

⁹ See also Cowart (1998), for the elaboration of a methodology that can tell us whether the variation within a given speaker is significant or not.

regional variation, a topic that has received some attention in the literature on Haitian Creole (e.g., Fattier, 1998; Orjala, 1970; Tinelli, 1970). A classic example of this type of variation in Haitian Creole involves the manifestation of case on the possessor phrase in the nominal structure. While the northern part of Haiti uses a prenominal possessive particle, the central and southern parts of the country have a phonologically null case marker. This is shown in (16), from Lumsden and Lefebvre (1994, p. 109), based on e.g., Goodman (1964), Lumsden (1991), and Orjala (1970, p. 36).

- (16) a. liv a Jan NORTHERN HAITIAN
 book PART John
 ‘John’s book’
- b. liv Jan \emptyset CENTRAL AND SOUTHERN HAITIAN
 book John GEN
 ‘John’s book’ (=(1) in Lumsden & Lefebvre, 1994)

Given the data in (16), it is to be expected that speakers of Haitian will have different grammaticality judgments with respect to case marking of the possessor phrase.

Given all these dimensions of potential variation in the data, it is to be expected that data provided by some speakers may, in some cases, contradict those of other speakers of the language. Researchers have to take variation into account when they are collecting their data. Moreover, their choice of basilectal, mesolectal, or acrolectal speakers as informants is determined by their research questions, as mentioned in the first section of this column.

Can data on a given language ever be complete?

Throughout this column, we have seen that data collection is not a neutral activity and that the type of data researchers gather is not independent from their research paradigms and questions. Establishing data on a given language is thus a collective enterprise. This is how with time data on a given language and patterns of variation in a language get established. Can

data on a given language ever be complete? In the light of the discussion in this column, the answer to this question cannot be positive. The task of establishing the data on any language is an endless one.

References

- Agbidinokoun, C. C. (1991). *Analyse contrastive des syntagmes nominaux du fongbe et du français*. Doctoral dissertation, Université Paris III (Sorbonne Nouvelle).
- Ans, A.-M. d'. (1968). *Le créole français d'Haïti*. The Hague: Mouton de Gruyter.
- Aoun, J. (1992). The search for facts. *Textes des séances plénières, 3e circulaire, XV^e Congrès international des linguistes* (pp. 77–88). Quebec: Université Laval.
- Avolonto, A. (1992). *De l'étude sémantico-syntaxique des marqueurs préverbaux à la structure de la phrase en fɔngbè*. Master's thesis, Université du Québec à Montréal.
- Baker, P., & Corne, C. (1982). *Isle de France Creole*. Ann Arbor, MI: Karoma.
- Bickerton, D. (1996). Why serial verb construction in Isle de France Creole can have subjects: A reply to Corne, Coleman and Curnow. In P. Baker & A. Sycia (Eds.), *Changing meanings, changing functions* (Westminster Creolistics Series 2) (pp. 155–169). London: University of Westminster Press.
- Bloomfield, L. (1887–1949). *A Leonard Bloomfield anthology*, edited by C. F. Hockett (1970) (Indiana University Studies in the History and Theory of Linguistics). Bloomington: Indiana University Press.
- Botha, R. P. (1976). "Gut feelings" in generative grammar. Bloomington: Indiana University Linguistics Club (Brochure).
- Bouquiaux, L., & Thomas, J. M. C. (1992). *Studying and describing unwritten languages*. Dallas, TX: Summer Institute of Linguistics.
- Brousseau, A.-M., & Lumsden, J. S. (1992). Nominal structures in Fongbe. *The Journal of West African Languages*, 22(1), 5–26.
- Burling, R. (1984). *Learning a field language*. Ann Arbor, MI: University of Michigan Press.
- Campbell, L. (1985). *The Pipil language of El Salvador* (Mouton Grammar Library 1). Berlin/New York: Mouton de Gruyter.
- Carden, G., & Stewart, W. A. (1988). Binding theory, bioprogram, and creolisation: Evidence from Haitian Creole. *Journal of Pidgin and Creole Languages*, 3(1), 1–68.
- Chansou, M. (1997). The methodology of corpus building. *Terminologies nouvelles*, 16, 33–36.
- Chaudenson, R. (1977). Toward the reconstruction of the social matrix of creole language. In A. Valdman (Ed.), *Pidgin and Creole linguistics* (pp. 259–277). Bloomington/London: Indiana University Press.
- Cowart, W. (1998). *Experimental syntax. Applying objective methods to sentence judgments*. Beverly Hills/London: Sage.
- DeGraff, M. A. F. (1992). *Creole grammars and acquisition of syntax: The case of Haitian*. Doctoral dissertation, University of Pennsylvania, Philadelphia.

