# Combining WordNets with Treebanks to study idiomatic language: A pilot study on Rigvedic formulas through the lenses of the Sanskrit WordNet and the Vedic Treebank

**Luca Brigada Villa,**° **Erica Biagetti,*** **Riccardo Ginevra,**[+] **Chiara Zanchi***

°University of Pavia/Bergamo, *University of Pavia, [+]Università Cattolica del Sacro Cuore (Milan)

## Abstract

This paper shows how WordNets can be employed in tandem with morpho-syntactically annotated corpora to study poetic formulas. Pairing the lexico-semantic information of the *Sanskrit WordNet* with morpho-syntactic annotation from the *Vedic Treebank*, we perform a pilot study of formulas including SPEECH verbs in the *RigVeda*, the most ancient text of the. Sanskrit literature.

## 1 Introduction

The *Sanskrit WordNet* (SWN; Hellwig 2017)[1] is currently under construction in the framework of an international project carried on by the University of Pavia, the UCSC of Milan, the University of Exeter, and the Center for Hellenic Studies at Harvard University, which aims to build a family of WordNets (WNs) for ancient Indo-European (IE) languages. The family additionally comprises WNs for Ancient Greek[2] and Latin (Biagetti et al. 2021). These WNs are designed to be interoperable with each other and with other WNs for modern languages, as well as linkable to external resources (see also Zanchi et al. 2021). Furthermore, these WNs bring together WN relational semantics with semantic theories of Cognitive Linguistics, while introducing a number of innovations to the WN architecture to account for the specificities of ancient IE languages (Biagetti et al. 2021).

By means of a case study employing the SWN, this paper shows how WNs can be employed in tandem with morpho-syntactically annotated corpora to study poetic formulas, and more generally idiomatic expressions of ordinary language. Building on the methodology by Zanchi et al. (2022), we develop a pilot study on the *RigVeda* (RV), the most ancient text of Sanskrit literature, composed in the so-called Vedic variety. [3] To extract formulas with different degrees of schematicity, we pair the lexico-semantic information of the SWN with the morpho-syntactic annotation of the *Vedic Treebank*. The Vedic Treebank (VTB, Hellwig et al. 2020) is a morpho-syntactically annotated corpus of Vedic literature, tagged according to the Universal Dependencies formalism (Nivre et al. 2016).

The paper is structured as follows. Sec. 2 introduces the background. Sec. 3 explains our methodology. In Sec. 4, we show and discuss our results. Sec. 5 concludes the paper and draws future lines of research.

## 2 Formulas as constructions

### 2.1 The path toward a constructionist approach to formularity

By investigating South Slavic oral epic poetry, M. Parry (1971[1928]) and A. B. Lord (1960) demonstrated that the *Iliad* and the *Odyssey* are examples of oral poetry: these poems result from online composition during bards' performances, and their written versions are secondary. Within this research, Parry gave a first definition of *formulas* in oral poetry as "traditional fixed expressions regularly employed in fixed metrical

---

conditions to express a given essential idea”. As later stressed by Lord, formulas are organized in larger scenes and narrative themes to be productively manipulated by mature bards to continuously re-build poetry in their performances.

Since Parry and Lord's seminal work, research on formularity has flourished. Different investigations have granted major emphasis to the semantic aspects or the formal constrains of formulas (see, among many others, Nagy 1974, Nagler 1976, Watkins 1976, 1995, Russo 1963, 1966, Hainsworth 1968). Notably, all the studies mentioned so far look at formulas as a phenomenon *sui generis*. Kiparsky (1976) first proposed a unified account for formulas and idioms of ordinary language. He distinguished flexible/deep-structure formulas (1)a vs. bound phrases (1)b (or idioms/ready-made surface formulas):

(1)  a. *The X-er, the Y-er*
    b. *It takes one to know one*

Admittedly, Kiparsky did not prove that flexible formulas and bound phrases belong to two discrete categories, but meanwhile, from his generative perspective, it was not possible to settle these types along a *continuum*.

