

Ref

# Essentials of Language Documentation

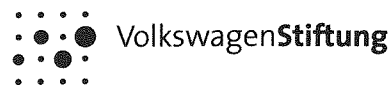
*edited by*

Jost Gippert  
Nikolaus P. Himmelmann  
Ulrike Mosel

Mouton de Gruyter  
Berlin · New York

Mouton de Gruyter (formerly Mouton, The Hague)  
is a Division of Walter de Gruyter GmbH & Co. KG, Berlin.

Published with support of VolkswagenStiftung, Hannover, FRG.



The hardcover was published in 2006 as volume 178  
of the series *Trends in Linguistics Studies and Monographs*.

© Printed on acid-free paper which falls within the guidelines  
of the ANSI to ensure permanence and durability.

*The Library of Congress has cataloged the hardcover edition as follows:*

Essentials of language documentation / edited by Jost Gippert, Nikolaus P. Himmelmann, Ulrike Mosel.  
p. cm. — (Trends in linguistics. Studies and monographs ; 178)  
Includes bibliographical references and index.  
ISBN-13: 978-3-11-018864-6 (cloth : alk. paper)  
ISBN-10: 3-11-018864-3 (cloth : alk. paper)  
1. Linguistics — Documentation. 2. Language and languages — Documentation. I. Gippert, Jost. II. Himmelmann, Nikolaus P., 1959– III. Mosel, Ulrike IV. Series  
P128.D63E85 2006  
025.06'41—dc22

2006001315

ISBN-13: 978-3-11-018406-8

ISBN-10: 3-11-018406-0

*Bibliographic information published by Die Deutsche Bibliothek*

Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie;  
detailed bibliographic data is available in the Internet at <<http://dnb.ddb.de>>.

© Copyright 2006 by Walter de Gruyter GmbH & Co. KG, D-10785 Berlin  
All rights reserved, including those of translation into foreign languages. No part of this  
book may be reproduced or transmitted in any form or by any means, electronic or mechanical,  
including photocopy, recording or any information storage and retrieval system, without  
permission in writing from the publisher.  
Cover design: Martin Zech, Bremen.  
Printed in Germany.

## Editors' preface

Language documentation is concerned with the methods, tools, and theoretical underpinnings for compiling a representative and lasting multipurpose record of a natural language or one of its varieties. It is a rapidly emerging new field in linguistics and related disciplines working with little-known speech communities. While in terms of its most recent history, language documentation has co-evolved with the increasing concern for language endangerment, it is not only of interest for work on endangered languages but for all areas of linguistics and neighboring disciplines concerned with setting new standards regarding the empirical foundations of their research. Among other things, this means that the quality of primary data is carefully and constantly monitored and documented, that the interfaces between primary data and various types of analysis are made explicit and critically reviewed, and that provisions are taken to ensure the long-term preservation of primary data so that it can be used in new theoretical ventures as well as in (re-)evaluating and testing well-established theories.

This volume presents in-depth introductions into major aspects of language documentation, including a definition of what it means to “document a language,” overviews on fieldwork ethics and practicalities and data processing, discussions on how to provide a basic annotation of digitally-stored multimedia corpora of primary data, as well as long-term perspectives on the preservation and use of such corpora. It combines theoretical and practical considerations and makes specific suggestions for the most common problems encountered in language documentation.

The volume should prove to be most useful to students and researchers concerned with documenting little-known languages and language varieties. In addition to linguists and anthropologists, this includes students and researchers in various regional studies and philologies such as African Studies, Indology, Turkology, Semitic Studies, or South American Studies. The book presupposes familiarity with the basic concepts and terminology of descriptive linguistics (for example, basic units such as *phoneme* or *lexeme*), but most chapters will also be accessible and useful to non-specialists, including educators, language planners, politicians, and government officials concerned with linguistic minorities.

*stapled, stuck, taped, tethered, tied, welded.* In English the central variable seems to be the kind of material creating the attachment.

11. There are proposals from linguistics itself about a "Natural Semantic Metalanguage" through which definitions of complex notions can be framed in terms of simpler, allegedly universal (hence 'natural') semantic primes. See <http://www.une.edu.au/arts/LCL/disciplines/linguistics/nsmpage.htm>, where one can find a bibliography of the many publications of Anna Wierzbicka.
12. Faced with Pearson's challenge, Reed College senior Chris Haulk "promptly came up with, 'oh, you mean – wrap a string around a cylinder; versus, wrap a string around a cone'" (Albyn Jones, personal communication, March 1, 2005) – proving that mathematicians can be lexicographers, too.
13. Note that "Donor" here is a single entity, defined in Framenet as "The person that begins in possession of the **Theme** and causes it to be in the possession of the **Recipient**."
14. Visit <http://www.cs.arizona.edu/icon>.
15. The program symbolizes glottalized or ejective consonants and long vowels as capital letters, and a 0 is used to signal the absence of medial *j*.
16. See the descriptions of various stimulus kits developed by the Language and Cognition Group at the Max Planck Institute for Psycholinguistics at <http://www.mpi.nl/world/data/fieldmanuals>.
17. See Levinson et al. (2003) for an unashamedly extensional, comparative approach.
18. A short video used to elicit descriptions for Tzotzil 'inserting' actions is available on the book's website.
19. Samples of the sort of cartoon I have found useful for such tasks are available at <http://www.wdrmaus.de/lachgeschichten/mausspots> in streaming video format.
20. The expression is not confined to English; both Italian *pollice verde* (according to Elena Collavin) and German *grüner Daumen* (according to Nikolaus Himmelmann) have exactly the same metaphorical and literal meanings as 'green thumb', i.e., someone good at gardening. Similarly, Italian *senza fegato* 'without a liver' suggests a meaning similar to 'lily-livered.'
21. I ignore basic syntactic issues here: for example, in the expression *miil waarril* the word *miil* 'eye' is the syntactic subject of *waarril* 'fly.' In *miil bagal* 'eye' is syntactic object of *bagal* 'poke.'
22. In the Tzotzil of nearby Larrainzar, the equivalent ritual doublet is at once humble and literal: *ach'elal, takopal* 'mud, body.'
23. See Zgusta's discussion of polysemy (1971: esp. 77 ff.); also Evans and Wilkins (2000, 2001), Evans (1992).
24. See Jane Hill's discussion of the Hopi dictionary project in Chapter 5.

