



# How to do things with HoDeL 2.0

Chiara Zanchi, University of Pavia chiara.zanchi01@unipv.it

# **Table of Contents**

1.	What HoDeL is and who created it2
2.	How HoDeL was built and what data it contains
3.	Preliminaries: How to type and display the Greek fonts in HoDeL (or
chc	oose not to)8
4.	Getting started and performing a simple query 10
5.	Typing and looking for your own verb(s)15
6.	Additional incorporated filters 23
Ref	erences and external links

#### 1. What HoDeL is and who created it

The **Ho**meric **De**pendency **L**exicon – HoDeL is a new linguistic resource that greatly facilitates investigating Homeric verbs and their dependents. 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* (TASF, 2015 PRIN call, grant no. 20159M7X5P), coordinated by Michela Cennamo and Silvia Luraghi. Chiara Zanchi and Paolo Ruffolo are the main responsible for its creation and worked under Silvia Luraghi's supervision. In various ways and at different times many people contributed to HoDeL, to whom I am greatly thankful: (in alphabetical order) Federico Boschetti, Giuseppe G. A. Celano, Giulia D'Agostino, Marco Forlano, Francesco Mambrini, Nahumi Nugrahaningsih, Marco Passarotti, Edoardo M. Ponti, and Eleonora Sausa.

HoDeL is a corpus-based lexicon automatically induced from the Homeric texts treebanked at AGLDT 2.0 (*Ancient Greek and Latin Dependency Treebank*). HoDeL allows users to search for Homeric verbs and their dependents that, according to the guidelines of the analytical layer of AGLDT 2.0 (see a-AGLDT 2 guidelines in the references), are included in verbal valency. In the next sections, we will briefly touch upon how HoDeL was built (Section 2) and thoroughly explain how to effectively use it (Sections 3##).

An exhaustive explanation of the functionalities of HoDeL is crucial for users: as Anthony (2013: 142) emphasized, data may yield different results not only because of the researchers' subjectivity but also due to idiosyncrasies of the tool or interface through which data can observed by researchers. Furthermore, acknowledging that (i) annotated corpora and linguistic resources induced from them inevitably contain errors and that (ii) annotations are based on an underlying linguistic theory should be followed by carefully instruct users about both of them. This is crucial, as linguistic resources induced from annotated corpora distance the

researcher from the researched data at least at three different levels: (a) the interface-level; (b) the data-extraction level; (c) the annotation-level.

### 2. How HoDeL was built and what data it contains

In this section, I cursorily touch upon a number of issues that we have faced while building HoDeL, which we invite users to keep in mind when using the lexicon for their own research purposes. For a thorough discussion of such issues, we encourage readers to have a look at Zanchi et al. (2018), Luraghi & Zanchi (2020), and Zanchi (submitted).

HoDeL has been modelled on IT-VaLex (*Index Thomisticus Valency Lexicon*), documented in McGillivray & Passarotti (2009) and Passarotti (2011).

HoDeL is closely connected with the Homeric texts treebanked at AGLDT 2.0 too, as it is a corpus-based lexicon automatically induced from the syntactic or analytical annotation layer of AGLDT 2.0. The guidelines of the analytical layer of AGLDT 2.0 rely on those of the *Prague Dependency Treebank* (PDT) with some *addenda* aimed to increase descriptiveness and precision by incorporating Smyth's *Greek Grammar for Colleges* (Smyth 1920; Celano 2019: 285–286). From the analytical layer of AGLDT 2.0, HoDeL automatically extracts all Homeric finite and non-finite verbal forms together with their dependents labelled as Sbj, Obj, OComp, and PNom, which are illustrated in Table 1.

Tag	Gloss	Example
Sbj	subject	hề murí' Akhaioîs álge' éthēke 'which brought
		countless woes upon the Achaeans' (II. 1.2)
Obj	object	mênin áeide theà '(The) wrath sing, goddess' (ll. 1.1)
		hề murí' <b>Akhaioîs álge'</b> éthēke 'which brought countless
		woes upon the Achaeans' (ll. 1.2)
OComp	object	autoùs dè <b>helốria</b> teûkhe kúnessin oiōnoîsí te pâsi 'and
	complement	made them themselves <b>spoil</b> for dogs and every bird' (II.
		1.4)
PNom	predicate	hòs nûn pollòn <b>áristos</b> Akhaiôn eúkhetai eînai 'who
	nominal	now claims to be far ( <b>the best</b> ) of the Achaeans' ( <i>II</i> . 1.91)

Table 1. Extracted dependents with examples (bold marks the dependent tagged with the corresponding label)

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 a-AGLDT 2 guidelines). All aforementioned dependents may either be direct child nodes of a verbal form or be attached to the verbal head via one of the bridge nodes, glossed in Table 2.

Tag	Gloss
AuxP	preposition
AuxC	conjunctions
Coord	coordinator, including coordinative conjunctions and particles
Apos	apposing elements, such as commas

Table 2. Bridge nodes

According to the guidelines of the analytical layer of AGLDT 2.0, the dependents in Table 1 should correspond to verbal arguments. Therefore,

the automatic extraction of these dependents from the Homeric poems treebanked at AGLDT 2.0 was expected to yield a data-driven valency lexicon for Homeric Greek.

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 or adjectives that may morphologically agree with their head noun, but qualify the whole event denoted by the verb) that the AGLDT 2.0 guidelines explicitly do not consider as belonging to the verbal valency.

Dependents in Table 1 have been extracted using a series of SQL queries and then recorded in a spreadsheet, which works as a backend for HoDeL interface. The original query algorithm and its implementation were conceived to build IT-VaLex (McGillivray and Passarotti 2009). To induce HoDeL, the queries have been adapted to the AGDT 2.0 tagset.