Bozzone (2014) and Pagán Cánovas and Antović (2016; see also Antović and Pagán Cánovas 2016) found a solution to this issue, by identifying usage-based linguistics, and Construction Grammar in particular, as a theoretical framework that allows providing a definition of formulas that accounts for their functional and formal components and handles their gradience. In Construction Grammar (e.g., Fillmore and Kay 1993, Goldberg 1995), constructions are understood as learned pairings of form and function, just as formulas. In this view, lexicon and syntax arrange along a *continuum*, varying for their degree of abstractedness and complexity. Lexically filled formulas, partially filled formulas, lexically empty formulas, and fully schematic syntactic structures (such as the transitive construction) are all constructions, which can be arranged along the lexicon-syntax *continuum*.

This definition of formulas, accounting for their semantic and formal flexibility, suits well the Rigvedic formulaic style: the form of the hymns relies on the tradition of preceding poets, but at the same time Vedic poets stress the novelty of their poems. As Biagetti (forthc.) puts it, “this tension between tradition and innovation is mirrored in continuous and conscious variations in expressing traditional themes” (see Sec. 4.1).

## 2.2 A case study on Ancient Greek

Zanchi et al. (2022) adopted this approach to perform a case study on the Iliadic KILL and SPEECH formulas. They enhanced F. Mambrini's *Universal Dependency* conversion of the *Ancient Greek Dependency Treebank,* [4] containing the Homeric poems, with the *Ancient Greek WordNet* [5] synsets for KILL and SPEECH. Specifically, they automatically annotated the relevant verbal lemmas with the synsets v#00903723 “cause to die; put to death”, v#00652168 “use language”, v#00554194 “reply or respond to”, v#00608227 “address a question to and expect an answer from”, and v#00696790 “greet by a prescribed form”. Then, by means of a Python script employing the Udapi package,[6] they extracted the relevant pattern from the enhanced treebank: a transitive construction with some additional restrictions concerning the relative position of its elements and their occurrence within a single Homeric verse: $\text{obj}_{\text{accusative}}$ ptc X $\text{verb}_{\text{finite}}$ $\text{atr}_{\text{nominative}}$ $\text{subj}_{\text{nominative}}$.[7] The analysis of the extracted occurrences confirmed that this syntactic and metrical configuration is frequently – but not exclusively – employed to express two basic ideas, that is, KILL and SPEECH. The output verses make up a family of formulas, whose members share some – but not necessarily all – functional and/or formal features with the other members of the family, as exemplified by (2). The verses in (2)a-b share their basic idea, SPEECH, but their formal realization is different: the verb in (2)a occupies the 4th position in the verse, whereas the verb in (2)b occurs in the third place. Instead, (2)a and (2)c convey two distinct basic ideas, SPEECH and KILL, but are formally more similar: the initial accusative is followed by a particle and a connective; then a third person singular aorist

form occurs; the nominative subject modified by two attributes concludes the verse. Finally, the verse in (2)d is formally closer to (2)a and (2)c than to (2)b (the verb occurs in exactly the same position as in (2)a and (2)c, but is preceded by a participle and not by a connective), but conveys a further basic idea: THINK. Traditionally, the verses in (2)a-d are not treated as belonging to a single family of formulas, despite their evident similarities.

(2)  Il.24.668, 1.121, 22.376, 11.599

|   | obj | ptc | X | verb | atr | subj |
|---|-----|-----|---|------|-----|------|
| a. | *tòn* | *d'* | *aûte* | *proséeipe* | *podárkēs dîos* | *Akhilleús* |
| b. | *tòn* | *d'* | -- | *ēmeíbet'* | *podárkēs dîos* | *Akhilleús* |
| c. | *tòn* | *d'* | *epeì* | *exenárixe* | *podárkēs dîos* | *Akhilleús* |
| d. | *tòn* | *dè* | *idṑn* | *enóēse* | *podárkēs dîos* | *Akhilleús* |