## Chapter 7

### Prosody in language documentation

Nikolaus P. Himmelmann

#### Introduction

Prosodic aspects of a linguistic message such as intonation and lexical accent are essential elements of its formal make-up. To date, the basics of analyzing prosodic features have not yet become an integral part of linguistic fieldwork training, and, accordingly, a reasonably detailed and comprehensive documentation and description of prosodic features is not yet part of standard linguistic fieldwork practices. This chapter is specifically concerned with the *documentation* of prosodic features, i.e. with the question of what kind of data a language documentation has to contain so that a thorough analysis of prosodic features is possible. In order to be able to productively apply the suggestions discussed in this chapter, a basic understanding of the core units and procedures of prosodic analysis is necessary. For a more comprehensive introduction to basic prosodic fieldwork focusing on issues of analysis and description, see Himmelmann and Ladd (forthcoming).

Given that a language documentation includes a large corpus of recordings of communicative events of different types, it may well be questioned whether there is any need to pay special attention to prosody when compiling it. Provided that the recordings are of a reasonable quality,<sup>1</sup> there can be no doubt that such a corpus can be used for prosodic analyses even when no particular attention was paid to prosodic features at the time of compiling the corpus.<sup>2</sup> However, there are essentially three reasons why some special attention for prosodic features is necessary when compiling a corpus of primary data so that it becomes really useful for prosodic purposes:

1. Prosodic phenomena are highly variable and susceptible to contextual influences. This makes it difficult to recognize basic distinctive patterns. Prosodic pattern recognition is much facilitated by having the same utterance produced by a number of different speakers (or at least to have multiple versions of the same utterance). See further Section 2.

2. Words produced in isolation are minimal utterances showing both lexical and utterance-level (post-lexical) features. Hence, the widespread practice of recording words in isolation when recording a wordlist is of limited use for prosodic purposes. See further Section 3.
3. Acoustic and auditory data (i.e. recordings of spontaneous and elicited utterances) do not provide direct evidence with regard to the perception of native speakers, i.e. what native speakers actually perceive as relevant prosodic contrasts (conversational material may provide indirect evidence, though; see further below). The most straightforward way to obtain perception data is to run perception experiments, as further discussed in Section 5.

Before these points are further elaborated, Section 1 provides a bit more detail on what exactly the term *prosody* is intended to refer to here. Furthermore, when discussing points (1) and (2), it will be repeatedly suggested that elicitation may provide useful materials to complement the data found in recordings of spontaneous speech. However, eliciting prosodic data is not an easy task, as discussed in Section 4.

### 1. Prosodic phenomena

Table 1 lists the major prosodic phenomena according to the different domains in which they are manifest, i.e. the recordable sound wave (acoustic), the perceptual impression (auditory), and as a component of the language system (phonological category). The rightmost column lists the most widely attested functions which may be conveyed by prosodic features (but of course can also be conveyed by non-prosodic means).

Table 1. Prosodic phenomena according to domain

Acoustic	Auditory	Phonological category	Function/meaning
<ul style="list-style-type: none"> <li>– fundamental frequency</li> <li>– duration</li> <li>– intensity</li> <li>– spectral characteristics<sup>3</sup></li> <li>– pauses/silence</li> </ul>	<ul style="list-style-type: none"> <li>– pitch</li> <li>– length</li> <li>– loudness</li> <li>– stress/prominence</li> <li>– rhythm/tempo</li> <li>– grouping</li> <li>– voice quality (creaky, etc.)</li> </ul>	<ul style="list-style-type: none"> <li>– tone</li> <li>– quantity</li> <li>– (lexical) accent</li> <li>– intonation</li> <li>– levels in prosodic hierarchy (syllable, foot, etc.)</li> </ul>	<ul style="list-style-type: none"> <li>– delimiting units</li> <li>– distinguishing lexical units</li> <li>– grammatical categories</li> <li>– speaker attitude</li> <li>– sentence modality</li> <li>– information structure</li> <li>– interactional tasks</li> </ul>

In discussing prosody, it is important to keep the different domains distinct and to be aware of the fact that there is no unambiguous mapping relation between features in different domains. To take just pitch as an example, regular correspondences exist between changes of fundamental frequency (F0) observed in the acoustic signal, changes in pitch perceived by the human ear, and tonal or intonational distinctions. But these correspondences do not consist of simple and direct mapping relations between the domains. Thus, there are changes in fundamental frequency which are generally not perceived as such by the human ear. These are known as “microprosodic perturbations” and include phenomena such as the lowering of F0 regularly induced by voiced consonants.<sup>4</sup> Furthermore, while it is true that tonal and intonational categories are primarily marked by changes in pitch, other auditory parameters such as length, loudness, and voice quality often also play a role in the marking of these categories.