However, HoDeL resulted not to be a valency lexicon as we expected, due to at least three kinds of issues. First and foremost, the AGLDT 2.0 is structured after the PDT (**P**rague **D**ependency **T**reebank). Crucially, the theory of valency of the *Functional Generative Description* (Sgall et al. 1969; Panevová 1994: 223-244), which underpins PDT and AGLDT 2.0, and which is described in the semantic/pragmatic or 'tectogrammatical' layer of PDT, contains a number of idiosyncrasies, including, but not limited to, the following. (i) Passive agents are annotated as Obj, whereas passive voice is generally acknowledged as a valency-decreasing strategy, by removing agents from argument structure (cf. e.g. Shibatani 1985, 1988; Dixon & Aikhenvald 2000: 7 ff.; Kulikov et al. 2006: vii-xvii). (ii) The notion of obligatoriness, despite being crucial to distinguish arguments from adjuncts, is explicitly said to be problematic and, as such, not much elaborated in the guidelines. Such rather vague notion of obligatoriness raises a number of issues in the annotation of AGLDT 2.0 and results, for

example, in an inconsistent annotation of Beneficiary and Instrumental dative dependents or of genitive Source and Stimulus dependents with perception verbs (please, check Zanchi & Luraghi 2020 and Zanchi [submitted] for relevant examples).

Second, the analytical layer of AGLDT 2.0 does not include empty nodes for referential null subject and objects (in the treebank architecture, these should be integrated at the tectogrammatical annotation layer, which, however, is not available for the Homeric poems). Ancient Greek is a prodrop language, thus null subjects represent the default solution with topical subjects. Null referential objects have been shown to be preferable or obligatory under certain syntactic and pragmatic conditions (Luraghi 2003; Keydana & Luraghi 2012). 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.

Third, the usage of the Pnom label is rather inconsistent in the treebank: for example, in a sentence such as *laoi* d' **ein agorêi** ésan athrói 'The folk, gathered together, was **in the place of assembly**.' (Il. 18.497), the prepositional phrase in bold is at times tagged as Pnom, despite the verb *eimí* 'be' functioning as an existential/locative verb and not as a copula in similar contexts.

Fourth, a further class of issues is due to the inherent difficulty in interpreting the syntactic status of certain events participants, such as Location or Direction with posture verbs (cf.  $hiz\bar{o}$  'sit', e.g. Il. 9.87). This matter is further complicated by the theory of valency of the *Functional Generative Description*. In example (1), the Origin participant, encoded by ek+GEN, is tagged as Obj in dependence of a verb of breaking, *ágnumi* 'break':

(1) 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)

This annotation, however, is problematic for multiple reasons: first, it regards Origin as an argumental participant of a verb such as 'break'; second, it treats the initial local particle (see Zanchi 2019: 82-86 on this terminology) *ek* as a preposition governing the genitive *astragálon*, which is not necessarily the case. The actual syntactic relation holding between the local particle and the genitive case may well be less tight, and ek could 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 it takes the prepositionless genitive. As is well-known, in Homeric Greek, preverbs could occur in the so-called 'tmesis' positions and thus be 'split' by various linguistic material from the verbs that they semantically modify, retaining much of the freedom of their original adverbial status (cf. Zanchi 2019: ch. 3 for a diachronic interpretation of this preverb positioning). This syntactic ambiguity is reflected in an inconsistent annotation: in Od. 10.559-560, a quasi-identical formulaic expression (ek dé hoi aukhen astragálon eáge, in which only the external possessor dative changes) occurs, but ek is tagged as AuxZ. The latter is the label usually assigned to local particles in 'tmesis' position. In AGLDT 2.0, the tag AuxZ 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 is inadequate to account for this peculiarity of the Homeric language. Since Homeric Greek is a language with free word order, there is no easy and automatic way to find all examples similar to (1) and, therefore, solving this kind of problem at the annotation level does not seem possible.

Finally, given that HoDeL takes its data directly from AGLDT 2.0, it inherited a number of annotation errors contained in the treebank. We worked and are working to fix some of them: For example, we have manually re-annotated voice information of all verbal forms, which are now tagged according to strict morphological criteria. Besides, we are trying to correct the lemmatization of some verbal forms and dependents. For example, the verb *horáō* 'see' has a paradigm consisting of three stems, specifically, the present *horáō*, the aorist *eîdon*, and the perfect *ópopa*, which are, however, lemmatized as two different entries, the former including the forms based on the themes *horáō* and *ópopa*, and the latter including the forms based on the theme of the aorist *eîdon*. However, the preverbed verb *eis-oráō* 'look', whose paradigm features the same three themes, is lemmatized as a single entry.

# **3. Preliminaries: How to type and display the Greek** fonts in HoDeL (or choose not to)

In order to type the desired Greek lemmas, HoDeL users should employ Beta Code, as in the *Perseus Project Library* and in the TLG (**T**hesaurus **L**inguae **G**raecae). The correspondences between Greek and Beta Code characters are reported in Table 3 for users' convenience. The least intuitive Greek-Beta Code correspondences are highlighted in grey.

Greek	а	β	Y	δ	з	ζ	η	θ	Т	к	λ	μ	v	ξ	0	п	ρ	σ,ς	т	U	φ	Х	Ψ	ω
Beta Code	а	b	g	d	е	z	h	q	i	k	-	m	n	с	0	р	r	s	t	u	f	х	у	w
Greek	ά	à	ã	ά	ά	Ó	ά	ά		č	Ì	Ő	ì	Ó	á	ò	Ϋ́	à	ć	ź –	ų			
Beta Code	a/	a∖	a=	a)	a(	a	)/	a(	7	a	)\	a(	()	a)	)=	a(	=	a	a)	/	a)=			
	An * should be used to type capital letters (e.g. Borgnic 🗲 *brishi/s)																							

Table 3. Greek and Beta Code characters correspondences

In HoDeL, users can choose to visualize data in Greek fonts or transliterated, by flagging **greek** or **trans** in the **Display** box at the top of HoDeL homepage, as shown in Figure 1 and Figure 2.

