

## Transcription, orthography and other levels of representation

Friederike Lüpke  
F12@soas.ac.uk



### Some distinctions

- Transcription
  - Records (all? some?) properties of spoken language
- Writing system
  - Refers to the structural and functional characteristics of the system
- Script
  - Refers to the set of symbols used (roughly corresponding to characters)
- Orthography
  - Determines how symbols are used (= spelling)

### Why is a writing system not a transcription?

- Speech and writing are different in fundamental aspects.
- Writing tends to be more conservative and prescriptive than speech.
- Writing is always 'reducing' a language (Pike 1947).
- (Native) readers don't need transcriptions.
- It is questionable whether (advanced) readers phonologically recode what they read.

### Differences between speech and writing (Coulmas 2003: 11)

- |                           |                    |
|---------------------------|--------------------|
| ■ Speech                  | ■ Writing          |
| – Continuous              | – Discrete         |
| – Bound to utterance time | – Timeless         |
| – Contextual              | – Autonomous       |
| – Evanescent              | – Permanent        |
| – Audible                 | – Visible          |
| – Produced by voice       | – Produced by hand |

## Transcription

5

### How to represent (what of) speech

- How can speech be transcribed and represented?
- How do we represent prosodic properties of speech?
- How do we decide on word boundaries and clause boundaries?
- (How) do we represent gesture, direction of eye gaze and other non-verbal signs complementing the speech signal?

6

## Transcription

Jaminjung (Mindi family, N. Australia, data from Eva Schultze-Berndt):

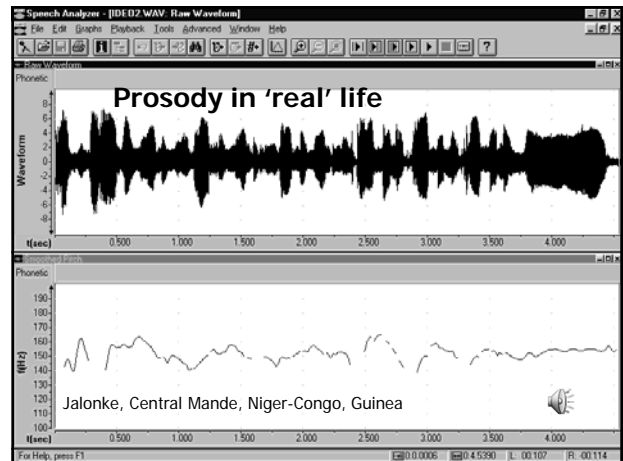
Orth: bugu mulurrng ganamany

Pros: i \n \# † e ' \z \- Q  
= n h , h ' h Ì # c

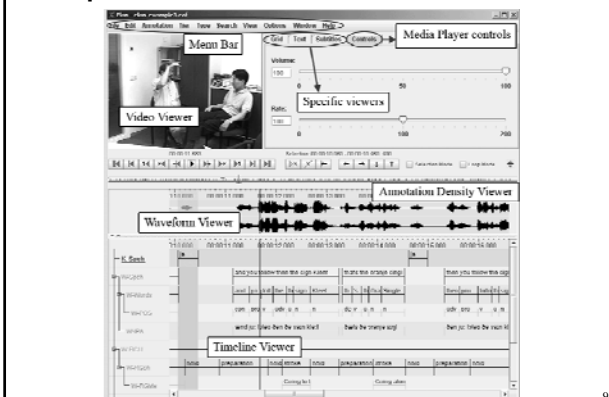
bugu dibard garumany  
yinthuwurlawung

i \n \# # † e k p i h ¾  
= n h ô Ì ' h Ì # † e

q L , W \ ^ E , z h = ^ \ Q c  
it just crashed on the ground  
(and he) just came jumping over here'



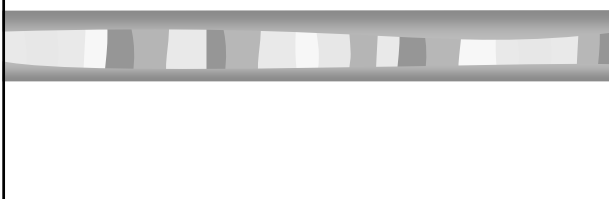
## Multiple tiers for different levels of information



## Segmentation

- For transcription of spoken language, a division into intonation units may be more appropriate than a division into clauses and sentences “as in written language”
- Intonation unit – preliminary definition:
  - coherent intonation contour
  - contains at least one pitch accent
  - delimited by boundary intonation (a rising or falling pitch movement)
  - prototypically, but not always, delimited by a pause (cf. e.g. Cruttenden 1997, Halliday 1985, Chafe 1994, Ladd 1996)

## Graphisation



## Reading and writing are social practices



Students of a Q'uranic school in Bamenda, Northwestern Cameroon

- The linguist's illusion: only linguistic considerations are underlying writing systems, scripts and orthographies.
- The reality: many different and conflicting factors influence writing system choice
  - Sociolinguistic considerations
  - Historical considerations
  - Religious considerations
  - Political considerations
  - Psycholinguistic considerations

## Di- and multigraphia

### ■ Digraphia:

- Sometimes used to describe the existence of more than one script for one language
- Sometimes used preserving the hierarchical relationship contained in the related term diglossia.
- My solution: use of digraphia in analogy to diglossia.



