Dataset containing binominal lexemes in Harakmbut (isolate, Peru), for "The derivational use of classifiers in Western Amazonia" and "When the alienability contrast fails to surface in adnominal possession: Bound nouns in Harakmbut"

An Van linden, University of Liège & KU Leuven

Department of Modern Languages: Linguistics, Literature & Translation, University of Liège
Place Cockerill 3 - 5, 4000 Liège, Belgium
an.vanlinden@uliege.be
ORCID ID: 0000-0001-7369-9850

1. PURPOSE OF THE DATA

The table uploaded here was compiled to supply data to Steve Pepper's PhD project "Binominal word-formation in the world's languages", which resulted in Pepper (2020). It contains data from Harakmbut, a Peruvian Amazonian language spoken in the regions of Cusco and Madre de Dios, more specifically of the Amarakaeri/Arakmbut variety. These were partly collected in the field, and partly collected at a distance with the help of one of my language consultants, Yesica Patiachi Tayori, from the Puerto Luz community. For a description of the language, the reader is referred to Van linden (2023).

Subsequently, the table was used to provide quantitative data on binominal constructions for Harakmbut in the article "The derivational use of classifiers in Western Amazonia" by Rose & Van linden (2022).

Finally, the quantitative data on one specific type of binominal lexeme in Harakmbut, viz. noun-noun compounding constructions, that were used in Rose & Van linden (2022) were also used in Van linden's (Forthcoming) article "When the alienability contrast fails to surface in adnominal possession: Bound nouns in Harakmbut". Specifically, the latter article reports on statistical analyses of the 15 noun-noun compounding constructions in the dataset. The table is hence intended primarily for use as a data file accompanying this latter article.

2. CONTENTS AND ORGANIZATION OF THE FILE

The table contains one sheets, called HAR_binominals, comprising the data supplied to Steve Pepper. Its first five datafields (columns A to E) were requested by Pepper, for which he provided the following instructions and examples.

Meaning

This is the meaning to be translated. Also given in Russian, Spanish and French.

Translation equivalent

1. For each meaning give the canonical **translation equivalent** (TE) using the Latin script (or IPA); if no equivalent exists, leave the field blank
2. Choose the **most common** translation equivalent
3. If two translation equivalents are equally common, supply either one but prefer one that is **analysable** to than one that is mono-morphemic

TE (non-Latin script)

4. For non-Latin writing systems, provide the word in the **native script** (see the RUSSIAN example)

Gloss (complex words only)

5. For TEs consisting of more than one morpheme provide a gloss.
6. If the TE does not contain any **polymorphemic words**, simply supply the gloss: e.g. for FRENCH 'railway' (*chemin de fer*) enter "**way of iron**"
7. For TEs in which **one or more words are polymorphemic**, repeat the translation with word-internal morpheme breaks indicated by a period, and add the gloss in square brackets: e.g. for GERMAN 'railway' (Eisenbahn) enter "**eisen.bahn [iron.way]**"
8. Only words that are synchronically analysable should be glossed
9. Use recommended abbreviations from the **Leipzig Glossing Rules** wherever possible, except:
10. Use a period instead of a hyphen to mark morpheme breaks (see the examples)
11. Use a colon instead of a period when a single object-language element is rendered by several metalanguage elements (see the BEZHTA example)

Notes

12. Put any comments regarding the source of loans, calques, etc. in this column

EXAMPLES

Meaning (ENG)	Translation equivalent	TE (non-Latin script)	Gloss (complex words only)	Notes
railway	kil.os hinu		iron.OBL:GEN way	BEZHTA
railway	chemin de fer		way of iron	FRENCH
railway	eisenbahn		eisen.bahn [iron.way]	GERMAN
railway	železnaja doroga	железная дорога	žel.naja doroga [iron.ADJZ road]	RUSSIAN
railway	reli		-	SWAHILI
railway	železnica		žel.n.ica [iron.ADJZ.NMLZ]	SLOVAK

Further datafields in HAR_binominals:

Column	Header	Description
F	AVL	My personal notes; T1995 refers to Tripp's 1995 dictionary
G	analysis into classes of morphemes	Schematic structure of the translation equivalent in terms of classes of morphemes, such as bound noun, nominalizing prefix, etc (see abbreviations below); an underscore indicates a word boundary; a hyphen indicates a morpheme boundary
H	type of binominal lexeme in Rose & Van linden 2022	Classification of the translation equivalent according to the nine-way typology in Pepper (2020: 145-169) as analysed by Rose & Van linden 2022, which is sometimes different from Pepper's (2020) analysis (column K)
I	type of N-N	For N-N compounds, it is indicated whether the component nouns are bound nouns (b) or independent nouns (i)
J	counted by Pepper 2020 as binominal lexeme	See Pepper (2020)
K	Suptype of binominal lexeme in Pepper (2020: 479)	See Pepper (2020)

The table contains 78 translation equivalents for 76 entries of the list of 100 complex concepts designed by Pepper (2020: 391-392). Entries for which two translations were analysed were given a second row in the table. 29 translation equivalents are binominal lexemes, 15 of which are N-N compounds.

Values used in the type of binominal lexeme in Rose & Van linden 2022 category (column H):

clf	"thing-root with a classifier, where the denotatum of the binominal is different from that of the base (the classifier is used to derive a new meaning rather than for classification)" (Rose & Van linden 2022: 241)
cmp	"compounding of two thing-roots in a single word" (Rose & Van linden 2022: 241)
der	"derivation from a thing-root with a thing-affix that contributes some semantic content" (Rose & Van linden 2022: 241)
gen	"head and modifier are independent lexemes, with an additional word-class preserving morpheme attached to the modifier" (Rose & Van linden 2022: 241)
loan	Loan word
simple	Monomorphemic word
V-based	Verb-based lexeme

Abbreviations used in the glosses (column D) and analyses into classes of morphemes (column G):

1PL	first person plural
ADJ	adjective
ADV	adverbial (as a syntactic label; function of adjunct)
AN	animate
BN	bound noun
CLF	classifier
DIM	diminutive
GEN	genitive
LOC	locative
N	independent noun
NMLZ	nominalizing prefix
PAPA	past participle
SOC	sociative causative
SPAT	spatial linking element
TRNS	transitivizer
V	verb base (can be larger than just V root)
VBZ	verbalizer
VPL	verbal plural

REFERENCES

Pepper, Steve. 2020. *The typology and semantics of binominal lexemes. Noun-noun compounds and their functional equivalents*. Oslo: University of Oslo PhD dissertation.

Rose, Françoise & An Van linden. 2022. The derivational use of classifiers in Western Amazonia. In Steve Pepper, Francesca Masini & Simone Mattiola (eds.), *Binominal lexemes in cross-linguistic perspective: Towards a typology of complex lexemes* [Empirical Approaches to Language Typology 62], 237-276. Berlin: De Gruyter Mouton. [https://doi.org/10.1515/9783110673494-008]

Tripp, Robert. 1995. *Diccionario amarakaeri-castellano*. Yarinacocha: Min. de Educación & Instituto Lingüístico de Verano.

Van linden, An. 2023. Harakmbut. In Patience Epps and Lev Michael (eds.), Amazonian Languages, An International Handbook. Language Isolates, Volume 1: Aikanã to Kandozi-Shapra [Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 44/1], 441-481. Berlin: de Gruyter Mouton. [https://doi.org/10.1515/9783110419405-010]

Van linden, An. Forthcoming. When the alienability contrast fails to surface in adnominal possession: Bound nouns in Harakmbut. Special Issue "Re-assessing the explanatory potential of alienability contrasts", guest-edited by Françoise Rose & An Van linden. *Linguistics – An Interdisciplinary Journal of the Language Sciences*. [https://doi.org/10.1515/ling-2022-0039]