Our algorithm provides a transliteration that conforms to the most common academic standards. Example (2) shows how the first line of the Odyssey looks like in Greek characters (2)a, Beta-Code (2)b, and transliteration (2)c.

- (2) a. ἄνδρα μοι ἕννεπε μοῦσα πολύτροπον ὃς μάλα πολλὰ
  - b. a)/ndra moi e)/nnepe mou=sa polu/tropon o(\s ma/la polla\
  - c. ándra moi énnepe moûsa polútropon hòs mála pollà

'Tell me, Muse, about the wily man who (wandered) long and far' (Od. 1.1)



Figure 1. How to visualize the Greek characters

# HoDeL

greek ✓ trans		Order By : [ʌ]lemm	a []rev. lemma []freque	Filter :
	List of Verba	l Head Lemma	S	
uery	Next Deere			
	Next Page			
Verbal Head Lemmas: 2482	(8)	ágnumi (26)	akéomai (9)	aléomai (36)
000000000000000000000000000000000000000	aáō (19)	ágō (308)	akheúō (109)	alēteúō (5)
Jccurrences. 40695	abakéō (1)	agoráomai (27)	akhlúō (2)	aletreúō (1)
	abrotázō (1)	agoreúō (166)	ákhthomai (5)	aléxō (20)
as Number	adéō (5)	agréō (6)	akontízō (35)	algéō (4)
<b>3</b>	aeídō (40)	agróssō (1)	akostáō (2)	alitaínō (7)
	aeikízō (8)	agurtázō (1)	akouázomai (2)	allophronéō (2)
as Order	aeírō (67)	aidéomai (42)	akoúō (182)	aloáō (1)
	aelptéō (1)	aikhmázō (1)	akrokelainiáō (1)	alogéō (2)
	áēmi (12)	ainéō (10)	alálēmai (15)	alphánō (4)
Args Lemmas: 4219	áesa (2)	ainízō (2)	álalke (13)	álthomai (1)
Occurrences: 49137	aēthéssō (1)	aínumai (14)	alalúktēmai (1)	alúō (5)
occurrences. 49107	aéxō (20)	aíō (22)	aláomai (27)	aluskázō (3)
	agaíomai (1)	aióllō (1)	alaóō (2)	alúskō (27)
guments	agállō (7)	aírō (43)	alapázō (11)	alússō (3)
	ágamai (27)	aiskhúnō (12)	alastéō (2)	amaldúnō (3)
	agapáō (2)	aíssō (61)	aldaínō (2)	amáō (5)
	agapázō (6)	aísthō (2)	aldéskō (1)	amathúnō (1)
	ageírō (62)	aistóō (2)	aleeínō (26)	ameíbō (168)
	aggéllō (27)	aitéō (14)	alegízō (6)	ameléō (4)
	aginéō (6)	aíthō (22)	alégō (11)	amélgō (5)
	agkázomai (1)	aitiáomai (7)	alegúnō (5)	amenēnóō (1)
	ágkhō (1)	aitízō (10)	aleíphō (10)	amérdō (6)
	aglaízō (1)	akakhízō (2)	aléō (7)	ampékhō (1)
	agnoéō (7)	akēdéō (2)	aleómai (1)	amphagapázō (2)

Figure 2. How to visualize the transliteration

## 4. Getting started and performing a simple query

Figure 3 below represents HoDeL homepage. As you can see, HoDeL by default shows a list of all Homeric verbal lemmas (in orange), alphabetically ordered. After each lemma, its frequency in the poems is provided. If preferred, users can choose to visualize Homeric verbs by *reverse* alphabetical order or by increasing frequency, by flagging either the [^]<u>rev. lemma</u> or the [^]<u>frequency</u> box on top right of the home page.

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.

Note that the token frequency of dependent lemmas is higher than the token frequency of verbal heads (i.e., 49,137 > 40,693). This is not surprising: if two dependent lemmas are taken by a certain verb *in the same occurrence*, that occurrence is listed twice in the dependent occurrence count, that is, one for each dependent. When users add filters to their queries, HoDeL always provides these and other frequency counts.

Furthermore, users can visualize the entire array of lemmas that depend on Homeric verbs, by clicking on <u>+ Args Lemmas: 4219</u>. By doing so, they obtain the screenshot of Figure 4. Note that a number of verbs, such as *aeírō* `lift', *hairéō* `take', *aírō* `raise', etc., appear in the dependent list of Figure 4: these verbs are there as they function as main verbs in dependent Sbj or Obj clauses.

# HoDeL

ay trans 🗘		Order By : [^]lemm	a []rev. lemma []freque	Filter :
	List of Verb	al Head Lemma	S	
ery	Next Page			
erbal Head Lemmas: 2482	(8)	ágnumi (26)	akéomai (9)	aléomai (36)
10602	aáō (19)	ágō (308)	akheúō (109)	alēteúō (5)
Occurrences: 40693	abakéō (1)	agoráomai (27)	akhlúö (2)	aletreúō (1)
	abrotázō (1)	agoreúō (166)	ákhthomai (5)	aléxō (20)
Number	adéō (5)	agréō (6)	akontizō (35)	algéō (4)
	aeidō (40)	agróssö (1)	akostáō (2)	alitaínō (7)
	aeikizō (8)	agurtázō (1)	akouázomai (2)	allophronéö (2)
Drder	aeírō (67)	aidéomai (42)	akoúō (182)	aloáō (1)
	aelptéö (1)	aikhmázō (1)	akrokelainiáö (1)	alogéō (2)
	áēmi (12)	ainéō (10)	alálēmai (15)	alphánō (4)
Lemmas: 4219	áesa (2)	ainízō (2)	álalke (13)	álthomai (1)
urrences: 49137	aēthéssö (1)	ainumai (14)	alalúktēmai (1)	alúõ (5)
	aéxõ (20)	aiö (22)	aláomai (27)	aluskáző (3)
	agaíomai (1)	aiólló (1)	alaõõ (2)	alúskō (27)
ents	agállō (7)	airō (43)	alapázō (11)	alússō (3)