In the present chapter, the above distinctions and the corresponding terminology will be observed rather strictly. Many of the terms are widely used in the literature in the sense they are used here, but it may be worth pointing out that the strict distinction also applies to the terms (*lexical*) *accent* and *stress*, which are used in many different and often somewhat confusing ways in the literature. Both terms refer to the phenomenon that a given syllable is in some sense more prominent than neighboring ones, but *lexical accent* here designates this property with reference to the phonological structure of lexical items (i.e. as a phonological category), while *stress* refers to an

auditory impression (which may or may not have clear acoustic or phonological correlates). In this usage, then, “lexical accents” can be realized in different ways, including “stress” or a fixed change in pitch (so-called melodic or pitch accent as found, for example, in Japanese; cf. Beckman [1986] and Gussenhoven [2004] for further discussion).

There is no space and need here to discuss in detail all the prosodic phenomena and functions listed in Table 1. The main purpose of this table is to give an extensional definition of the range of phenomena referred to with the term *prosody* in this chapter. A detailed introduction to the phonetics (both acoustic and auditory) of prosodic features can be found in Laver (1994: 431–546; see also Ladefoged 2003: 75–103). The major phonological categories are discussed in Ladd (1996), Cruttenden (1997), Hirst and di Cristo (1998), Hyman (2001), Yip (2002), Gussenhoven (2004), and Jun (2005), among others. These works also provide useful information regarding the crosslinguistic variability of prosodic features.

The discussion in this chapter in principle applies to all the prosodic features listed in Table 1. However, intonation and accent will usually be mentioned as the main examples and often be singled out for extra comment because these are the two categories that have been most widely neglected in linguistic fieldwork, as opposed to tone, for example, which is a standard topic in linguistic fieldwork.

## 2. The need to work with several speakers

Linguistic fieldwork often involves the close cooperation with just one or possibly two native speakers who are the main contributors or “informants” in the sense that a) they provide most of the elicited information on the language (texts are often recorded with a broader range of speakers); and b) all data provided by other speakers is processed and checked with them. This procedure is based on the fact that with regard to core grammatical features the information provided by different speakers tends not to differ (or to differ only minimally). Thus, for example, if one speaker states that the definite article has to precede the noun and cannot be postponed, this will in all likelihood be confirmed by all other speakers in the community.

While this set-up works reasonably well for the most basic structural features of a language, it becomes more and more problematic when more variable and complex linguistic features are being investigated. The phonetics of prosodic features are highly variable and depend on a complex set of

factors, including speaker variables and context. There are very few, if any absolute values. What is high with regard to pitch for one speaker, may be low for another; what is loud in one context, is just normal in another; and so on. Furthermore, the perception of prosodic features tends to be heavily influenced by the investigator’s own native prosodic system, which further distorts the data and complicates the analysis. In the early stages of an investigation of the prosody of a language, it thus tends to be extremely difficult to recognize a basic pattern in the recorded data. This problem is particularly pressing in the case of intonation, which for this reason serves as the major example in this section, but it may also occur with regard to lexical accent or tone.

The easiest way to solve the pattern-recognition problem is to have several speakers “do the same thing,” i.e. to produce the same utterance in the same context with the same intention. Figure 1 illustrates the problem and the suggested solution. It shows fundamental frequency tracings of the segment (*was für große Ohren du hast* ‘(what big) ears you have’, taken from the recordings of the folktale *Little Red Riding Hood* by five German speakers.<sup>5</sup> All speakers produce a rise on the initial accented syllable *Oh* and then a continuous fall until the end of the utterance. Note how variable the initial rise is (shaded area of left-hand column). For speaker JH it is quite long, starts steep but then becomes flatter, while for speaker NF it is steep and short. Speaker JN’s rise is very minor indeed and it could be argued that there is no rise at all in this syllable. Nevertheless, as the five speakers are doing the same thing, i.e. producing the same utterance in the same context (of reading the story aloud) with the same intention (of expressing surprise at the radical changes in the grandmother’s appearance), it is also legitimate to assume that the different rise-falls in F0 seen in these tracings are in fact realizations of the same category, i.e. the nuclear fall of Northern Standard German (symbolized with  $H^*+L$  in ToBI notation). Or, viewed from the point of view of someone trying to detect a basic pattern, the fact that one may reasonably assume that the five performances of the utterances are “the same” on the level of the language system allows one to recognize a common pattern, rise on the accented syllable plus continuous fall until the end of the intonation unit.<sup>6</sup>

