The *Homeric Dependency Lexicon*

What it is and how to use it

Chiara Zanchi | ORCID: 0000-0001-9216-9986

University of Pavia, Pavia, Italy

chiara.zanchi@unipv.it

Abstract

This paper presents the *Homeric Dependency Lexicon* (HoDeL), a new resource with a user-friendly interface facilitating the study of Homeric verbs and dependents. HoDeL was induced from the analytical layer of AGDT 2.0, extracting all dependents tagged as SBJ, OBJ, PNOM, and OCOMP with a set of SQL queries. The paper illustrates HoDeL functionalities and shows how they can be employed by researchers to answer specific research questions about the Homeric language. Introducing the uses of HoDeL offers the opportunity to reexamine some crucial, though frequently underestimated, methodological challenges concerning annotated corpora and resources derived from them that relate to the linguistic theories underlying annotations and error propagation. It is argued that the careful documentation of how linguistic resources were created, what data they contain, and how they can be queried through their dedicated interfaces is essential to lay the groundwork for users' investigations.

Keywords

Homeric Greek – HoDeL – dependency treebanks – AGDT 2.0 – valency lexica

1 Introduction

I present here HoDeL (*The Homeric Dependency Lexicon*), a new linguistic resource facilitating the investigation into Homeric verbs and their dependents. Introducing the uses of HoDeL also offers an opportunity to reexamine some crucial, though frequently underestimated, methodological challenges that arise when one uses electronic linguistic resources to study ancient and modern languages.

Thus, my aim here is two-fold: (a) to fully document the construction of HoDeL and explain its basic functionalities with concrete examples; (b) to thoroughly account for the ways in which methodological challenges raised by electronic linguistic resources may affect HoDeL users. On the one hand, a detailed explanation of the functionalities of these resources is crucial for users: as Anthony (2013: 142) has emphasized, data may yield different results owing both to the observers' subjectivity and also to the idiosyncrasies of the tool or interface through which data is observed. On the other hand, it is important to acknowledge that (i) annotated corpora and linguistic resources induced from them inevitably do contain errors and that (ii) annotations based on particular underlying linguistic theories should be followed by careful notes alerting users to the theoretical foundations that underlie the annotations. This is critical, as linguistic resources induced from annotated corpora distance the researcher from the researched data on at least at three different levels: (a) the interface level, (b) the data-extraction level, and (c) the annotation level.

I structure my presentation as follows. Section 2 contains the background framing HoDeL: it discusses the notions of corpus, treebanks, and valency lexica, and reviews the linguistic resources available for the study of Ancient Greek (henceforth, AG). Section 3 presents HoDeL: after providing details on the architecture of dependency treebanks, it accounts for HoDeL's construction and describes its basic functionalities. Section 4 illustrates the types of theoretical and practical issues that HoDeL users may encounter with specific exemplars and it provides some suggestions for how the lexicon can be used productively. Section 5 contains concluding remarks.

2 Background: linguistic resources and valency lexica

Ancient languages such as AG enjoy a long tradition of literary, philological, and linguistic research, based on the manual collection of data from written texts, in many cases preserved up to the present by accidents of history (Joseph & Janda 2003: 15–19). Because this data derives from historical records of ancient languages, they are known as *Korpussprachen* 'corpus languages' (see, among others, Mayrhofer 1980; Untermann 1983; Haug 2015: 187; Eckhoff et al. 2018b: 300); in this context, the term *corpus* is broadly understood as "a body of naturally occurring language" (McEnery et al. 2006; for a narrower definition of *corpus*, see, e.g., Sinclair 2005). In such a long tradition of corpus studies, the digital turning point premiered with Father Busa's *Index Thomisticus*, a pioneering electronic collection of all words contained in Thomas Aquinas' *opera omnia* (Busa 1980; Nyhan & Passarotti 2019). From the second half of the 1960s,

Father Busa undertook the task of systematically collecting Thomas Aquinas' words, initially on punched cards and later on magnetic tapes. The printed version of the *Index* appeared in 1980 and consisted of 56 volumes, and a CD-ROM version was released later in 1989. Today, a morphosyntactically annotated version of Thomas Aquinas' corpus is maintained at Università Cattolica del Sacro Cuore of Milan and is being integrated with other linguistic resources for Latin within the *LiLa* Project. Since Father Busa's time, the efforts to digitize various linguistic and non-linguistic materials have increased, and they contribute to expanding the field of what has come to be known as Digital Humanities (see Schreibmann et al. 2004, among many others).

Large-sized electronic corpora of AG texts are now available online. The list of such corpora is long and includes, among others, AGDT 2.0 (*Ancient Greek Dependency Treebank*), *The Chicago Homer* (a database for Early Greek epic), DĀMOS (a database of Mycenaean; Aurora 2015), DFHG (*Digital Fragmenta Historicorum Graecorum*), *The Diorisis Ancient Greek Corpus* (Vatri & McGillivray 2018), EAGLE (*Electronic Archive of Greek and Latin Epigraphy*), the *Perseus Digital Library* (Bamman & Crane 2008), PROIEL (*Pragmatic Resources of Old Indo-European Languages*; cf. Haug & Jøndal 2008),¹ SEMANTIA (which contains 313 papyrological texts from the Duke Databank of Documentary Papyri; cf. Vierros 2018), TITUS (*Thesaurus Indogermanischer Text- und Sprachmaterialien*), and TLG (*Thesaurus Linguae Graecae*, fully available online for subscribers only).

In some of these corpora, the digitized versions of AG texts are also enriched with mark-up and/or annotation. The term "mark-up" pertains to the textual document as a whole and can be stylistic, philological and/or archaeological in nature; for example, dates and literary genres are annotated in *The Diorisis Ancient Greek Corpus*. Information concerning the place of discovery, editions, textual issues, and hand attribution is associated with the digitized version of the Mycenaean tablets contained in DĀMOS (on the necessity of including stylistic, philological and archeological information in annotated corpora of ancient languages, see Eckhoff et al. 2018b: 302).

For the aims of this paper, the most relevant type of metadata is the multi-layered annotation that provides linguistic information on words and/or text chunks at different linguistic levels: (a) morphological annotation, including POS-tagging, lemmatization, and inflectional morphology (see e.g. *The Chicago Homer*; *The Diorisis Ancient Greek Corpus*, and to some extent, derivational

1 Additional morphosyntactically annotated corpora have stemmed from the PROIEL project, specifically, ISWOC for old Romance and Germanic languages and TOROT for Old Slavic languages (Eckhoff & Berdičevskis 2015; Eckhoff et al. 2018a).

morphology also is annotated in the corpora of the PROIEL family); (b) syntactic annotation, according to which syntactically parsed sentences are represented and stored as syntactic trees in “treebanks” (see e.g. AGDT 2.0; PROIEL); (c) semantic and pragmatic annotation, i.e. metadata concerning semantic roles, other semantic information (e.g. animacy of event participants, semantic class of verbs), and information structure (cf. AGDT 2.0, documented in Celano & Crane 2015; PROIEL); and (d) genre-specific annotation (e.g. *The Chicago Homer*, in which formulas of the Early Greek epic texts are annotated).²

Corpora that contain multi-layered linguistic annotation can be exploited to create other treebank-based linguistic resources, such as valency lexica. The term “valency” was first borrowed from chemistry into linguistics by Lucien Tesnière (see Ágel & Fischer 2015 for precursors), within the framework of Dependency Grammar (Tesnière 1959). Here, valency refers to the extent to which verbs combine with and determine the form of a fixed and predictable number of participants. These participants are called *actants* by Tesnière. In Tesnière (1959), *actants* are contrasted with *circumstants*, free modifiers that also depend on verbs but do so instead by virtue of their modifying potential. The English terms *complement* and *adjunct* are meant to replicate Tesnière’s *actant-circumstant* distinction (Matthews 1981). Frequently, *actants/complements* are also called *arguments*, especially in the US linguistic tradition of the last quarter of the 20th century. The latter term, originally borrowed from logic (Frege 1891), holds an ambivalent status between semantics and syntax: it can either indicate all inherent roles that occupy a place in the semantic relationality of a concept or be used to name constituents headed by other verbal constituents. This interaction—and, often, overlap—between semantic and syntactic valency is fundamental to grasping the theory of valency underlying the data of HoDeL (on which, see Section 4).

Valency theories occupy a core position within the study of human language, as human beings perceive events as consisting of entities, chiefly expressed by nouns, and relationships between entities, chiefly expressed by predicates. Thus, to speak about events, human beings create predicate-argument structures (e.g., Ágel & Fischer 2015). Accordingly, since Tesnière, several linguistic theories have put valency and argument structure at the very core of their

2 To my knowledge, no AG corpus is currently annotated for metrical structure. One linguistic resource that contains metrical annotation is *VedaWeb* (<https://vedaweb.uni-koeln.de>), an online morphologically tagged corpus of the *Rigveda*, the most ancient Indo-Aryan textual record.

research agenda but have disagreed about the very nature of these phenomena: valency has been variably regarded as a concept fundamentally syntactic, semantic, or both. However, until the relatively recent development of Construction Grammar (see Fried & Boas 2005 for an overview), there has been at least one point of substantial agreement: argument structure is a property of verbs, and as such, its information can be stored in (verbal) lexica.

