Quantitative approaches to Ancient Near Eastern Languages: The case study of Hittite clitics[1]

Ekaterina Lyutikova – Lomonosov Moscow State University, Interdisciplinary Scientific and Educational School of Moscow University "Brain, Cognitive Systems, Artificial Intelligence"
Andrei Sideltsev – Institute of Linguistics, RAS

[The paper deals with methodological issues of the study of Ancient Near Eastern languages in the perspective of modern theoretical linguistics. So far the data of the languages – particularly the major ones like Sumerian, Akkadian and Hittite – despite their being among the best documented languages of the world are very conspicuously absent from theoretical linguistics. The reason is uncertainty whether the observed characteristics of data like correlations between specific parameters or lack of a form or a meaning are significant or accidental. In this paper, on the basis of a fragment of Hittite grammar – argument clitics and their cooccurrence in the clitic chain – we suggest that applying quantitative approaches to the languages of the Ancient Near East can solve the problem and bring their data into theoretical linguistics.]

1. *Introduction*

The languages of the Ancient Near East are among the best documented and best studied languages of the world. However, the use of their data in theoretical linguistics is painfully limited and sharply disproportionate to their use in historically and philologically oriented studies.

The main problem effectively blocking their relevance in theoretical linguistics is the inability of a researcher to determine whether the absence of some linguistic object (form, meaning, combination of parameters) is significant or incidental in the available data. The same holds for the tentative correlations of linguistic parameters observed in the surviving texts.

Quantitative approaches to linguistic analysis in the meanwhile have managed to cope with the problem. Employing statistical techniques, they provide objective estimation of whether a generalization drawn from the available data is incidental or significant with sufficient degree of confidence and reliability.

Statistical analysis of linguistic form for a dead language can only be based on corpus data.

Creation and use of corpus data has become fairly mainstream for live modern languages and is quite widespread for the major languages of the Ancient Near East – Sumerian and Akkadian, see, e.g. http://bdts.filol.csic.es for administrative Neo-Sumerian texts (end of 3rd millennium) and http://etcsl.orinst.ox.ac.uk for Sumerian literary texts; http://www.archibab.fr for non-literary Old Babylonian texts (XX-XVII cc. BC) (Charpin 2014a), https://seal.huji.ac.il for Akkadian literary texts from the 3rd and 2nd millennia BC, http://oracc.museum.upenn.edu/saao/corpus for the Neo-Assyrian period and www.achemenet.com for the texts of the Achaemenid period of VII-IV cc. BC (including the Akkadian ones). Various corpora are accumulated at http://oracc.museum.upenn.edu. See also (Streck 2010) for the relative volume of various Ancient Near Eastern textcorpora and Charpin (2014b) on Assyriological internet resources. Corpus-based studies also regularly appear, their topics range from orthography to syntax and semantics, e.g., (Alexandrov 2019; Arkhipov, Loesov 2019; 2020; Cohen 2005; 2012; Hernáiz 2019; Jagersma 2010; Kalinin, Loesov 2017; Kouwenberg 2017; Loesov 2004; 2005; 2006; de Ridder 2018; Woods 2008). However, more advanced digital resources like databases and consequent statistical evaluation of the data apart from simple counts of attested forms and meanings are conspicuously lacking.

This is all the more conspicuous in view of the fact that other dead ancient languages are very actively studied within corpus and big data linguistics. These are primarily Latin and Ancient Greek as well as medieval Germanic, Romance or Slavic languages. A very good illustration is the recent book of van Gelderen (2018) based on Old, Middle and Early Modern English data specifically devoted to the quantitative study of dead languages; let us mention also numerous corpus-based papers such as Walkden & Morrison (2017) on Jespersen's Cycle in Early Modern English or Walkden & Rusten (2016) on null subjects in Middle English, as well as many others.

This paper attempts to demonstrate that the advanced statistical analysis of quantitative data obtained from representative databases for the languages of the Ancient Near East and not reducible to simple form counts is not just a fashionable trick, but provides an indispensable tool for the study of these languages. They help obtain the generalizations that cannot be received with the use of more traditional methodology.

We will use Hittite as the show case, namely one fragment of its grammar – the system of enclitic personal pronouns and the cooccurrence of different enclitic pronouns within the enclitic chain. We will demonstrate that the application of modern digital methods to data collection and processing enables us to establish the substantive basis for developing and supporting theoretical hypotheses.

Hittite is a particularly good illustration of the advantages of these modern tools of research for several reasons. Out of the ancient languages of the Near East it has a much more compact corpus. Still, it does not possess a sizeable on-line corpus with search tools. There are several on-going projects aiming at creating such corpora (www.hittitecorpus.ru or www.hittitetexts.com; the indispensable tool for Hittitological research, www.hethport.uni-wuerzburg.de, does host a much more sizeable body of Hittite texts which can be searched, but the search engine is different from that standard in corpus linguistics and finds only particular word forms written in a particular way which impedes its use as a proper corpus), but their present scope does not allow to use them for proper statistically significant conclusions. In this light creation of databases which is a much more feasible and less time consuming task is a very realistic way out of the lack-of-corpus situation.

It must be explicitly stated at this point that, just like studies of the other languages of the Ancient Near East, studies on Hittite *are* frequently based on the data from a selected corpus of texts, either limited to a particular chronological period or covering all the three periods of the development of the Hittite language. See, just for example, Goedegebuure's (2014) masterful study

of Hittite demonstrative pronouns or Inglese, Rizzo and Pflugmacher (forthcoming) on a corpus study of definite referential null objects in Old Hittite. The texts which are sourced for forms/usages are exhaustively listed, the forms/usages within the selected body of texts are counted in relation to each other. Works of this type represent the best of Hittite scholarship and establish the descriptive data beyond any reasonable doubt. However, such works do not even set the task of establishing whether the selected body of texts is statistically sufficient to extrapolate from them to the entire collection of texts or whether the forms occurring within the selected texts are sufficient to draw reliable conclusions which would stand the test of theoretical linguistics. In other words, the question whether absence of a certain form or meaning or a combination of forms and meanings is significant or simply incidental is not even posited. True exceptions are available, the only one that comes to mind is Goedegebuure (2018), but in the interest of the discipline where the necessity for statistical research is simply not acknowledged it is worth laying out the methodology of modern quantitative linguistics.

Thus, for Hittite in view of both its fairly limited general number of texts, on the one hand, and lack of on-line digital corpora with search tools, on the other hand, the eternal problem of data integrity in a dead language is particularly acute and impedes linguistic research that would be acceptable in theoretical linguistics.

In this methodologically oriented study whose aim is not limited to Hittite proper we will show how a representative database and advanced statistical evaluation of its resources furnish the all-important answer even in the absence of a full searchable corpus. It is important that we claim no new methodology from the point of view of general statistics or corpus linguistics. The methods we are making use of are standard and universally accepted. What we do attempt is to introduce the understanding of the necessity of such methods into the studies of Ancient Near Eastern languages which is currently totally lacking.

The paper is organized as follows. In section 1 we lay out an overview of the traditional Hittitological treatment of argument clitics. Section 2 presents our database: it describes the corpus which provides the database entries and informs the reader on the parameters of the database. Section 3 contains descriptive statistical data. In section 4, we proceed to the inferential statistics, which allows us to infer the relationships from the descriptive measures. Section 5 concludes.

2. *Argument clitics in Hittite: an overview*

Like the majority of Hittite clitics (save focal -*pat* and occasionally locative clitics), argument clitics in a clause must appear in a clitic chain and have fixed positions within the chain. In traditional Hittitology the positions are described as slots, and the relevant part of the clitic chain (other slots are available for the quotative particle (slot 1), reflexive particle (slot 5) and locative adverbs (slot 6)) is represented as follows in standard grammars (see most recently Hoffner and Melchert 2008: 410).[2]

---

2. The fragment of the Hittite grammar is also studied in our recent paper (Lyutikova, Sideltsev 2020), but from a very different perspective than the present paper. The aim of (Lyutikova, Sideltsev 2020) is to put the Hittite data into the cross-linguistic perspective of constraints on the cooccurrence of several personal enclitic pronouns in the enclitic chain and to show that Hittite sheds new light on the theoretical aspects of this cooccurrence in the natural human languages.