Multilingual and –graphic inscription in Kaba, Mali.

### ■ Multigraphia:

- Used by me to designate the existence and use (synchronously or diachronically) of more than one script for one language.

13

## Exographia



Letters in the Jalonke village Saare Kindia in Guinea are always written in French or Fula Ajami, never in Jalonke.

- "I use the term exographia to designate writing in a speech community taking place exclusively in another language. Exographia is very widespread in endangered and minority languages, for which often no written variety is available at all. It is often the case that an official language (often the ex-colonial one) occupies formal writing contexts and a regional lingua franca is used for writing in semiformal and informal contexts, such as adult literacy campaigns, the writing of personal letters, etc." (Lüpke, in press 2010)

14

## Should endangered languages be written?

### ■ PRO:

- "Graphocentric" (Blommaert 2004) ideologies entail that writing carries enormous prestige and symbolic value, independently of practical use.
- Seifart (2006): argues that only an orthography makes a documentation accessible to the community.

### ■ CONTRA:

- Graphisation and the creation of a written environment and written genres are enormous tasks of uncertain outcome in multilingual and or exographic communities.
- In the age of audio- and video documentation, communities do not need to rely on written documentation for accessibility.

15

## Orthography choice

16

## Common assumptions

- "The place of writing systems in the study of language planning and language policies is often seen as secondary. The various questions related to writing, such as the choice of writing systems, the type or orthography, etc., are often understood as being obvious, based on two main assumptions: first, that the Latin script is the most suitable to form the base of a new writing system; and second, that a writing system should be phonemic. However, these answers are mainly based on linguistic observations, without much concern for the place and role of a writing system in society. (Grivelet 2001: 1)

17

## Linguistic considerations

- How close to the phoneme inventory aims the orthography to be?
- Should it preserve morpheme identity (German <reisen> vs. <reist>) or output of morphophonological processes (Dutch <reizen> vs. <reijst>)?
- Should it reflect suprasegmental properties of speech?

These questions can only be answered based on properties of individual languages, i.e. the number of phonemes, of morphophonological processes, its agglutinative vs. inflectional nature, the number and functional load of tone, etc...

18

## A note on the representation of tone

- A number of studies (Bird 1999a, 199b, 2001) investigate the impact of tone marking on reading and writing fluency in the Grassfields Bantu language Dschang (Cameroon) and in African languages in general.
- The main finding:
  - Tone is either radically under-represented or marked in a shallow way based on phonemic analysis and following predefined regional orthographic conventions.
  - These conventions are not generally mastered by readers and writers of the languages in question.
- His recommendations:
  - Identify the function and depth of tone.
  - Practice selective under-representation.
  - Endeavour to maintain fixed word-images while using only accessible information.
  - Test the options and analyse errors.

19

## Other design considerations

- Accommodation of variation vs. standardisation
  - Selection of a standard
  - Underspecification to allow for the neutralisation of (certain) dialectal differences (see Seifart 2006 for details)
- What are the proportions of reading and writing that are anticipated?
  - Are people going to read more than write, or vice versa?
  - What media are going to be made available for reading?
  - What technologies are anticipated for writing?
- How can an existing script be adapted to a new language?
  1. Creation of new graphemes;
  2. Neutralisation of contrasts of the spoken language in writing;
  3. Appropriation of existing graphemes.

20

## Identity issues

- The choice of a script may be crucial for expressing a particular identity:
  - Speakers of the African lingua franca Fula have a longstanding writing tradition using Ajami.
  - However, writers of Fula in the Futa Tooro region of Senegal prefer a Latin-based script to write Fula, in order to differentiate themselves from the very successful Wolof Ajami tradition (Wolofal).
- Just like writing systems and scripts, particular graphemes may play an important role for communities, as they symbolically reflect their distinct or shared identity.



Tifinagh letter yaz.

21

## Practical matters



Participants of the Tostan UNICEF Jokko initiative in Senegal.

- Despite seemingly unlimited technological possibilities for encoding fonts, the practical obstacles remain huge.
- In many cases, the only high-tech context in which minority languages are used in writing are mobile phones.

22

## Resources and impact

- A successful orthography requires a large investment, including:
  - Selection of a writing system and set of graphemes;
  - Establishment of rules specifying the relationships between sounds and graphemes;
  - Determinations of rules specifying word boundaries and spelling;
  - Production of a dictionary listing spellings and materials for learning and later independent reading.



Page from a Jalonke primer.

Hasty orthography developments have a lasting impact and can be detrimental to future written language use.

23

## Good documentation has great potential for the future

- A good documentation with a transparent transcription whose rules are documented has a greater potential for future language and corpus planning than orthography development that lacks the necessary manpower and resources.

24

**A radical new method: BOLD**

- Basic Oral Language Documentation (BOLD, Bird 2010):

Stage I: Audio capture

10 h of oral texts from a variety of genres and contexts, captured on digital recorder.

Stage II: Oral annotation

1 h of selected recordings respoken and orally translated onto another recorder (10h).

Stage III: Transcription

0.1 hour of selected recordings transcribed and translated into notebook (10 hours).