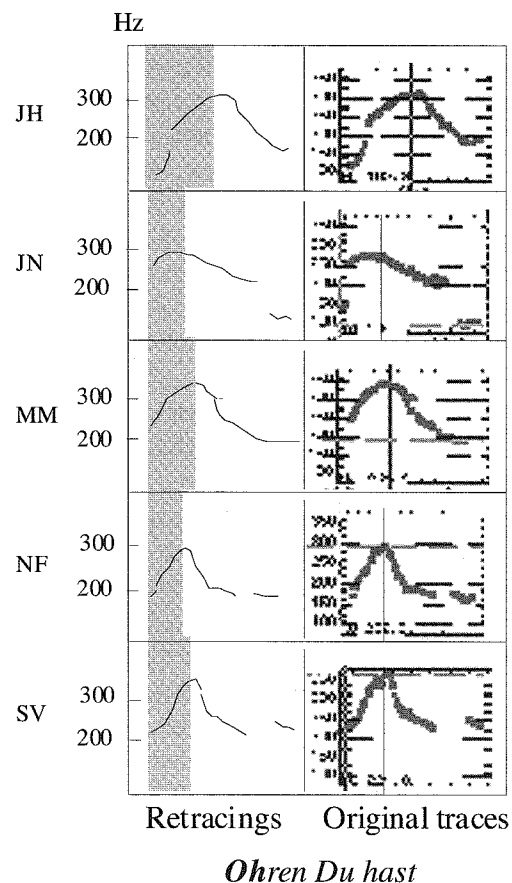


Figure 1. Multiple performances of the same utterance (from Grabe 1998: 245, Appendix C)

“Doing the same thing” here importantly involves three aspects. First, the utterances have to be segmentally identical (or at least very similar), because different segments have different microprosodic effects and it is not a straightforward task to filter out these effects in an attempt to recognize a basic pattern. Second, the utterances have to convey the same meaning and, most importantly, they have to be performed with the intention of achieving the same illocutionary act. As is well known, segmentally identical utterances can be used to ask a question, give a command, make an ironic comment, express surprise, etc. All of these different functions affect the prosodic packaging and hence have to be controlled for when searching for

prosodically identical utterances. Third, the utterances have to be produced in identical (or very similar) circumstances, e.g. as casual remarks between adolescents, in a working environment between people of different status, etc.

With regard to the number of “same” utterances needed for a detailed prosodic analysis, there are the following rough guidelines. The absolute minimum for recognizing a pattern with some degree of reliability is three instances, because with only two versions of the same utterance it is difficult, if not impossible, to decide what is distinctive and what coincidental with regard to those aspects where they diverge. A good start with a detailed analysis can be made with four versions of the same utterance, ideally two by male speakers and two by female speakers. With eight different versions, statistical analyses become more viable and useful. With 10–12 speakers, the sample size approaches that which is found in much work on well-documented languages such as English or Japanese.

There is no principled upper limit for the sample size and, depending on the phenomena being investigated, larger samples may become necessary which also take into account variables such as age, register, and local dialect. But to repeat, in the typical field setting of a hitherto undocumented language spoken by a small number of speakers, samples of four to ten versions of the same utterance will provide a good basis for a detailed prosodic analysis and will thus greatly improve the databasis for prosodic research.

Note also that, while preferable, it is not absolutely necessary that the different versions are produced by different speakers. They could also have been produced by the same speaker(s) on different occasions. Importantly, more or less immediate repetitions of the same utterance (such as when asking the speaker to repeat something she just said or to say something twice) usually do not produce multiple versions of the same utterance in the intended sense, because repetition usually has some impact on prosody.

It should be obvious that even in a very large corpus of recordings of more or less spontaneous speech it will be difficult to find a set of four to ten versions of the same utterance in the intended sense. There may be hundreds or even thousands of utterances one may reasonably safely identify as polar questions (e.g. *Is he coming tonight?*). But how many of these will be segmentally identical or at least very similar? Furthermore, the circumstances in which the question is asked may not be really comparable. All of which makes it difficult to determine those aspects in the prosodic packaging that are related to categorical distinctions. To be sure, in the case of polar questions, it may be possible to determine these aspects with a rea-

sonable degree of certainty on the basis of a sufficiently large sample from spontaneous speech. But it is more cumbersome to do this only on the basis of such a sample and it may become more and more difficult to do it when investigating more complex issues. In particular, when investigating problems in the prosodic packaging of information structure (focus, contrast, deaccenting, etc.), the number of variables to be controlled and accounted for may become so high that all results remain speculative.

Ideally, then, a comprehensive language documentation should contain sets of different versions of the same utterance, each set representing a different major function where prosody may be of relevance (i.e. one set for polar questions, one for all-new utterances, one for polite commands, and so on). While such sets may happen to occur in a sufficiently large corpus of spontaneous recordings without paying particular attention to the topic of prosodic analysis, there are three ways to ensure that they are in fact represented in the documentation.

First, work with prompting tools such as video clips, a picture story, or matching games where one speaker instructs another to identify an object among a set of similar objects or to find a path through an imaginary landscape (the so-called “map task”) will produce similar, if not truly identical utterances.<sup>7</sup> Particularly useful are games where speakers engage in different types of speech acts (e.g. asking a question, giving directions, confirming a suggested solution), provided that the structure of the game forces speakers to talk about the same “world” (i.e. to use the same lexical items) so that the utterances become reasonably similar with regard to their segmental make-up.