Table 1. Argument clitic cluster in Hittite

Slot 2	Slot 3	Slot 4
1PL.DAT/ACC -*naš*, 2PL.DAT/ACC -*šmaš*, 3PL.DAT –*šmaš*	3SG.NOM.C -*aš* 3SG.ACC.C -*an*, 3SG/PL.ACC.N/C -*at*, 3PL.ACC.C -*uš*, 3PL.ACC.N -*e* 3SG/PL.NOM.N/C -*at*, 3PL.NOM.C/N –*e*	1SG.DAT/ACC -*mu*, 2SG.DAT/ACC -*tta*, -*du*, 3SG.DAT –*šši*

Table 1 shows the structure of the argument clitic cluster in Hittite. Slot 2 is occupied by dative-accusative 1-2 person plural clitics, slot 3 is occupied by nominative and accusative singular / plural 3rd person clitics, while slot 4 is occupied by dative-accusative 1-2 person singular clitics and the dative 3rd person singular clitic pronoun. Only one clitic can occupy a certain slot. Besides, the pronouns of slots 2 and 4 are mutually exclusive and never co-occur (Ibid.: 411). Thus, there can be only one accusative/nominative and only one dative clitic simultaneously and a clause can contain at most two argument clitics, one dative and one nominative/accusative.

We will illustrate this distribution with two possible argument clitic combinations:

(1) =accusative 3rd person singular=dative 1st person singular:
   n=at=mu=kan   *mān šanna-tti*
   CONN=**it.ACC.SG.N=me.DAT.SG**=LOCP   if   conceal-2SG.PRS
   'And if you conceal **it from me** …' (eNH (CTH 42.A) KBo 5.3+ obv. i 28).

(2)   =dative 2nd person plural=accusative 3rd person plural:
   nu=šmaš=at   *peḫḫi*
   CONN=**you.DAT.PL=them.ACC.PL**   give-1SG.PRS
   'And I shall give **it to you (pl.)**.' (eNH (CTH 42.A) KBo 5.3+ obv. ii 8).

The traditional Hittitological description of argument clitics, although observationally adequate and easy-to-use in assessing particular clauses containing enclitic pronouns, leaves open a number of questions relating to the argument clitics system as a module of grammar. Specifically, the above description does not address the following questions:

— whether the argument clitics system is diachronically stable or undergoes changes throughout the written history of Hittite;
— which deviations from the general pattern can be found, and how frequent they are;
— how the argument clitics interact with other components of the clitic chain;
— how the argument clitics interact with other parameters of the clause (i.e. transitivity);
— whether argument clitics demonstrate restrictions on cooccurrence;
— which grammatical features expressed by argument clitics (gender, number, person, case) are subjects to such restrictions.

In what follows we provide answers to these questions and, importantly, show that they are crucial for the explanatory grammar of Hittite argument clitics. In the next section we present our research tool — the database of Hittite argument clitics.

3. *The corpus and the database*

The database of Hittite argument clitics contains 1362 clauses originating from the corpus of 33 texts. The total number of clauses in the corpus is 4643. Every clause of the corpus containing at least one argument clitic was included into the database.

3.1. *The corpus*

Our corpus was based exclusively on New Hittite originals, i.e., on the texts composed and written down in the New Hittite period of the development of the Hittite language. For the purity of the experiment no exhaustive coverage of the New Hittite texts was sought. However, texts of different genres and different periods inside New Hittite were selected to make the corpus representative. Texts were sourced from different major genres – prayers, treaties, letters, instructions/oaths as well as the apology of Hattusili III; all the subperiods within New Hittite were covered – early New Hittite (texts from the reign of Suppululiuma I), classic New Hittite (texts from the reigns of Mursili II, Muwatalli II, Hattusili III), late New Hittite (texts from the reigns of Tudhaliya IV, Suppululiuma II).

The following texts were included into the selection:
Early New Hittite (texts composed during the reign of Suppiluliuma I) 324 clauses:

CTH 42 (treaty of Suppiluliuma I with Huqqana of Hayasa, ex. A KBo 5.3+ restored and supplemented from ex. B KBo 19.44+) 304 clauses,
CTH 253.1 (instructions of Suppiluliuma I for the Military and a corresponding Oath, ex. A KUB 21.41 restored and supplemented from ex. B KUB 26.57) 20 clauses;
New Hittite (texts composed during the reigns of Mursili II, Muwatalli II, Hattusili III) 2774 clauses:
CTH 62 (Mursili II, treaty of Mursili II with Tuppi-Tessup of Amurru, KBo 5.9+, restored and supplemented from duplicates) 209 clauses,
CTH 63 (Mursili II, Mursili II's dictate to Tuppi-Tessup's Syrian antagonists, KBo 3.3+, restored and supplemented from duplicates) 92 clauses,
CTH 69 (Mursili II, treaty of Mursili II with Manapa-Tarhunta von Seha, KUB 19.49+, restored and supplemented from duplicates) 212 clauses,
CTH 81 (Hattusili III, Apology, KUB 1.1+, restored and supplemented from duplicates) 362 clauses,
CTH 176 (Hattusili III, a letter from Puduhepa to Ramses II (?), KUB 21.38) 182 clauses,
CTH 191 (Muwatalli II, a letter to Muwatalli II from Manapa-Tarhunta of Seha, KUB 19.5+) 42 clauses,
CTH 181 (Hattusili III, a letter from Hattusili III to the king of Ahhiyawa, KUB 14.3) 313 clauses,
CTH 254 (Hattusili III, Oath of the Men of Hattusa to Hattusili III and Puduhepa, KUB 21.46) 5 clauses,
CTH 377 (Mursili II, prayer to Telipinu, ex. A KUB 24.1+ restored and supplemented from exx. B and C) 63 clauses,
CTH 378.1 (Mursili II, first plague prayer, ex. A KUB 14.14+ restored and supplemented from duplicates) 148 clauses,

CTH 378.2 (Mursili II, second plague prayer, ex. A KUB 14.8 restored and supplemented from duplicates) 184 clauses,

CTH 378.3 (Mursili II, second plague prayer, ex. A KUB 14.12) 40 clauses,

CTH 378.4 (Mursili II, fourth plague prayer, ex. A KUB 14.8 restored and supplemented from duplicates) 123 clauses,

CTH 378.5 (Mursili II, a plague prayer, ABoT 2.23) 4 clauses,

CTH 378.6 (Mursili II, a plague prayer, KBo 51.20) 11 clauses,

CTH 378.7 (Mursili II, a plague prayer, KBo 54.7) 12 clauses,

CTH 378.8 (Mursili II, a plague prayer, KBo 54.8) 12 clauses,

CTH 379 (Mursili II, fifth plague prayer to the assembly of gods, KUB 31.121+) 57 clauses,

CTH 381 (Muwatalli II, prayer to the assembly of gods, KUB 6.45+, restored and supplemented from duplicates) 223 clauses,

CTH 382 (Muwatalli II, prayer to the Storm-god of Kummanni, KBo 11.1) 153 clauses,

CTH 383.1 (Hattusili III, prayer to the Sun-goddess of Arinna) 164 clauses,

CTH 384.1 (Hattusili III, prayer of Puduhepa to the Sun-goddess of Arinna) 163 clauses.