# 3   Data and methods[8]

## 3.1   The Vedic Treebank

Our initial data comes from the Rigvedic section of the VTB, [9] which is currently only partially annotated for syntax. Since elements of the formulas are linked to each other by syntactic relations, we needed a fully annotated treebank to extract the relevant patterns. Thus, we matched the syntactically annotated portion of the treebank with silver annotation produced by an automatic parser for Vedic, and obtained a fully annotated version of the RV.[10]

## 3.2   Enhancing the VTB with synsets

To check whether it is possible to extract formulas as pairings of form and function/basic idea, we further annotated the treebank with synsets. Similarly to Zanchi et al. (2022), we chose three synsets for SPEECH (CALL, ASK, SAY) and automatically assigned one of them to each relevant verbal lemma occurring in the treebank.[11] Furthermore, since Rigvedic hymns are mainly devoted to praising the gods of the Vedic pantheon,

we automatically added the synset DEITY to proper names of all such gods, to check whether they constitute the main addressees of the SPEECH verbs under investigation (see the Appendix for the list of synsets and associated lemmas).

## 3.3   Extraction of the formulas

The extraction consisted of two phases: initially, we focused on trigrams involving at least a SPEECH verb. We noticed that most trigrams involved an obj, an adverbial clause modifier in the dative case (advcl), and optionally a subj, in addition to the SPEECH verb. We thus focused on patterns involving these four elements: verb, obj, advcl, and optionally subj.

We further enriched the treebank with metric information of all the sentences in which an advcl modifier in the dative case occurred. To do so, we added a feature "PositionInVerse" to the MISC field of the conllu file, which can take one of two values: Initial or Final.[12] To extract the patterns, we used UDeasy (Brigada Villa 2022), a tool for querying treebanks.

| Nodes | verb | upos=VERB |
|-------|------|-----------|
|  | obj | deprel=obj |
|  | subj (optional) | deprel=nsubj |
|  | advcl (optional) | deprel=advcl\| advcl:fin |
| Relations | verb governs all the other nodes in the query | |

Table 1: Query employed for data extraction

As shown in Table 1, we extracted patterns consisting of four nodes, in which subj and advcl were optional elements and, together with obj, had to depend syntactically on the verb. In addition, we restricted the results to those patterns involving a verb whose synset was CALL, ASK or SAY.

---

[8] Data employed for this study are available at the following GitHub repository: https://github.com/unipv-larl/rv-formulas.
[9] https://github.com/OliverHellwig/sanskrit/tree/master/papers/2020l rec/treebank.
[10] The automatic parsing of the RV was performed by Oliver Hellwig and can be found at the following GitHub repository: https://github.com/OliverHellwig/sanskrit/tree/master/dcs/data/conl lu/files/Ṛgveda. In order to recognize sentences annotated by the parser, we added a feature SyntaxAnnotation=silver to the MISC field of the conllu file.

[11] Since formulas convey a "given essential idea", in this case study we were not interested in capturing all the different senses of each verb, but rather in detecting all formulas conveying the basic idea of SPEECH. Therefore, we assigned one single synset to each verb based on its first meaning in the Monier-Williams Sanskrit Dictionary.
[12] Rigvedic verses (*ślokas*) are divided into text lines (*pādas*); different verses can be distinguished based on the number of *pādas* they contain and on the number of syllables of each *pāda*. When taking metric information into account, in this phase we did not focus on the number of syllables nor on syllable lengths, but simply on the position of verb, obj, advcl and subj in each *pāda*.

# 4 Results

## 4.1 Rigvedic constructions

Composed and transmitted orally for centuries, the RV did not follow the same principles of oral composition as we know it from Homeric epic: its compositional technique makes little use of the metrically defined and invariant formulas (ready-made surface formulas; Kiparsky 1976:83) that are common in Homeric poetry. As our results confirm, the RV rather consists of a texture of schematic (deep-structure) formulas, which are variously instantiated in the text due to, e.g., lexical or grammatical substitution and metrical variation (Jamison and Brereton 2014:14, cf. Jamison 1998).