The second method to produce relevant data sets is to try direct elicitation by asking speakers to produce utterances or, more precisely, mini-discourses prepared in advance. The major problem here is how to present the target utterances in such a way that the prosody is not influenced by the prompt. We will look at the prompting problem more closely in Section 4. Here are a few examples of the kind of sentences one may want to try to elicit with an indication of the prosodic function they target given in square brackets:

- (1) Has X arrived? No, I haven't seen seen him/her/them yet.  
[polar question-answer pair]
- (2) (In the market:) What are you looking for? (I am looking for)  
vegetables. [question word – question-answer pair]

- (3) Have a seat, please! [polite command]
- (4) (Group of people standing at road side, obviously agitated.  
What happened?):  
A bus turned over!  
or: The dog killed a pig! [all-new utterances]
- (5) I like the **blue** shirt, not the **red** one. [contrastive focus]
- (6) Have you ever eaten a black snake? No, I don't eat snakes.  
[deaccenting]<sup>8</sup>
- (7) (Surprise:) How big you are already! [speaker attitude]

This list of examples is not complete and should be expanded and adapted as required by the project setting and make-up. However, since eliciting such examples will usually not be an easy task and not something which native speakers will be very eager to do, one should plan to spend considerable time on drafting the right set of examples and to test all of them with one close collaborator before approaching a larger number of speakers for a recording.

One consideration in drafting the examples is segment structure. Examples should include as few fricatives as possible and in general should avoid voiceless consonants of all manners of articulation. The ideal example in fact consists only of like vowels and nasals, which of course is an ideal that will hardly ever be attainable when attempting to construct examples which make sense and are culturally appropriate. Having semantically and pragmatically well-formed and culturally appropriate utterances will in general be the more important concern since otherwise the elicitation will not work at all.

The third way of getting comparable data sets for prosodic analysis is to make sure that the corpus of recordings contains a sufficient number of utterances using stylized intonation. A typical example of an everyday use of stylized intonation is a calling or vocative contour (Ladd 1996: 88, 136 f.). There may be different calling contours, for example, one for calling someone (“Peter!”), one for market cries, one used by street-vendors for advertising fish, and so on. In many languages, listing items (e.g. *they had lots of cows, goats, chicken, and dogs*) also involves a special, somewhat stylized intonation (“listing intonation”; see also next section). Otherwise, stylized intonation is a common feature of many forms of ritual speech, in particular of the so-called chanted speech.

For purposes of prosodic analysis, the main advantage of stylized intonation contours consists in the fact that by its very nature, intonational contrasts are more stable and more marked than in non-stylized contours. Consequently, patterns are generally much more easily recognizable. In fact, while native speakers often do not have very clear intuitions about non-stylized intonation patterns, they often know about stylized contours and can readily imitate them.

Obviously, patterns used in stylized intonations differ from those used in non-stylized ones and similarly, it may be the case that intonation patterns in elicited examples differ quite clearly from those found in spontaneous speech (compare the phenomenon of “reading intonation” found in many European languages). In this regard, it should be clearly understood that elicited and stylized data sets have the function of allowing one to get started on prosodic, and specifically intonational, analysis. They enable the investigator to get a basic idea of what kind of contrasts are being made in the language and thus to develop hypotheses that have to be tested with the spontaneous material. A comprehensive prosodic analysis of course has to be able to account for the full range of phenomena found in a corpus of spontaneous recordings.

### 3. Recording words

It is a widespread practice in linguistics to record lists of elicited words in order to be able to check one’s transcriptions and to document the basic sound structure of lexical items. The format usually used in such recordings is first to give the translation equivalent of the word in the contact language being used (or the number of the word in a word list) which is then followed by the word in the documented language, often repeated once or twice. In this way, words are recorded “in isolation,” which is often understood to mean “in their most basic form, free from any ‘contaminating’ contextual influences.” This, however, is a misconception, since uttering an isolated word always constitutes a minimal utterance, which is of particular import for prosody. Importantly, “words in isolation” do not only display whatever lexical prosodic features they might have (lexical tone or accent) but also features of (usually declarative) utterance prosody. This may appear to be a rather trivial point, but even in the specialist literature this distinction has not been made consistently until fairly recently.<sup>9</sup>

As an example, compare Figures 2 and 3. Figure 2 shows the waveform and F0 tracing for a single Waima’a word, *kaluha* ‘cloud’, recorded in isolation. Figure 3 shows the waveform and F0 tracing for a short Waima’a utterance, *kii baa ini* ‘there are people fighting’ (lit. ‘people hit each other’; an all-new response to a *what’s-going-on* type of question). Note that the F0 tracing is essentially identical in both figures: it starts out flat at mid-level,<sup>10</sup> rises and begins to fall again on the penultimate syllable, and continues to fall on the last syllable. Hence, the question arises whether the rise on the penultimate syllable in *kaluha* is part of the lexical make-up of this item, reflecting at least in part a regular lexical accent on the penultimate syllable. Alternatively, this rise-fall on the last two syllables – which can be observed for practically all Waima’a lexical items uttered in isolation – is due exclusively to the fact that uttering a Waima’a word in isolation also involves the utterance-level features of a standard Waima’a declarative utterance. (At the time of writing this chapter, I believe that the latter option is correct, but this needs further research and testing. For current purposes, it is not relevant which of the two options turns out to be correct. The point to be clearly understood is that words in isolation always and by necessity display features of utterance-level intonation.)