Late New Hittite (texts composed during the reigns of Tudhaliya IV, Suppiluliuma II) 1545 clauses:

CTH 105 (treaty of Tudhaliya IV with Sausgamuwa of Amurru) 181 clauses,

CTH 106.A (Tudhaliya IV, treaty between Tudhaliya IV and Kurunta of Tarhuntassa, Bo 86/299) 308 clauses,

CTH 123 (Suppiluliuma II, treaty of Suppiluliuma II (?) with an unknown partner, KBo 4.14) 321 clauses,

CTH 141 (Suppiluliuma II, treaty of Suppiluliuma II with Alasiya, KBo 12.39) 49 clauses,

CTH 178.1 (Tudhaliya IV, Tudhaliya IV's letter to Babu-aha-iddina, ex. A KUB 23.103 restored and supplemented from duplicates) 139 clauses.

CTH 182 (Tudhaliya IV, a letter from Tudhaliya IV to Tarkasnawa of Mira, KUB 19.55+) 94 clauses,

CTH 255.1 (Tudhaliya IV, Tudhaliya IV's Instructions and Loyalty Oath Imposition for Lords, Princes and Courtiers, ex. A KUB 21.42+, restored and supplemented from duplicates) 179 clauses,

CTH 255.2 (Tudhaliya IV, Tudhaliya IV's Instructions and Oath Imposition for Courtiers, ex. A KUB 26.1+ restored and supplemented from duplicates) 219 clauses,

CTH 256 (Suppiluliuma II, Suppiluliuma II's Instructions and Oath Imposition for the Men of Hattusa, ABoT 1.56) 55 clauses.

3.2. *The database*

The relational database is created in MS Access. The entry is a clause containing one or more argument clitics. The 29 fields accumulate the following information for each clause:
— Clause
— Translation
— Source
— Period
— Number of argument clitics

— Clitics in slots 2, 3 and 4 and their form
— Transitivity of the clause
— Clitic doubling
— Clitic reduplication
— Reflexive clitic (slot 5) and its argumental status
— Locative clitic (slot 6) and its form
— Nominative-accusative clitic: its presence, form, grammatical features (person, number) and its anaphoric antecedent
— Dative clitic: its presence, form, grammatical features (person, number) and its anaphoric antecedent; additionally, dative clitics governed by postpositions are tagged.

The database design allows to easily access the complex combinatorial information about argument clitics distribution: e.g., filter out those examples where the 2p SG DAT clitic occurs in a passive clause, or find all the Early New Hittite clauses with clitic reduplication, etc. In the next two sections we present descriptive and inferential generalizations drawn from the database.

4. *Descriptive statistics*

In this section we present the frequency characteristics of various parameters characterizing the records in our database.

The database contains 1362 records. Each record corresponds to a separate clause that contains one or two argument clitics. There are 1260 clauses containing one argument clitic and 102 clauses containing two argument clitics. In clauses with two argument clitics, one clitic is dative and the other is nominative or accusative. No clauses with two dative clitics or two nominative/accusative clitics have been found. The frequency distribution tables present the quantitative information for clauses with one argument clitic (Table 2) and two argument clitics (Table 3).

Table 2. Grammatical features of clitics in clauses with one argument clitic

1SG ACC	86	143	159	764	1260
2SG ACC	57				
1PL ACC	9	16			
2PL ACC	7				
3SG ACC	311	477	605		
3SG NOM	166				
3PL ACC	98	126			
3PL NOM	28				
3SG/PL NOM/ACC[3]	2	2			
1SG DAT	169	257	321	496	
2SG DAT	88				
1PL DAT	24	64			
2PL DAT	40				
3SG DAT	148		175		
3PL DAT	27				

3. Two clauses contain an accusative/nominative argument clitic which is ambiguous with respect to its number and case features values.

Table 3. Grammatical features of clitics in clauses with two argument clitics

1SG ACC	0	0	0	102
2SG ACC	0			
1PL ACC	0	0		
2PL ACC	0			
3SG ACC	44	81	102	
3SG NOM	37			
3PL ACC	12	21		
3PL NOM	9			
1SG DAT	57	64	70	102
2SG DAT	7			
1PL DAT	1	6		
2PL DAT	5			
3SG DAT	26	32	32	
3PL DAT	6			

The total number of argument clitics in the database is 1464. Their frequency distribution is given in Table 4.

Table 4. Grammatical features of clitics in all types of clauses

1SG ACC	86	143	159	866	1464
2SG ACC	57				
1PL ACC	9	16			
2PL ACC	7				
3SG ACC	355	558	707		
3SG NOM	203				
3PL ACC	110	147			
3PL NOM	37				
3SG/PL NOM/ACC	2	2			
1SG DAT	226	321	391	598	
2SG DAT	95				
1PL DAT	25	70			
2PL DAT	45				
3SG DAT	174	174	207		
3PL DAT	33	33			

The first look at the data suggests several possible correlations between various characteristics of argument clitics. For accusative clitics,[4] 3p clitics occur much more often than 1p and 2p clitics, in both singular and plural (see the histogram in Figure 1). On the contrary, for dative clitics, the number of 1p and 2p clitics compares to the number of 3p clitics (Figure 2).

---

4. Nominative clitics are excluded from the comparison because of the absence of 1p and 2p nominative clitics in Hittite.
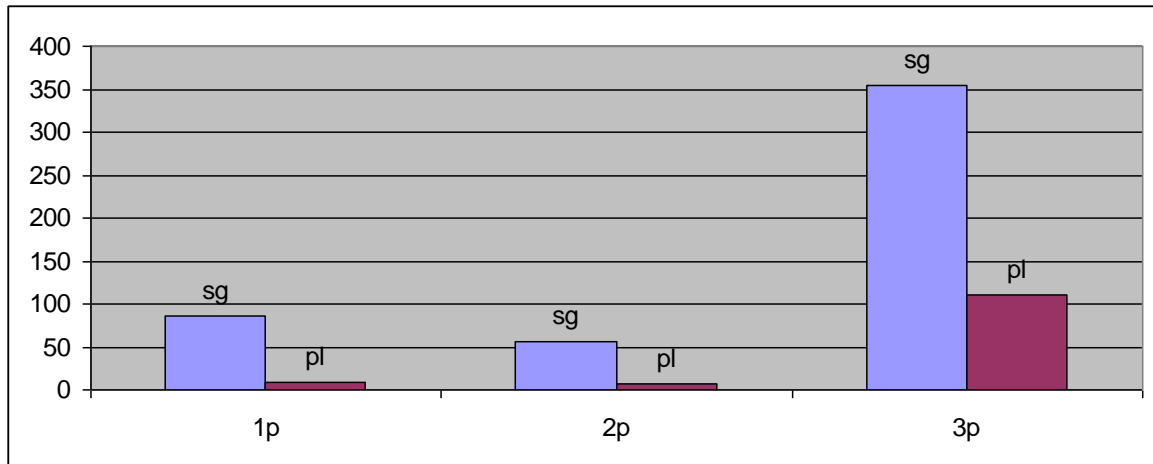
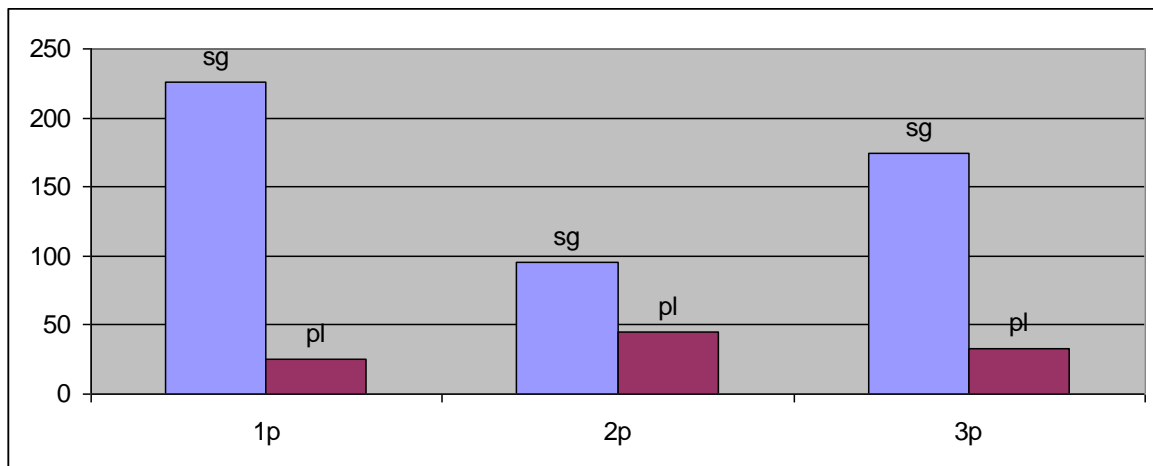Figure 1. Frequency distribution of accusative clitics: person and number

