

Naturalness and Spoken Data

Martin Weisser

The following paper will deal with the concept of 'naturalness'. I shall first attempt to explain and classify what I understand by this notion and then try to describe to what extent it can or should possibly be applied to both the collection and investigation of spoken data. To some readers the latter may seem like rather a strange question to ask since it seems to (and should) be one of the ultimate goals in researching language to use as much authentic material as only possible, but unfortunately research methods and technology do not always fully allow for this.

1. What Is Naturalness?

Although we all have at least a general idea about what naturalness is (or may be), there does not seem to be any proper definition for it. Like 'fluency' or '(oral) proficiency', it is one of those terms a lot of researchers use without anybody really having any clear idea of what it stands for, although at least for the former two, attempts at defining them can be found in linguistic dictionaries.

It seems to me the best way to try and define naturalness may be to give a negative definition. Thus theoretically - and especially if one is a great believer in binary systems - it might be rather tempting to define it as anything that is not in any way influenced by any element of artificiality or any outside constraints.

Can this definition, however, actually work in practice? Probably not, because who is to say what exactly actually constitutes artificiality or outside constraints? And to what extent would they have to be present to make language in general or the idiolect of any individual unnatural? I think, rather than trying to define naturalness as one abstract entity, one probably has to allow for various levels of naturalness, closely tied to levels of register or situation, and therefore also to the notion of appropriateness.

Thus a highly formalised context - such as the conversation at a(n official) dinner party - may and will most probably, because of the social constraint of formality, require a higher degree of 'unnaturalness' than for example a casual chat between students. Another example of 'unnatural', formalised language is literary language in general and, in particular, the kind of language used for poetry or some types of grandiloquent modern writing that bear a distinctive resemblance to Eupheism as far as

structure and vocabulary are concerned. On the other hand, naturalness may also be determined less by the social, but rather by what could be called the ‘medial’ context, i.e. spoken vs. written. Certain language phenomena are simply more or less common according to what the medium is. Hughes for example, in his discussion of recordings for listening tests points out:

“If recordings are made especially for the test, then care must be taken to make them **as natural as possible** [my emphasis]. There is typically a fair amount of redundancy in spoken language: people are likely to paraphrase what they have already said [...], and to remove this redundancy is to make the listening task unnatural.” (Hughes: 136)

He then goes on to cite an example, which, incidentally, shows that the social and ‘medial’ contexts may very well go hand in hand in producing the effect of (un/)naturalness:

“In particular, we should avoid passages originally intended for reading, like the following, which appeared as an example of a listening comprehension passage for a well-known test:

She found herself in a corridor which was unfamiliar, but after trying one or two doors discovered her way back to the stone-flagged hall which opened onto the balcony. She listened for sounds of pursuit but heard none. The hall was spacious, devoid of decoration: no flowers, no pictures.” (ibid.)

In this example, the formalised characteristics of this type of literary language are partly responsible for a rather ‘eloquent’ wording, using expressions like *devoid* which clearly are not part of a ‘natural’, everyday vocabulary, and partly for the kind of more elaborate, paratactic structure which is on the one hand so typical of literature, but on the other also of written language in general. Thus what may be regarded as perfectly natural in one type of situation or register may not be regarded as natural in another at all. Therefore, naturalness in language may be loosely defined as the degree of appropriateness for a given situation. This appropriateness in turn depends on whether the speaker’s/writer’s behaviour (with regard to situation and medium) is sufficiently close to what is expected of him or her, in other words some implicit norm.

This implicit norm not only exists for appropriateness which is, after all, only one of the criteria for naturalness; it applies equally well to the notion of grammaticality. For example, it is a well known, though seldom appreciated fact, that even native speakers occasionally tend to produce ‘ungrammatical’ utterances. This is perfectly

natural if we bear in mind that speakers do not always have time to prepare exactly what they are going to say beforehand - and thus may fail to achieve subject-verb agreement, produce false starts, etc. - or in some cases may genuinely not know the correct form as can be the case especially with 'hard words', either because a given speaker may not know the correct plural form or because he/she does not know that a 'plural word' like *data* is commonly used with a singular verb. In general, therefore, it should also be permitted to the non-native speaker to produce such 'ungrammaticalities'; the question is only: What is still permissible and what should count as a genuine mistake? As can be seen from the above examples, the hidden norm in this case is partly related to the notion of usage (and, again, appropriateness) and as such may change in time, but partly also to genuine errors or slips, where the level of acceptance is less easy to account for.

