THEORETICAL ESSAY

“Natural” and “elicited” language data: why we don’t need these terms

Kristina BALYKOVA
University of Texas at Austin (UT)

ABSTRACT

Language documentation discourse commonly divides language data into two large types: natural(istic) vs. elicited. The goal of this paper is to put this dichotomy under critical scrutiny. By examining key publications on linguistic fieldwork, I show that the two terms seldom receive any clear definition and are often used inconsistently, giving rise to evident contradictions. The analysis reveals that the terms are typically distinguished by two parameters – linguistic unit (texts vs. not texts) and method of obtaining the data (controlled vs. uncontrolled) – but the distinction is virtually never thoroughly maintained. I argue that the dichotomy natural(istic) vs. elicited is insufficient to capture the complexity of possible scenarios and forms under which language is produced. Building upon previous literature, I propose a more detailed classification of language data, which abandons the notion of ‘natural(istic)’ and ‘elicited’ altogether. The paper concludes by discussing the gains of a more careful reflection on language data types.

RESUMO

O discurso da documentação linguística comumente divide os dados linguísticos em dois grandes tipos: naturais (ou naturalísticos) vs. elicitados. O objetivo deste artigo é pôr essa dicotomia sob escrutínio crítico. Ao examinar publicações importantes sobre trabalho de campo linguístico, demonstro que os dois termos raramente recebem uma definição clara e são frequentemente usados de forma inconsistente, dando origem a contradições evidentes. A análise revela que os termos são tipicamente distinguidos por dois parâmetros – unidade linguística (textos vs. não textos) e método de obtenção de dados (controlado vs. não controlado) – mas a distinção quase nunca é rigorosamente mantida. Argumento que a dicotomia natural (ou

OPEN ACCESS

EDITED BY
- Maria Filomena Spatti Sandalo (Unicamp)

EVALUATED BY
- Karin Camolese Vivanco (Unicamp)

DATES
- Received: 14/10/2025
- Accepted: 21/12/2025
- Published: 16/06/2026

HOW TO CITE
Balykova, Kristina. (2026). “Natural” and “elicited” language data: why we don’t need these terms. *Revista da Abralin*, v. 24, n. 2, p. 22-39, 2025.

naturalístico) vs. elicitado é insuficiente para capturar a complexidade dos possíveis cenários e formas sob os quais a linguagem é produzida. Com base na literatura prévia, proponho uma classificação mais detalhada dos dados linguísticos, que abandona por completo a noção de 'natural (ou naturalístico)' e 'elicitado'. O artigo conclui discutindo os ganhos de uma reflexão mais cuidadosa sobre os tipos de dados linguísticos.

KEYWORDS
Language documentation. Natural data. Elicitation.

PALAVRAS-CHAVE
Documentação linguística. Dados naturais. Elicitação.

RESUMO PARA NÃO ESPECIALISTAS
No trabalho com línguas ameaçadas, normalmente supõe-se que podemos coletar dois grandes tipos de dados. Os dados naturais (ou naturalísticos) seriam aqueles que se assemelham à língua falada na vida real. Normalmente, seriam textos (como histórias pessoais) e sofreriam pouco controle por parte do pesquisador. Já os dados elicitados seriam aqueles forçados à existência pelo pesquisador, como, por exemplo, as frases em português que o pesquisador pede para o falante traduzir para a língua indígena. Neste artigo, eu defendo que a realidade da documentação linguística é mais complicada do que isso e que a dicotomia entre 'natural' e 'elicitado' não faz jus a essa complexidade. Argumento que precisamos de uma nova classificação de dados linguísticos, na qual os termos 'natural' e 'elicitado' sequer deveriam ter um papel.

Introduction

Collection of natural language data and elicitation are two commonly recognized pillars of linguistic fieldwork, albeit they do not share the same status. Elicitation is seen, at best, as a necessary tool that should be handled with great caution and several reservations. The true core of language documentation lies in the collection of natural language data, which allow the researcher to apprehend the *real* language in use.

These considerations appear to be uncontroversial and even commonplace until we stop for a moment and ask: what is the exact difference between 'natural' and 'elicited'? Surprisingly, no one seems to have come up with a coherent answer.

In the last 30 years, many publications (manuals, book chapters, journal articles) have discussed best practices in linguistic fieldwork and, in particular, best ways to collect reliable language data. All of them have touched on the question of 'natural' and 'elicited' data, but few have proposed clear definitions of these terms. Even when such definitions are present, they are often contradicted in other parts of the same work.

Furthermore, as becomes clear from both examining these sources and conducting actual linguistic fieldwork, the dichotomy 'natural' vs. 'elicited' fails fundamentally in capturing the diversity of language documentation scenarios, and its very existence seems unjustified. The goal of this paper is to defend this point.

In Section 1 and Section 2, I analyze the interpretations of the terms 'natural' and 'elicited', respectively, found in the literature on linguistic fieldwork. In Section 3, I discuss other terms that can complement and, most importantly, replace the dichotomy 'natural' vs. 'elicited' in classification of language data types. Section 4 is dedicated to the advantages of a more sophisticated way of looking at different language data types. Section 5 contains the conclusions.