Figure 2. Frequency distribution of dative clitics: person and number

This contrast is even more evident if we compare frequencies of accusative and dative clitics within the three person categories. The data are represented in Figure 3. We see that 1p and 2p clitics are more likely to appear as dative arguments, and 3p clitics — as accusative arguments. Given that dative arguments are mostly indirect objects, and accusative arguments are direct objects, the generalization is that 1-2p participants tend to be indirect objects, whereas 3p participants — direct objects.
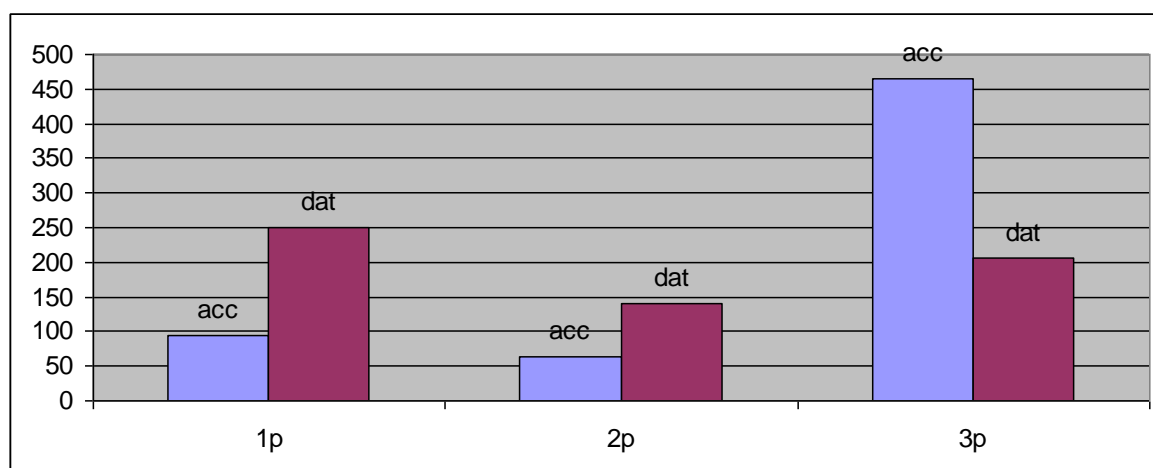
Figure 3. Frequency distribution of clitics: person and case

The observed correlation of person and grammatical relation is widely attested cross-linguistically (see Haspelmath 2004 for an overview). Functionally oriented studies explain this correlation assuming harmonic alignment between the two feature hierarchies — animacy hierarchy and thematic role hierarchy (Croft 1990):

(3) Animacy hierarchy
1-2p > 3p pronouns > proper nouns > human > animate > inanimate

(4) Thematic role hierarchy
Agent > Recipient > Theme

The higher is the participant on the animacy scale, the more is it likely to be high on the thematic role hierarchy, and vice versa, the more prominent is the thematic role, the more likely it is to be realized by a highly animate participant. Taking into account the standard thematic role — grammatical relation correspondence where Recipients are indirect objects and Themes are direct objects, we expect to find exactly the pattern we observe in our data.

Another general tendency supported by the Hittite data is the prevalence of the singular over the plural. This result is in accordance with the markedness hierarchy for number ((3), Croft (1990)) which predicts that the less marked value of a feature should be at least as frequent as the more marked one in both cross-linguistic and intralinguistic aspects.

(5) Markedness hierarchy: number
SG < PL < DU < PAUCAL

It should be noted, however, that only 3p pronouns can be claimed to realize the nominal number category; the "plural" forms of 1-2p pronouns are semantically different in that they lack referential homogeneity ('we' cannot be represented as 'many speakers' but rather as 'the speaker and others'). This is the reason for treating the pronominal number as a separate category or even replace it with a system of privative features (Corbett 2000, Harley, Ritter 2002, Daniel 2005).

For 3p argument clitics, plural forms are much rarer than singular ones, both in total and within each case group, see Table 5 and Figure 4.

Table 5. Number forms of 3p argument clitics

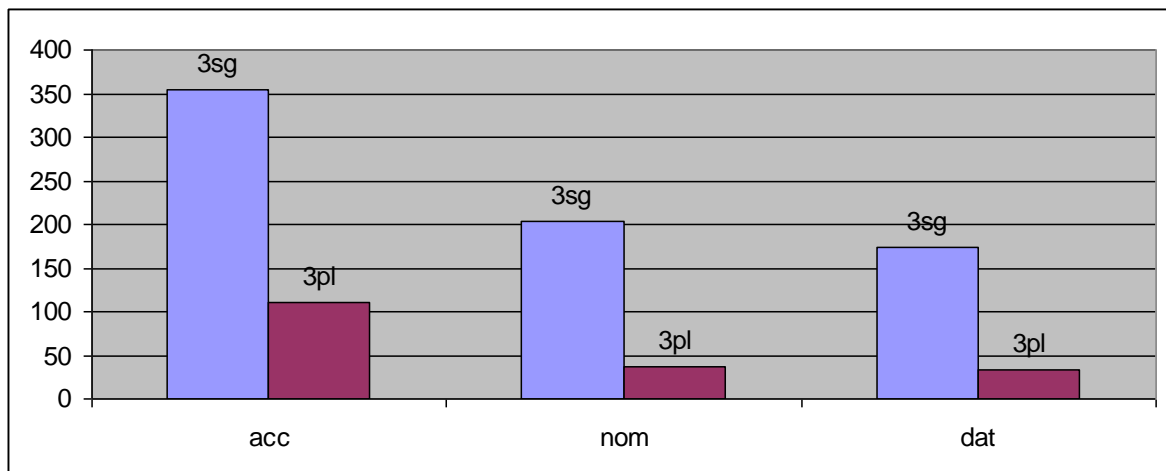|	3SG	3PL
ACC	355	110 (30%)
DAT	174	33 (19%)
NOM	203	37 (18%)
Total	732	180 (24%)

Figure 4. Frequency distribution of 3p argument clitics: number and case

There seems to be a weak correlation between grammatical relation and proportion of plural forms: in direct objects, plural forms are more represented than in subjects and indirect objects. Again, the explanation might come from the general principle of harmonic alignment. Plural objects are claimed to be lower than singular ones on several referential scales, such as Object individuation (Hopper, Thompson 1980) or Definiteness / Specificity (Bickel, Nichols 2007). Accordingly, they are more expected to occur with lower thematic roles (Theme) than with the higher ones (Recipient).

A significant number of clauses contain not only argument clitics, but also a reflexive clitic (slot 5) and/or a locative clitic (slot 6).

Reflexive clitic occurs in 91 clauses. Reflexive clitic has three major types of uses in Hittite (Boley 1993; Hoffner, Melchert 2008: 357-363; Inglese 2017): argumental, where it roughly corresponds to an accusative or dative reflexive pronoun, cf. example (6a), detransitivizing, where it marks valency decreasing of the verb (6b), as well as reciprocal (Inglese 2017). Besides, reflexive clitic appears in specific syntactic constructions, e.g., in copular clauses with nominal predicates (6c).

(6) a. DINGIR-*LUM=mu=za=kan*   GAŠAN=*YA*   *ḫūmandaza=pat*   *daškiši*
        god=me=REFL=LOCP   lady=my   all.ABL=FOC   take.IPF.2SG.PRS
   'you goddess, My Lady, choose me (lit. take me for yourself (-*za da-*)) among all (the others; lit. from all)' NH (CTH 81.A) KUB 1.1+ obv. i 50).

b.　　[*U*]L=at=za　　*kikkištari*
　　　　NEG=it= REFL happen.IPF.3SG.PRS.MED

'Will it not happen?' (lNH (CTH 123) KBo 4.14 obv. ii 4) following F. Fuscagni (ed.), hethiter.net/: CTH 123 (INTR 2014-02-17). Cf. Stefanini (1965: 39).

c.　　*kinun=ma=wa=tu=za*　　　　*UL*　ÌR^MEŠ
　　　　now=but=QUOT=you.DAT=REFL　NEG subjects

'Now (we are) no (longer) your subjects' (lNH (CTH 105.A) KUB 23.1+ obv. ii 32).

Out of 91 clauses where reflexive clitic -*za* occurs, only 15 represent argumental -*za*. It corresponds to the dative Recipient ('take for oneself'); outside of our corpus, however, -*za* might also represent the reflexive direct object (example (7)).

(7)　*kinun=a=šmaš=za*　　　　LUGAL-*uš*　*labarnaš*　　　*ula-nun*
　　　now=but=you.DAT.PL=REFL　king.NOM.SG labarna.NOM.SG　unite-1SG.PST

'Now I, the labarna, the king, have united myself to you.' (OH/NS (CTH 414.1.A) KUB 29.1 obv. i 33-34 following Hoffner, Melchert (2008: 358); S. Görke (ed.), hethiter.net/: CTH 414.1 (TX 11.06.2015, TRde 13.03.2015)).

Locative clitics cooccur with argument clitics in 278 clauses (20%); they are mainly represented by -*kan* (268 clauses); -*šan* occurs in 10 clauses, -*ašta* is not attested in the database. Example (8) shows the clause with the locative clitic -*šan* and the dative argument clitic -*mu* 'I.DAT'.

(8)　*nu=mu=ššan*　　*mān* ÉRIN^MEŠ-*it*　ANŠE.KUR.R[A^ḪI]·A-*it*　*lammar*　　*UL*
　　　CONN=me=LOCP　if　troops-INSTR　horses-INSTR　　　　immediately　NEG
　　　ārti
　　　come.2SG.PRS

'If you do not come to me immediately with infantry and chariotry' (eNH (CTH 42.A) KBo 5.3+ obv. ii 28) cf. G. Wilhelm (ed.), hethiter.net/: CTH 42 (INTR 2013-02-24).

Finally, let us discuss other phenomena related to argument clitics which are represented in our database. Those are clitic doubling and clitic reduplication.

Clitic doubling is a rare phenomenon in Hittite. Normally, arguments are expressed either by argument clitics or by stressed pronouns or nominals; additionally, anaphoric *pro* can appear in subject — and, less frequently, object —positions. That is, argument clitics are not used for indexing verbal arguments, and clitic doubling is not generally found (Sideltsev 2011, 2016). However, Hittite attests a nominal construction similar to English "we linguists" construction (Postal 1966, Déchaine, Wiltschko 2002) where personal pronouns function as determiners and precede a noun, cf. (9):

(9)　*nu*　　　*mān zik*　　　^m ^dLAMMA-*aš*　　*kī*　　　　*tuppiyaš*
　　　CONN　　if　you.SG.NOM　Kurunta.NOM.SG　this.ACC.PL.N tablet.GEN.SG
　　　uttār　　　　*UL* *paḫḫašti*
　　　word.ACC.PL.N　　NEG observe.2.SG.PRS

'And if you, Kurunta, do not observe these words of the tablet, ...' (lNH (CTH 106.A) Bo 86/299 rev. iv 5).

It turns out that the pronominal component of this construction can be represented by clitic pronouns in Hittite; in this case, the noun phrase splits when the clitic pronoun raises to join the clitic chain and leaves the nominal part *in situ*. This might superficially look like clitic doubling, cf. (10):

(10) a.   EN=*KUNU*=*ma*=*aš*=*kan*   *kuit* BA.ÚŠ
           lord=your=but=he= LOCP   as   died
'Because your (pl.) lord died' (lNH (CTH 178.1.A) KUB 23.103 rev. 9 following segmentation in Mora, Giorgieri (2004: 168), cf. Hoffner (2009: 325).

b.   *paḫši*=*ya*=*an*   ᵈUTU-*Š*=*I*
     protect.2SG.IMP=and=him   majesty=my
'And protect him, My Majesty' (eNH (CTH 42.A) KBo 5.3+ obv. i 16) cf. G.Wilhelm (ed.), hethiter.net/: CTH 42 (INTR 2013-02-24).

The overall number of such cases in our database is 14 (out of 1362 clauses, i.e. approximately 1%); various clitic pronouns are attested (3p SG/PL ACC — 6 clauses, 3p SG/PL NOM — 6 clauses, 3p SG DAT — 1 clause; 2p SG ACC — 1 clause).

The term "clitic reduplication" refers to the internal reorganization of the clitic complex when a second copy of the accusative or nominative clitic is attached to the right of the clitic array, after the slot 4 or 5. Examples are given in (11): (11a) shows the copy of the nominative clitic after the slot 4 (dative singular clitic) and (11b) shows the copy of the accusative clitic after the slot 5 (reflexive clitic).

(11) a.   *n*=*at*=*ši*=*at*ˈ   *NIŠ* DINGIR-*LÌ*   GAM   GAR-*r*[*u*]
           CONN=it=him=it   oath deity   down   lie.3SG.IMP
'It shall be placed under oath for him' (lNH (CTH 255.1.A KUB 21.42+ obv. i 26')) following Miller (2013: 284-5). The actual sign which is used to write the second -*at* is AB.

b.   *n*=*an*=*za*=*an*   ᵐNIR.GÁL-*iš*   *taraḫta*
     CONN=him= REFL=him   Muwatalli. NOM.SG.C   conquer.3SG.PST
'Muwatalli defeated him' (lNH (CTH 105.A) KUB 23.1+ obv. i 36-7).

Our database contains 23 clauses with clitic reduplication. In 6 clauses, the copy attaches to the 3p SG DAT clitic (4 times -*at*-*ši*-*at* and twice -*aš*-*ši*-*aš*), in 16 clauses, the copy attaches to the reflexive clitic -*za* (all of them of the form -*an*-*za*-*an*). One clause exhibits a different form of the copy (-*an*=*za*=*aš* instead of -*an*=*za*=*an*), cf. (12).

(12) *n*=*an*=*za*=*aš*   *šaḫešneški*
      CONN=him=REFL=them   fortify.IPF.2SG.IMP
'Fortify it continuously' (lNH (CTH 105.A) KUB 23.1+ l.e. 1) following Kühne, Otten (1971: 17), F. Fuscagni (ed.), hethiter.net/: CTH 105 (TX 07.05.2013, TRde 07.05.2013).

It can be understood as a purely scribal mistake *n*=*an*=*za*=*an*ˈ, as suggested by CHD (Š: 10).
To sum up, we have described major quantitative parameters of our database. In what follows we use the data presented in this section for inferential statistics, essentially for correlational

studies. This is where the core of our proposal becomes clear: we do not only catalogue the data and identify putative generalizations, proportions and correlations, but also employ statistics to check whether they are reliable and not just a matter of chance.

5. *Inferential statistics*

Inferential statistics applies statistical techniques to test hypotheses using experimental or observational data. Obviously, the only source of data for a dead language is the existing corpus of texts, so only observational studies are feasible. The basic idea of statistical analysis in this case is to regard the corpus (or its subcorpus) as a randomly selected sample representing the entire population, i.e., all possible linguistic expressions of the language studied, and to extrapolate from the sample to the entire population. Statistical analysis allows us to estimate the parameters of the population with the required degree of confidence basing on sample statistics.

In what follows we deal with two types of challenges posed by our data, and, accordingly, we solve statistical problems of two kinds. The first type of problems is the extrapolation of measures obtained on the sample (in our case — the database) to the characterization of the Hittite language in general. Having calculated, say, the proportion of clauses containing argument clitics, we cannot claim that this is the true proportion for the whole collection of Hittite texts (including those irretrievably lost or still not discovered); indeed, we cannot say what the proportion for this collection is. What we can do, however, is to calculate a range of values within which we can have a certain degree of confidence that the true proportion lies. The span of this range is calculated based on the value observed on the sample (in our case — proportion of clauses with argument clitics in our corpus) and the size of the sample (in our case — the corpus our database is built on).

The second type of problems is the statistical hypothesis testing. The hypotheses tested in this paper are about the significance of the observed differences in the two (or more) sets of data. Let us suppose that we have measured the proportion of clauses with clitics in treaties and prayers, and it equals 31% and 29%, respectively. The question is whether this difference can reasonably be expected to occur 'by chance', or we should infer that treaties differ from prayers in employing more argument clitics in general. To answer this question we take the following path. We formulate the null hypothesis that the two sets of data do not differ with respect to the parameter studied (in our case — that treaties and prayers do not differ in employing clitics). The alternative hypothesis is that there exist differences between the two data sets. Our strategy will be to try to accumulate enough evidence to reject the null hypothesis. We will do it by using specific statistical tests (in this paper, only chi-square test is applied). Statistical tests work as follows: they provide the instruction for calculating a specific variable — test statistic — based on the characteristics of the sets of data. The more drastic the differences between the sets of data are, and the larger the sets themselves, the bigger is the test statistic's value. The test statistic's value allows us to find the probability of obtaining the results at least as extreme as the results we have observed on the assumption that the null hypothesis is true. This probability is traditionally denoted as p-value. The larger is the value of the test statistic, the smaller is the p-value. If the p-value is reasonably small as compared to the significance level (usually, the threshold 0,05 is conventionally accepted), the null hypothesis has to be rejected: we conclude that the observed differences between the sets of data (in our case — treaties and prayers) are not accidental but reflect the properties of these genres of texts. If, on the other hand, the p-value is higher than the significance level, than the data contains no evidence to refute the null hypothesis, and the differences between the two sets of data can be accidental.

In this section we employ the techniques outlined above in correlational studies identifying argument clitic distribution with respect to the following factors: diachronic change (5.1), association with other parameters of the clause (5.2) and cooccurrence of argument clitics with specific feature values (5.3).

5.1. *Argument clitics diachronically*

The whole corpus (4643 clauses) contains 1362 clauses attesting argument clitics and 102 clauses attesting two of them. Thus, in our sample, about 29,3% of clauses contain at least one argument clitic and about 2,2% — two argument clitics. The major question we address here is to what extent we can extrapolate these proportions to the whole population, i.e. to the whole collection of Hittite texts.

Provided that the sample is large, or the observed proportion is quite close to 50%, or both, we can assume that the sampling distribution of the proportion is approximately normal. The conventional threshold is N×p>5 (and N×(1—p)>5), where N stands for the sample size and p — for the observed proportion. In our case, N=4643, proportion of clauses with at least one argument clitic p_1=0,293, and proportion of clauses with two argument clitics p_2=0,022, respectively. This enables us to consider the proportions in our sample as observations of normally distributed variables. Consequently, we can calculate the 95% confidence interval estimate of the proportions as proportion in the sample ±1.96×standard error. This gives us the confidence limits (α≤0,05) for the proportion of clauses with at least one argument clitic as 29,3±1,3% (28%<p_1<30,6%) and for the proportion of clauses with two argument clitics as 2,2±0,4% (1,8<p_2<2,6%). These numbers might be extremely valuable to estimate the corpus size required to obtain the desired number of argument clitics.

The calculation above relies heavily on the assumption that our sample is randomly selected. However, our estimation can be skewed if the sample is biased with respect to some parameter affecting the variable being measured. In our case, we might suppose that dating of texts in the sample has impact on the proportion of clitics. We might wonder whether the tendency to employ argument clitics varies diachronically.

Accordingly, we can try to check whether the proportions calculated for the whole database change through time. The texts of the database are assigned to one of the following periods: Early New Hittite (eNH), New Hittite (NH) and Late New Hittite (lNH). The frequency distribution of clauses with clitics in these periods is given in Table 6; Figure 5 shows a diagram presenting the composition of the corpus.

Table 6. Clauses with argument clitics diachronically

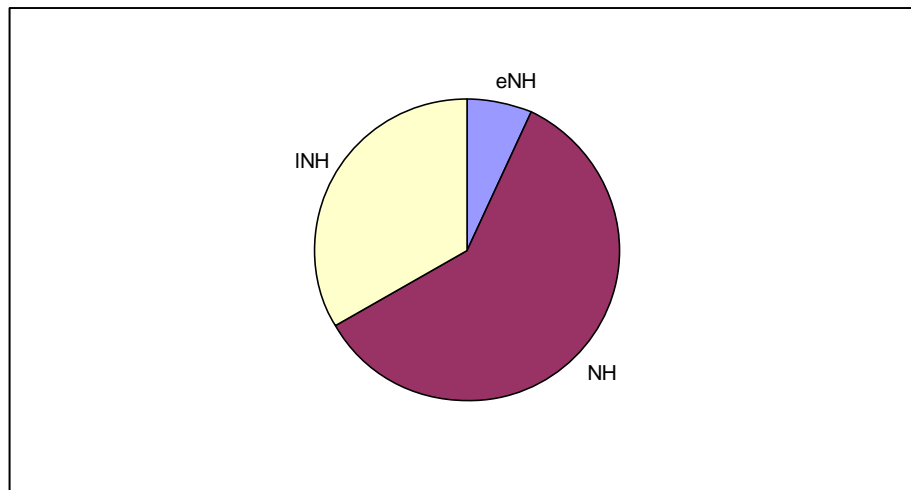|	Corpus size, clauses	Clauses without argument clitics	Clauses with at least one argument clitic	Clauses with two argument clitics
eNH	324	204	120 (37%)	13 (4,0%)
NH	2774	1991	783 (28%)	56 (2,0%)
lNH	1545	1086	459 (30%)	33 (2,1%)
Total	4643	3281	1362 (29%)	102 (2,2%)

Figure 5. Pie diagram of the corpus composition

The question is, therefore, whether eNH, NH and lNH subcorpora differ with respect to the proportion of clauses with argument clitics. Were this the case, then our proportion estimates could be biased because the subcorpora are of different size.

To answer the question we use the chi-square test of independence/association of the two nominal variables characterizing a clause: text dating (eNH, NH, lNH) and presence of argument clitic / two argument clitics (yes/no).

The results are as follows. For the proportion of clauses with two argument clitics, the chi-square test statistic is lower than the critical value for 2 degrees of freedom (it corresponds to p-value=0,057684, which is higher than the significance level $\alpha \leq 0,05$); this means that, according to our data, the proportion of clauses with two argument clitics does not change diachronically. However, for the proportion of clauses with at least one argument clitic, the results are different: the chi-square test statistic value corresponds to p-value=0,003904, which is lower than the significance level ($\alpha \leq 0,05$), thus, the hypothesis of independence of the two variables should be rejected, and the data show differences between subcorpora in the proportion of clauses with clitics.

Returning to Table 6 we can observe that the data for eNH subcorpus differ significantly from the two other subcorpora; moreover, the observed proportions in the eNH subcorpus (p_1=37% and p_2=4,0%) lie outside the interval estimates ($28\%<p_1<30,6\%$ and $1,8<p_2<2,6\%$, respectively). We might suggest that the eNH corpus is simply not large enough for the chi-square test to reveal the association for the proportion of clauses with two argument clitics as well. The question now is if the data support the diachronic change in the proportion of clauses with argument clitics.

We think that there is not enough evidence for this claim. Note that the eNH subcorpus is substantially smaller than the two other, cf. Figure 5. More importantly, it is represented by only two texts: CTH 42 (treaty of Suppiluliuma I with Huqqana of Hayasa, ex. A KBo 5.3+ restored and supplemented from ex. B KBo 19.44+) 304 clauses, CTH 253.1 (instructions of Suppiluliuma I for the Military and a corresponding Oath, ex. A KUB 21.41 restored and supplemented from ex. B KUB 26.57) 20 clauses. Therefore, we can state that the eNH subcorpus is unbalanced for text

genres, and attribute its peculiarities to the properties of a specific genre, or even a specific text.[5] Importantly, it is the specificity of the eNH subcorpus that produced the statistically significant effect in our calculation of the diachronic stability of proportions. Indeed, if we group eNH together with NH subcorpus, no difference between the two samples (eNH+NH vs. lNH) is found (chi-square test, p-value=0,68146).

5.2. *Argument clitics and other parameters of the clause*

In this section we are going to check eventual association between presence of argument clitics and other parameters characterizing the clause to be statistically relevant.

We start with the eventual correlation between transitivity of the clause and presence of argument clitics of various kinds.

Although Hittite attests both subject (NOM) and direct object (ACC) clitics, only one of them can be present in the clause. Their distribution relies on transitivity: in transitive clauses, only direct object (ACC) clitics can appear, subject (NOM) clitics are not licit. In intransitive clauses, we expectedly encounter only subject (NOM) clitics. The complementary distribution of subject (NOM) and direct object (ACC) clitics is captured by the traditional Hittitological templatic model as the ban on the multiply filled slot — both subject (NOM) and direct object (ACC) clitics belong to the slot 3. However, this model cannot explain why the distribution is the way that it is: the cooccurence ban cannot filter out a transitive clause with a non-clitic direct object and a clitic subject.

The principled explanation is given by Garrett (1990; 1996) who observes that subject clitics are only licit as unaccusative and passive subjects. Building on this generalization, Lyutikova and Sideltsev (2020) hypothesize that argument clitics are only licensed as internal arguments. If so, the distribution of subject (NOM) and direct object (ACC) clitics follows: they originate in the same structural position of the direct internal argument, and then become direct objects in transitive clauses and subjects otherwise (in unaccusative and passive clauses). Therefore, transitivity affects the distribution of subject (NOM) and direct object (ACC) argument clitics in a trivial manner: it determines whether the internal argument surfaces as a subject (NOM) or direct object (ACC). Intransitive clauses lacking internal argument (unergatives) can contain neither subject (NOM) nor direct object (ACC) clitic. The distribution is summarized in Table 7.

Table 7. The distribution of NOM and ACC argument clitics

	Transitive	Intransitive		
		Passive	Unaccusative	Unergative
NOM	——	+	+	——
ACC	+	——	——	——

The generalizations presented in Table 7 reveal the grammatical constraints on appearance of argument clitics in a certain structural position, namely, the external argument position. However, it does not show whether transitivity affects the likelihood of the internal argument to be realized by the clitic: we do not know whether clitic internal arguments are more likely to be subjects or

---

5. To verify this hypothesis we should create the subcorpus of treaties with various datings and check whether the proportions in this subcorpus differ significantly from the rest of the corpus. We leave this task for the future research.

direct objects. The research question then can be put as follows: does (in)transitivity correlate with the clitic realization of the corresponding argument?

In order to answer this question we need to compare the proportion of clitic arguments in transitive and intransitive (passive and unaccusative) clauses, respectively. To do so we have to know the distribution of transitive and intransitive clauses in the whole corpus of texts that fed our database. Unfortunately, since only those clauses that contain at least one argument clitic are included in the database and tagged for transitivity parameter, this way requires additional markup of the corpus.

To avoid the extra work we can use clauses without NOM/ACC clitics in our database as a baseline. So we compare the proportion of transitive clauses in two samples: the first sample (Sample 1) including clauses with a NOM/ACC clitic and the second sample (Sample 2) including clauses containing only one dative clitic.

An additional restriction should be that for Sample 1, we should select only clauses with a 3p NOM/ACC clitic. The reason for this is the fact that Hittite lacks nominative 1-2p clitics. Consequently, inclusion of 1-2p accusative clitics would skew the proportion towards transitive clauses. As for Sample 2, we can use clauses with dative clitics of whatever person or with dative clitics of the 3p only; however, since the proportion of transitive clauses does not change drastically, we prefer the first option, since it gives us larger numbers. The contingency table (Table 8) shows the data of the two samples.

Table 8. Transitivity and clitic internal argument

	transitive	intransitive	clauses in total
Sample 1. Clauses with a single NOM/ACC 3p clitic	411 (67,9%)	194 (32,1%)	605 (100%)
Sample 2. Clauses with a single DAT clitic	221 (44,5%)	275 (55,5%)	496 (100%)

It is easy to see that clauses in Sample 1 show a much higher proportion of transitive clauses than in the baseline Sample 2 (67,9% vs. 44,5%). To make sure that this difference is statistically relevant we can compare the distribution in Sample 1 with the expected distribution based on the proportion provided by the baseline Sample 2. If the proportion of clitic internal argument were the same in transitive and intransitive clauses, we would expect the following distribution for Sample 1 (Table 9):

Table 9. Transitivity and clitic internal argument, expected

	transitive	intransitive	clauses in total
Sample 1. Clauses with a single NOM/ACC 3p clitic, expected	270 (44,5%)	335 (55,5%)	605 (100%)
Sample 2. Clauses with a single DAT clitic	221 (44,5%)	275 (55,5%)	496 (100%)

The chi-square test shows that the difference between the observed and expected distribution in Sample 1 is statistically significant (p-value<0,001). We can interpret this as direct objects being significantly more often expressed by a clitic than intransitive subjects are. This difference is partially due to the absence of subject clitics with unergatives. However, unergative clauses are very rare in Hittite (about 10% among all intransitive clauses, according to our data based on the

corpus of Hittite verbs not discussed in this paper). So the general tendency remains: direct objects are more likely to be expressed by weak pronouns than intransitive subjects.

The next potential correlation that can be checked in our database is the association of the dative clitic with the presence of the locative phrase. Cross-linguistically, dative clitics are often related to prepositional / postpositional locative phrases, like the so called *dessus*-datives in French discussed extensively in Kayne 1975 and Postal 1990, cf. (13).

(13) *Elle le   leur_i   a   jeté   [_{PP} *dessus*   t_i].
     she   him.ACC them.DAT   AUX   throw.PP   onto
     'She threw him on top of them.'

In Hittite, dative clitics may raise out of the postpositional phrase as well, cf. (14). The postposition is stranded in situ, and its argument clitic joins the clitic chain. Note that postpositions in New Hittite govern dative-locative, and, moreover, dative and locative became morphologically undistinguishable (Hoffner, Melchert 2008: 66, 68, 74, 257-262).

(14) *našma=mu=za   mān   LÚ ^{URU}Ḫatti   namma   kuiški   idālu*
     or=me= REFL   if   Hittite   then   some.NOM.SG.C   evil.ACC.SG.N
     menaḫḫanda *šarā   dāi*
     against   up   take.3SG.PRS
     'Or further, if some Hittite undertakes evil **against** <u>me</u>' (eNH (CTH 42.A) KBo 5.3+ obv. ii 32-33).

Another interesting issue is the alleged association of dative and locative clitics. Hoffner and Melchert (2008: 371) note that dative clitics are frequently accompanied by the locative clitics *-kan* and *-šan*. So we might want to test these qualitative generalizations against quantitative data. The corresponding frequencies for dative clitics are given in Table 10, data on clauses without dative clitics serving as a baseline. Expectedly, the baseline clauses do not exhibit postposition stranding.