A third criterion for naturalness is that of authenticity. This criterion obviously applies less to spontaneously produced speech than to material(s) used for teaching or general descriptions of language, such as grammar or text books. Not too long ago, many linguists still relied on their intuition as to what should be regarded as proper examples of language and therefore sometimes invented them in order to prove a particular grammatical point. This came to be known as 'armchair linguistics' and can be observed in its extreme form in many books written by linguists following the Chomskyan tradition of transformationalism:

"the man who persuaded John to be examined by a specialist was fired"
(Chomsky: p. 130)

Now, while there can certainly be no doubt as to whether the above example is grammatically well-formed, it is highly questionable whether such an utterance would ever occur naturally, simply because it does not make much sense, or at least it is very difficult to imagine a context in which it would. It is almost obvious that it can only be an invented example. Today, with both increasing interest and progress in Corpus Linguistics and the availability of numerous computer corpora, the need to invent examples has become almost obsolete. Nevertheless, even corpora can to a certain extent be artificial, even if they contain only genuinely attested data. A corpus which consists only of literary texts, for example, will (for the reasons noted earlier on) almost definitely be more formal, and thus artificial, than a corpus of spontaneous

spoken conversation. On the other hand, it is also possible that something occurring in a corpus may not necessarily be very natural if a speaker starts ‘playing with the language’, creating ad hoc constructions that are not part of the common usage as for example neologisms, or that certain language phenomena may never occur although they are quite natural and grammatically perfectly correct. In the latter case, of course, the question of authenticity presents us with a dilemma that cannot easily be resolved.

Having given a brief account of what can be understood by naturalness, I shall now move on to discuss how this concept can be applied to the collection and analysis of spoken data and how much of a compromise between naturalness and ‘artificiality’ has to be made by the researcher to obtain valid data.

2. Analysability.

One of the main factors in analysing spoken data is the quality of the recordings. The degree of quality required depends, of course, on the type of analysis one wants to perform. If it is purely a matter of transcribing classroom or conversational data used for semantic, gender or discourse analysis, then the quality need in most cases only be high enough to understand what is being said, whereas for a ‘full-blown’ phonetic analysis it has to be as high as only possible.

Everyone who has ever tried to transcribe a piece of spoken discourse from a tape even only orthographically knows that it can be very difficult to make out just what exactly a speaker is actually saying. Background noise, loudness or clarity of enunciation can have a major influence on the listener’s perception and therefore the quality of the transcription. In trying to produce a phonetic/phonemic transcription, this becomes even more difficult as one has to perceive not only what is said, but also, and especially, how. Having one’s informants produce the same text makes the task of ‘transcribing the what’ a little easier by somewhat ‘normalising’ the output, but even if one supposedly knows what the speakers are producing, one still has to allow for a certain degree of variability on the part of each individual speaker. Thus a particular speaker may show a strong preference for certain fillers or produce a large amount of hesitation phenomena, and those, too, have to be faithfully included in any transcription for phonetic or discourse analysis purpose.

But those are only the problems the ‘human interpreter’ encounters when trying to analyse the data. Yet the problem of analysability is far greater for the non-human interpreter, i.e. the computer or rather those special programs that should enable the researcher to detect and isolate certain features of speech that a purely auditory analysis would fail to recognise. However, while the human listener is able to make a great deal of allowance for irregularities in speech and can on the basis of inference still determine the content of an utterance, computer programs usually need an extremely high quality input in order to be able to interpret and display the data correctly and may sometimes fail completely if they encounter a high degree of background noise:

“The signal-to-noise ratio should be maximized. That is, the signal should be at least 20 dB above the level of background noise. In most cases analysis procedures will not work unless the voice of a single speaker can be recorded in a quiet environment. For example, pitch extractors that derive the fundamental frequency of phonation need a recorded signal that contains the speech signal of only one speaker. The pitch extractor cannot track the voice of one speaker when a second voice simultaneously occurs, as is often the case when one records a normal conversation. Nor will pitch extractors work when competing periodic noises occur, for example, when the tape recording contains the low frequency, low amplitude sounds produced by some central air-conditioning systems.” (Lieberman/Blumstein: p. 74)

Therefore, although one might argue that the use of read material may lead to a kind of precision in the delivery of the informants that would not occur in natural unscripted speech, it is to a certain extent necessitated by the need for qualitatively analysable data. However, I believe that one can and should always try to alleviate the negative effect of scripting by making the scripted situation as natural and close to real life as only possible.

3. Comparability.

A second major factor, especially in a contrastive analysis like mine, is the one of comparability. In order to be able to draw any valid conclusions from individual realisations of spoken material, its instances have to be as similar to one another as possible since, even more so than it is the case in written language, realisations in spoken language are highly context dependent. While in written language there is only a limited degree of variability as far as both (syntactic) context and realisation are concerned - usually words in written texts are delimited by spaces or punctuation marks and there are normally very few spelling variants, if any, for each word -, in spoken language no word is ever pronounced exactly the same way twice even by the same

speaker and there is a large amount of variability according to the speaker's regional and social background.