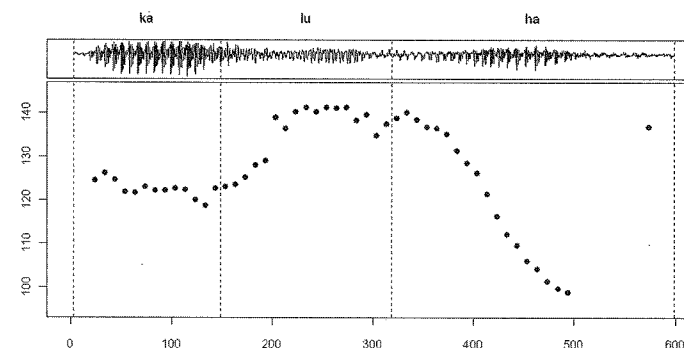


Figure 2. Waveform and F0 for Waima’a word in isolation (*kaluha* ‘cloud’)



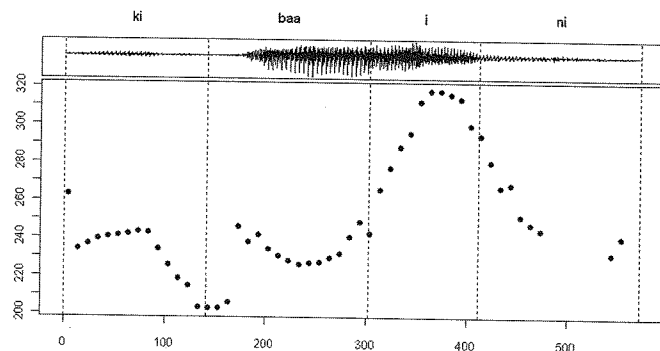


Figure 3. Waveform and F0 for Waima'a short utterance (*kii baa ini* 'there are people fighting')

In order to be able clearly to separate lexical and post-lexical (utterance-level) prosodic features, it is now a common practice in research on prosody (but also in many segmental phonetic studies) not to record words in isolation even when lexical features are the primary concern. Instead, the ideal is to record the target word(s) in different positions in a carrier phrase, as in the following English examples:

(8) The target word *America* in different positions in a carrier phrase

- a. "America" is a word I know. [initial position]
- b. I said "America" once. [phrase-internal position]
- c. She said "America." [final position]

As seen in these examples, the different position will usually involve different information structural implications, which may, but do not have to, correlate with post-lexical prosodic distinctions. Furthermore, since the purpose of these recordings is to compare characteristics of different lexical items, the carrier phrase usually involves very general items, in particular verbs such as "say", "hear", or "know (a word)" which in principle are compatible with all lexical items.

Figures 4 and 5 illustrate the effect of carrier phrase position with another example from Waima'a. Here, the target word *aboo* 'grandparent, old/respected person' occurs at the end of a carrier phrase (*ne ehe aboo* 'she said *aboo*') and at the beginning of another one (*aboo aku de nau* '[the word] *aboo* I don't know').<sup>11</sup> Note how the change in position correlates with a clear change in pitch (rise-fall on *boo* in final position, late rise on

*boo* in initial position). But note also what remains constant in both positions. Most importantly, in both instances *boo* is roughly twice as long as the initial syllable *a*. Consequently, it may be hypothesized that *boo* contains a long vowel as part of its lexical make-up and that the fact that this syllable is long in both recordings is not due to an utterance-level effect.

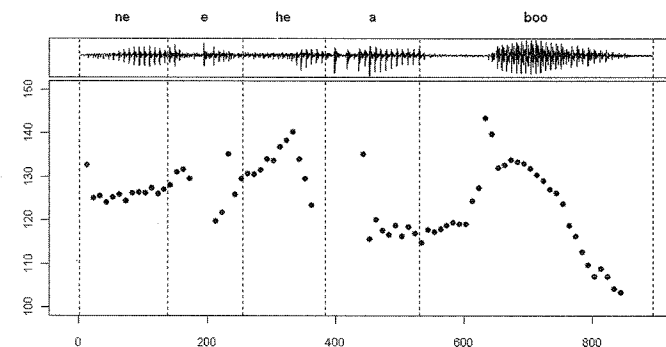


Figure 4. Waima'a carrier phrase with final target word (*ne ehe aboo* 's/he said *aboo*')

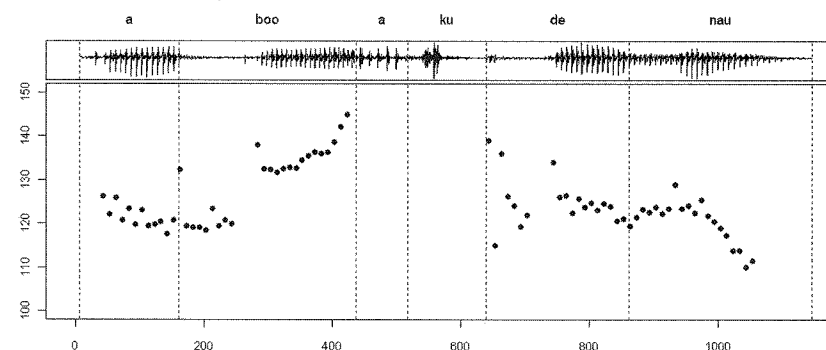


Figure 5. Waima'a carrier phrase with initial target word (*aboo aku de nau* 'aboo I don't know')

If working with carrier phrases proves to be too cumbersome or does not work for some other reason (see next section), one may try to record words in mini-lists of three to four items, alternating the position of the words contained in the list, as in (9).

## (9) Mini-lists with alternating orders

- a. *America*, Africa, Antarctica
  - b. Africa, Antarctica, *America*
  - c. Antarctica, *America*, Africa
- etc.

