

## Readme explaining contents and structure of dataset:

### Dinka\_LuanyjangDialect\_Jan-May2007\_controlled\_nounsSGandPL

**Language:** Dinka

**Dialect:** Luanyjang variant within the Rek cluster

**Date and place of recordings:** January to May 2007, Edinburgh

**Equipment:** Marantz PMD solid-state recorder; Shure SM10A headset-mounted microphone

**Data types:** 1096 sound files in wav format at a sampling frequency of 44.1 kHz. Also 1 pdf file: nounnumber\_dataset.pdf

**Speaker:** CAM (Luanyjang speaker number 13)

**Speaker details:** male; about 41 years old at the time of the recording; originally from the village of Makuac, not far from Wuncuei; he spent the first 20 years of his life primarily in Makuac; after that he has lived mostly in Khartoum, but also periods in Egypt and Britain; he is married to another speaker of the same dialect (Luanyjang); Dinka is his first language; he also speaks Arabic and English, both fluently.

**Recorded by:** Bert Remijsen (University of Edinburgh)

**Permissions:** The speaker has agreed for the speech data to be used for research, and to be disseminated. Agreement is recorded through signature on an consent form.

**Description:** The data are elicited Dinka utterances: collected through controlled elicitation, using English to convey the utterances to be translated to the speaker. In most cases, the target word is a noun, elicited in one or more of a small number of sentence frames. These noun data were recorded so that listening to them and inspecting their acoustic representations could be used to help in determining the correct phonological transcription. The resulting transcriptions of noun forms are recorded in the table in nounnumber\_dataset.pdf. This pdf is included in this dataset, and also includes nouns for which the transcription was determined purely based on auditory impressions. The nouns in nounnumber\_dataset.pdf form the basis of a published paper: Ladd, D. Robert, Bert Remijsen & Caguor A. Manyang (2009). On the distinction between regular and irregular inflectional morphology: evidence from Dinka. *Language* 85(3), 659-670. This archived dataset also includes evidence a) the construct state of some nouns, as it presents itself on nouns modified by a possessor or a demonstrative; b) forms from the paradigm of the verb {mel} 'soil, make dirty'.

### **Structure of the file name:**

The file name encodes several pieces of information, through fields that are separated by an underscore (\_). Fields 1,2,3 invariably contain the same type of information, as does the final field. The information in fields 4 to n-1 depends primarily on the value of field 3.

Field 1: dialect code. In this case, L = Luanyjang (dialect variety level); R = Rek (cluster level)

Field 2: speaker number. In this case, all data come from C.A. Manyang, coded as Luanyjang speaker number 13.

Field 3: utterance type. The information in fields 4 to n-1 depends primarily on the value of this field.

EXIST stands for existential predicate. /a-dee X/ 'There is/are X.', with X being a noun. If field 3 is EXIST, then fields 4 and 5 specify the English translation of the noun and its inflection, respectively. In field 5 S or SG stands for singular and P or PL for plural.

ISOL stands for the citation or isolation form. If field 3 is ISOL, then fields 4 and 5 specify the English translation of the noun and its inflection, respectively. In field 5 S or SG stands for singular and P or PL for plural.

VA stands for verb+argument. If field 3 is VA, then fields 4 and 5 specify the English translation of the verb and its inflection, respectively, and fields 6 and 7 specify the English translation of the internal-argument noun and its inflection, respectively. Following a verb in field 4, field 5 is most commonly PAST, the past-tense infinitive construction, or ZERO, the unmarked present-tense form. But for {mel} 'soil', I also elicited verb forms inflected for subject (1S, 2S, 3S, 1P, 2P, 3P), the infinitive construction for negation (NEG), the non-topical subject in present tense (NTS), the non-topical subject in past-tense infinitive (NTSPAST), the passive (PASS), and an infinitive form with continuative or habitual meaning (CONT).

Field n: the last field in the file name is the repetition number.