- DeGraff, M. A. F. (1993). A riddle on negation in Haitian. *Probus*, 5, 63–93.
- DeGraff, M. A. F. (2000). Empirical quicksand: Probing two recent articles on Haitian Creole. *Journal of Pidgin and Creole Languages*.
- De Guiraudon, T. G. (1894). *Manuel de la langue foute*. Paris/Leipzig: H. Welter.
- Dejean, Y. (1993). Comments from M. Yves Dejean on the articles published in *Travaux de recherche sur le créole haïtien* (Addressed to me by the author), Université du Québec à Montréal.
- Déprez, V. (1992a). Is Haitian Creole really a pro-drop language? *Travaux de recherche sur le créole haïtien* (Université du Québec à Montréal), 11, 23–40.
- Déprez, V. (1992b). Raising constructions in Haitian Creole. *Natural Language & Linguistic Theory*, 10(2), 191–231.
- Déprez, V. (1994). Haitian Creole: A pro-drop language? *Journal of Pidgin and Creole Languages*, 9(1), 1–24.
- Déprez, V. (1997). Two types of negative concord. *Probus*, 9(2), 103–142.
- Donaldson, B. C. (1993). *A grammar of Afrikaans* (Mouton Grammar Library 8). Berlin/New York: Mouton de Gruyter.
- Donohue, M. (1999). *A grammar of Tukang Besi* (Mouton Grammar Library 20). Berlin/New York: Mouton de Gruyter.
- Doran, M., Drapeau, L., & Lefebvre, C. (1982). Le projet Centre-Sud: Corpus et méthodologie. In C. Lefebvre (Ed.), *La syntaxe comparée du français standard et populaire: Approches formelle et fonctionnelle* (pp. 73–125). Montreal: Office de la langue française.
- Étienne, G. (1974). *Le créole du nord d'Haïti: Étude des niveaux de structure*. Doctoral dissertation, Université des sciences humaines de Strasbourg.
- Faine, J. (1937). *Philologie créole: études historiques et étymologiques sur la langue créole d'Haïti*. Port-au-Prince: Imprimerie de l'État.
- Fattier, D. (1998). *Contribution à l'étude de la genèse d'un créole: Atlas linguistique d'Haïti, cartes et commentaires*. Doctoral dissertation, Université de Provence, Marseilles.
- Férère, G. A. (1974). *Haitian Creole sound-system, form-classes, texts*. Doctoral dissertation, University of Pennsylvania, Philadelphia.
- Fournier, R. (1977). *N ap fè yun ti-kose su la (La grammaire de la particule la en créole haïtien)*. Master's thesis, Université du Québec à Montréal.
- Gacic, M. (1994). The methodology of compiling a terminological dictionary. *Filologija*, 22–23, 375–386.
- Goddard, Y. (1987). Leonard Bloomfield's descriptive and comparative studies of Algonquian. *Historiographia Linguistica*, 14(1–2), 179–217.