As noted by Nagy (1974: 196), metrical patterns seem to result from the crystallization of phraseology, i.e., idiomatic expressions, which are known to display restricted syntax (Croft and Cruse 2004:290). We thus started our inquiry by looking at the most common orders for the elements obj, verb and advcl, and then analyzed each pattern with respect to the position of its elements in the verse. We found three patterns to be the most frequent ones:

1. obj, verb, advcl (25x)
2. obj, advcl, verb (24x)
3. verb, obj, advcl (16x)

For reasons of space, we exclusively discuss pattern 1. We arrange constructions along a *continuum* from more schematic morpho-syntactic structures to metrically- and lexically-fixed formulas, with the latter inheriting formal and semantic properties from the former (on inheritance, see Goldberg, 1995: 70-81).

## 4.2 Formulas with different degrees of schematicity: obj, verb, advcl constructions

We found the syntactic order obj, verb, advcl to occur 25x with verbs for CALL/SAY, always with an animate object referring to the addressee, as in (3)a. Most of these occurrences (21x) are instances of a metrically-fixed construction, in which advcl is always found in verse-final position, as in (3)b. This construction may be further analyzed according to two lexico-semantically specified subtypes: a more frequent pattern (19x) with a DEITY as obj (addressee) and forms of *hvā-/brū-*[13]

as verb, as in (3)b1, and a less common pattern (2x) with a 1.Sg/Pl pronoun referring to POETS as obj and forms of *vac-/ah-* as verb, as in (3)b2. The former construction deserves further attention.

| (3) | obj | verb | advcl |
|---|---|---|---|
| a | ANIMATE | CALL/SAY | Dat |
| b | ANIMATE | CALL/SAY | Dat, verse-final |
| b1 | DEITY | *hvā-/brū-* | Dat, verse-final |
| b2 | 1Sg/Pl.POET | *vac-/ah-* | Dat, verse-final |

For the construction b1 with a DEITY as obj and a verse-final advcl, we observed three more metrically- and lexically-fixed patterns, as displayed in (4); in all three, the obj may be both preceded and followed by an optional slot (X) of $n$ syllables ($\sigma$).

(4) Constructions inheriting from b1

| | X | obj | X | verb | X | advcl |
|---|---|---|---|---|---|---|
| b1.1 | $n_\sigma$ | INDRA/ DEITY | $n_\sigma$ | *hvā-/brū-* $(2/3_\sigma)$ | -- | *ūtaye*$(3_\sigma)$, verse-final |
| b1.2 | $n_\sigma$ | DEITY | $n_\sigma$ | *hvā-*, verse-final | $n_\sigma$ | *ūtaye/ somapītaye/ svastaye*, verse-final |
| b1.3 | $n_\sigma$ | INDRA | $n_\sigma$ | *hvā-* $(2/3_\sigma)$, verse-initial | $n_\sigma$ | ACQUISITION$(3_\sigma)$, verse-final |

In construction b1.1, which occurs 9x in lines such as (5), the obj may have INDRA or another DEITY as referent. The construction is characterized by a bi- or tri-syllabic form of the verb *hvā-* or *brū-* directly preceding the advcl *ūtaye* 'for help', which occupies the last 3 syllables of the verse.