1. What is natural?

The term 'natural' is ubiquitously used to describe certain types of language data. For example, Aikhenvald (2007, p. 9) argues that data coming from "video stimuli, picture stories, and focused elicitation through the language itself" are secondary to "natural texts and conversations". Everett (2001, p. 185) advises: "The linguist also wants to be sure that the data collected is natural", and Himmelmann (1998, p. 162) admits that "Naturalness of the data is one of the methodological issues in the data collection".

It is, therefore, ever more remarkable that the term receives no clear definition in most literature on the topic. Although the meaning of this term might appear obvious and even *natural*, it turns out that different authors can mean different things by it. There seem to exist two main interpretations to the term 'natural' in language documentation discussions: one associates 'natural' data with texts, whereas the other one focuses on the (relative) lack of the researcher's influence over 'natural' data.

The first view can be illustrated by the following citations: "Looking for **natural discourse** is at the heart of the matter. Any serious investigation into the syntax and pragmatics of a language should involve the collection of **a corpus of oral or written texts**"[1] (Dimmendaal, 2001, p. 71), and "By **text collection** I am referring to the practice of compiling and analyzing **naturally occurring speech and narratives** in the language under study" (Chelliah, 2001, p. 152). The idea that natural data are the same as 'texts' has its counterpart in the idea that elicited data are the same as 'sentences'. Although this dichotomy is a very common rule of thumb in classifying different types of language data,

[1] Here and throughout the paper the bold is mine.

it is completely contradicted by the diversity of scenarios in which language can be produced and documented (see Section 2 below)[2].

The second view, that of 'natural' data being uninfluenced by the researcher, is best represented in Himmelmann (1998, p. 184-185). Arguably, it is also the most detailed discussion of what 'natural' means in language documentation

> For evaluating the quality of data, it may again be useful to sketch a typology of communicative events with respect to their "naturalness." The basic parameter of such a typology is the degree to which speakers are linguistically self-aware, ranging from complete unawareness to paying full attention to linguistic form as in a metalinguistic evaluation of a given form or construction. Linguistic self-awareness on the part of the contributors is influenced in various ways by the compiler(s) of a language documentation.

Chelliah; de Reuse (2011, p. 423) express a similar view: "Data from naturally occurring speech is reliable in that they have not been corrupted by priming or by other translation or elicitation effects, since speakers concentrate on the stories rather than on the constructions they are producing". Thus, a speaker who perceives that the linguistic form of their utterances is the focus of the researcher's interest and attention cannot use their language 'naturally'. This idea is, in fact, related to the idea that 'natural' data are texts, given that the longer a speech event is, the greater the chances are for the speaker to lower their guard and decrease their linguistic self-awareness.

Another aspect of the same idea is that 'natural' data are in no way prompted by the researcher's suggestions or questions. 'Natural' speech just happens to be produced in a normal flux of life, hence the use of the expressions, such as "everyday, natural speech" by Himmelmann (2012, p. 203) and "speech in its natural function, in spontaneous interaction among speakers" by Mithun (2001, p. 51)[3]. And here lies the main paradox of language documentation: truly 'natural' speech does not happen while language documentation is carried out.

The observer's paradox is normally derived from the inevitable status of the researcher as an outsider, a foreign presence difficult to ignore. Another common consideration is that speakers tend to adjust their speech to the researcher's level of understanding. But even if one imagines a researcher integrated into the community as much as possible, casual speaking simply does not involve directing audio recorders and video cameras at one another. The core of this unavoidable contradiction is best expressed in Himmelmann (1998, p. 185), who proposes different types of communicative events and states the following:

> *Natural communicative events*: communicative events unaffected by any external interference into the conventional communicative routines of the participants. Such events are, in principle, not amenable to documentation since the documentation process itself constitutes an extraordinary factor in the communicative situation.

---

[2] In a broader terminological sense, this dichotomy also fails to consider that "'text' includes not only narratives and traditional stories, but any piece of language which has been produced by a native speaker" (Bowern, 2008, p. 227).

[3] Note that the term 'spontaneous' is used as opposed to 'planned', and not precisely 'unnatural', in Himmelmann (1998) and Bowern (2008).

Interestingly, even if it is not normally discussed that collecting truly 'natural' language data is a self-contradictory enterprise, many authors implicitly agree with this idea. One piece of evidence is that the term 'natural' is often used as referring to a gradient property, rather than a binary one. Here are some examples from the relevant literature:

- "[l]anguage used in **more natural** settings such as narratives, conversations and procedural texts" (Meakins; Green; Turpin, 2018, p. 160),
- "discourse data (i.e., **more or less natural** data documenting linguistic behavior)" (Himmelmann, 2012, p. 202),
- "With free narrative collecting, it is important to get **as natural** a response from speakers **as possible**." (Chelliah; de Reuse, 2011, p. 428),
- "Eliciting **natural-sounding** conversations is difficult." (Chelliah; de Reuse, 2011, p. 429),
- "to ensure **some amount of naturalness** in the conversation, I usually left the room after I turned the tape recorder on" (Chelliah, 2001, p. 155).