Figure 3. HoDeL homepage

HoDeL				
The Homeric Dependency Lexico	on			
Display trans 🗘		Order By : [ʌ] <u>lemma</u>	] ]rev. lemma [ ]frequ	Filter :
Query	List of Argu:	ment Lemmas		
+ Verbal Head Lemmas: 2482 + Occurrences: 40693	(49) aáatos (1) áatos (1)	aeírō (1) aékōn (1) áella (11)	Agélaos (3) agélē (9) Agéleōs (2)	aglaíző (1) aglaós (1) ágnőstos (2)
Args Number	Áblēros (1) Ábudos (1)	aér (25) aesíphrön (1)	agénōr (1) agēnoría (3)	ágnumi (3) ágō (25)
Args Order	Abudóthen (1) adaémōn (2) adaēmonía (1) adákrutos (1)	aëtës (4) aéthlion (3) aéthlios (2) aetós (5)	agēraos (7) agérastos (1) aggelía (30) aggéllō (1)	agonos (1) agorá (52) agorênde (4) agorêtês (3)
+ Args Lemmas: 4219 + Occurrences: 49137	Ádamas (1) Adámas (1) adeês (1)	ágalma (6) ágamai (4) Agamémnön (120)	ággelos (14) ággos (6) aginéō (1)	agorêthen (1) agorêtús (1) agoreúő (16)
Arguments	Ádmētos (1) Adrásteia (1)	aganóphrön (1) aganophrosúnē (1)	Agkhíalos (2) agkhímolos (2)	agón (13) ágra (2)

Figure 4. Visualizing the dependent lemmas

In both lists of Figure 3 and 4, each verbal or dependent lemma is in turn clickable. What appears after these clicks is explained in what follows and visualized in Figures 5 and 6.

By clicking on a verb – say,  $akou\bar{o}$  'hear' –, users get (i) all its forms contained in the Homeric poems, (ii) their ordered contexts of occurrences (automatically chunked by an algorithm that exploits strong punctuation marks), (iii) a number of syntactic subtrees representing the queried verb and its dependents tagged as Sbj, Obj, Pnom, Ocomp, if any (Figure 5). To see more results, users can simply click on <u>Next Page</u>.

HoDeL summarizes the selected query filters in the grey box 'Query' and at the top of the output page (see the green circles in Figure 5) and provides users with frequency information: the verb  $akou\bar{o}$  '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. Again, note that the token frequency of argument lemmas is higher than the token frequency of  $akou\bar{o}$  (210 > 182). This is not surprising: if two argument lemmas are taken by  $akou\bar{o}$  in the same occurrence, that occurrence is listed once in the verb frequency count, but twice in the argument frequency count, that is, one time for each argument lemma.

By clicking on a dependent lemma – say, *Aíguptos* 'Egypt'–, users get the list of verbs that take that dependent (Figure 6). In turn, by clicking on each of the resulting verbs in Figure 6, for example on *hiknéomai* 'come', users obtain the relative contexts and subtrees for that verb (Figure 7).

HoDeL	
The Homeric Dependency Lexicon	
Display greek ‡	Occurrences and Contexts - Lemma: ἀκούω
Query	Next Page
verb: ἀκούω, active drop all	liad;1.380-1.382 τοῖο δ΄ Απόλλων εὐξαμένου ἦκουσεν, ἐπεὶ μάλα οἱ φίλος ἦεν,
+ Occurrences: 182	ήκε δ έπ Άργείοισι κακὸν βέλος • 🗔
Args Number	ήκουσεν
Args Order	рвл
+ Args Lemmas: 86 + Occurrences: 210	τοίο
Arguments	lliad;1.396-1.400 πολλάκι γάρ σεο πατρός ένὶ μεγάροισιν <b>ἁκουσα</b>
	εύχομένης οτ έφησθα κελαινεφέϊ Κρονίωνι
	οίη ἐν ἀθανάτοισιν ἀεικέα λοιγὸν <b>ἀμῦναι ,</b>
	ό <b>ππ</b> ότε μιν <b>ξυνδῆσαι</b> Όλύμ <b>π</b> ιοι <b>ἡθελον</b> ἀλλοι
	Ηρη τ΄ ήδὲ Ποσειδάων καὶ Παλλὰς Ἀθήνη• 🗔
	άκουσα
	рвл
	σεο

Figure 5. A simple query from the verb list

HoDeL							
The Homeric Dependency Lexicon							
Display trans	List of Ve - Argume	Order By : [^]le rbal Head Lemr ent Lemma: Aíg	ernma []rev. lemma []frec nas uptos	iuency Filter :			
Argument Lemma: Aiguptos drop all     + Verbal Head Lemmas: 4     + Occurrences: 5	eîmi (1)	ékhō (1)	epelaúnō (1)	hiknéomai (2)			
Args Number Args Order							
+ Args Lemmas: 1 + Occurrences: 5							
Arguments							

Figure 6. A simple query from the dependent list



Figure 7. Query: Aíguptos + hiknéomai

# 5. Typing and looking for your own verb(s)

So far, we performed a couple of simple queries, and we did that starting either from the verb list or the dependent list, as provided by HoDeL (Section 4). Needless to say, the lexicon allows users to directly type in the verbal and dependent lemmas they are interested in. To do that, let us go back to HoDeL homepage, by clicking either on *The Homeric Dependency Lexicon* or on <u>drop all</u>, both framed in green in Figure 8. As

an aside note, it is worth pointing out here that, by flagging either constraint summarized in the grey **Ouerv** box and clicking on **drop unchecked**, users can go back to previous steps of their query without jumping directly to the homepage.