Given its importance for linguistic expression, it is unsurprising that valency-related information is contained in traditional dictionaries of AG. For example, the Liddell-Scott-Jones dictionary (LSJ), under the entry *manthánō* 'learn', provides the different meanings of this verb, along with the form and semantic features of the participants associated with these meanings (i.e., *manthánō*+ACC 'learn something', *manthánō*+DAT 'understand somebody', *manthánō*+INF 'learn something', *manthánō*+apó/ek/pros/pará+GEN 'learn from something/ somebody'). Valency-dedicated lexicography began with the pioneering work of Helbig & Schenkel's, (1991[1969]) *Wörterbuch zur Valenz und Distribution deutscher Verben* (on German, see also Schumacher et al. 2004; on English, see Herbst et al. 2004; for a contrastive valency dictionary of Dutch-French-English, see Coleman et al. 2004). Non-automatic valency lexica are also available for ancient Indo-European languages; for instance, Happ (1976) contains a non-exhaustive list of Latin verbs and their valency patterns as evidenced in a corpus of 800 sentences sampled from Cicero's *Orationes*. In her dissertation, Frigione (2015) collected and published the materials for a verbal lexicon of Old Church Slavic.

Manually collected and thus intuition-based valency lexica are necessarily partial in coverage, not containing comprehensive frequency information on the collected valency patterns, and are extremely time-consuming to build. For these reasons, scholars have increasingly exploited existing annotated corpora to automatically extract verbal valency patterns and their frequencies from texts. In principle, automatic valency lexica are able to capture not only verbal valency as a potential of a verbal lexeme, but also valency as the textual-situational realization of such potential. The first attempts in this direction were methodologically hybrid: for example, PropBank (Kingsbury & Palmer 2002), FrameNet (Ruppenhofer et al. 2010), and PDT-Vallex (Hajič et al. 2003) were initially built with an intuition-based method and later refined with corpus-driven data.

Primarily automatic approaches to valency lexica were first carried out for modern languages (e.g., VALEX, Korhonen et al. 2006 on English; LexShem, Messiant et al. 2008 on French). However, automatically derived valency lexica exist nowadays for ancient Indo-European languages as well. For example, Bamman & Crane (2008) and McGillivray (2013: 31–60) used the mor-

phosyntactically annotated texts in the *Perseus Digital Library* and in the *Latin Dependency Treebank* (LTD; Bamman & Crane 2006), respectively, to induce valency lexica for Latin. Similarly, the syntactic subcategorization lexicon for Thomas Aquinas' Latin, IT-VaLex (McGillivray & Passarotti 2009), was derived from the *Index Thomisticus Treebank* (IT-TB; McGillivray & Passarotti 2015). McGillivray & Vatri (2015) used exactly the same computational method documented in McGillivray & Passarotti (2009) and McGillivray (2013: 31–60) to induce a valency lexicon from AGDT 1.0 (to my knowledge, these lexica have never been made available online). Finally, a semantic valency lexicon of Latin, *Latin Vallex*, was derived from a semantically-annotated subset of LTD and IT-TB (Passarotti et al. 2016). Currently, the PROIEL-style treebanks, which can be consulted through the *Syntacticus* interface, also come with generated dictionaries that include a comprehensive list of valency frames extracted from treebanks.³

In the next section, I present HoDeL—a lexicon of Homeric Greek verbs that has also been built upon the IT-VaLex model—and some of its functionalities.

3 Presenting HoDeL

In this section, I first document how HoDeL was created and then illustrate its basic usage by showing a number of simple incorporated filters for queries with HoDeL's online interface.⁴

3.1 *The building of HoDeL*

HoDeL is closely connected with the Homeric texts treebanked at AGDT 2.0. The structure of AGDT 2.0 is modelled on that of the *Prague Dependency Treebank* of Czech, the groundbreaking project in the field of Dependency Grammar (PDT 3.0; Hajič et al. 1999).

Like other dependency treebanks, PDT 3.0 and AGDT 2.0 are predicate-centred, contain the same number of nodes as the number of words and allow for non-binary-branching trees (on the differences underlying constituent vs. dependency treebanks and among different annotation styles of dependency

3 Syntacticus can be used at this link: <http://syntacticus.org>. The following link contains an AG example: <http://syntacticus.org/lemma/syntacticus:20180920:grc:%CE%B2%CE%AC%CE%BB%CE%BB%CF%89:V->. By clicking on each valency frame, users may find occurrences instantiating it.

4 HoDeL's guidelines can be accessed online at https://su-lab.unipv.it/tasf/wp-content/uploads/2021/01/HoDeL_guidelines.pdf.

treebanks, see the summary in Biagetti 2018: 9–37). PDT 3.0, and in principle AGDT 2.0, are multi-layered dependency treebanks: metadata is structured and stored in separated but interlinked morphological, analytical, and tectogrammatical layers. The analytical (i.e., syntactic) layer contains dependency syntactic trees and works as a basis for the tectogrammatical (i.e., semantic) layer, which stores complex semantic role-labelling, information structure, and anaphora/ellipsis resolution, annotated in the framework of the Praguian linguistic tradition of the *Functional Generative Description* (Sgall et al. 1986).

At its first release, the *Ancient Greek and Latin Dependency Treebank* was the first treebank for AG and Latin (Bamman & Crane 2011). Currently, the AG section of the resource, AGDT 2.0, contains 557,922 tokens and the following works: *Iliad*, *Odyssey*, *Hymn to Demeter*, Aeschylus' tragedies, Sophocles' *Ajax*, *Antigone*, *Electra*, *Oedipus Tyrannus*, *Trachinae*, Hesiod's *Theogony*, *Works and Days*, *Shield of Heracles*, Plato's *Euthyphro*, Lysias' *On the Murder of Erasthosthenes*, *Against Alcibiades* 1 and 2, *Against Panceleon*, Plutarch's *Alcibiades* and *Lycurgus*, and sections of Aesop's *Fables*, Athenaeus' *Deipnosophists*, Diodorus Siculus' *Library*, Herodotus' *Histories*, Polybius's *Histories*, Pseudo-Apollodorus's *Library*, and Thucydides' *Histories*. All works are annotated for the morphological and analytical layers, whereas the tectogrammatical annotation is available only for subsections of Aesop's *Fables* and Diodorus Siculus' *Library* (Celano 2019: 283–284, 288). Thus, no tectogrammatical annotation is available for the Homeric poems.

HoDeL has been automatically induced from the analytical layer of the *Iliad* and *Odyssey* treebanks at AGDT 2.0. The guidelines of the analytical layer of AGDT rely on those of PDT with some *addenda* aimed to increase descriptiveness and precision by incorporating Smyth's *Greek Grammar for Colleges* (Smyth 1920; Celano 2019: 285–286). Figure 1 shows the dependency tree of *Iliad* 1.1–7, reported in (1).⁵

- (1) *mênin* *áeide* *theà* *Pēlēiádeō* *Akhilēos*
 wrath.ACC sing.IMPV.PRS.3SG goddess.VOC Peleus'_son.GEN A.GEN
ouloménēn *hè* *muri'* *Akhaioís* *álge'*
 destructive.ACC REL.NOM countless.ACC.PL A.DAT.PL woe.ACC.PL
éthēke, *pollás* *d'* *ipthímous* *psukhàs* *Áidi*
 put.AOR.3SG many.ACC.PL PTC valiant.ACC.PL soul.ACC.PL H.DAT

5 Glosses follow the Leipzig Glossing Rules, to which the following labels were added: AOR = aorist, IMPF = imperfect, MID = middle, M/P = medio-passive, PTC = particle. Mood is glossed only if not indicative, tense if not present, voice if not active, and number if not singular. Gender is omitted.

FIGURE 1 *Iliad* 1.1–7. A dependency tree in the analytical annotation layer

proíapsen hērōōn autoūs dē helória teúkhe
 send.AOR.3SG hero.GEN.PL 3PL.ACC PTC spoil.ACC.PL make.IMPF.3SG
kúnessin oiōnoísí te pásí, Diòs d'
 dog.DAT.PL bird.DAT.PL and INDF.DAT.PL Z.GEN PTC
eteleíeto boulé, ex hoú dē tà
 fulfil.IMPF.3SG.M/P plan.NOM out_of REL.GEN PTC DEM.ACC.PL
prôta diastētēn erísante te
 first.ACC.PL part.AOR.3DU fight.PTCP.AOR.NOM.DU and
Atreídēs áanax andrôn kai díos
 Atreus'_son.NOM king.NOM men.GEN and divine.NOM
Akhilleús
 Achilles.NOM

'The wrath sing, goddess, of Peleus' son, Achilles, that destructive wrath which brought countless woes upon the Achaeans, and sent forth to Hades many valiant souls of heroes, and made them themselves spoil for dogs and every bird; thus, the plan of Zeus came to fulfillment, from the time when first they parted in strife Atreus' son, king of men, and brilliant Achilles.' (*Il.* 1.1–7)

From this layer, HoDeL incorporates queries extracting all verbal forms and their dependents labeled as:

- SBJ (Subject), as in *hē murí Akhaioís álge' éthēke* '... which brought countless woes upon the Achaeans' from (1);
- OCOMP (Object Complement), as in *autoūs dē helória teúkhe kúnessin oiōnoísí te pásí* 'and made them themselves spoil for dogs and every bird' from (1);
- PNOM (Predicate Nominal), as in *hòs nún pollòn áristos Akhaiòn eúkhetai eínai* '... who now claims to be far (the best) of the Achaeans' (*Il.* 1.91);

- OBJ (Object), as in *mênin áeide theà* ‘(The) wrath sing, goddess’, and in *hè muri’ Akhaioîs álge’ éthēke* ‘which brought countless woes upon the Achaeans’ from (1).

The OBJ label comprises all verbal arguments except SBJ and arguments labeled as OCOMP and PNOM and hence includes accusative, dative, genitive nouns or pronouns, prepositional phrases, infinitive verbs, accusative+infinitive constructions, and other types of subordinate clauses that can function as verbal objects (for details, see Celano 2019: 286–287 and https://github.com/PerseusDL/treebank_data/blob/master/AGDT2/guidelines/Greek_guidelines.md#obj). All these dependents may either be direct child nodes of a verbal form or be attached to the verbal head via one of the bridge nodes, specifically, AUXP (preposition), AUXC (conjunction), COORD (coordinator, including coordinative conjunctions and particles), APOS (apposing elements, such as commas). For example, the phrase *Atreídēs te áanax andrôn kai díos Akhilleús* ‘Atreus’ son, king of men, and brilliant Achilles’ in Figure 1 is tagged with the bridge COORD (*kai*) and contains two subjects, coordinated and hence tagged as SBJ_CO (*Atreídēs* and *Akhilleús*).⁶ All extracted dependents are considered part of verbal valency according to the guidelines of AGDT 2.0.