Thus, language documentation is not a matter of collecting either 'natural' or 'unnatural' data, but rather that of navigating communicative events which vary as to their level of 'naturalness'.

Another piece of evidence for implicit recognition of the 'naturalness' paradox is the widespread use of the term 'naturalistic'. In fact, it is as common as the term 'natural' and clearly preferred by some authors. In other cases, both terms are used interchangeably within the same work. I have not come across any publication in which both terms are used but the authors clarify any conceptual difference between them. For example, Himmelmann (2012) gives Section 3.2 the title "On dividing resources between work on **naturalistic** data and grammar-targeted elicitation", but only uses the term 'natural' and its derivations elsewhere.

The popularity of the term 'naturalistic' in the literature on linguistic fieldwork may be explained by its semantics. By dictionary definitions, 'naturalistic' is 'similar to what exists in nature'[4]. Thus, using the term 'naturalistic' is a way to recognize that data of this sort *look like* natural speech produced in the absence of the researcher but, ultimately, *are not* natural speech. The question, however, is to what extent and in what manner language data should resemble the ideal 'natural' data in order to be considered 'naturalistic'. As discussed above, it seems that for most scholars the measure is the degree of the researcher's control over the form and content of a given speech event. Thus, the term 'naturalistic' is associated with the method(s) of obtaining data, but at the same time completely obscures the methodological side of language documentation. First, because the term

[4] https://dictionary.cambridge.org/dictionary/english/naturalistic.

'naturalistic' refers to what the data are considered to be in their essence, not to what the researcher did to obtain them. Second, because it completely erases the gradable nature of the researcher's influence.

These considerations allow us to challenge the status of 'natural' data as an objectively real pure entity which should be grasped by the researcher. 'Natural' is a complex term loaded with several connotations and coexisting with the tacit agreement on the practical unattainability of truly 'natural' data. Furthermore, the term 'natural(istic)' appears to serve as a blanket term, under which various methods of language documentation hide without proper characterization. After having put 'natural' data into question, the next move is to do the same with its commonly assumed opposite, elicitation.

2. What is elicitation?

The first observation to be made about the term 'elicitation' is that, despite being well established in language documentation discourse, it is hardly felicitous. Given that the general dictionary meaning of the verb 'elicit' is 'to obtain something, especially information or a reaction'[5], virtually any recording session qualifies as 'elicitation' in this sense. Below, I cite a few excerpts from the relevant literature, where the authors describe collecting what one might call 'natural(istic)' data, but use either the term 'elicit' itself or other vocabulary compatible with the general sense of the term, i.e., 'to seek information from the other':

- "The personal narrative is easy to collect. **It can be elicited by asking a speaker to talk about something** exciting that has happened to them." (Chelliah; de Reuse, 2011, p. 428),
- "There are several ways to **elicit texts**. I have never had a problem **getting people to tell stories** when the speakers have been fluent." (Bowern, 2008, p.116),
- "[...] I found it difficult to **elicit spontaneous textual data**, so my description was based for the most part on data translated from English prompts." (Crowley, 2007, p. 183),
- "**Tell consultants that you'd like them to tell a short story** in the language [...]" (Dixon, 2007, p. 22),
- "A **request** for the word for 'tree' can be less daunting than a **request** for an eloquent speech to a microphone." (Mithun, 2001, p. 36).

[5] https://dictionary.cambridge.org/dictionary/english/elicit

Hence, it seems reasonable to conclude that "[b]asically, the meaning of "elicitation" boils down to "data collection", or even "whatever one does to get the consultant to say something" (Chelliah; de Reuse, 2011, p. 361).

Most linguists, however, would argue that the term 'elicitation' has a more specific meaning in language documentation. What is it, then? It turns out, in fact, that the interpretations of the term 'elicitation' found in the literature on linguistic fieldwork are a mirror image of those already discussed for the term 'natural(istic)'. Implicitly, elicitation is often associated with the collection of words, phrases and sentences, rather than texts. As an explicit definition, it is often stated that elicited data are in some way influenced by the researcher. I will now analyze the two interpretations.

Although none of the authors I have checked for this research states explicitly that texts are never elicited, this appears implied from the way the argument is often laid out, especially when the two terms are placed side by side in coordinated structures suggesting a contrast between the two. Below, some examples are given:

- "[...] we provide steps in how to develop an audio-visual corpus through the collection of a range of linguistic data including **texts** (narrative, procedural, conversation etc.) and **elicitation**." (Meakins; Green; Turpin, 2018, p. 7),

- "Typically, there will be a lot of **elicitation** and little **text collection** at the beginning, but the proportions will be reversed – i.e. little **elicitation**, and a lot of **text collection** – at the end." (Chelliah; de Reuse, 2011, p. 359),

- "[...] one should base the study on **texts** and on participant observation, but this must always be augmented by judicious **elicitation** in the language, to fill in gaps and also to check generalisations" (Dixon, 2007, p. 24),

- "While various strategies of direct elicitation can certainly help in the search for new vocabulary, it is extremely important that **elicitation** be supplemented by an extensive collection of **spoken texts**." (Crowley, 2007, p. 108),

- "In the initial phase of fieldwork, the compiler will be able to handle only **elicited materials** and simple **texts**." (Himmelmann, 1998, p. 171-172).