HoDeL	
The Homeric Dependency Lexicon	
Display trans +	Occurrences and Contexts - Lemma: <b>hiknéomai</b> - Argument Lemma: <b>Aíguptos</b>
Constraints: verb: hiknéomai, middle Argum r Lemma: Aiguptos drop all drop unchecked + Occurrences: 2 Args Number Args Order	Odyssey;14.257-14.258 " pemptaîoi d Aígupton eürreítēn hikómestha , stêsa d en Aigúptōi potamôi néas amphielíssas. hikómestha DBJ Aigupton
+ Args Lemmas: 1 + Occurrences: 2 Arguments	Odyssey; 17.447-17.448 stêth ohútōs es mésson, emês apáneuthe trapézēs, mề tákha pikrền Aígupton kaì Kúpron híkēai · híkēai kai BJ_COBJ_CO Kúpron

Figure 8. Going back in HoDeL

Let us now move on and type in a Greek verb. For example, let us look for the verbal lemma *akoúō* 'hear'. By clicking on the grey box **Ouerv**, users can open the query window, where, using Beta-Code (cf. Section 4), they can type the requested lemma, a)koúw, in the **Verbal Head Lemma** box, as shown in Figure 9. After clicking on the **Submit** button, similarly to what we have seen before, HoDeL gives back all forms of *akoúō* contained in the Homeric poems, their contexts, and the syntactic subtrees representing the queried verb and its supposedly argumental dependents. This output is captured in Figure 10: overall,  $akou\bar{o}$  occurs 182 times in the poems and takes 86 different dependent lemmas. According to the AGLDT 2.0 annotation, the first form of  $akou\bar{o}$  occurring in the poems, **ékousen** (in the 3<sup>rd</sup> person singular of the indicative aorist), showing up in *II*. 1.380-382, takes one Obj dependent, inflected in the genitive singular. In the output passages and subtrees, the verbal forms are framed in orange; the dependents are circled highlighted in blue; the subtree additionally contains the prepositional bridge-nodes when present (see the a-AGLDT guidelines for this terminology), which are on the other hand not circled in the output passages.

HoDeL							
The Homeric Dependency Lexicon							
play greek \$		Order By : [^]lemr	na [ ]rev. lemma [ ]freque	ncy Filter :			
	List of Verba	l Head Lemma	ıs				
Query	Next Page						
Poem 🗘	(8)	<b>άείδ</b> ω (40)	άιστόω (2)	ຕ່າຍ (6)			
Verbal Head Lemma	(ອ) ໃຮ່ນປະບົດ (1)	άεικίζω (8)	aiorow (2)	άλενύνω (5)			
a)kou/w	άάω (19)	άείοω (67)	aitéw (14)	άλένω (11)			
Notes and the second seco	άβακέω (1)	άελπτέω (1)	αίτιάομαι (7)	άλεείνω (26)			
Voice ÷	άβροτάζω (1)	άέξω (20)	αίτίζω (10)	άλείφω (10)			
	άναίομαι (1)	άεσα (2)	αίχμάζω (1)	άλέξω (20)			
Exact Sequence	άνάλλω (7)	άζομαι (8)	άίω (22)	άλέομαι (36)			
Exact Cardinality	άγαμαι (27)	άζω (1)	άκαχίζω (2)	άλεόμαι (1)			
	άγαπάζω (6)	άηθέσσω (1)	άκέομαι (9)	άλετρεύω (1)			
Argument Lemma	άγαπάω (2)	ăημι (12)	άκηδέω (2)	άλέω (7)			
	άγγέλλω (27)	άθερίζω (3)	άκοντίζω (35)	άλητεύω (5)			
	άγείρω (62)	άθλεύω (4)	άκοστάω (2)	άλθομαι (1)			
Relation ÷	άγινέω (6)	άθλέω (2)	άκουάζομαι (2)	άλιόω (3)			
Case/Mood \$	άγκάζομαι (1)	άθρέω (5)	άκούω (182)	άλίσκομαι (18)			
	άγλαίζω (1)	άθύρω (1)	άκροκελαινιάω (1)	άλιταίνω (7)			
Prep. +	άγνοέω (7)	αίδέομαι (42)	άλάλημαι (15)	άλλομαι (28)			
Coni.	άγνυμι (26)	αἴθω (22)	άλαλκε (13)	άλλοφρονέω (2)			
	άγοράομαι (27)	αίνέω (10)	άλαλύκτημαι (1)	άλοάω (1)			
Position 🗘	άγορεύω (166)	αίνίζω (2)	άλάομαι (27)	άλογέω (2)			
	άγρέω (6)	αἴνυμαι (14)	άλαόω (2)	άλυσκάζω (3)			
	άγρώσσω (1)	αἰόλλω (1)	άλαπάζω (11)	άλύσκω (27)			
Add another Argument	άγυρτάζω (1)	αίρέω (415)	άλαστέω (2)	άλύσσω (3)			
	άγχω (1)	αἴρω (43)	άλγέω (4)	άλύω (5)			
Submit Reset	άγω (308)	άίσθω (2)	άλδαίνω (2)	άλφάνω (4)			
	άδέω (5)	άίσσω (61)	άλδήσκω (1)	άμαθύνω (1)			

Figure 9. Looking for the verb a)kou/w

HoDeL						
The Homeric Dependency Lexicon						
Display trans 🗘	Occurrences and Contexts - Lemma: <b>akoúō</b>					
Query	Next Page					
Constraints: ✓ verb: akoúō, active drop all	<sup>Hiad;1.380-1.382</sup> toîo d'Apóllōn <b>euxaménou Ékousen</b> , epeì mála hoi phílos <b>êen</b> ,					
+ Occurrences: 182	hêke d ep Argeíoisi kakòn bélos • 🗟					
Args Number	ékousen					
Aras Order	рвл					
+ Args Lemmas: 86	toîo					
+ Occurrences: 210						
	lliad; 1.396-1.400					
Arguments	eukhoménēs hót éphēstha kelainephéï Kroniōni					

Figure 10. Output of the query shown in Figure 9

If users pass over whatever word in the given outputs with their pointer, they obtain morphological information regarding that word, as annotated in AGLDT 2.0. As shown in Figure 11, for example, HoDeL tells that *Argeíoisi* is the dative plural of the masculine noun *Argeîos* 'Argive'.