In contrast, we did not extract dependents that are tagged as ADV (adverbials, which provide the event with background information), ATR (NP modifiers) and ATV/ATVV (non-governed complements, i.e., predicative noun phrases / adjectives which may morphologically agree with their head noun, but qualify the whole event denoted by the verb) that the AGDT 2.0 guidelines do not consider belonging to the verbal valency.

Argument dependents have been extracted using a series of SQL queries and then recorded in a spreadsheet, from which a relational database has been built. The relational database in turn interacts with the user interface. The original query algorithm and its implementation were conceived to build IT-VaLex (McGillivray & Passarotti 2009). To induce HoDeL, the queries have been adapted to the AGDT 2.0 tagset. An earlier version of HoDeL, released in 2016 (Zanchi et al. 2018), was based on a previous version of the treebank (AGDT 1.0) and lacked transliteration and English translation.

Transliteration of the Greek script (conforming to the most common current academic standards) and links to English translation have been added to the current version. In order to type Greek characters, HoDeL users should employ Beta Code, as in the *Perseus Project* and in TLG. The correspondences between

6 Coordinated dependents are marked by _CO, apposed dependents are indicated by _AP, whereas coordinated apposed dependents are given the label _AP_CO. See Figure 7 for examples.

Greek	α	β	γ	δ	ε	ζ	η	θ	ι	κ	λ	μ	ν	ξ	ο	π	ρ	σ,ς	τ	υ	φ	χ	ψ	ω
Beta Code	a	b	g	d	e	z	h	q	i	k	l	m	n	c	o	p	r	s	t	u	f	x	y	w
Greek	ά	ά	ά	ά	ά	ά	ά	ά	ά	ά	ά	ά	ά	ά	ά	ά	ά	ά	ά	ά	ά	ά	ά	ά
Beta Code	a/	a\	a=	a)	a(	a)/	a(/	a)\	a(\	a)=	a(=	a	a)/	a)=										
An * should be used to type capital letters (e.g. Βρισην → *brish/s)																								

TABLE 1 Greek characters-Beta code correspondences

Greek fonts and Beta Code are reported in Table 1. The least intuitive Greek-Beta Code correspondences are highlighted in grey.

Example (2) shows how the first line of the Odyssey looks in Greek characters (in 2a), Beta-Code (in 2b), and in transliteration (in 2c).

- (2)
- a. ἄνδρα μοι ἔννεπε μούσα πολύτροπον ὃς μάλα πολλὰ

b. a)/ndra moi e)/nnepe mou=sa polu/tropon o(\s ma/la polla\

c. ándra moi énnepé mou̓sa polútropon hòs mála pollà
- 'Tell me, Muse, about the wily man who (wandered) long and far' (*Od.* 1.1)

3.2 *HoDeL basic incorporated queries and filters*

The HoDeL home page shows a list of Homeric verbs ordered alphabetically. After each lemma, its frequency is provided (Figure 2). Note that users can choose to visualize either the Greek script or its transliteration by flagging ‘greek’ or ‘trans’ in the ‘Display’ box at the top of HoDeL homepage.

By default, HoDeL gives frequency information concerning verbal lemmas and their dependents tagged as SBJ, OBJ, PNOM, and OCOMP, and specifically:

- 2,482 = type frequency of verbal heads

– 40,693 = token frequency of verbal heads

– 4,219 = type frequency of dependent lemmas

– 49,137 = token frequency of dependent lemmas.

When users add filters to their queries, HoDeL always provides these and other frequency counts.

Users can also visualize all lemmas that depend on Homeric verbal heads, as shown in Figure 3. A number of verbs appear in this list: these verbs function as main verbs in dependent SBJ or OBJ clauses.

Both lists in Figures 2 and 3 contain clickable lemmas. For example, by clicking on a verbal lemma in the list in Figure 2, e.g., *akoúō* ‘hear’, users obtain (i) all its forms in the Homeric poems, (ii) the ordered contexts of these occurrences (automatically chunked by an algorithm that exploits punctuation marks), and (iii) syntactic subtrees representing the queried verb and its dependents tagged as SBJ, OBJ, PNOM, and OCOMP (Figure 4).

HoDeL

The Homeric Dependency Lexicon

FIGURE 2 HoDeL home page

HoDeL summarizes the selected query filters in the grey box 'Query' and at the top of the output page and provides users with frequency information: the verb *akoúō* 'hear' occurs 182 times in the Homeric poems and takes 86 different argument lemmas. In turn, the argument lemmas have a token frequency of 210. Note that the token frequency of argument lemmas is higher than the token frequency of *akoúō* (210 > 182). This is not surprising: if two argument lemmas are taken by *akoúō* in the same occurrence, that occurrence is listed once in the verb frequency count, but twice in the argument frequency count, once for each argument lemma.⁷

⁷ HoDeL currently does not provide a list of valency frames for each verb as, e.g., the *Syntacticus* interface of the PROIEL family treebanks does (cf. Section 2 and fn. 3). We made this choice because we preferred to route users directly to Homeric examples. In HoDeL, given some theoretical issues underlying the annotation scheme (see Section 4 for details), a list of valency frames with frequencies could be misleading. For example, AGDT 2.0 does not contain nodes for null referential arguments, including null subjects. Therefore, a valency frame with zero

HoDeL

The Homeric Dependency Lexicon

Copyright (c) CIRCSE

FIGURE 3 List of dependent lemmas

The lexicon also allows directly typing in verbal and dependent lemmas. By clicking on the grey box, ‘Query’, a window opens in which the requested lemma can be typed using Beta Code (Figure 5; cf. Table 1). In Figure 5, I typed the verb *akouō* a)kou/w in the ‘Verbal Head Lemma’ box. Its relative subtrees can be seen by clicking ‘Submit’, the button that launches queries.

arguments in principle could refer both to impersonal verbs and to intransitive verbs with null subjects. For the present, we have decided not to integrate empty nodes for null subjects: the absence of empty nodes is a structural feature of the analytical layer of dependency tree-banks modelled on PDT. Moreover, AG also allows for null referential objects (Luraghi 2003; Haug 2012). Thus, if one decides to integrate null subjects by manually disambiguating impersonal verbs, they must also integrate null objects for the sake of consistency. Null referential objects are, however, far more difficult to detect than null subjects (Section 4). Without integrating null objects, an automatically induced valency frame with a single argument could in principle refer to intransitive usages of transitive verbs or to transitive verbs taking null referential objects.

HoDeL

The Homeric Dependency Lexicon

FIGURE 4 Frequency information and syntactic subtrees of *ἀκούω* ‘hear’

In the output passages and subtrees (see Figure 4 and Figure 6), the verbal forms are circled in orange, whereas the dependents are highlighted in blue. By pointing at a word in the output contexts, users obtain morphological annotation as stored in the morphological layer of AGDT 2.0 (Figure 6). For example, Figure 6 shows that the form *euxaménou* is the genitive masculine singular of the aorist middle participle of the verb *eúkhomai* ‘pray’. Furthermore, if users click on the blue folder after the Greek text, HoDeL provides the corresponding English translations. The latter have been automatically aligned with the Greek text using an algorithm that exploits punctuation marks and text chunks contained in the texts provided at the *Perseus Digital Library*. The automatic alignment has been manually checked and, when necessary, modified according to the translation available at *The Chicago Homer* (Figure 6).⁸

HoDeL

The Homeric Dependency Lexicon

Display greek Order By : [^]lemma [rev_lemma] [frequency] Filter :

Query

Poem ↓

Verbal Head Lemma
akou/w

Voice ↓

Exact Sequence ☐

Exact Cardinality ☐

Argument Lemma

Relation ↓

Case/Mood ↓

Prep. ↓

Conj. ↓

Position ↓

Add another Argument

Submit Reset

List of Verbal Head Lemmas

[Next Page](#)

(8)	ἀείδω (40)	αἰστών (2)	ἀλεγίζω (6)
ἔνψιός (1)	ἀεικίζω (8)	αἰσύνω (12)	ἀλεγύνω (5)
ἀάω (19)	ἀείρω (67)	αἰτέω (14)	ἀλέγω (11)
ἀβακέω (1)	ἀελλπτέω (1)	αἰτιάσμαι (7)	ἀλεεῖνω (26)
ἀβροτάω (1)	ἀδέω (20)	αἰτίξω (10)	ἀλειφω (10)
ἀγασίμαι (1)	ἀεσα (2)	αἰχμῶζω (1)	ἀλέξω (20)
ἀγάλλω (7)	ἀζομαι (8)	αἰώ (22)	ἀλέσμαι (36)
ἀγασμαι (27)	ἄζω (1)	ἀκαγίζω (2)	ἀλέομαι (1)
ἀγαπάω (6)	ἀηθέσσω (1)	ἀκείσμαι (9)	ἀλετρεύω (1)
ἀγαπάω (2)	ἀημι (12)	ἀκηδέςω (2)	ἀλέω (7)
ἀγγέλλω (27)	ἀθερίζω (3)	ἀκοντιζέω (35)	ἀλητεύω (5)
ἀγείρω (62)	ἀθαεύνω (4)	ἀκοαστάω (2)	ἀλθομαι (1)
ἀγνέω (6)	ἀθλέω (2)	ἀκουάζομαι (2)	ἀλιός (3)
ἀγκάζομαι (1)	ἀθρέω (5)	ἀκούω (182)	ἀλίσκομαι (18)
ἀηλαίω (1)	ἀθύρω (1)	ἀκροκελαινιάω (1)	ἀλίσταίνω (7)
ἀγνοέω (7)	αἰδέσμαι (42)	ἀλάλημαι (15)	ἀλλομαι (28)
ἀγνυμι (26)	αἰθω (22)	ἀλάλκε (13)	ἀλλοφρονέω (2)
ἀγοράσμαι (27)	αἰνέω (10)	ἀλαλῶκτημαι (1)	ἀλσάω (1)
ἀγορεύω (166)	αἰνίζω (2)	ἀλάσμαι (27)	ἀλσάω (2)
ἀγρέω (6)	αἰνυμαι (14)	ἀλασώ (2)	ἀλυσκάω (3)
ἀγρώσσω (1)	αἰώλλω (1)	ἀλαπάω (11)	ἀλύσσω (27)
ἀγυρτάω (1)	αἰρέω (415)	ἀλαστέω (2)	ἀλύσσω (3)
ἄγχω (1)	αἶρω (43)	ἀλγέω (4)	ἀλύω (5)
ἄγω (308)	αἰσθω (2)	ἀλδαίνω (2)	ἀλφάνω (4)
ἀδέω (5)	αἰσσω (61)	ἀλδήσκω (1)	ἀμαθύνω (1)