However, the idea that texts and elicitation are necessarily different and mutually exclusive types of data does not hold up to scrutiny. Firstly, words and sentences are, of course, not always elicited. They can be produced without any direct eliciting, as when the speaker thinks of a word or even a list of related words that might be unknown to the researcher or utters a comment provoked by something remarkable happening around.

The dichotomy breaks down further when we consider that texts themselves are often elicited, which forces us to examine the second interpretation of the term 'elicitation': data produced under some form of the researcher's influence. Sometimes, this influence is defined quite vaguely. For example, Hyman (2001, p. 18) argues that "in elicitation, the researcher necessarily plays **an active role** in generating the data." Meakins; Green; Turpin (2018, p. 122) observe that "elicited data is often not considered 'real' language because it is **constructed** and relatively context free". Himmelmann (1998, p. 185) defines the investigator's control over the interaction as a characteristic feature of language elicitation:

> The mere presence of a person known to be investigating linguistic behavior may already have some influence on the contributors' linguistic behavior. This influence increases in accordance with the degree to which the investigator dominates or controls the interaction between the contributors and him- or herself. The control is particularly strong in the case of communicative events that have been "invented" for research purposes. The best-known communicative event of this kind is the (scientific) interview, the linguistic variant of which is called *elicitation*.

Other definitions delineate specific practices which qualify as 'elicitation'. Often, elicitation is understood as a method of data collection based on the use of some auxiliary material prepared or selected by the researcher beforehand. Such understanding can be seen in the following statements:

- "[...] "elicitation (formal) – the act of obtaining language data from another person which involves the use of questionnaires, stimuli or translation equivalents." (Meakins; Green; Turpin, 2018, p. 316),

- "[...] grammatical elicitation – going through a battery of sentences in the lingua franca and asking for their translation into the native language" (Dixon, 2007, p. 23),

- "[...] by elicitation I mean either the use in language analysis of native-speaker intuitions or translations of decontextualized utterances from a contact language to the language being studied." (Chelliah, 2001, p. 152).

Hence, for most linguists, elicitation involves a certain level of the researcher's control over the emergence and the unfolding of a speech event, and the control often manifests itself in the usage of prompts for language production. No element of this definition suggests that texts cannot be elicited. In fact, many authors recognize elicitation as a valid method for collecting texts, even if they use different terminology. For example, Chelliah; de Reuse (2011, p. 427) argue that narratives can be obtained in a controlled activity, "where the fieldworker can predict something about the lexical or grammatical content of the resulting text, because the prompt has been carefully prepared to elicit a specific type of text." Bowern (2008, p. 123) uses an even more explicit term, "manufacturing":

> If you cannot record naturalistic discourse data, there are some ways to manufacture it. Of course, using manufactured data is not the same as spontaneous speech, and should be a last resort. For example, your consultants could translate dialogues about some subject. This is still planned speech, but it is more likely to have topic chaining and other features than elicited speech.

Note that the author considers the manufactured texts as naturalistic (albeit, planned) data, and not elicited data.

In general, the idea that texts can be elicited with help of prompts is not, of course, a novelty to linguists, given the common use of storyboards for text collection[6]. A more radical idea is that traditional narratives and conversations are also often elicited. In this case, the researcher's control is carried out not directly through prompts but in a more subtle way, i.e., by the very forcing into existence of a speech event that might not have happened had the researcher not asked for it.

Mannheim; van Vleet (1998)[7] explore the differences between two versions of a pan-Andean traditional story that features the flooding of a lake. The first version, classified as "conversational narrative", was told during a conversation between van Vleet and the storyteller about their travel plans. The second version was "formally elicited" by Mannheim. Although the authors do not clarify what they mean by "formally elicited", the speaker was probably asked to narrate "the flooded city story". The authors point out the structural peculiarities of the second version: "it is less detailed, uses less reported speech, and is less closely connected to the specific circumstances in which it was told" (Mannheim; van Vleet, 1998, p. 334).

Thus, researchers who are particularly tuned to the ethnographic context of speech production may consider 'formal elicitation' what for others would be a textbook example of 'natural' text collection. In summary, the question of what 'elicitation' is seems to reduce to the question of where a particular researcher feels necessary to draw the line between elicited and not elicited data.

Ultimately, the term 'elicitation' appears as vague for describing language data as the term 'natural' discussed above. The next logical question is: do we need them at all?

3. Beyond natural and elicited

It should become clear by now that the dichotomy 'natural(istic)' vs. 'elicited' is too simple to adequately capture the diversity of possible scenarios and forms under which language is produced. A crucial problem with this dichotomy is that it implicitly works with two parameters which can be very schematically summarized as *language unit* (texts vs. non-texts) and *method of obtaining language data* (controlled vs. not controlled). Even if one assumed that these parameters are binary, their application would result in four possible types of language data, not just two. In fact, several authors have sought, if not to replace, at least to supplement the dichotomy in question with more

---

[6] Such as found at https://totemfieldstoryboards.org/.