(5) Instances of the b1.1 construction
a. ***tám tvā**ᵒᵇʲ havíṣmatīr víśa úpa **bruvata**ᵥₑᵣᵦ **ūtáye**ₐdᵥcl*
'Upon you the clans, offering oblations, call for help.' (RV 8.6.27ab)
b. ***indravāyū́**ᵒᵇʲ manojúvā víprā **havanta**ᵥₑᵣᵦ **ūtáye**ₐdᵥcl*
'Indra and Vāyu, mind-swift, do the inspired poets call for help.' (RV 1.23.3ab)

Construction b1.2 and b1.3 both occur in sequences composed of two verses. Construction b1.2 occurs 5x in examples like (6). The former verse has any DEITY as the obj and ends with a form of the verb *hvā-*, whereas the latter verse always ends with one of the three advcl *ūtaye* 'for

---

[13] The citation form for Vedic verbs is the root followed by a hyphen (cf. the root *hvā-* 'call' and the 3Pl form *havanta* 'they call'). The

citation form for nouns is the stem followed by a hyphen (cf. *ūtí-* 'help' with the dative form *ūtáye* 'for help').

help', *somapītaye* 'for the drinking of soma', and *svastaye* 'for well-being'.

(6) Instances of the b1.2 construction
a. **víśvān devā́n**<sub>obj</sub> **havāmahe**<sub>verb</sub>
   *marútaḥ* **sómapītaye**<sub>advcl</sub>
   'The All Gods we call, the Maruts, for soma-drinking.' (RV 8.23.10ab)
b. *ihá_**indrā́ṇī́m**<sub>obj</sub> úpa **hvaye**<sub>verb</sub>
   *varuṇānī́ṁ* **suastáye**<sub>advcl</sub>
   'Here I call upon Indrāṇī, Varuṇānī for well-being.' (RV 1.22.12a)

Construction b1.3 occurs 3x in two-verse sequences like (7). In the former verse the obj always has INDRA as one of its referents (lexically realized either by a pronoun, as in (7)a, or by a specialized epithet, as in (7)b), whereas the latter verse starts with a bi- or tri-syllabic form of the verb *hvā-* and ends with a trisyllabic word for ACQUISITION as advcl.

(7) b1.3
a. *índrāvaruṇa* **vām**<sub>obj</sub> *aháṁ*
   **huvé**<sub>verb</sub> *citrā́ya* **rā́dhase**<sub>advcl</sub>
   'Indra and Varuṇa, I invoke you two for brilliant bounty. (RV 1.17.7ab)
b. **ugrám**<sub>obj</sub> *pūrvī́ṣu pūrvyáṁ*
   **hávante**<sub>verb</sub> **vā́jasātaye**<sub>advcl</sub>
   'They call on (you) the strong, foremost among the many (peoples), for the winning of prizes.' (RV 5.35.6cd)

| LEMMA | SYNSET | N |
|---|---|---|
| *ūtí-* (22), *ávas-* (8), *adhivāká-* (1), *gopīthá* (1) | PROTECTION n#00522858 | 32 |
| *vā́jasāti-* (6), *sātí-* (3), *dhánasāti-* (1), *rā́dhas-* (1), *gṛbh* (1) | ACQUISITION n#00045827 | 13 |
| *sakhyá-* (10) | FRIENDSHIP n#10038317 | 10 |
| *svastí-* (6), *saúbhaga-* (1) | WELL-BEING n#10366086 | 7 |
| *sómapīti-* (3), *pītí-* (2) | DRINKING n#00540820 | 5 |
| *rayí-* (2) | WEALTH N#9614312 | 2 |
| *mṛḍīká-* (1), *sumná-* (1) | FAVOUR n#05575676 | 2 |
| *sadhástuti-* (1) | PRAISE n#05018478 | 1 |
| *śvetanā́-* (1) | WHITENING n#00176075 | 1 |
| *nirṇíj-* (1) | RAIMENT n#02212047 | 1 |

Table 2: Synsets of lemmas exployed as advcl.

### 4.3 Many expressions, same basic ideas

We analyzed all lemmas employed as advcl and observed that most are synonyms sharing the same synset (see Table 2). The most frequent synset is PROTECTION (n#00522858 "the activity of protecting someone or something"), mostly instantiated by the lemma *ūtí-* 'help, protection' (22x), followed by *ávas-* 'assistance, protection' (8x). Expressions with either term may thus be considered the core of this construction, whereas expressions with *adhivāká-* 'advocacy, protection' and *gopīthá-* 'protection', both occurring only once, seem to belong to its periphery.