FIGURE 5 Typing in the lemma *akouō* 'hear'

The purple box 'Args Number' shows (i) frequency information concerning the number of arguments taken by verbal heads, and (ii) frequency information concerning the syntactic relations (SBJ, OBJ, PNOM, OCOMP) of arguments taken by verbal heads. For example, as shown in Figure 7, *akouō* 'hear' can take from zero (37 occurrences) to three arguments (2 occurrences); when this verb is the head of two dependents, the latter can have different syntactic functions, called 'Subcat.(egories)' in the resource. By flagging one of these categories ('No. Args') and subcategories (i.e., argument number *and* functions), users obtain filtered passages and subtrees. The categories and subcategories suggested by the system are not pre-established but rather corpus-induced for each selected verb.

The incorporated filter 'Args Order' allows users to investigate constituent order in Homeric Greek. As shown in Figure 8, at a lower level of granularity, attested relative orders of verbs (*akouō* 'hear' in this case) and OBJ depen-

8 The translation available at the *Perseus* project are those by Murray (1919, 1924). *Chicago Homer's* translations are those by Lattimore (1951, 1967).

HoDeL

The Homeric Dependency Lexicon

FIGURE 6 Visualizing morphological metadata and English translations

dents, together with their frequencies (attested verb-OBJ orders are labelled as ‘Cat.(egories)’ in the interface), are provided for users. At a higher level of granularity, for each attested verb-OBJ order, the relative positioning of other argument dependents, such as SBJs, can be taken from the lexicon (in this case, attested orders are labelled as ‘Subcat.(egories)’). As seen in the ‘Args Number’ example, these orders are given by the system based on patterns attested in the corpus. Both categories and subcategories of orders can be flagged to obtain filtered contexts and subtrees.

In addition, users can search the Homeric verbs by argument relation and case/mood using the blue box ‘Arguments’. In Figure 9, the attested functions and forms of arguments taken by the verb *akoúō* ‘hear’ are shown, and each attested ‘Cat.(egory)’ and ‘SubCat.(egory)’ can be flagged to obtain filtered examples and relative subtrees.

Additional filters are incorporated in the ‘Query’ box as illustrated in Figure 10. For example, users can work on one or more poems and search for verbs

HoDeL

The Homeric Dependency Lexicon

FIGURE 7 Frequency information on argument number

inflected in a specific morphological voice. Users can further search for specific argument lemmas and filter their outputs based on the syntactic and morphological features of these lemmas, including relation, case/mood, preposition, conjunction, and position with respect to the verb.

Finally, HoDeL allows users to combine all these parameters to search for more than one argument at a time by means of the ‘Add another argument’ option. Multiple arguments can be either filtered per sequence or per cardinality by flagging the corresponding box.

HoDeL

The Homeric Dependency Lexicon

FIGURE 8 Word order information on verbal arguments

4 HoDeL: issues and perspectives

In this section, I discuss a number of theoretical and practical issues of HoDeL, occasionally framing them within a more general discussion about the pros and cons of annotated corpora. Then, I detail how HoDeL functionalities can be used to answer specific research questions.

4.1 HoDeL: theoretical issues

As illustrated in Section 3.1, HoDeL has been automatically induced from a dependency treebank. Yet, HoDeL (and similar valency lexica, such as those described in McGillivray & Vatri 2015) is not corpus-driven nor theory-free. On one hand, the series of SQL queries underlying HoDeL has the advantage of making possible the automatic virtual retrieval of all instantiations of all valency frames of all Homeric verbs. On the other hand, queries based on annotation schemes, including those of HoDeL, are limited by only being capable of retrieving whatever has been annotated according to such annotation schemes (cf., e.g., Tognini-Bonelli & Sinclair 2006: 214). Annotation schemes are never theory-free: they necessarily rely on a source theory, as annotating

FIGURE 9 Filtering arguments by relation and morphological information

means “adding interpretative information into a collection of texts” (Hovy & Lavid 2010: 1). Therefore, issues may arise when (a) annotation guidelines are either too fine-grained or too coarse-grained in applying and exemplifying the source theory; or (b) there is a mismatch between the theory of annotators and that of linguists using the annotated corpus for theoretical research and/or applied purposes (McGillivray & Vatri 2015: 103).

HoDeL is no exception in this respect. In principle, annotation schemes should be designed according to general linguistic theories that meet general consensus within the linguistic community and can be easily ported to several different languages (Haug 2015). However, the theory of valency underlying AGDT 2.0—and PDT before it—is the theory of valency of the *Functional Generative Description* (cf., e.g., Panevová 1994), which is linked to the Praguian linguistic tradition and Czech strategies of argument encoding.

The *Functional Generative Description* regards valency as the ability of linguistic elements to open positions for other linguistic elements, called *complementations*. Complementations are grouped in two ways: (i) inner participants vs. free modifications; (ii) obligatory vs. optional complementations. Inner par-

The Homeric Dependency Lexicon

The screenshot displays the 'The Homeric Dependency Lexicon' web interface. On the left is a 'Query' panel with various filters: 'Poem' (dropdown), 'Verbal Head Lemma' (text input with 'ajkou/w'), 'Voice' (dropdown), 'Exact Sequence' and 'Exact Cardinality' (checkboxes), 'Argument Lemma' (text input), 'Relation' (dropdown), 'Case/Mood' (dropdown), 'Prep.' (dropdown), 'Conj.' (dropdown), and 'Position' (dropdown). There are also buttons for 'Add another Argument', 'Submit', and 'Reset'. On the right is the 'Occurrences and Contexts' section for the lemma 'ἀκούω'. It shows two examples from the Iliad. The first example is from Iliad 1.380-1.382, showing the text 'τοῖο δ' Ἀπόλλων εὐξαμένου ἤκουσεν, ἐπεὶ μάλα οἱ φίλος ἦεν, ἤκε δ' ἐπ' Ἀργείοισι κακὸν βέλος.' with a dependency tree for the verb 'ἤκουσεν' (he heard) showing a relation 'OBJ' to the object 'τοῖο' (this). The second example is from Iliad 1.396-1.400, showing the text 'πολλάκι γὰρ σεο πατὴρ ἐνὶ μεγάροισιν ἀκουσα εὐχομένης δ' ἔφησθα κελαϊνεφέϊ Κρονίωνι οἷη ἐν ἄθανάτοισιν ἀεικέα λοιγὸν ἀμῦναι, ὁππότε μιν ξυνδῆσαι Ὀλύμπιοι ἤθελον ἄλλοι ἦρη ἢ ἡδὲ Ποσειδάων καὶ Παλλὰς Ἀθήνη.' with a dependency tree for the verb 'ἀκουσα' (I heard) showing a relation 'OBJ' to the object 'σεο' (your).

FIGURE 10 Additional filters on verbs and verbal dependents

ticipants (e.g., actor, patient, addressee, origin) are verb-specific and thus can occur only once per verb, whereas free modifications (e.g., time, place, goal, instrument, etc.) are not verb-specific and can be added freely to the sentence. The distinction between inner participants and free modifications seems to be semantically based, while that between obligatory and optional complementations is grounded in the syntactic notion of obligatoriness. These two classifications do not overlap: there can exist *optional inner participants*, such as origin with motion verbs, and *obligatory free modifications*, such as instrument with verbs of cutting. In PDT, this valency theory is not accounted for at the analytical layer, but at the tectogrammatical one, which also contains anaphora and ellipsis resolution (Section 3.1). Overall, then, the theory of valency of the *Functional Generative Description* seems to be based on semantic criteria rather than on syntactic ones. The only notion that seems to distinguish syntactic arguments (i.e., OBJ) from adjuncts (i.e., ADV), which is mentioned in the PDT guidelines, is obligatoriness.

As shown in Section 3, there is no tectogrammatical layer available for the Homeric poems treebanked at AGDT 2.0. Naturally, it follows that the valency theory is discussed in the guidelines of the analytical layer of AGDT. Those guidelines, though, explicitly state that AGDT 2.0 inherits its valency theory

from PDT, but the critical notion of obligatoriness is not dealt with at a sufficient level of granularity. Moreover, the underlying treebank architecture, the lack of the tectogrammatical layer of AGDT 2.0, and the semantic nature of the theory of valency of PDT produce a number of mismatches in the argument vs. adjunct annotation with respect to the most widely accepted theories of syntactic valency.