[7] I am very grateful to Anthony Webster for pointing out the existence of this paper to me.

precise distinctions. As a rule, these distinctions refer to the second parameter, i.e., the method of obtaining the data. Below, I will discuss some relevant proposals.

Hyman (2001, p. 18) distinguishes two ways of collecting language data based on the role of the researcher in generating the data. The researcher's active role in this process corresponds to *elicitation*, the lack of it corresponds to *observation*. According to the author, the ideal situation, i.e., the purest form of observation, is "where the fieldworker unobtrusively records the linguistic event: a spontaneous interaction between speakers, a narrative, a political speech, etc." (Hyman, 2001, p. 18). An advantage of this distinction is that it abandons the elusive notion of 'natural' in data collection and focuses on two concrete models of the researcher's behavior: they either actively *elicit* language information or *observe* a speech event happening independently from them. A disadvantage, however, is that it is still a dichotomy, and as such fails to capture the diversity of real scenarios.

It is not uncommon to find participant observation among the methods of language data collection. Himmelmann (1998) lists *observed* communicative events as the closest possible to *natural* ones (inherently impossible to be documented, as discussed above). The author defines observed events as "communicative events in which external interference is limited to the fact (known to the communicating parties) that the ongoing event is being observed and/or recorded." (Himmelmann, 1998, p. 185).

In other interpretations, participant observation appears to be limited to the situations in which the researcher witnesses the language being used but does not document the observed events by audio and video recording. Himmelmann (1998, p. 162) also uses the term in this sense when he states: "[t]he methods used in putting together this sample [a sample of primary data] may include participant observation **(jotting down overheard utterances)** […]". Aikhenvald (2007) and Dixon (2007) consider this type of participant observation as an indispensable part of what they call 'immersion fieldwork'. This is illustrated by the following citations:

- "Linguistic fieldwork ideally involves observing the language as it is used […] One records texts, working one's way through them, and at the same time learns to speak the language and observes how it is used by native speakers […]" (Aikhenvald, 2007, p. 5),

- "Texts are the most important part of a field linguist's database but they can never be the full story. They must be supplemented by what the linguist hears around them – by what people say and by what they tell the linguist to say." (Dixon, 2007, p. 23).

Another term to refer to this type of data is *overheard*, used by Chelliah; de Reuse (2011). The authors define it as "language data gathered outside of a formal field elicitation session" (Chelliah; de Reuse, 2011, p. 213). Arguably, observed/overheard speech events can also occur during recording sessions, if they arise as a result of "distraction" from the recording context, i.e., short dialogues on unrelated topics with neighbors or family members, or comments about the environment.

If we move to the language data produced within and because of language documentation activities, the active role of the researcher can also be less than generally assumed. Without doubt, every language documenter knows that far from everything a person says is directly prompted by the researcher. To describe such data received without or with only minimal prompting, the authors usually use one of the two following qualifiers: 'volunteered' and 'spontaneous'.

I agree with Himmelmann (1998) that *spontaneity*, i.e., the degree to which verbal behavior is planned, is a general characteristic of communicative events. All of them can be placed on a continuum of spontaneity depending on whether they are more or less planned. Furthermore, the degree of spontaneity does not necessarily correlate with the degree of the researcher's control. For example, ritual speech events are often highly unspontaneous, even when they are not (and even must not be) documented.

The term 'volunteered' seems to be more suitable to describe speech events whose content and form arise from the speaker's deliberate choice in a documentation setting. Such speech events are conditional on the researcher's general interest in the language but not on any direct request. Volunteered events include the information shared by the speaker when not explicitly asked, e.g., a story, or a wordlist from a particular semantic field. In more specific cases, volunteered language data are produced when the speaker has created an idea of what the researcher wants with a particular linguistic discussion and decides to contribute more relevant information than what has been asked for. For example, after the researcher has gone through a manual of local birds to record their names, the speaker might want to volunteer names for other birds, which did not occur in the book. Or when the speaker perceives the researcher's interest in a particular type of construction, e.g., the difference between conditionals and counterfactuals, they might want to volunteer a few sentences to help the researcher's understanding.

Many language data are produced with the researcher directly asking for a speech event of a certain form and/or content to take place. Although this type of language data is often generically called 'elicitation', more fine-grained distinctions can be made. Himmelmann (1998) distinguished *staged communicative events* from *elicitation*. In the former, the researcher either gives general instructions (i.e., asking for a particular traditional narrative to be told) or uses more specific prompts (i.e., pictures or video stimuli). Arguably, however, the difference between these two subtypes of staged events is more significant, such that the former (based on general instructions) would normally be classified as 'natural(istic)' data and the latter as 'elicited' data in the traditional language documentation discourse. I propose that one can distinguish between *staged* events, those based on more generic requests, and *stimuli-based* ones, those based on auxiliary non-verbal material (usually some sort of visual stimuli) prepared or selected by the researcher beforehand.