Further frequently recurring synsets are ACQUISITION, FRIENDSHIP, WELL-BEING and RITUAL, with ACQUISITION attesting to a high degree of lexical variation: *sātí-* and its compounds *vā́ja-sāti-* and *dhána-sāti-* belong to the core, whereas *rā́dhas-* and *gṛbhá-* are more peripheral.

Notably, *pītí-* 'drink' and its compound *sóma-pīti-* 'soma drinking', together with *sadhá-stuti-* 'joint praise' and *śvetanā́-* 'whitening (of dawn)' instantiate WN's well-known "tennis problem", that is, the impossibility to capture semantic solidarity between lemmas sharing membership in the same topic of discourse (Fellbaum, 1998: 10-11). In this specific case, the ritual drinking of soma and the joint praise were part of a Vedic ritual taking place at dawn. Thus, in the constructions under investigation, the four lemmas employed as advcl all have the function of calling the gods to take part in the ritual.

## 5 Conclusion and future work

With this case study, we showed the potential of employing WNs in tandem with other language resources to study idiomatic expressions. Pairing the lexico-semantic information of the SWN with morpho-syntactic annotation contained in the VTB, we were able to extract poetic formulas involving a SPEECH verb in the RV, and to detect recurring pairings of form and meaning at various levels of schematicity. In the future, as the SWN grows, we intend to add semantic annotation to the entire VTB. Furthermore, the same approach may be applied to the study of idiomatic expressions in everyday language by combining information contained in WNs and treebanks of modern languages.

## Acknowledgements

## References

Mihailo Antović and Cristóbal Pagán Cánovas (eds.). 2016. *Oral Poetics and Cognitive Science*. Berlin/Boston, de Gruyter.

Erica Biagetti, Chiara Zanchi and William M. Short. 2021. Toward the creation of WordNets for ancient Indo-European languages. In *Proceeding of the 11th Global WordNet Conference*. University of South Africa (UNISA): Global Wordnet Association, pages 258–266. https://aclanthology.org/2021.gwc-1.30

Erica Biagetti. Forthc. Integrare Sanskrit WordNet e Vedic TreeBank: uno studio pilota sulla formularità del Rigveda tra semantica e sintassi. In Isabella Bossolino and Chiara Zanchi, *Prospettive sull'antico. Decennalia dei Cantieri d'Autunno*. Pavia, Pavia University Press.

Chiara Bozzone. 2014. *Homeric Constructions*. PhD thesis, University of California, Los Angeles.

Luca Brigada Villa. 2022. UDeasy: a Tool for Querying Treebanks in CoNLL-U Format. In *Proceedings of the Workshop on Challenges in the Management of Large Corpora* (CMLC-10), pages 16–19.

William Croft and D. Alan Cruse. 2004. *Cognitive Linguistics*. Cambridge, Cambridge University Press.

Christiane Fellbaum (ed.). 1998. *WordNet: An electronic lexical database*. MIT Press, Cambridge, MA.

Charles J. Fillmore and Paul Kay. 1993. *Construction grammar coursebook*. Berkeley, University of California.

Adele E. Goldberg. 1995. *Constructions: a Construction Grammar Approach to Argument Structure*. Chicago, Chicago University Press.

John B. Hainsworth. 1968. *The Flexibility of the Homeric Formula*. Oxford, Clarendon.

Oliver Hellwig. 2017. Coarse semantic classification of rare nouns using cross-lingual data and recurrent neural networks. In *IWCS 2017-12th International Conference on Computational Semantics-Long papers*, Montpellier, France.

Oliver Hellwig, Salvatore Scarlata, Elia Ackermann, and Paul Widmer. 2020. The Treebank of Vedic Sanskrit. In *Proceedings of The 12th Language Resources and Evaluation Conference* (LREC 2020), pages 5137–5146.