Moreover, if users click on the small blue folder following the Greek text, they can see or hide the English translation of the output passages (Figure 12). The translations have been aligned automatically with the Greek text through an algorithm that exploits punctuation marks and the text chunks contained in the text provided at the Perseus Project. The automatic alignment has later on been manually checked and, when necessary, integrated/modified according to the translation available at *The Chicago Homer*.

HoDeL	
The Homeric Dependency Lexicon	
Display trans 🗘	Occurrences and Contexts - Lemma: <b>akoúō</b>
Query	Next Page
Constraints: ♥ verb: akoúõ, active drop all	<sup>lliad</sup> ;1.380-1.382 toîo d'Apóllōn <b>euxaménou <mark>ékousen</mark>,</b> epeì mála hoi phílos <b>êen</b> ,
+ Occurrences: 182	hêke d ep Argeíoisi kakòn bélos • 🗟 and Apollo heard his praye Alicetica infloction to him, and sent against the Argives an evil shaft.
Args Number	ékousen dative
Args Order	OBJ masculine plural
+ Args Lemmas: 86 + Occurrences: 210	toio
Arguments	<sup>lliad</sup> ;1.396-1.400 polláki gár <mark>seo</mark> patròs enì megároisin <mark>ákousa</mark>

Figure 11. Morphological information

HoDeL	
The Homeric Dependency Lexico	n
Display trans 🗧	Occurrences and Contexts - Lemma: <b>akoúō</b>
Query	Next Page
Constraints: ✓ verb: akoúō, active drop all	lliad;1.380-1.382 toîo d'Apóllōn <b>euxaménou <mark>ékousen</mark>, epeì mála hoi phílos <b>êen</b>.</b>
+ Occurrences: 182	hêke d ep Argeioisi kakòn bélos • 🗔
Args Number	ékousen
Args Order	DBJ
+ Args Lemmas: 86 + Occurrences: 210	toio
Arguments	<sup>lliad</sup> ;1.396-1.400 polláki gár <mark>seo</mark> patròs enì megároisin <mark>ákousa</mark>

Figure 12. Translation of the output passages

Besides the frequencies that we have discussed in Section 4, HoDeL gives users additional quantitative information on searched verbs. As represented in Figure 13, by clicking on the purple box Args Number, users can easily see how many times *akoúo* takes 0-3 arguments. The number of possible arguments is not fixed in the form, but data-driven: if another verb – say,  $\dot{a}z\bar{o}$  'dry' – occurs only once in the poems and takes 0 arguments, the only available result for  $\dot{a}z\bar{o}$  is **No. Args; 0** in the Args Number box. Going back to the outputs of *akoúo*, in addition, by clicking on one of the available results - say No. Args 1 (103) -, users get a list of functions that that single dependent carries out, together with the relative number of occurrences for each function; in this case: Obj (84), Obj Ap (2) = appositional object, Obj Co (8) = coordinated object, Sbj (9); a thorough list of these labels can be found in the a-AGLDT 2.0 guidelines). By flagging one of these categories (i.e. argument number) and subcategories (i.e. argument number and functions), users obtain results filtered by these chosen parameters.



Figure 13. Number of arguments

As captured in Figure 14, users can also use HoDeL to get information regarding constituent order in Homeric Greek. At a lower level of granularity, by clicking on the purple box **Arg Order**, users can visualize the possible relative orderings of Obj(s) and verbs, as well as the frequency of each order; these are called 'Cat.(egories)' in the lexicon. At a higher level of granularity, by clicking in turn on one of the previously obtained categories – say, **OBJ;V** –, users can access the frequencies of the relative ordering of Obj(s), verbs, and Sbj(s); these are called 'SubCat.(egories)' in HoDeL. Again, as seen above, these categories (Obj-verb orders) and subcategories (Obj-verb-Sbj orders) can be flagged to get filtered contexts and subtrees.



Figure 14. Word order information on verbal arguments

Figure 15 illustrates that users can also obtain frequency information and filter their outputs by argument relation and argument case/mood, clicking and opening the blue box **Arguments**. For example, users can decide to visualize only the occurrences in which *akoúō* takes coordinated objects by flagging **OBJ\_AP\_CO** (appositional coordinated Objs) and **OBJ\_CO** (coordinated Objs) in the **Arg By Relation** box. The **Arg by Case/Mood** box is instead useful if users are interested in visualizing arguments that only feature certain morphosyntactic characteristics, i.e. case if we are dealing with nominal arguments, mood if we are dealing with Sbj/Obj clauses. For example, by flagging **Mood Indicative** and **Mood Infinitive**, users get all argumental subordinate clauses taken by the verb *akoúō*. As in the case of the number of arguments we have discussed above, the possible categories in the **Arg by Relation** and in the **Arg by Case/Mood** boxes are also data driven.