Finally, there are language data whose form and/or content are specified by the researcher in detail. This is normally understood by 'elicitation' in its most widely agreed sense. The activities that give rise to such data are often based on the speaker's metalinguistic skills, such as when the researcher offers a linguistic form for the speaker to provide contexts for its usage, to judge its acceptability, to correct it, or to translate it. We can call this type of language data 'controlled'.

Thus, I propose the following types of documentable language data based on the method by which they were obtained:

- *observed* (produced independently from the documentation setting),
- *volunteered* (produced within the documentation setting, but without a direct request),
- *staged* (based on generic requests),
- *stimuli-based* (prompted by material non-verbal stimuli),
- *controlled* (highly specified in their form and/or content).

These types are complementary to the classification by the linguistic unit (e.g., text, sentence, and wordlist). For example, a text can be obtained by any of the abovementioned documentation methods. Hence, simply stating that a study is based on a text corpus does a poor job at clarifying the degree and the nature of the researcher's influence over the data.

Of course, this classification is not intended to be the last word in understanding different types of language data. Rather, it is an invitation to recognize that the dichotomy is fundamentally deficient given the complexity of the language documentation enterprise.

4. Why does it matter?

A question that might arise from this discussion is: if the numerous problems of the 'natural(istic)' vs. 'elicited' dichotomy seem to lie on the very surface, why has it not been a target for more critical revisions? In my opinion, one reason is that, fortunately, most linguists still work with languages used, at least to some extent, in daily communication and in other traditional domains. For those who record a narrative, or a conversation, or a song in a relatively vital language, the term 'natural' may not seem problematic. Those speech events can appear to be natural insofar as 'natural' is often associated with 'fluent', i.e., produced with full confidence, without visible effort, without frequent hesitation in search of a correct word and frustration when the retrieval fails. This perception, however, can be very different for those who, like the author of this paper, work with critically endangered languages.

My experience in documenting the Guató language (gta), currently remembered by two elderly native speakers, has consistently revealed the inadequacy of the term 'natural'. The speakers of Guató have not used the language in their daily life for years now. The 'natural' context of their current lives does not presuppose speaking anything other than Brazilian Portuguese. At the same time, several traditional and personal narratives have already been recorded in the language. On the one hand, I

should feel satisfied by being able to check the box for 'natural data' as an indispensable item for grammar description. On the other hand, it seems that the Guató narratives documented so far should be considered valid and reliable data, without the necessity of making claims as to their 'naturalness'.

As increasingly more under-described languages cease being spoken daily, no speech event one can eventually document in those languages is 'natural', in the sense of something that could have taken place during the researcher's absence or even in the sense of something produced with total ease. At the same time, these speech events can be complete and sophisticated, providing material for detailed descriptions of morphosyntax, clause linkage and information structure. How can we determine if they resemble hypothetical 'natural' data for the same language and does it really matter as long as the data reflect a complex and internally consistent system?

Another issue is that the current practices in classifying language data are not sufficient for the purpose of accountability in language documentation projects. For example, the IMDI metadata standard used in *Lameta*[8] (Hatton *et al.*, 2021) covers language data types under the 'genres' category, where each speech event can be attributed only one principal genre. Instead of the term 'natural', names for different genres normally associated with 'natural' data are made available: narrative, song, conversation, etc. Elicitation is also one of the options available under the 'genres' category. But framing the choice as "either narrative/song/conversation or elicitation" is illogical, since the former option refers to what the speech event is in its structure and the latter one refers to how it was obtained.

Furthermore, 'elicitation' is too vague a term, almost functioning as a wastebasket for everything that does not fit into a 'natural(istic)' genre. For example, sessions organized around lexical fields often end up classified as 'elicitation', mostly by not conforming to any other genre in the list. But, as we have already seen, the lexicon (e.g., bird names) can be obtained by various methods. It can be volunteered by the speaker (e.g., 'I remembered a few bird names last night'), staged by the researcher (e.g., 'Can you name all the birds common in this area?'), or stimuli-based (e.g., supported by a field guidebook). The need to put such data under the generic 'elicitation' label all but erases the complexity of scenarios in which they are often collected.

Additionally, the bias against 'elicited' data can affect how the researcher's performance in a documentation project is evaluated. It is commonplace to prioritize the collection of natural(istic) data over elicited ones, as the former are assumed to be more significant scientifically and culturally. But since one classifies lexical data as 'elicitation', this label ends up covering not only artificial sentences from Eurocentric questionnaires, but also kinship terminology, personal names and toponyms, flora and fauna nomenclature, among other domains of traditional knowledge. It seems indisputable that such information is as relevant as, say, myths of creation. However, if documented lexicon falls under elicited deliverables, its collection may be counterintuitively assessed as a drawback, rather than a highlight, of a documentation project.

Finally, the natural(istic) vs. elicited dichotomy prevents a realistic gauge of the potential outcomes when a project is aimed at documenting a critically endangered language. The reason is that

[8] www.lameta.org

the researcher complies with delivering a certain amount of 'natural(istic)' speech (normally much more than the proposed amount of elicitation), but the actual production of data considered 'natural(istic)' is largely out of the researcher's control.