Stephanie W. Jamison 1998. Rigvedic *viśvátaḥ sīm*, or, Why syntax needs poetics. In Jay Jasanoff, H. Craig Melchert and Lisi Oliver, *Mir curad: Studies in honor of Calvert Watkins*, Innsbruck, Innsbrucker Beiträge zur Sprachwissenschaft, pages 291–299.

Stephanie W. Jamison and Joel P. Brereton. 2014. *The Rigveda. The Earliest Religious Poetry of India*. Oxford, Oxford University Press.

Paul Kiparsky. 1976. Oral Poetry: Some Linguistic and Typological Considerations. In Benjamin A. Stolz and Richard Stoll Shannon (eds.), *Oral Literature and the Formula*. Ann Arbor, Center for Coordination of Ancient and Modern Studies, pages 73-106.

Albert B. Lord. 1960. *The Singer of Tales*. Cambridge, MA, Harvard University Press.

Michael N. Nagler. 1967. Towards a Generative View of the Homeric Formula. *Transactions of the American Philological Association* 98:269–311.

Gregory Nagy. 1974. *Comparative Studies in Greek and Indic Meter*. Cambridge, MA, Harvard University Press.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter et al. 2016. Universal Dependencies V1: A Multilingual Treebank Collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, Portorož, Slovenia. European Language Resources Association (ELRA), pages 1659–1666.

Cristóbal Pagán Cánovas and Mihailo Antović. 2016. Formulaic creativity: Oral poetics and cognitive grammar. *Language & Communication* 47:66–74.

Parry, Milman. 1971 [1928]. The Traditional Epithet in Homer. In Adam Parry (ed.), *The Making of Homeric Verse: The Collected Papers of Milman Parry*. Oxford, Oxford University Press, pages 1–190.

Joseph Russo. 1963. A Closer Look at Homeric Formulas. *Transactions of the American Philological Association* 94:235–247.

Joseph Russo. 1966. The Structural Formula in the Homeric Verse. *Yale Classical Studies* 20: 217-240.

Calvert Watkins. 1976. Answer to P. Kiparsky's Paper: Oral Poetry: Some Linguistic and Typological Considerations. In Benjamin A. Stolz and Richard S. Shannon (eds.), *Oral Literature and the Formula*. Ann Arbor, Center for Coordination of Ancient and Modern Studies, pages 107–111.

Calvert Watkins. 1995. *How to Kill a Dragon: Aspects of Indo-European Poetics*. New York and Oxford, Oxford University Press.

Chiara Zanchi, Silvia Luraghi and Erica Biagetti. 2021. Linking the Ancient Greek WordNet to the Homeric Dependency Lexicon. In *Computational Linguistics and Intellectual Technologies*. *Papers from the Annual International Conference "Dialogue"*, Vol. 20:729-737.

Chiara Zanchi, Luca Brigada Villa, and Andrea Farina. 2022. *Toward combining Ancient Greek WordNet and AGDT2 for linguistic research: A pilot study on formulas of Iliad*. Paper presented at the 3rd International Colloquium on Ancient Greek Linguistics, Universidad Autónoma de Madrid, Spain, 16-18 June 2022.

## Appendix

Table 3 contains synsets and their respective lemmas as they were added to the VTB.

| SYNSET | LEMMAS |
|---|---|
| v#00501506 "utter in a loud voice or announce" | *hvā-, vac-, brū-* |
| v#00608227 "address a question to and expect an answer from" | *yāc-, pracch-* |
| v#00652168 "use language" | *vad-, ah-* |
| n#06861622 "any supernatural being worshipped as controlling some part of the world or some aspect of life or who is the personification of a force" | *deva-, indra-, agni-, varuṇa, aśvin-, vāyu-, marut-, mitra-, savitṛ-, sūrya-, uṣas-, aditi-, rudra-, viṣṇu-* |

Table 3: Synsets and their associated lemmas.