- Goodman, M. F. (1964). *A comparative study of Creole French dialects*. The Hague: Mouton de Gruyter.
- Hall, R. A. (1953). *Haitian Creole: Grammar, texts and vocabulary*. Menasha, WI: Georges Banta.
- Isquierdo, A. N. (1998). The issue of the constitution of the corpus: Limits and possibilities. *Estudos Linguisticos*, 27, 272–277.
- Issoufi, O. A. (1998). L'enquête sur les pratiques et les représentations linguistiques. *Cahiers de l'Institut de linguistique et des sciences du langage* (Université de Lausanne), 10, 141–154.
- Joseph, F. (1988). *La détermination nominale en créole haïtien*. Doctoral dissertation, Université Paris VII (Denis Diderot).
- Kibrik, A. K. (1977). *The methodology of field investigation in linguistics*. The Hague: Mouton de Gruyter.
- Koopman, H. (1982). Les constructions relatives. In C. Lefebvre, H. Magloire-Holly & N. Piou (Eds.), *Syntaxe de l'haïtien* (pp. 167–203). Ann Arbor, MI: Karoma.
- Kouwenberg, S. (1994). *A grammar of Berbice Dutch Creole* (Mouton Grammar Library 12). Berlin/New York: Mouton de Gruyter.
- Kuo, C. H. (1997). A corpus-based approach to linguistic analysis: Applications and features. *Studies in English Literature and Linguistics*, 23, 113–125.
- Labov, W. (1972a). *Sociolinguistic patterns*. Philadelphia, PA: University of Pennsylvania Press.
- Labov, W. (1972b). *Language in the Inner City: Studies in the Black English vernacular*. Philadelphia, PA: University of Pennsylvania Press.
- Labov, W. (1975). *What is a linguistic fact?* Lisse: The Peter De Ridder Press.
- Law, P. (1992). What is a null subject language? *Travaux de recherche sur le créole haïtien* (Université du Québec à Montréal), 11, 41–58.
- Law, P. (1994). *Resumptives in Haitian and relexification*. Paper presented at *The MIT Symposium on the role of relexification in creole genesis: The case of Haitian Creole*.
- Lefebvre, C. (1976). Linguistic survey of Cuzco Quechua: Sampling procedures and data collection. *Anthropological Linguistics*, 18(7), 328–339.
- Lefebvre, C. (1982). L'expansion d'une catégorie grammaticale □ Le déterminant *la*. In C. Lefebvre, H. Magloire-Holly & N. Piou (Eds.), *Syntaxe de l'haïtien* (pp. 21–63). Ann Arbor, MI: Karoma.
- Lefebvre, C. (1990). On the interpretation of predicate cleft. *Linguistic Review*, 6, 169–194.
- Lefebvre, C. (1996). The tense, mood, and aspect system of Haitian Creole and the problem of transmission of grammar in creole genesis. *Journal of Pidgin and Creole Languages*, 11(2), 231–311.

- Lefebvre, C. (1997). Relexification in creole genesis: The case of demonstrative terms in Haitian Creole. *Journal of Pidgin and Creole Languages*, 12(2), 181–201.
- Lefebvre, C. (1998a). *Creole genesis and the acquisition of grammar: The case of Haitian Creole* (Cambridge Studies in Linguistics 88). Cambridge: Cambridge University Press.
- Lefebvre, C. (1998b). Multifunctionality and variation among grammars: The case of the determiner in Haitian and in Fongbe. *Journal of Pidgin and Creole Languages*, 13(1), 93–150.
- Lefebvre, C. (2000). On the empirical reliability of some Haitian data. *Journal of Pidgin and Creole Languages*.
- Lefebvre, C. (forthcoming). The interplay of relexification and levelling in creole genesis and development.
- Lefebvre, C., & Fournier, R. (1976). Le corpus du créole haïtien à Montréal. Unpublished corpus, Université du Québec à Montréal.
- Lefebvre, C., & Lumsden, J. S. (1994). Le rôle central de la relexification dans la genèse des langues créoles. *Plurilinguismes*, 8, 47–93.
- Lefebvre, C., & Massam, D. (1988). Haitian Creole syntax: A case for DET as head. *Journal of Pidgin and Creole Languages*, 3(2), 213–243.
- Levin, B. (1989). Towards a lexical organization of English verbs. Unpublished paper, Northwestern University, Evanston, IL.
- Longacre, R. E. (1964). *Grammar discovery procedures: A field manual*. The Hague: Mouton de Gruyter.
- Lumsden, J. S. (1989). On the distribution of determiners in Haitian Creole. *Revue québécoise de linguistique*, 18(2), 64–93.
- Lumsden, J. S. (1990). The bi-clausal structure of Haitian clefts. *Linguistics*, 28, 741–760.
- Lumsden, J. S. (1991). On the acquisition of nominal structures in the genesis of Haitian Creole. *Travaux de recherche sur le créole haïtien* (Université du Québec à Montréal), 4, 37–61.
- Lumsden, J. S. (in press). Aspect and lexical semantics in Haitian Creole. *The Linguistic Review*.
- Lumsden, J. S., & Lefebvre, C. (1994). The genesis of Haitian Creole. In C. Lefebvre & J. S. Lumsden (Eds.), *The central role of relexification in creole genesis: The case of Haitian Creole* (Research report prepared for SSHRCC on the project *La genèse du créole haïtien*—un cas particulier d'investigation sur la forme de la grammaire universelle, Université du Québec à Montréal).
- McDaniel, D., McKee, C., & Cairns, H. S. (Eds.). (1996). *Methods for assessing children's syntax*. Cambridge, MA: The MIT Press.