For HoDeL users, this means keeping in mind that the valency theory of the *Functional Generative Description* can result in a number of unexpected consequences, the most relevant of which are discussed here. To begin with, agents of passive verbs, such as *ek Diós* ‘by Zeus’ in (3), are annotated as OBJ, as part of the verbal valency:

- (3) *ēdè philēthen ek Diós*
 and love.AOR.3PL.PASS out_of Z.GEN
 ‘... and they were loved **by Zeus**.’ (*Il.* 2.668–669)

This is because, at a semantic level, agent participants are *inner* participants of transitive verbs. However, passive voice is usually acknowledged to be a syntactic valency decreasing strategy, which removes agents from argument structure and makes them *optional* (e.g., Siewierska 2005).

Beneficiary and instrumental datives are inconsistently tagged at times as OBJ, i.e., as syntactic arguments (4), and at times as ADV, i.e., as adjuncts (5):

- (4) *keînon d’ ou pote pámpan egôn idon ophthalmôisi*
 DEM.ACC PTC NEG ever wholly 1SG.NOM see.AOR.1SG eye.DAT.PL
 ‘... but I never ever saw **with (my) eyes** him ...’ (*Od.* 11.528)
- (5) *epei hekàs ophthalmôisi gaîan egôn idómēn*
 since afar eye.DAT.PL land.ACC 1SG.NOM see.AOR.1SG.MID
 ‘... since I myself saw **with (my) eyes** the land far away.’ (*Od.* 5.358–359)

In (4) and (5), the same dative plural, *ophthalmoîsi* ‘with (my) eyes’, is inconsistently tagged as to its dependence on forms of the same verb, *eîdon*, the aorist suppletive form of *horáo* ‘see’. Again, this inconsistency results from the fact that ‘eyes’ are *optional inner* participants in the event of seeing: from a semantic standpoint, ‘eyes’ are an argument dependent, whereas from a syntactic point of view, they are not. This mismatch generates errors in the annotation.⁹

9 An anonymous reviewer suggests that this mismatch is a mere inconsistency of manual anno-

Other instructive examples of inconsistent tagging are genitive dependents taken by *akoúō* ‘hear’: these genitive participants can either play the role of stimuli and function as syntactic arguments or play the role of sources of information and function as adjuncts (on the argument structure of *akoúō* ‘hear’, see Luraghi 2020: 85, 127–135). Irrespective of the actual syntactic status of such genitives, they are all tagged as OBJ in AGDT 2.0, as shown in (6) and (7). Thus, in this case, the annotation is consistent but does not account for a fundamental difference in the valency frame of this verb.

- (6) *autàr Odussêos talasíphronos ou pot' éphasken zōoû oudè*
 PTC O.GEN steadfast.GEN NEG ever say.IMP.3SG alive.GEN NEG
thanóntos epikhthoníōn teu akoúsai
 die.PTCP.AOR.GEN earthly_one.GEN.PL INDF.GEN hear.AOR.INF
 ‘Yet **concerning Odysseus**’ steadfast heart, whether living or dead, he said
 he had heard **from no man** on earth.’ (*Od.* 17.114–115)

- (7) *hós té pou è autòs pareòn è*
 as PTC anywhere or self.NOM be_present.PTCP.PRS.NOM or
állou akoúsas
 other.GEN hear.PTCP.AOR.NOM
 ‘As though you had been present yourself or had heard (it) **from someone else**.’ (*Od.* 8.491)

In (7), the genitive plays the role of source of information and thus is syntactically an adjunct, whereas the stimulus role is played by a null referential object (cf. Luraghi 2003; Haug 2012 and the discussion below). Once more, this mismatch is due to the fact that the source of information is an *optional inner* participant in the event of hearing/coming to know.

Example (7) directs our focus to another feature of AGDT 2.0 that also raises issues for automatic argument structure detection. Specifically, the analytical layer of AGDT 2.0 does not comprise empty nodes for null arguments (Section 3.1): in the dependency treebanks modelled on PDT, null arguments are integrated at the tectogrammatical layer. However, AG is a pro-drop language,

tation and does not reflect theoretical issues in the annotation scheme. However, by checking the annotation of all the Homeric passages in which a form of *eídon* and a prepositionless dative plural of *ophthalmós* co-occur, one notices that this dative is annotated as OBJ in 4 out of 26 passages. Moreover, the status of the ‘eyes’ participant in events of seeing must be particularly problematic, given that even *en ophthalmóisin* ‘with her eyes’ is annotated as OBJ in *Od.* 8.459.

which means that by default it omits topical subjects, which are indexed on verbs through personal endings (cf. also fn. 9). Moreover, AG, as well as other ancient Indo-European languages, preferably or obligatorily selects null referential objects in certain syntactic and pragmatic contexts, including conjunct participles, coordinated verbs and clauses, and yes/no questions (Luraghi 2003; Haug 2012; Keydana & Luraghi 2012; Sausa & Zanchi 2015). Both null subjects and null referential objects appear frequently in the Homeric poems and, crucially, fill slots of verbal valency. The fact that they are not included in the syntactic trees of the analytical layer of AGDT 2.0 results in an incomplete account of the valency of a number of verbs.¹⁰ Integrating these types of arguments in HoDeL would be problematic with respect to the structural consistency of AGDT 2.0: the architecture of AGDT 2.0 is modelled on PDT, and thus, in principle, the analytical layer of AGDT 2.0 does not allow for empty nodes. Furthermore, only an extremely careful linguistic analysis would allow distinguishing among instances of referential null arguments, which are assuredly part of verbal valency, instances of non-referential null arguments, whose status is less straightforward, and complete absence of arguments. Example (8) is syntactically represented in the treebank as shown in Figure 11:

- (8) *óphra min autòn iónta lókhēsomai ēdè*
 so_that 3SG.ACC DEM.ACC come.PTCP.PRS.ACC await.FUT.1SG PTC
phuláxō
 keep.guard.PRS.3PL
 ‘... so that I will be able lie in wait **for him** as he comes and keep guard.’
 (*Od.* 4.760)

In (8), the future form of *lókháo* ‘lie in wait for’, *lókhēsomai*, takes an accusative second argument, *min autòn* ‘him’. In contrast, it is unclear whether the coordinated verb *phuláxō* is used intransitively and means ‘I will keep guard’ or takes a null object coreferential with *min autòn*. This structural ambiguity is not mirrored in the way coordinative structures are represented in the treebank, as shown in Figure 11: all coordinated verbs and their dependents, whether shared by the coordinants or not, are annotated as child nodes of the coordinating element.

10 The problematic issues regarding passive agents, null objects, and others were noted by the creators of the PROIEL family treebanks (Haug & Jøhndal 2008; Haug 2012; Eckhoff & Berdičevskis 2015; Eckhoff et al. 2018a). Although based on a version of Dependency Grammar, the tagset of the PROIEL treebanks includes additional labels and relations to improve descriptiveness, such as the label AG for passive agents and specific relations for elliptic structures.

FIGURE 11 Coordination in AGDT 2.0

A further class of issues arises from the inherent difficulty of interpreting the syntactic status of certain events participants, such as location or origin with posture and motion verbs (cf. *hízō* ‘sit’, e.g., in *Il.* 9.87). This matter is further complicated by the valency theory of the *Functional Generative Description*, which treats origin as an *optional inner* participant. In (9), the origin participant, encoded by *ek*+GEN, is tagged as OBJ in dependence of a verb of breaking, *ágnumi* ‘break’:

- (9) *ek dé moi aukhēn astragálōn eágē*
out_of PTC 1SG.DAT neck.ACC vertebra.GEN.PL break.AOR.3SG.PASS
‘My neck (lit. ‘the neck to me’) was broken away from the vertebrae.’ (*Od.* 11.64–65)

This annotation, however, is problematic for multiple reasons: first, it regards origin as a syntactic argument of *ágnumi* ‘break’, which it is not; second, it treats the initial local particle *ek* (see, e.g., Zanchi 2019: 82–86 on this terminology) as a preposition governing the genitive *astragálōn*, which is not necessarily the case either. The latter point is elaborated in Section 4.2, which deals with practical issues related to annotation errors and to specific features of the Homeric language.

4.2 *HoDeL: practical issues*

Like other manually annotated corpora, AGDT 2.0 contains some errors. These may result from annotators' insufficient training, lapses in their attention, their implicit theories of how language works (on implicit linguistic theories, see, e.g., Iannàccaro 2000: 53), and/or methodological choices underlying corpus collection (for a general overview of errors in annotated corpora, see, e.g., Dickinson 2005: 4–27 and references therein). The errors and inconsistencies of AGDT 2.0 have been inevitably inherited by HoDeL. Some examples are highlighted in what follows. A more detailed discussion may be found in Zanchi & Luraghi (2020).

For instance, there may be inconsistencies concerning lemmatization. The AG verb *horáō* 'see' has a paradigm consisting of three stems, specifically, those in the present *horáō*, the aorist *éidon*, and the perfect *ópopa*, which are, however, lemmatized as two different entries, the former including the forms based on the themes in *horáō* and *ópopa*, and the latter including the forms based on the theme of the aorist *éidon*. In contrast, the preverb + verb combination *eis-oráō* 'look', whose paradigm features the same three themes, is lemmatized as a single entry. Given that this issue is not limited to this pair of simplex and complex verbs, resolving this inconsistency would be complex, as it would involve restructuring the underlying lemmatization system of the treebank.

Another issue is the inconsistent tagging of voice information of verbs, which is in some passages tagged according to morphological criteria and in others according to semantic criteria. The HoDeL team has worked to address this issue, which would crucially affect a valency lexicon. In particular, we re-annotated voice information in our relational database induced from AGDT 2.0 according to strict morphological criteria. In addition, the lemmas of 185 forms were not annotated in the treebank: we also added this omitted morphological information, directly modifying our relational database.