Using a prefabricated text to be produced by the informants can eliminate some of those problems by doing two closely related things. The first one is to impose at least a loose structure upon the utterance(s) of the speaker(s), so that relevant details may later be more easily compared and identified by the researcher. In the case of my data this is particularly important in order to be able to observe the (both native and non-native) speakers' turntaking behaviour and text comprehension abilities, the latter of the two obviously being more important where non-native speakers are concerned. The other one is to be able to elicit specific types of more performance-related information like the pronunciation of individual words, features of assimilation, intonational structures, etc. In any non-scripted situation, eliciting this type of information would be rather difficult as there could only be very little direct control over which words the individual speaker would be using to describe a particular incident or situation and thus to create a comparable context.

However, as convenient and necessary it may be to impose this kind of structure upon the data, there are some problems one constantly needs to bear in mind when analysing the material thus collected later. For example, some of the phenomena that typically occur in natural nonscripted spoken language like hesitation phenomena may be governed by punctuation rather than the speaker's overall understanding of the text. The same may also be true of intonational contours. Apart from this, when creating such material, one always runs the risk of creating something 'constructed' or artificial that may not be used much in everyday spoken language, something similar to the degree of artificiality that is unfortunately to be found in most textbooks for foreign learners. Thus certain phrases or expression may end up being highly suitable for eliciting certain phonetic detail, but may sound rather odd to the native speaker's ear. This risk is especially high when the material is created by researchers who are non-native speakers themselves, as they may sometimes not possess enough knowledge of idiomaticity or current usage, especially if they themselves have only or mainly learned their English from the aforementioned textbooks. In order to avoid any danger of this happening, I created the dialogue used for my study with the help of a native speaker friend of mine. However, even this did not prevent some strange

reactions to parts of the dialogue by other native speakers, which seems to indicate that there is no absolutely foolproof way of creating any such material.

4. Familiarity.

The final important advantage in using prefabricated material is that the speakers can gain a certain degree of familiarity with the text they are later supposed to produce. Even though this might in the first instance rather seem like a disadvantage because it could lead to a high level of precision or a tendency for the informants to act out the text instead of producing it naturally, it may nevertheless be a necessity for both native and non-native speakers for two different reasons. The first of those is that the informants need to know what the situation is so that they can actually produce something natural since otherwise they would have to employ certain strategies for understanding the text in the first place, which would lead to a style of delivery that is closer to reading than to speaking, and may not use appropriate intonational structures because they have to process the content first. The second reason is more important for non-native speakers (although it could technically apply to native speakers as well) and relates to knowledge of the vocabulary used in the text. If an informant comes across a word he/she does not actually know or understand fully, this will not only introduce hesitation phenomena like extended pauses that will render the speech rhythm more unnatural, but may also lead to 'distorted' intonation patterns, for example when a given speaker starts 'questioning him/herself' and thus intonationally turns a statement into a question.

5. Conclusion.

As can be seen from the criteria listed above, there are certainly both advantages and disadvantages in using prefabricated material such as scripted dialogues for collecting and analysing spoken data. For a proper phonetic analysis it is, for some of the above discussed reasons, almost essential to use such material as otherwise the risk of collecting unanalysable data is too great. Regarding the question of naturalness, however, this remains a tricky question... As I hope to have shown in the beginning of this paper, it is very difficult to define naturalness itself as it largely depends on contextual factors that vary all the time and thus make it very difficult to establish a valid set of criteria for achieving it.

However, the big question still remains: Is it at all possible to create materials for the analysis or teaching of a language that are natural or will there always be a certain (indeterminate) degree of artificiality? I personally believe the latter to be true and as a consequence think that it is necessary to adopt a slightly different perspective towards the creation and use of such material. Researchers should stop trying to aim at creating something ‘perfectly natural’ or to expect others to do so and to accept that there will always have to be a compromise between naturalness and usability. On the other hand, they should try as much as possible to bear in mind and alleviate the effect(s) of having to establish the right research conditions. Thus in dealing with spoken data phonetically, the researcher should always attempt to aim at as close an approximation of natural speech as only possible by setting up ‘near real-life situations’ using techniques such as non-scripted role play (where language learners can try to freely express themselves according to their abilities and then learn through feedback on their performance), which is, incidentally, as close to natural speech as one can ever get in language teaching classroom. Having accepted that perfect naturalness is not achievable, one should then aim at setting up a framework for comparison of one’s findings with real life data via the process of abstraction. Thus having found out more about how spoken language actually works, it should be possible, for example, to create a realistic set of criteria for evaluating (and teaching) it, rather than relying purely on those old-fashioned, impressionistic methods still commonly used in the assessment of foreign language learners.

Bibliography:

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, Massachusetts: MIT Press.

Crystal, D. ³1991. *A Dictionary of Linguistics and Phonetics*. Oxford: Blackwell.

Hughes, A. 1989. *Testing for Language Teachers*. Cambridge: CUP.

Lieberman, P. & Blumstein, S. 1988. *Speech Physiology, Speech perception, and Acoustic Phonetics*. Cambridge: CUP.

Richards/Platt/Weber. 1985. *Longman Dictionary of Applied Linguistics*. London: Longman.

Naturalness and Spoken Data

Pros and Cons of Using Prefabricated Material:

