README for GaVal, represented in a text version and an Excel version, with the same content.

GaVal is a valence lexicon for Ga, a Kwa language spoken in the area of Accra, Ghana. The lexicon is a curated version of a verb valence dictionary developed by Prof. Mary Esther Kropp Dakubu, through a valence-extended version of a Toolbox file underlying the dictionary Dakubu (2009). The classification system used in the verb valence dictionary, called ‘Construction Labeling’ (CL), was presented in Hellan and Dakubu (2010). CL was also used for the labeling of verb entries in the lexicon for a computational grammar of Ga, called ‘GaGramm’ (cf. Hellan 2020), which uses the LKB platform designed for grammars based on HPSG (cf. Copestake 2002, Pollard and Sag 1994), and where the conversions of the verb valence Toolbox files to LKB format were made by Bruland 2011. Off from the LKB lexical format, the verb entries were in turn mapped to the current format of GaVal, now as a verb valence catalogue, as described in Hellan (2023).

GaVal covers valence entries for 498 verb lexemes. A verb lexeme can in general have more than one frame, and GaVal identifies each such frame hosted by a given verb by an entry specific to that frame and verb, called a 'lexically instantiated valence frame, abbreviated as 'lexval'; the lexval entries represented in GaVal constitute in total 1980 entries.

Each lexval entry in GaVal consists of a syntactic valence frame, semantic role specifications for the arguments in the frame, a situation type licensing the roles, and a morphologically annotated example. The entry also specifies the standard written form of the lexeme instantiated by the lexval and the standard phonological representation of the lexeme. As entry identifier for each lexval is entered the written form of the lexeme together with a number indicating the place of the entry in an enumeration of the lexvals. An English gloss is provided in each lexval, mostly indicating the lexeme meaning in general, but also indicating subsenses when these are instantiated by the frames. In the morphological glossing of the accompanying example sentence, the semantic gloss of a verb may correspond to a contextually induced meaning, and the free English translation of the sentence may even more reflect the contextually appropriate English rendering of the sentence rather than maintaining a word-by-word rendering.

The items of a lexval are represented under the following labels, both it in a text string in the text version and in the ordering of columns in the excel-version:

lexval-id

gramm-valence

sem-val (roles and situation type)

gramm-valence and-sem-val combined

lexeme-with-phon

Engl-lex-gloss

example-sentence

morph-gloss

free-transl

Further notes

GaVal uses the same notion of 'lexval' as the Norwegian valence catalogue NorVal (cf. Hellan 2021, 2023). The main differences from NorVal are:

(1) The number of lexemes in GaVal is only about 6% of the number of lexemes in NorVal.

(2) GaVal includes specification of semantic roles and situation types, none of which appears in NorVal except for location and directionality.

(3) GaVal provides both an English gloss for each lexval and a glossed sentence example. NorVal provides an example sentence, but it is not glossed or translated, nor is there an English word gloss for a lexval or a lexeme.

(4) GaVal so far does not formally represent a 'valpod' (i.e., the set of lexvals defined relative to a lexeme), as NorVal does.

In content, what GaVal perhaps comes closest to is the English VerbNet (cf. Korhonen and Briscoe 2004). Compared to VerbNet, GaVal has a richer inventory of semantic roles, and its inventory of situation types is about the size of VerbNet’s ‘verb classes’, although not quite with the same content.

References

Dakubu, M. E. Kropp, 2009. *Ga-English Dictionary with English-Ga Index.* Accra: Black Mask Publishers

Bruland, T. (2011). Ga\_verb\_dictionary\_for\_digital\_processing. Accessible at: https://typecraft.org/tc2wiki/Ga\_Valence\_Profile.

Hellan, L. and M.E. Kropp Dakubu. 2010. *Identifying verb constructions cross-linguistically*. In *Studies in the Languages of the Volta Basin* 6.3. Legon: Linguistics Department, University of Ghana.

Hellan, L. (2020). A computational grammar of Ga. LREC 2020, workshop RAIL.

Hellan, L. (2022). A Valence Catalogue for Norwegian. In Loukanova, R. (ed) *Natural Language Processing in Artificial Intelligence - NLPinAI 2021*. Springer. <https://doi.org/10.18710/8U3L2U>.

Hellan, L. (2023). A Unified Cluster of Valence Resources. In: *Logic and Algorithms in Computational Linguistics 2021* *(LACompLing2021)*. Springer Nature 2023 ISBN 978-3-031-21780-7.

Korhonen, A., & Briscoe, T. (2004). Extended Lexical-Semantic Classification of English Verbs. In *Proceedings of the HLT/NAACL Workshop on Computational Lexical Semantics,* Boston, MA

Pollard, C. & Sag, I. (1994) *Head-Driven Phrase Structure Grammar*. Chicago University Press.