Other errors may derive from annotators misinterpreting the annotation scheme. For example, the label PNOM should be reserved for predicate nominals of copular verbs. However, in (10), the verb *eimí* 'be' is not used as a copula, but as a location/existential verb: accordingly, the dependent *ein agorêi* 'in the place of assembly' should be tagged as OBJ and not as PNOM.

- (10) *laoi* *d'* *ein agorêi* *ésan* *athrói*
 men.NOM.PL PTC in assembly.DAT be.AOR.3PL in_crowds.NOM.PL
 'The folk, gathered together, were **in the place of assembly**.' (Il. 18.497)

A different set of issues stems from peculiarities of the Homeric language. As is widely acknowledged, in Homeric Greek, preverbs could occur in the so-called

“tmesis” positions and thus be “split” from the verbs that they semantically modify by various linguistic forms, retaining much of the syntactic freedom of their original adverbial status (cf. Zanchi 2019: ch. 3 with references therein for a diachronic interpretation of this preverb positioning). One of these tmesis positions is shown in example (9) above, which is repeated below in (11):

- (11) *ek dé moi aukhèn astragálōn eágē*
 out_of PTC 1SG.DAT neck.ACC vertebra.GEN.PL break.AOR.3SG.PASS
 ‘My neck (lit. ‘the neck to me’) was broken away from the vertebrae.’ (*Od.* 11.64–65)

In (11), the local particle *ek* occupies the sentence-initial position. In spite of this positioning, in AGDT 2.0, *ek* is assigned the label AUXP and functions as a head node of the genitive plural *astragálōn* ‘from the vertebrae’. However, the actual syntactic relation holding between *ek* and *astragálōn* may well be less rigid, so *ek* might just as easily function as an adverbial modifier with respect to the relation conveyed by the construction *ágnumi*+GEN. Alternatively, *ek* could modify the meaning of *ágnumi* in such a way that this verb takes the prepositionless genitive.

This syntactic ambiguity is reflected in an inconsistent annotation: in *Od.* 10.559–560, an almost identical formulaic expression occurs (*ek dé hoi aukhèn astragálōn eágē*, which differs from (11) only in the dative external possessor), but *ek* is a verbal dependent tagged as AUXZ. However, not only is AUXZ assigned to local particles in “tmesis” positions, but it is also employed for logical operators that are undoubtedly independent adverbs, such as those meaning ‘not’, ‘as well’, and ‘also’. This analysis is at odds with the function of preverbs, which, by forming a semantic unit with the verb, can occasionally modify its argument structure. Thus, the annotation scheme of AGDT 2.0 inadequately accounts for this peculiarity of the Homeric language. Since Homeric Greek is a language with free word order (thus, “split” preverbs do not always occur in sentence-initial position) and the label AUXZ is ambiguous and frequent, there is no easy, automatic way to find all examples similar to (11). Therefore, to date, solving this problem at the annotation level has been beyond the scope of the HoDeL project.

4.3 Some examples of how to use HoDeL

Having explained how HoDeL was built (Section 3.1), its basic functionalities (Section 3.2), and the precise kinds of data it contains (Sections 4.1–4.2), we now can show some examples of how the lexicon can help researchers to operationalize specific research questions.

The main advantage of using HoDeL lies in the fact that it allows users to carry out corpus-based quantitative studies on Homeric Greek without learning the complex formalisms necessary to directly query the treebanks of AG. Currently, AGDT 2.0 can be queried online from the web-repository of *Universal Dependencies* using a language called PML-Tree Query. However, this method has the disadvantage of not allowing specific texts to be singled out from the rest of the treebank: so, for example, Homeric texts cannot be investigated separately from later diachronic varieties of AG.¹¹ To focus on the Homeric texts independently from the rest of the treebank, one has to download the whole treebank in the .xml format, separate the Homeric texts, convert them into another format (e.g. to the .pml format), download a query tool (to my knowledge, the *Tree Editor—TrEd* is one of the few supporting the .pml format) and finally query the texts using the supported formalism. The other main treebank of AG, the PROIEL, can likewise be queried via *Universal Dependencies* or via INESS Search (a reimplement of Tiger Search), but it does not contain Homeric Greek. Thus, HoDeL fills a gap among the available linguistic resources in that it offers an extremely user-friendly interface to perform corpus-based research on the Homeric poems.

To begin with, HoDeL can be used to automatically retrieve all relevant examples of the construction under investigation. For example, the ‘Args Order’ option (Figure 8) can be employed to obtain the frequency distribution of sentences attesting to the VSO, SVO, and SOV orders in Homeric Greek. This data could shed light on a number of still-open issues regarding Homeric word order and information structure (on which, cf., e.g., Beschi 2018 with references therein).

The functionality ‘Arguments’ (Figure 9) can be used to extract all coordinated subjects and objects by selecting the relevant argument relations, specifically, SBJ_CO, OBJ_CO, SBJ_AP_CO, and OBJ_AP_CO (on bridge nodes, see Section 3.1, fn. 6). If the outputs of this filter are cross-checked with those of the ‘Args Order’ filter, researchers can effortlessly obtain frequency information on positioning patterns of coordinated subjects and objects with respect to verbs: do coordinated elements tend to surface in the same position, be it preverbal or postverbal, or do coordinants tend to be “split” by verbs? How do these ordering patterns correlate with verbal agreement in the case of coordinated subjects? What do these ordering patterns reveal about verbal government of coordinated objects?

11 On the GitHub page of AGLDT 2.0, it is stated that the treebanks can also be queried online via *Structural Search* and *Tindra*. Currently, however, neither of the two links given seems to work (http://perseusdl.github.io/treebank_data/; last access: 2021-05-08).

HoDeL

The Homeric Dependency Lexicon

FIGURE 12 Accusative and dative dependents taken by *bállō* ‘throw, hit’

Beyond facilitating word order queries, HoDeL can also be useful for detecting passages containing infrequent patterns in the Homeric language, which without HoDeL would require a time-consuming, manual reading of the poems to be detected. For example, by searching in the ‘Query’ box for a specific verbal lemma with a preverb and combining it with the prepositional phrase headed by the same local particle, one can easily find attested instances of preverb repetition outside the preverbal context. This information can be used to account for the different paces of grammaticalization or lexicalization paths undergone by different AG preverbs (on which, see, e.g., Zanchi 2017): the local particles that allow for repetition are more lexicalized or grammaticalized into preverbs and prepositions.

The option ‘Add another argument’ can be employed to investigate ditransitive verbs that feature argument structure alternations, such as the transfer verb *bállō* ‘throw, hit’ (Figure 12). This verb could mean ‘throw something (ACC) toward something else / someone (DAT)’, as in *Il.* 1.245–246, or ‘hit someone (ACC) with something (DAT)’, as in *Il.* 7.11–12. Both passages are shown in Figure 12.

Note that in *Il.* 7.11–12, the instrumental dative is labelled as OBJ, in spite of its uncertain argument status in the domain of syntactic valency (cf. Section 4.1, in which the similar case of *ophthalmoîsi* ‘with (my) eyes’ is discussed). Thus, the OBJ tag may well be imprecise from a theoretical standpoint, but this analysis has the welcome advantage that it demonstrates the suitability of HoDeL for this study and similar ones. Indeed, HoDeL is richer than a strictly syntactic valency lexicon and allows investigations into the behavior of event participants whose argument status is controversial, such as those regarded as *optional inner complements* in the view of the *Functional Generative Description*.

5 Conclusions

I have here presented HoDeL, a new linguistic resource intended to ease and refine the researching of Homeric verbs and their dependents. The building of HoDeL was fully documented and framed within the larger picture of morphosyntactically annotated corpora and valency lexica of ancient and modern Indo-European languages and AG. The basic functionalities and incorporated filters of the HoDeL online interface were illustrated, accompanied by suggestions about how to interpret frequency counts.

The presentation of HoDeL also provided an opportunity to re-examine a number of methodological challenges, of both a theoretical and a practical nature, that emerged while creating new resources from the existing ones. HoDeL was originally intended to be a syntactic valency lexicon of Homeric Greek. However, due to a theoretical mismatch between the source theory of annotation of AGDT 2.0 and the most widely accepted theories of syntactic valency, it evolved to include a greater number of event participants than those commonly acknowledged as arguments, such as passive agents and instrumental datives. Thus, the data contained in HoDeL is richer than that of a strict syntactic valency lexicon.

This paper also addressed some of the limitations of AGDT 2.0. It was noted that the constraints and the tagset of the analytical layer of AGDT 2.0 are not adequate for representing some peculiarities of the Homeric language, including null argument participants and tmesis. For the present, the authors have opted to leave the annotation as it is in this respect, as modifying these features would require massively rethinking and reannotating the treebank. It was also noted that HoDeL inherited a number of errors contained in the treebank, owing to its close connection with the Homeric texts treebanked at AGDT 2.0. Thus far, voice information on verbs and gaps in lemma information have been corrected in HoDeL.

Finally, after explaining to potential users how HoDeL was built, its functionalities, and what data it contains, I have shown how the lexicon can be employed to easily operationalize diverse research questions concerning Homeric verbs and Homeric syntax, and how its user-friendly interface and incorporated filters and queries allow scholars with basic computational skills to perform advanced corpus-based studies on the Homeric language. In addition, I have demonstrated how HoDeL may be used to search morphological information, transliteration and aligned translations of the AG passages, which also greatly facilitate the interpretation of the output results.

In the future, we plan to continue improving the quality of the base data contained in AGDT 2.0, for example, by correcting cases of misuse of the PNOM label. In addition, the HoDeL team is working to link the lexicon with other lexical resources of AG, such as the growing *Ancient Greek WordNet* (Biagetti et al. 2021). As shown in Zanchi et al. (2021), the enhanced access to data and the extreme user-friendliness of HoDeL can be exploited to integrate sentence frames in the metadata associated with each verbal entry of the *Ancient Greek WordNet*.