While not as useful as recordings in carrier phrases, such mini-lists often allow one to make at least a distinction between final and non-final utterance prosody, provided that the speakers actually use listing intonation and do not simply produce three isolated utterances in rapid sequence. As in the carrier-phrase example, prosodic features which remain the same across different positions in the list can be hypothesized to be lexical rather than post-lexical.

#### 4. The prompting problem

Most of the procedures discussed in the preceding two sections involve the elicitation of prosodic data by asking speakers to produce various kinds of utterances or mini-discourses. While elicitation quite generally may involve problems with regard to the naturalness and reliability of the data thus obtained, elicitation of prosodic data is particularly prone to major distortions since prosodic features are highly susceptible to contextual influences. Thus, there is little use in presenting the items to be recorded simply by having native speakers repeat what the researcher or one of her local co-workers says. In almost all circumstances, this will produce highly distorted utterances which will largely imitate features of the presented utterance or display the prosodic characteristics of a repeated utterance.

The most widely-used procedure in prosodic research on languages with a well-established writing tradition is to have speakers *read* the target utterances. This procedure, while not directly influencing the prosody by providing a model for imitation, may encounter a number of other problems. Most importantly, the reading tasks require that the speakers actually enact the intended utterance type. Obviously, there is little use in having someone read a question or a surprised exclamation in a rather flat, non-engaged monotonous voice. Not all speakers are capable or willing to engage in such a performance. Successful reading prompts also presuppose that the speakers are reasonably fluent in reading the language. This will often not be the case even in those communities where speakers are literate in a

dominant language but not used to seeing their own language written (reading in such circumstances will be slow and in a word-by-word style). Another complication may arise from the fact that reading intonation differs significantly from conversational intonation.

In non-literate societies, written prompts obviously will not work at all. The main alternative here is to try various kinds of role-playing or the experimental tasks involving video clips, etc., already mentioned above in Section 2. Role-plays may work when carefully prepared with a local team member. They involve speakers pretending to be in a given situation and reacting with an appropriate short utterance rehearsed in advance. Thus, for example, one may ask a pair of speakers to pretend meeting one another in the market, one asking *what's happening there?* and the other responding with the target utterance *people are fighting*. In the best of circumstances, the speakers engaged in this role-play will actually engage in a short conversation, continuing this imagined question-answer pair with a short sequence of further utterances. It will often not be possible to make them use exactly the target utterance prepared in advance, but minor variations in its segmental make-up will usually not cause major problems for comparability. The more realistic the role-playing is, the better the quality of the prosodic data produced in this way will be.

In preparing role-plays and experimental tasks it should be kept in mind that these will in all likelihood be very strange kinds of communicative events for native speakers who are not familiar with the basic idea of role-playing, experiments, or interviews. Thus, one has to be prepared to face quite a few obstacles when trying to collect data in this way. Continuous laughing or giggling because of the unusualness and unnaturalness of the situation is one very common problem. Speakers may also change the speech act, i.e. rather than responding with a statement (“He has gone to the market”) they may produce a command (“Go to the market!”). Furthermore, it is not uncommon that speakers who are asked to retell a short action sequence in a video clip comment on the kind of dress people are wearing or the color of the sky visible in the clip instead of engaging with the given task. Considerable time and ingenuity in developing appropriate prompts may thus be required in order to make experimental tasks work or to develop useful forms of role-playing in a given community. But this effort will be well spent because the data generated in this way will be very useful not only for prosodic analyses, but often also for other types of analysis.

## 5. Perception experiments

The procedures presented so far in this chapter all focus on production data, i.e. sets of utterances which can be analyzed acoustically and auditorily. But production data do not provide any basis for determining which components of the complex signal are actually perceived as prosodically distinctive by native speakers. It is well known from research on European languages that not all the distinctive information available in the acoustic data is perceived as such by native speakers. Consequently, there is a need for data to answer questions such as: Is this clearly observable prominence (e.g. a change in pitch direction, or increased loudness or duration) actually perceived by native speakers? Is it perceived at the location where it is observed in the signal? Which of the major phenomena observed for lexical accents is perceived as distinctive: pitch, duration, length, or vowel quality? The most straightforward way to answer such questions is to run perception experiments. In such experiments, the prosodic parameters observed in a set of utterances are modified and sets of modified utterances (or sets of modified and unmodified utterances) are then evaluated by native speakers. For example, loudness on a lexically accented syllable could be reduced and then it could be tested whether the syllable is still perceived as prominent. Or, the final rise in a question utterance could be reduced or shifted to an earlier syllable and then tests could be run to determine whether the utterance is still perceived as a question.

It is not an easy task to prepare and run perception experiments of this type and to date, very few perception experiments have been reported for languages outside Europe and Japan.<sup>12</sup> In some ways, the easiest part is the preparation of stimuli since speech analysis tools such as *EMU*, *PRAAT*, *Wave Surfer*, or *Speech Analyzer* allow for a relatively easy and straightforward modification of pitch and other prosodic parameters in digitized utterances. The more difficult part is to find a way of how to run the tests, especially in societies which have little or no experience with experiments.