Query	Next Page
Constraints: ✓ verb: ἀκούω, active drop all	Hiad;1.380-1.382 τοῖο ὁ Ἀπόλλων <b>εὐξαμένου ἦκουσεν ,</b> ἐπεὶ μάλα οἱ φίλος ἦ <b>εν ,</b>
+ Occurrences: 182	<b>ἦκε</b> δ΄ ἐπ΄ Άργείοισι κακὸν βέλος • 🗔
Args Number	ἤκουσεν
Args Order	DBJ
+ Args Lemmas: 86 + Occurrences: 210	τοῖο
Arguments	lliad;1.396-1.400 πολλάκι γάρ <mark>σεο</mark> πατρὸς ἐνὶ μεγάροισιν <b>ἀκουσα</b>
- Args By Relation OBJ (130) OBJ_AP (3) OBJ_AP,OBJ (1) OBJ_AP,CO (1) OBJ_CO (11) SBJ (39) SBJ_AP (1) SBJ_CO (2) - Args By Case/Mood Case Nominative (40) Case Accusative (63) Case Accusative (74) Mood Indicative (6) Mood Infinitive (4)	εύχομένης ὀf ἐφησθα κελαινεφέϊ Κρονίωνι οἰη ἐν ἀθανάτοισιν ἀεικέα λοιγὸν ἀμῦναι, ὀππότε μιν ξυνδῆσαι Ὁλύμπιοι ἦθελον ἀλλοι Ἡρη f ἡδὲ Ποσειδάων καὶ Παλλὰς Ἀθήνη・ ακουσα BJ σεο

Figure 15. Filtering arguments by relation and by case/mood

# 6. Additional incorporated filters

In Section 6, we have seen that, by clicking on the grey <u>Query</u> box, users can open a search form, type verbal lemmas in Beta-Code, and perform specific queries on them. As represented in Figure 16, in this search form, additional filters are incorporated.

Р	oem 🔶
	Verbal Head Lemma
	Voice +
E	xact Sequence
	Argument Lemma
	Relation 🗘
	Case/Mood 🗘
	Prep.
	Conj. 🛟
	Position 🗘
ſ	Add another Argument
ſ	

Figure 16. Additional filters

First, users can work only on a single Homeric poem, by using the dropdown menu **Poem**. Similarly, they can search for verbs in a specific morphological **Voice** (available options = active : passive : middle : mediopassive).

In addition to verbal lemmas, users can also search for specific argument lemmas, by typing them in the **Argument Lemma** box. Users can also filter their output contexts and subtrees based on some features of the argument lemmas, by using the drop-down menus provided under the **Argument Lemma** box. Specifically, they can search by **Relation** (available options = Sbj : Obj : Pnom : Ocomp), by **Case/Mood** (available options for Case = nominative : genitive : dative : accusative : vocative; available options for Mood = Indicative : Subjunctive : Infinitive : Imperative : Participle : Optative), by **Preposition** (a data driven list of the Ancient Greek lemmas that are annotated as prepositions in AGLDT 2.0 is automatically given by the resource, both in Greek and in Latin scripts), by **Conjunction** (a data driven list of the Ancient Greek lemmas that are annotated as conjunctions in AGLDT 2.0 is automatically given by the resource, both in Greek lemmas that are annotated as conjunctions in AGLDT 2.0 is automatically given by the resource, both in Greek and in Latin script), and by **Position** (an argument can occur before verb : after verb: b./a. verb). All these parameters can be combined with one another and can be associated with a typed verbal and/or argument lemma. As seen before, users should remember to click the **Submit** button to run their queries.

HoDeL also allows users to search for more than one argument at one time: to do this, one should employ the **Add another Argument** button (framed in green in Figure 16). By clicking on it, an additional **Argument Lemma** box appears, together with the related drop-down menus for choosing argument features (Figure 17). Each additional argument can be deleted using the **Delete this argument** button (circled in green).

When searching for more than one argument, the options **Exact Sequence** and **Exact Cardinality** become useful (these are framed in green in Figure 17): the former searches for the exact sequence of arguments as in the form below; the latter searches for the exact number of arguments as listed in the form below, regardless of their order. The button **Reset** clears the form.

Query	
Poem 🗘	
Verbal Head Lemma	
Voice 🔶	
Exact Sequence  Exact Cardinality	
Argument Lemma	
Relation 🗘	
Case/Mood 🗘	
Prep.	
Conj. 🗘	
Position 🔶	
Argument Lemma	
Relation 🗘	
Case/Mood 🗘	
Prep.	
Conj. 🗘	
Position 🔶	
Delete This Argument	
Add another Argument	
Submit Reset	

Figure 17. Adding an argument lemma box

#### 7. Some examples of how to effectively use HoDeL

Once learnt about the building of HoDeL, the data it contains, and its basic functionalities, we are now in the position to see some examples of how the lexicon can help researchers to operationalize specific research questions. The main advantage of using HoDeL relies in the fact that it allows users to carry out corpus-based quantitative studies on Homeric Greek without the hassle of learning the complex formalisms necessary to directly query the treebanks of AG (currently, the two main treebanks of Ancient Greek, AGLDT 2.0 and PROIEL, can be directly interrogated using e.g. PML-Tree Query and INESS Search, respectively).

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 14) can be employed to obtain the frequency distribution of sentences attesting to the VSO, SVO, and SOV orders in Homeric Greek. This data could contribute to enlightening a number of still open issues on Homeric word order and information structure.

The functionality 'Arguments' (Figure 15) 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. If the outputs of this filter are cross-checked with those of the 'Args Order' filter, researchers could effortlessly get 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 the verb? How do these ordering patterns correlate with verbal agreement in the case of coordinated subjects? What do these ordering patterns tell about verbal government of coordinated objects?

Besides word order, HoDeL can be useful to detect passages containing infrequent patterns in the Homeric language, which manually can only be found with a thorough and time-consuming reading of the poems. For example, by searching for a specific preverbed verbal lemma in the 'Query' box 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 possibility can be interesting to account for the different paces of grammaticalization or lexicalization paths undergone by different AG preverbs: 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 interesting argument structure alternations, such as the transfer verb  $b\dot{a}ll\bar{o}$  'throw, hit' (Figure 18). 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.



Figure 18. Accusative and dative dependents taken by bállō 'throw, hit'

Note that in Il. 7.11-12, the instrumental dative is labelled as Obj (cf. Section 2). This tag may well be imprecise from a theoretical standpoint, but this analysis has the welcome consequence that it makes HoDeL suitable for this and similar kinds of studies. Indeed, HoDeL is richer than

a strict valency lexicon and allows investigating the behavior of events participants whose argumental status is controversial.