Acknowledgements

HoDeL was created at the Department of Humanities of the University of Pavia and funded by the Italian Ministry of Education and Research in the framework of the project *Transitivity and Argument Structure in Flux* (2015 PRIN call, grant no. 20159M7X5P), coordinated by Michela Cennamo and Silvia Luraghi. Chiara Zanchi and Paolo Ruffolo are the main people responsible for its creation, and we worked under Silvia Luraghi's supervision. In various ways and at different times many people contributed to HoDeL, to whom I am greatly thankful: Federico Boschetti, Giuseppe G.A. Celano, Giulia D'Agostino, Marco Forlano, Francesco Mambrini, Nahumi Nugrahaningsih, Marco Passarotti, Edoardo M. Ponti, Eleonora Sausa. The insightful comments of two anonymous reviewers contributed to improving several aspects of this paper. Needless to say, all remaining errors are mine.

Websites

- AGDT 2.0 (Ancient Greek Dependency Treebank) = https://perseusdl.github.io/treebank_data/
- The Chicago Homer* = <https://homer.library.northwestern.edu>
- DAMOS (Database of Mycenaean at Oslo University) = <https://damos.hf.uio.no/1>
- DFHG (Digital Fragmenta Historicorum Graecorum) = <http://www.dfhg-project.org>
- The Diorisis Ancient Greek Corpus* = <https://www.turing.ac.uk/research/publications/diorisis-ancient-greek-corpus>
- EAGLE (Electronic Archive of Greek and Latin Epigraphy) = http://www.edr-edr.it/it/Link_it.php
- HoDeL (Homeric Dependency Lexicon) = <https://hodel.unipv.it/hodel-res/>
- Ancient Greek WordNet* = <https://greekwordnet.chs.harvard.edu>
- ISWOC (Information Structure and Word Order Change in Germanic and Romance Languages) = <http://www.hf.uio.no/ilos/english/research/projects/iswoc/>; <http://iswoc.github.io>
- IT-TB (Index Thomisticus Treebank) = <https://itreebank.marginalia.it/itvalex>
- IT-VaLex (Index Thomisticus Valency Lexicon) = <https://itreebank.marginalia.it/itvalex>; <https://github.com/CIRCSE/ITVALEX>
- LiLa* (Linking Latin) = <https://lila-erc.eu/#page-top>
- LSJ (Liddell-Scott-Jones Greek Lexicon) = <http://www.perseus.tufts.edu/hopper/resolveform?redirect=true&entry=fe/rw>
- LDT (Latin Dependency Treebank) = https://perseusdl.github.io/treebank_data/
- Perseus Digital Library* = <http://www.perseus.tufts.edu/hopper/collections>
- PDT 3.0 (Prague Dependency Treebank) = <https://ufal.mff.cuni.cz/pdt3.0>
- PML-Tree Query = <https://ufal.mff.cuni.cz/pmltq>
- PROIEL (Pragmatic Resources of Old Indo-European Languages): <https://www.hf.uio.no/ifikk/english/research/projects/proiel/>
- SEMANTIA = <https://papygreek.hum.helsinki.fi>
- Syntacticus* = <http://syntacticus.org>
- TITUS (Thesaurus Indogermanischer Text- und Sprachmaterialien) = <http://titus.fkidl.uni-frankfurt.de/framee.htm?texte/texte2.htm>
- TLG (Thesaurus Linguae Graecae) = <http://stephanus.tlg.uci.edu>
- TOROT (The Tromsø Old Russian and OCS Treebank) = <http://torottreebank.github.io/>; <https://nestor.uit.no>
- TrEd (Tree Editor) = <https://ufal.mff.cuni.cz/tred/>
- Universal Dependencies* = <https://universaldependencies.org>
- VedaWeb (Online Research Platform for Old Indic Texts) = <https://vedaweb.uni-koeln.de>

References

- Ágel, Vilmos & Klaus Fischer. 2015. Dependency Grammar and Valency Theory. *The Oxford Handbook of Linguistic Analysis* (2nd edn.), ed. by Bernd Heine & Heiko Narrog, 223–256. Oxford: Oxford University Press.
- Anthony, Laurence. 2013. A critical look at software tools in corpus linguistics. *Linguistic Research* 30(2): 141–161.
- Aurora, Federico. 2015. DAMOS (Database of Mycenaean at Oslo). Annotating a fragmentarily attested language. *Procedia—Social and Behavioral Sciences* 198: 21–31.
- Bamman, David & Gregory Crane. 2006. The design and use of a Latin dependency treebank. *Proceedings of the Fifth Workshop on Treebanks and Linguistic Theories (TLT2006)*, ed. by Joakim Nivre & Jan Hajič, 67–78. Prague: ÚFAL.
- Bamman, David & Gregory Crane. 2008. Building a dynamic lexicon from a digital library. *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2008)*, ed. by Andreas Paepcke, Mor Naaman, & Jose Borbinha, 11–21. New York: Association for Computing Machinery.
- Bamman, David & Gregory Crane. 2011. The Ancient Greek and Latin Dependency Treebank. *Language Technology for Cultural Heritage*, ed. by Caroline Sporleder, Antall van Den Bosch, & Kalliopi Zervanou, 79–89. Berlin: Springer.
- Beschi, Fulvio. 2018. The Ancient Greek sentence left periphery. A study on Homer. *Journal of Greek Linguistics* 18: 172–210.
- Biagetti, Erica. 2018. A dependency treebank of Classical Sanskrit. MA thesis, University of Pavia.
- Biagetti, Erica, Chiara Zanchi, & William M. Short. 2021. Toward the creation of WordNets for ancient Indo-European languages. *Proceedings of the 11th Global WordNet Conference*, ed. by Piek Vossen & Christiane Fellbaum, 258–266. EACL/GWC: Global WordNet Association.
- Busa, Roberto. 1980. *Index Thomisticus: sancti Thomae Aquinatis operum omnium indices et concordantiae, in quibus verborum omnium et singulorum formae et lemmata cum suis frequentis et contextibus variis modis referuntur quaeque/consociata plurium opera atque electronico IBM automato usus digessit Robertus Busa SJ*. Stuttgart—Bad Cannstatt: Frommann—Holzboog.
- Celano, Giuseppe A. 2019. The Dependency Treebanks for Ancient Greek and Latin. *Digital Classical Philology. Ancient Greek and Latin in the Digital Revolution* [Age of Access? Grundfragen der Informationsgesellschaft 10], ed. by Monica Berti, 279–298. Berlin: De Gruyter.
- Celano, Giuseppe A. & Gregory Crane. 2015. Semantic role annotation in the Ancient Greek dependency treebank. *Proceedings of the Fourteenth International Workshop on Treebanks and Linguistic Theories (TLT14)*, ed. by Markus Dickinson, Erhard Hinrichs, Agnieszka Patejuk, & Adam Przepiórkowski, 26–34. Warsaw: Institute of Computer Science Polish Academy of Sciences.

- Colleman, Timothy, Bart Defrancq, Filip Devos, & Dirk Noël. 2004. *The Contragram Dutch–French–English Contrastive Verb Valency Dictionary*. Ghent: University of Ghent.
- Dickinson, Markus. 2005. Error detection and correction in annotated corpora. Doctoral dissertation, The Ohio State University, Columbus, OH (USA).
- Eckhoff, Hanne M. & Aleksandrs Berdičevskis. 2015. Linguistics vs. digital editions: The Tromsø Old Russian and OCS treebank. *Scripta & e-Scripta* 14–15: 9–25.
- Eckhoff, Hanne, Kristin Bech, Kristine Eide, Gerlof Bouma, Dag Trygve Truslew Haug, Odd Einar Haugen, & Marius Jøhndal. 2018a. The PROIEL treebank family: A standard for early attestations of Indo-European languages. *Language Resources and Evaluation* 52(1): 29–65.
- Eckhoff, Hanne M., Silvia Luraghi, & Marco Passarotti. 2018b. The added value of diachronic treebanks for historical linguistics. *Diachronica* 35(3): 297–309.
- Frege, Gottlob. 1891. *Funktion und Begriff*. Jena: Hermann Pohle.
- Fried, Mirjam & Hans C. Boas, eds. 2005. *Grammatical constructions: Back to the roots*. Amsterdam: John Benjamins.
- Frigione, Chiara. 2015. *Lexematik, Konstruktionsmuster und Argumentstruktur im altkirchenslavischen Verbum: Materialien für eine Lexikosyntax des Slavischen / Lessematica, modello costruzionale e struttura argomentale nel verbo paleoslavo: Materiali per una lessico-sintassi verbale paleoslava*. Doctoral dissertation, Universität zu Köln / Università per Stranieri di Siena.
- Hajič, Jan, Jarmila Panevová, Eva Buráňová, Zdeňka Urešová, & Alevtina Bémová (in cooperation with) Jiří Kárník, Jan Štěpánek, & Petr Pajas. 1999. Annotations at analytical level: Instructions for annotators. <https://ufal.mff.cuni.cz/pdt2.0/doc/manuals/en/a-layer/html/index.html> [last access: 2021-10-11].
- Hajič, Jan, Jarmila Panevová, Zdenka Urešová, Alevtina Bémová, Veronika Kolářová, & Petr Pajas. 2003. PDT-Vallex: Creating a large-coverage valency lexicon for treebank annotation. *TLT 2003: Proceedings of The Second Workshop on Treebanks and Linguistic Theories*, ed. by Joakim Nivre & Erhard W. Hinrichs, 57–68. Växjö: Växjö University Press.
- Happ, Heinz. 1976. *Grundfragen einer Dependenz-Grammatik des Lateinischen*. Göttingen: Vandenhoeck & Ruprecht.
- Haug, Dag T.T. 2012. Syntactic conditions on null arguments in the Indo-European Bible translations. *Acta Linguistica Hafniensia* 44(2): 129–141.
- Haug, Dag T.T. 2015. Treebanks in historical linguistics research. *Perspectives on historical syntax*, ed. by Carlotta Viti, 187–202. Amsterdam: John Benjamins.
- Haug, Dag T.T. & Marius L. Jøhndal. 2008. Creating a parallel treebank of the Old Indo-European Bible translations. *Proceedings of the Second Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2008)*, ed. by Caroline Sporleder & Kiril Ribarov, 27–34. Marrakech: ELRA.