Getting back to the classification of data types proposed in Section 3, the widely used term 'natural(istic)' typically covers observed, volunteered, and staged texts. In a critically endangered language like Guató, documenting observed speech is impossible since the language is not used in daily communication. The speaker may volunteer stories, but the extent depends heavily on their self-confidence and their rapport with the researcher. Crucially, the volunteered data are, by definition, out of the researcher's immediate control. The researcher can stage text production, e.g., asking for traditional and personal narratives to be told. However, success will essentially rely on the memory and willingness of the few remaining elderly speakers. This situation is very different from documentation of a vital language, where the range of potential contributors and the pool of potential texts is incomparably larger.

Thus, the question "How many hours of naturalistic speech can you document in this critically endangered language?" can only be answered with very rough estimates. This is not to say that the speakers' performance in this situation will necessarily be limited. Rather, it is to recognize that, ultimately, the number and duration of texts produced without prompts will not depend solely on the researcher's diligence. More appropriate questions would be "Are the remaining speakers inclined to volunteer stories at all?" and "What proportion of your recording sessions will be dedicated to attempts at staging storytelling?", remembering that not every attempt will succeed.

In summary, I defend that the method by which the data were obtained (observed, volunteered, staged, stimuli-based, or controlled) is a separate parameter and it should be stated explicitly and independently from the linguistic unit and/or the genre of the speech event. Furthermore, I propose that the terms 'natural(istic)' and 'elicitation' should be abandoned altogether as imprecise and leading to unjustified biases in language documentation, especially when critically endangered languages are concerned.

5. Conclusion

In this paper, I have argued that the terms 'natural' and 'elicited' are much more problematic than normally recognized. Language data that could be considered truly 'natural' are not documentable given that the context of language documentation is inherently 'unnatural' for language production. Furthermore, the term 'natural' is generally used as denoting a gradable concept. As such, it fails to adequately describe the diversity of language data normally lumped together under such a vague label.

The dictionary definition of the term 'elicitation' is so general that, ultimately, it can be considered as a synonym for 'language documentation'. Consequently, elicitation-related vocabulary is often used to describe the collection of 'natural(istic)' data, even by the authors who advocate for a more specific usage of the term 'elicitation'. The whole idea that 'natural(istic)' data are uncontrolled

texts, whereas 'elicited' data are controlled sentences and words, does not withstand scrutiny and leads to evident contradictions within publications on linguistic fieldwork.

As a result, the dichotomy 'natural(istic)' vs. 'elicited' appears deeply unsatisfactory for capturing the diversity of documentable language data. Based on previous proposals, I have offered a more fine-grained classification of language data, which abandons the terms 'natural(istic)' vs. 'elicited' altogether. While this classification is open to refinement, it provides an immediate framework for increasing the accountability and precision of metadata.

As this paper has shown, the term 'natural(istic)' is meaningless for documentation of critically endangered languages which are no longer spoken in daily life. It can even be harmful if the impossibility of recording data that conform to the notion of 'natural' leads to the conclusion that no in-depth research can be done with such languages.

Finally, it should be recognized that the parameters of linguistic unit and/or genre, on the one hand, and of method by which the data are obtained, on the other hand, interplay with each other, giving rise to diverse types of language data. A narrative as well as a single sentence can be observed or volunteered or stimuli-based, and a researcher should be able to clarify this interplay when making their work publicly available, e.g., in a language archive. Ultimately, a more careful distinction between data types improves accountability and clarifies what, exactly, is being documented about endangered languages.

Additional information

Evaluation and author's answers

Evaluation: https://doi.org/10.25189/rabralin.v24i1.2382.R

Editor

Maria Filomena Spatti Sandalo
Afiliação: Universidade Estadual de Campinas
ORCID: https://orcid.org/0000-0003-4595-7765

EVALUATION ROUNDS

Evaluator 1: Karin Camolese Vivanco
Affiliation: Universidade Estadual de Campinas
ORCID: https://orcid.org/0000-0002-7582-2047

EVALUATOR 1

The paper offers an insightful discussion of the concepts of 'natural/naturalistic' and 'elicited' data, showing how these terms—despite their widespread use —are often only vaguely defined. The author shows that actual documentation practices typically intertwine features associated with both 'naturalistic' and 'elicited' data, and proposes alternative parameters that appear more informative than this dichotomy.

I recommend the paper for publication with minor revisions. Below, I outline a few points that the author may wish to consider.

"Naturalistic" data in highly threatened languages

The conclusion raises a crucial issue concerning the notion of 'naturalness' . In the case of highly endangered languages, the target language may no longer be part of speakers' daily linguistic practices. This point is illustrated by the Guató example. Toward the end of the manuscript, the author notes that treating 'naturalness' as the central goal of documentation projects can inadvertently suggest that no meaningful or in-depth documentation can be carried out for such languages. Therefore, besides the fact that unreflective characterizations of fieldwork practices fail to capture the complexity of linguistic events, such views may also be harmful to the study of minoritized languages. This point is especially relevant for authors who (sometimes tacitly) elect one methodology—often the 'naturalistic'approach—as the norm, while overlooking its limitations.