- Merlan, F. C. (1994). *A grammar of Wardaman* (Mouton Grammar Library 11). Berlin/New York: Mouton de Gruyter.
- Mulder, J. (1996). Methodology and description. *Linguistique*, 32(1), 17–34.
- Muysken, P. (1995). Studying variation in older texts: Negerhollands. *Journal of Pidgin and Creole Languages*, 10(2), 335–347.
- Nagao, M. (1994). A methodology for the construction of a terminology dictionary. In B. T. S. Atkins & A. Zampolli (Eds.), *Computational approaches to the lexicon* (pp. 397–441). Oxford: Oxford University Press.
- Orjala, P. R. (1970). *A dialect survey of Haitian Creole*. Doctoral dissertation, Hartford Seminary Foundation, CT.
- Payne, T. E. (1997). *Describing morphosyntax: A guide for field linguists*. Cambridge: Cambridge University Press.
- Pompilus, P. (1976). *Contribution à l'étude comparée du créole et du français à partir du créole haïtien*. Morphologie et syntaxe. Port-au-Prince: Éditions Caraïbes.
- Popper, K. R. (1972). *Objective knowledge*. Oxford: Oxford University Press.
- Press, I. J. (1986). *A grammar of modern Breton* (Mouton Grammar Library 2). Berlin/New York: Mouton de Gruyter.
- Rice, K. D. (1989). *A grammar of Slave* (Mouton Grammar Library 5). Berlin/New York: Mouton de Gruyter.
- Ritter, E. (1992). Cross-linguistic evidence for number phrase. *The Canadian Journal of Linguistics*, 37(2), 197–219.
- Samarin, W. J. (1967). *Field linguistics: A guide to linguistic field work*. New York: Holt, Rinehart and Winston.
- Sankoff, G. (Ed.). (1980). *The social life of language*. Philadelphia, PA: University of Pennsylvania Press.
- Schutze, C. T. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.
- Senft, G. (1986). *Kilivila. The language of the Trobriand Islanders* (Mouton Grammar Library 3). Berlin/New York: Mouton de Gruyter.
- Sterlin, M.-D. (1988). Les différentes caractéristiques de *pou* en créole haïtien. *Travaux de recherche sur le créole haïtien* (Université du Québec à Montréal), 3, 1–34.
- Sylvain, S. (1936). *Le créole haïtien : morphologie et syntaxe*. Wetteren, Belgium: Imprimerie De Meester/Port-au-Prince: By the author.
- Thoiron, P., Iwaz, J., & Zaouche, N. (1997). The methodology of analysis of results of surveys on the implantation of the official terminology. *Terminologies nouvelles*, 16, 48–54.

- Tinelli, H. (1970). *Generative phonology of Haitian Creole*. Doctoral dissertation, University of Michigan, Ann Arbor, MI.
- Valdman, A. (1970). *Basic course in Haitian Creole*. The Hague: Mouton de Gruyter.
- Valdman, A. (1978). *Le créole : Structure, statut et origine*. Paris: Klincksieck.
- Valdman, A. (1996). *A learner's dictionary of Haitian Creole*. Bloomington, IN: Indiana University Creole Institute.
- Valdman, A. *et al.* (1981). *Haitian Creole-English-French dictionary*. Bloomington, IN: Indiana University Creole Institute.
- Vaux, B., & Cooper, J. (1999). *Introduction to linguistic field methods*. Harvard: Lincom Europa.
- Weinreich, U. (1984). Lexicographic definition in description semantics. *Alpha*, 28, 103–118.