As a final example, let us turn onto evidentiality, which is a category still relatively under-investigated for AG. Luraghi (2020: 142 ff.), which focuses on experiential predicates, showed that certain perception verbs tend to instantiate different argument structure constructions based on the source of evidence of the reported information. Once again, the relevant examples of such argument structure alternations can be quite easily extracted exploiting HoDeL functionalities.

## Use HoDeL and let us know about it!

I believe that, once users are fully informed of how HoDeL resource was built, how its basic functionalities work, and what data and errors it exactly contains, HoDeL can be employed to easily operationalize diverse research questions concerning Homeric verbs and Homeric syntax. Its user-friendly interface and incorporated filters and queries allow scholars with low computational skills to perform advanced corpus-based studies on the Homeric language. In addition, HoDeL optionally shows morphological information, transliteration and translation of the Greek passages, which also much facilitate the interpretation of the output results.

In the future, we plan to continue improving the quality of the base data contained in AGLDT 2.0 and to link HoDeL with other resources of AG, such as *The Chicago Homer* and the growing project *Ancient Greek WordNet*.

If you have suggestions, please do not hesitate to contact us!

### Contacts

Chiara Zanchi, <u>chiara.zanchi01@unipv.it</u> Silvia Luraghi, <u>silvia.luraghi@unipv.it</u>

### **References and external links**

- a-AGLDT 2 guidelines =
  - https://github.com/PerseusDL/treebank data/blob/master/AGDT2/guid elines/Greek guidelines.md#prg ann
- AGLDT 2 = <u>https://perseusdl.github.io/treebank\_data/</u>
- Ancient Greek WordNet = <u>https://greekwordnet.chs.harvard.edu</u>
- Anthony, Laurence. 2013. A critical look at software tools in corpus linguistics. *Linguistic research* 30(2): 141-161.
- a-PDT guidelines = <u>https://ufal.mff.cuni.cz/pdt2.0/doc/manuals/en/a-</u> <u>layer/html/ch02.html</u>
- Celano, Giuseppe A. 2019. "The Dependency Treebanks for Ancient Greek and Latin." In *Digital Classical Philology. Ancient Greek and Latin in the Digital Revolution* [Age of Access? Grundfragen der Informationsgesellschaft 10], edited by Monica Berti, 279-298. Berlin: De Gruyter.
- Chicago Homer = <u>https://homer.library.northwestern.edu</u>
- Dixon, Robert M.V. & Alexandra Aikhenvald. 2000. *Changing valency. Case studies in transitivity*. Cambridge: CUP.
- IT-VaLex = <u>https://itreebank.marginalia.it/itvalex</u>
- Keydana, Götz & Silvia Luraghi. 2012. Definite referential null objects in Vedic Sanskrit and Ancient Greek. Acta Linguistica Hafniensia 44/2: 116-128.
- Kulikov, Leonid, Andrej Malchukov, & Peter de Swart P. 2006. *Case, Valency and Transitivity*. Amsterdam: Benjamins.

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.

- McGillivray, Barbara & Marco Passarotti. 2009. The Development of the Index Thomisticus Treebank Valency Lexicon. In *Proceedings of LaTeCHSHELT& R Workshop 2009*, 43-50.
- Passarotti, Marco. 2011. Language Resources. The State of the Art of Latin and the Index Thomisticus Treebank Project. In *Corpus anciens et Bases de données, «ALIENTO. Échanges sapientiels en Méditerranée», N°2,* Marie-Sol Ortola (ed), 301-320. Nancy, France: Presses universitaires de Nancy.
- Passarotti, Marco, Berta González Saavedra & Christophe Onambele. 2016. Latin Vallex. A Treebank-based Semantic Valency Lexicon for Latin. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16, Portorož, 23-28 May 2016), 2599-2606. [http://hdl.handle.net/10807/78696]
- Panevová, Jarmila. 1994. Valency Frames and the Meaning of the Sentence.
   In *The Prague School of Structural and Functional Linguistics*. A short introduction Luelsdorff Philip A. (ed), 223-243. Amsterdam: Benjamins.

Perseus Project = <u>https://www.perseus.tufts.edu/hopper/</u>

PROIEL = <u>https://www.hf.uio.no/ifikk/english/research/projects/proiel/</u>

- Sgall Petr. 1969. *A functional approach to syntax in generative description of language*. New York: American Elsevier.
- Shibatani, Masayoshi. 1985. Passive and related constructions: a prototype analysis. *Language* 61/4: 821-848.
- Shibatani, Masayoshi (ed.). 1988. *Passive and voice*. Amsterdam: Benjamins.

TASF = <u>https://sites.google.com/universitadipavia.it/tasf/home</u>

- t-PDT guidelines = <u>https://ufal.mff.cuni.cz/pdt2.0/doc/manuals/en/t-</u> <u>layer/html/index.html</u>
- Smyth, Herbert W. 1920. *Greek Grammar for Colleges*. New York: American Book Company.

- Zanchi, Chiara, Eleonora Sausa & Silvia Luraghi. 2018. HoDeL, a Dependency Lexicon for Homeric Greek: Issues and Perspectives. In *Formal Representation and the Digital Humanities*, Paola Cotticelli & Federico Giusfredi (eds), 221-246. Cambridge: Cambridge Scholars Publishing.
- 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. Paper accepted for the proceedings of Dialogue 2020 – 26<sup>th</sup> International Conference on Computational Linguistics and Intellectual Technologies (on-line), Moscow, Russia, 15-20 June 2020.
- Zanchi, Chiara. Submitted. The *Homeric Dependency Lexicon*: what it is and how to use it. Submitted to the *Journal of Greek Linguistics*.