- Helbig, Gerhard & Wolfgang Schenkel. 1991[1969]. *Wörterbuch zur Valenz und Distribution deutscher Verben*. Tübingen: Niemeyer.
- Herbst, Thomas, David Heath, Ian F. Roe, & Dieter Götz. 2004. *A Valency Dictionary of English*. Berlin: De Gruyter.
- Hovy, Eduard & Julia Lavid. 2010. Towards a “science” of corpus annotation: A new methodological challenge for corpus linguistics. *International Journal of Translation Studies* 22(1): 13–36.
- Iannàccaro, Gabriele. 2000. Per una semantica più puntuale del concetto di “dato linguistico”: un tentativo di sistematizzazione epistemologica. *Quaderni di Semantica* 21(1): 51–79.
- Joseph, Brian & Richard D. Janda, eds. 2003. *The Handbook of Historical Linguistics*. Oxford: Blackwell.
- Keydana, Götz & Silvia Luraghi. 2012. Definite referential null objects in Vedic Sanskrit and Ancient Greek. *Acta Linguistica Hafniensia* 44(2): 116–128.
- Kingsbury, Paul & Martha Palmer. 2002. From treebank to propbank. *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)*, ed. by Manuel González Rodríguez & Carmen Paz Suarez Araujo, 1989–1993. Las Palmas-Gran Canaria: ELRA.
- Korhonen, Anna, Yuval Krymolowski, & Ted Briscoe. 2006. A large subcategorization lexicon for natural language processing applications. *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*, ed. by Nicoletta Calzolari, Khalid Choukri, Aldo Gangemi, Bente Maegaard, Joseph Mariani, Jan Odijk, & Daniel Tapias, 1015–1020. Genoa: ELRA.
- Lattimore, Richmond A. 1951. *The Iliad of Homer*. Chicago: University of Chicago Press.
- Lattimore, Richmond A. 1967. *The Odyssey of Homer*. New York: Harper and Row.
- Luraghi, Silvia. 2003. Definite referential null objects in Ancient Greek. *Indogermanische Forschungen* 108: 169–196.
- Luraghi, Silvia. 2020. *Experiential verbs in Homeric Greek*. Leiden: Brill.
- Matthews, Peter H. 1981. *Syntax*. Cambridge: Cambridge University Press.
- Mayrhofer, Manfred. 1980. *Zur Gestaltung des etymologischen Wörterbuchs einer “Großcorpus-Sprache”*. Wien: Akademie der Wissenschaften. Phil-Hist. Klasse.
- McEnery, Tony, Richard Xiao, & Yuko Tono. 2006. *Corpus-based Language Studies: An advanced resource book*. London: Routledge.
- McGillivray, Barbara. 2013. *Methods in Latin Computational Linguistics*. Leiden: Brill.
- McGillivray, Barbara & Marco C. Passarotti. 2009. The development of the Index Thomisticus Treebank Valency Lexicon. *Proceedings of the EACL 2009 Workshop on Language Technology and Resources for Cultural Heritage, Social Sciences, Humanities, and Education* [LaTeCH—SHELT&R 2009], ed. by Lars Borin & Piroška Lendvai, 43–50. Athens: ACL.
- McGillivray, Barbara & Marco C. Passarotti. 2015. Accessing and using a corpus-driven

- Latin Valency Lexicon. *Latin Linguistics in the Early 21st Century. Acts of the 16th International Colloquium on Latin Linguistics, Uppsala*, June 6th–11th, 2011, ed. by Gerd V.M. Haverling, 289–300. Uppsala: Uppsala Universitet.
- McGillivray, Barbara & Alessandro Vatri. 2015. Computational valency lexica for Latin and Greek in use: A case study of syntactic ambiguity. *Journal of Latin Linguistics* 14(1): 101–126.
- Messiant, Cédric, Anna Korhonen, & Thierry Poibeau. 2008. LexSchem: A large subcategorization lexicon for French verbs. *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)*, ed. by Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, & Daniel Tapias, 533–538. Marrakech: ELRA.
- Murray, A.T. 1919. *Homer Odyssey*, 2 volumes. Translated by A.T. Murray. Revised by George E. Dimock. Loeb Classical Library 104. Cambridge, MA: Harvard University Press.
- Murray, A.T. 1924. *Homer. Iliad*, 2 volumes. Translated by A.T. Murray. Revised by William F. Wyatt. Loeb Classical Library 170. Cambridge, MA: Harvard University Press.
- Nyhan, Julianne & Marco C. Passarotti, eds. 2019. *One Origin of Digital Humanities: Fr Roberto Busa in his own words*. Cham: Springer.
- Panevová, Jarmila. 1994. Valency frames and the meaning of the sentence. *The Prague School of Structural and Functional Linguistics. A short introduction*, ed. by Philip A. Luelsdorff, 223–243. Amsterdam: John Benjamins.
- Passarotti, Marco C., Berta González Saavedra, & Christophe Onambele. 2016. Latin Vallex. A treebank-based semantic valency lexicon for Latin. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, ed. by Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk, & Stelios Piperidis, 2599–2606. Portorož: ELRA.
- Ruppenhofer, Josef, Michael Ellsworth, Miriam R.L. Petruck, Christopher R. Johnson, & Jan Scheffczyk. 2010. *FrameNet II. Extended theory and practice*. <https://framenet2.icsi.berkeley.edu/docs/r1.5/book.pdf> (last accessed: 2021.09.09).
- Sausa, Eleonora & Chiara Zanchi. 2015. Non-accusative null objects in the Homeric Dependency Treebank. *Proceedings of the Workshop on Corpus-Based Research in the Humanities*, ed. by Marco Passarotti, Francesco Mambrini, & Caroline Sporleder, 107–116. Warsaw: Institute of Computer Science of the Polish Academy of Sciences.
- Schreibman, Susan, Ray Siemens, & John Unsworth, eds. 2004. *A Companion to Digital Humanities*. Oxford: Blackwell.
- Schumacher, Helmut, Jacqueline Kubczak, Renate Schmidt, & Vera de Ruiter. 2004. *VALBU—Valenzwörterbuch deutscher Verben*. Tübingen: Narr.
- Sgall, Petr, Eva Hajičová, & Jarmila Panevová. 1986. *The Meaning of the Sentence in its Semantic and Pragmatic Aspects*. Dordrecht: D. Reidel.

- Siewierska, Anna. 2013. Passive constructions. *The World atlas of Language Structures Online*, ed. by Matthew S. Dryer & Martin Haspelmath. Leipzig: Max Plank Institute for Evolutionary Anthropology. <http://wals.info/chapter/107> (last accessed: 2021-05-10).
- Sinclair, John. 2005. Corpus and text: Basic principles. *Developing Linguistic Corpora: A guide to good practice*, ed. by Martin Wynne, 1–16. Oxford: Oxbow Books.
- Smyth, Herbert W. 1920. *Greek Grammar for Colleges*. New York: American Book Company.
- Tesnière, Lucien. 1959. *Éléments de syntaxe structurale*. Paris: Klincksieck.
- Tognini-Bonelli, Elena & John McH. Sinclair. 2006. Corpora. *Encyclopedia of Language and Linguistics* vol. 3 (2nd edn.), ed. by Keith Brown, 206–219. Amsterdam: Elsevier.
- Untermann, Jürgen. 1983. Indogermanische Restsprachen als Gegenstand der Indogermanistik. *Le lingue indoeuropee di frammentaria attestazione. Die indogermanischen Restsprachen. Atti del convegno della Società italiana di glottologia e della Indogermanische Gesellschaft, Udine, 22–24 settembre 1981*, ed. by Edoardo Vineis, 11–28. Pisa: Giardini.
- Vatri, Alessandro & Barbara McGillivray. 2018. The Diorisis Ancient Greek Corpus. *Research Data Journal for the Humanities and Social Sciences* 3(1): 55–65.
- Vierros, Marja. 2018. Linguistic annotation of the Digital Papyrological Corpus: *Semantia. Digital papyrology II*, ed. by Nicola Reggiani, 105–118. Berlin: De Gruyter.
- Zanchi, Chiara. 2017. New evidence for the Source-Goal asymmetry. Ancient Greek preverbs. *Space in Diachrony*, ed. by Silvia Luraghi, Tatiana Nikitina, & Chiara Zanchi, 147–178. Amsterdam: John Benjamins.
- Zanchi, Chiara. 2019. *Multiple Preverbs in Ancient Indo-European Languages*. Tübingen: Narr.
- Zanchi, Chiara & Silvia Luraghi. 2020. Presenting HoDeL—A new resource for research on Homeric Greek verbs. *Papers from the Annual International Conference “Dialogue” 2020*. <http://www.dialog-21.ru/media/5185/zanchicplusluraghis-075.pdf>.
- Zanchi, Chiara, Eleonora Sausa, & Silvia Luraghi. 2018. HoDeL, a Dependency Lexicon for Homeric Greek: Issues and perspectives. *Proceedings of Formal Representation and Digital Humanities*, ed. by Paola Cotticelli & Federico Giusfredi (a cura di), 230–256. Cambridge: Cambridge Scholars Publishing.
- Zanchi, Chiara, Silvia Luraghi, & Erica Biagetti. 2021. Linking the Ancient Greek WordNet to the Homeric Dependency Lexicon. *Computational Linguistics and Intelligent Technologies. Papers from the Annual International Conference “Dialogue” (2021)*, Issue 20: 729–737.