That is, perception experiments are also faced with the prompting problem. Problems here may already arise because speakers may refuse to put on a headset (which is the best way of ensuring that they can listen carefully to the stimuli). But the main challenge consists in defining a manageable task. It will usually not be possible to ask directly for the identification of prosodic properties with questions such as: Where is the major prominence? Is X higher than Y? etc. Instead, what may work are tasks which involve some kind of comparing and ranking two different items, asking

questions such as: Which of these two items is more natural/more appropriate/more often heard? Which item would you use when speaking in public? or the like. Otherwise, general comments on the stimuli (such as: “this sounds rather odd or foreign”; “that’s how people down south speak”; etc.) may also provide important clues, although they will make for a very heterogeneous and difficult to quantify dataset.

In this regard, it should be noted that non-experimental, conversational data may sometimes also provide important clues as to which prosodic features are perceived as relevant in a given speech community. A somewhat trivial, but nonetheless relevant example is the fact that a conversational corpus allows one to collect a set of examples of utterances which are treated by the interactants as questions and to compare these to utterances which are prosodically similar but are not taken up as questions by the listener. More complex are examples where a misplaced emphasis or wrong intonation contour produces a misunderstanding, leading to a repair sequence. See the contributions in Couper-Kuhlen and Selting (1996) and Couper-Kuhlen and Ford (2004) for relevant observations and examples.

## Acknowledgements

I would like to thank Bruce Birch and Bob Ladd for extensive discussion of the issues and ideas presented in this chapter. They, of course, are not responsible for the use I have made of their input. Thanks also to Jan Strunk for help with plotting the figures.

Very special thanks to Maurício C.A. Belo, my Waima’a collaborator, who has patiently suffered through various trial runs of the procedures discussed in this chapter as well as some other procedures which have been found not to be productive. Further information and full acknowledgements for the Waima’a project can be found at <http://www.mpi.nl/dobes/WebpageDobes1/SubpagesTeams/SubpageWaimaa/Frameset.htm>.

This work was made possible by a research professorship funded by the Volkswagen foundation and I am most grateful for this very generous support.

## Notes

1. Features defining a good recording are listed in Section 2.1 of Chapter 4.
2. Examples of what can be done – and what cannot be done – in terms of prosodic analysis on the basis of a corpus of recordings alone are King's (1994) and Bishop's (2002) theses on the intonation of Dyirbal and Bininj Gun-wok, respectively. King's thesis is exclusively based on tape recordings of narrative and procedural texts made by R.M.W. Dixon in the 1960s and 70s. On the basis of this material, King is able to make a proposal for some key features of Dyirbal intonation. However, at various points she has to take note of the fact that the available genres (mostly narrative) severely limit the scope of her analysis. Furthermore, she notes that much of her analysis remains speculative as long as it is not possible to test whether the prosodic distinctions she establishes on the basis of acoustic data alone are actually also perceived as significant distinctions by Dyirbal speakers. For perception, see also Section 5 below.
3. "Spectral characteristics" here refers to those aspects of the formant structure of speech sounds which reflect prosodic features, e.g. the energy distribution across the frequency spectrum, which may be an acoustic correlate of stress.
4. Figure 3 below includes a very clear illustration of this effect in that the /b/ of *baa* causes a noticeable "dip" in the F0 contour. Laver (1994: 452–456) provides a fuller discussion of microprosodic perturbations.
5. The tracings are given in two versions, the right-hand column presenting the original F0 extractions, the left-hand column a somewhat smoothed version of these. See Grabe (1998: Chapter 2) for further information on the procedures used in collecting and processing the data. This thesis is available at <http://www.phon.ox.ac.uk/~esther/thesis.html>.
6. The precise details of the analysis are of no concern here. Note that Grabe (1998: Chapter 3, Section 2) makes the proposal that the nuclear fall in Northern Standard German allows for two major alternative realizations, one with a clear rise on the accented syllable and one where pitch is more or less level in the accented syllable (as with JN in Figure 1). The distinction between these two (phonetic) realizations of the same phonological category is argued to be gradual.
7. Further references and links for prompting tools can be found in Chapter 6 and on the book's website.
8. Examples (4)–(6) all target distinctions of information structure, a rather complex topic which cannot be adequately dealt with here. See Lambrecht (1994) and Jacobs (2001) for a thorough discussion of some of the basic distinctions and issues involved, Ladd (1996) for the role prosody may play in marking information structure, Drubig (2003) for a typological survey, and Dimroth (2002) for an elicitation task targeting information structure.
9. Bruce (1977) is widely considered the first modern work where the distinction is fully and consistently applied.
10. As mentioned above, the initial "dip" in Figure 3 is a microperturbation caused by the /b/ in *baa*. The utterance in Figure 2 is by a male speaker, the one in Figure 3 by a female speaker and therefore overall considerably higher. Wavefiles containing the utterances of Figures 2–5 are available at the book's website.
11. The speaker, of course, knows the word *aboo*, but putting it in initial position and not using a negation (i.e. using the equivalent of 'aboo I know' as a prompt) was not felt to be appropriate.
12. Most recent work in this field has been done by researchers associated with the Phonetics Laboratory, Universiteit Leiden Centre for Linguistics, mostly on languages of Indonesia, in particular Malay. See Ebing (1997), Odé (1997, 2002), van Zanten et al. (2003), and Stoel (2005: 108–208) for examples and references. These works also provide detailed discussion as to how prosodic experiments can be devised and administered. There is also a fair amount of work being done on the perception of prosodic differences between Russian dialects by a group of researchers associated with the Bochum *Linguistic Lab* (<http://www.ruhr-uni-bochum.de/lilab/Index.htm>).