Given its importance to the area, I think this insight—currently developed in a single paragraph—could be expanded.

(As a minor suggestion, an example from Mithun's "Field Linguistics: Gathering Language Data from Native Speakers" could further strengthen this argument. Mithun discusses situations in which fieldwork must be conducted with a fluent speaker who is temporarily or permanently detached from his speech community. Most linguists would certainly recognize the value of working with such speakers, even though their practices may not be 'natural' as their daily communication within their original environment)

Proposed types and parameters

In the final pages of the manuscript, the author proposes "a more fine-grained classification of language data, which abandons the terms 'natural(istic)' vs. 'elicited' altogether", proposing two parameters: linguistic unit/genre and the context of production. Earlier, in Section 3, the author had introduced a set of "types of documentable language data": observable, volunteered, staged, planned, and controlled. I think these two proposals could be articulated more deeply, possibly in a separate section, because they seem to be independent of each other in the current version.

One additional suggestion would be to avoid the terms 'types' or 'categories', as they may imply fixed or discrete properties of data. I recommend using 'features' or 'traits' to capture the idea that different arrangements of these features give rise to a wide range of data.

Minor suggestions

Consider using boldface rather than underlining for emphasis.

Consider the following works to enrich the literature review:

Labov (1972), “Some Principles of Linguistic Methodology”;
Samarin (1967), Field Linguistics: A Guide to Linguistic Field Work;
Munro (2017), “Field Linguistics: Gathering Language Data from Native Speakers”.

Conflicts of interest

The author have no conflicts of interest to declare.

Link for *Preprint*

https://doi.org/10.20944/preprints202510.1141.v1

Research Protocol and Pre-Registration

After evaluating the guidelines proposed by the Equator Network, we consider that none of them are relevant to the research in question. We also inform you that the research developed has not been pre-registered in an independent institutional repository.

Data Availability Statement

The research data is contained within the manuscript itself.

Acknowledgments

I am very grateful to the Endangered Languages Documentation Programme (ELDP) for the Graduate Scholarship (IGS1010 – Documentation of Guató, an endangered isolate from the Brazilian Pantanal). This paper was motivated largely by the methodological challenges encountered during this project.

REFERENCES

AIKHENVALD, Alexandra Y. Linguistic fieldwork: setting the scene. **Language Typology and Universals**, v. 60, n. 1, p. 3-11, 2007.

BOWERN, Claire. **Linguistic fieldwork**: A practical guide. New York: Palgrave Macmillan, 2008.

CHELLIAH, Shobhana L. The role of text collection and elicitation in linguistic fieldwork. *In*: NEWMAN, Paul; RATLIFF, Martha. **Linguistic fieldwork**. Cambridge: Cambridge University Press, 2001, p. 152-165.

CHELLIAH, Shobhana L.; DE REUSE, Willem J. **Handbook of descriptive linguistic fieldwork**. Dordrecht: Springer Science & Business Media, 2011.

CROWLEY, Terry. **Field Linguistics**: A Beginner's Guide. New York: Oxford University Press, 2007.

DIMMENDAAL, Gerrit J. Places and people: field sites and informants. *In*: NEWMAN, Paul; RATLIFF, Martha. **Linguistic fieldwork**. Cambridge: Cambridge University Press, 2001, p. 55-75.

DIXON, R. M. W. Field linguistics: a minor manual. **Language Typology and Universals**, v. 60, n. 1, p. 12-31, 2007.

EVERETT, Daniel L. Monolingual field research. *In*: NEWMAN, Paul; RATLIFF, Martha. **Linguistic fieldwork**. Cambridge: Cambridge University Press, 2001, p. 166-188.

HATTON, John; HOLTON, Gary; SEYFEDDINIPUR, Mandana; THIEBERGER, Nick. 2021. Lameta [software] https://github.com/onset/laMETA/releases

HIMMELMANN, Nikolaus P. Documentary and descriptive linguistics. **Linguistics**, v. 36, n. 1, p. 161-196, 1998.

HIMMELMANN, Nikolaus P. Linguistic data types and the interface between language documentation and description. **Language Documentation & Conservation**, v. 6, p. 187-207, 2012.

HYMAN, Larry M. Fieldwork as a state of mind. *In*: NEWMAN, Paul; RATLIFF, Martha. **Linguistic fieldwork**. Cambridge: Cambridge University Press, 2001, p. 15-33.

MANNHEIM, Bruce; VAN VLEET, Krista. The Dialogics of Southern Quechua Narrative. **American Anthropologist**, v. 100, n. 2, p. 326-346, 1998.

MEAKINS, Felicity; GREEN, Jennifer; TURPIN, Myfany. **Understanding linguistic fieldwork**. New York: Routledge, 2018.

MITHUN, Marianne. Who shapes the record: The speaker and the linguist. *In*: NEWMAN, Paul; RATLIFF, Martha. **Linguistic fieldwork**. Cambridge: Cambridge University Press, 2001, p. 34-54.