Table 10. Cooccurrence of dative and nominative-accusative clitics with postpositions and locative clitics

	Total	Postposition stranding	Locative clitic	Postposition stranding + locative clitic
DAT clitic	598 (100%)	57 (9,5%)	121 (20,2%)	13 (2,2%)
NOM-ACC clitic (without DAT clitic)	764 (100%)	0	157 (20,5%)	0

The data provide answers to the following questions:
— which proportion of dative clitics originates from postpositional phrases?
— is there a correlation between the presence of dative clitic and the presence of locative clitic?
— is there a correlation between the presence of postpositional dative clitic and the presence of locative clitic?

First, let us calculate the interval estimate of the proportion of postpositional dative clitics among all dative clitics. The same technique that we employed in 5.1 gives us the 95% confidence intervals 9,5±2,4% (7,1%<p<11,9%), α≤0,05. This means that with the 95% probability, the proportion of postpositional dative clitics in the whole collection of Hittite texts is between 7,1 and 11,9%.

Secondly, let us check whether the presence of the dative clitic correlates with the presence of the locative clitic. To do so we should compare frequency of locative clitics in clauses with and without dative clitics. If the presence of the locative clitic is independent of the presence of the dative clitic and vice versa, we expect that the proportion of locative clitics would be the same in clauses with and without dative clitics. The chi-square test shows that this is indeed the case: the chi-square test statistic is lower than the critical value (it corresponds to p-value=0,83965, which is higher than the significance level α≤0,05); this means that, according to our data, the proportion of clauses with locative clitics is independent of the presence of dative clitics.

Finally, let us verify the independence of postpositional dative clitics and locative clitics. Out of 598 clauses containing the dative clitic, 57 clauses exhibit postpositional dative, 121 attests a locative clitic, and 13 contain both of them. The distribution data are rearranged in Table 11.

Table 11. Cooccurrence of postpositional dative clitics and locative clitics

	Postpositional dative clitic	Other dative clitics	Total
Locative clitic	13	108	121
No locative clitic	44	433	477
Total	57	541	598

Again, the chi-test shows that there is no evidence for the association of the two parameters (p-value=0,73226). Thus, postpositional dative clitics and locative clitics are distributed independently of each other.

5.3. *Argument clitics and their features*

In this section, we present our findings about various parameters concerning cooccurrence of two argument clitics. As already noted above, Hittite clause might contain at most two argument clitics, one dative and one accusative or nominative. The question we are interested in is whether argument clitics demonstrate other restrictions on cooccurrence, and if so, which grammatical features expressed by argument clitics (gender, number, person, case) are subjects to such restrictions.

We start by establishing the independence of appearance of dative and accusative/nominative clitics. The contingency table (Table 12) presents frequencies of clauses with a single dative clitic, with a single accusative/nominative clitic and with two argument clitics.

Table 12. Clauses with a single argument clitic and two argument clitics

	DAT clitic	no DAT clitic	Total
ACC/NOM clitic	102 (2,2%)	764	866 (18,7%)
no ACC/NOM clitic	496	3281	3777
Total	598 (12,9%)	4045	4643(100%)

It appears that the proportion of clauses with two argument clitics (0,022) is very close to the product of the proportions of clauses with a dative clitic (0,129) and the proportion of clauses with an accusative/nominative clitic (0,187), which equals 0,024. The chi-square test confirms the

independence of appearance of dative and accusative/nominative clitics (p-value=0,31064). Therefore, we conclude that dative clitics and accusative/nominative clitics appear independently from each other.

The next question is whether features of clitics are independent of the presence of the other clitic in the clause. In section 4 we observed that the person feature shows correlation with the case feature: 1-2p clitics tend to be dative, whereas 3p clitics — accusative/nominative. Now we can check whether the proportion of 1-2p and 3p in dative or accusative/nominative clitic depends on the presence of the second argument clitic.

To answer this question we compare the number of clauses containing 1-2p and 3p clitics under two conditions: (i) if the clause contains no other argument clitic, and (ii) if the clause contains a second argument clitic. We are going to make this comparison for dative and accusative clitics separately.[6]

Data for dative clitics are grouped in Table 13. We see that the proportion of 1-2p within dative clitics is roughly the same under both conditions. The chi-square test also shows that the conditions do not influence the distribution of person feature in dative clitics (p-value=0,49226).

Table 13. Person in dative clitics

	1-2p	3p	Total
Condition (i): clauses with one clitic	321 (64,7%)	175	496 (100%)
Condition (ii): clauses with two citics	70 (68,6%)	32	102 (100%)
Total	391	207	598

Data for accusative clitics look different. Under condition (i) that only clauses with a single clitic are taken into account, 1-2p accusative clitics occur less frequently than 3p accusative clitics, but still their proportion amounts to 28%. In clauses with two argument clitics, 1-2p accusative clitics do not occur at all. This distribution cannot be accidental, and chi-square test confirms that (p-value<<0,001).

Table 14. Person in accusative clitics

	1-2p	3p	Total
Condition (i): clauses with one clitic	159 (28%)	409	568 (100%)
Condition (ii): clauses with two citics	0 (0%)	56	56 (100%)
Total	159	465	624

We conclude that the presence of the second argument clitic does not affect the distribution of person in datives, but severely restricts 1-2p realization in accusative clitics. The Hittitological templatic model of the clitic complex explains this pattern by exploiting the ban on multiply filled

---

6. We exclude nominative clitics from this comparison, since 1-2p nominative clitics are not attested in Hittite.

slots. Since 1-2p accusative clitics occupy the same slots (2 and 4) that dative clitics, either dative clitic or 1-2p accusative clitic can appear in the clause. On the contrary, 3p accusative clitics are in slot 3, which enables them to cooccur with dative clitics in slots 2 and 4.

There remains an additional complication, though. The ban on multiply filled slots cannot filter out the combination of dative and accusative clitics when they belong to different slots — for example, 1-2p pl accusative clitic (slot 2) and dative sg clitic (slot 4) or 1-2p sg accusative clitic (slot 4) and dative pl clitic (slot 2). In our corpus, these combinations never occur; let us check whether this absence is statistically significant.

Provided that the general proportion of accusative and dative clitics is the same in clauses with one or two argument clitics (as confirmed above), we can check whether the specific occupancy of the slot 2 (1-2p pl accusative clitic) is dependent on the presence of the clitic in the slot 4, and the specific occupancy of the slot 4 (1-2p sg accusative clitic) is dependent on the presence of the clitic in the slot 2. To do so we should compare the observed frequency of clauses with 1-2p accusative clitic and a dative clitic (0) with the expected frequency of such clauses under the assumption that the probability of appearance of 1-2p accusative clitics and the probability of appearance of dative clitics is invariant, provided that the ban on multiply filled slots is respected.

Thus, we are interested in the event of the simultaneous occupancy of slots 2 and 4 which is not filtered out by the restrictions on the uniqueness of the accusative clitic and the uniqueness of the dative clitic in the clause, and does not violate the ban on multiply filled slots. This gives us the following scheme of the contingency table (Table 15):

Table 15. Scheme of the contingency table for checking restrictions on 1-2p accusative clitics

	slot 2 filled	slot 2 empty
slot 4 filled	the event we are interested in: — 1-2p PL ACC in slot 2 AND 1-3p SG DAT in slot 4 OR — 1-3p PL DAT in slot 2 AND 1-2p SG ACC in slot 4	— 1-3p SG DAT in slot 4, slots 2 and 3 empty OR — 1-2p SG ACC in slot 4, slot 2 empty
slot 4 empty	— 1-3p PL DAT in slot 2, slots 3 and 4 empty OR — 1-2p PL ACC in slot 2, slot 4 empty	— clauses without clitics

Using observational data represented in Tables 2 and 3, we can calculate the frequencies of the events in the cells of Table 14. This gives us Table 16.

Table 16. Observed frequencies

	slot 2 filled	slot 2 empty	Total
slot 4 filled	0	548	548 (14%)
slot 4 empty	107	3281	3388 (86%)
Total	107 (2,7%)	3829 (97,3%)	3936

What we are going to check now is the probability to obtain the distribution in Table 16 under the assumption of independence of the appearance of clitics in slots 2 and 4, provided that the

restrictions on the uniqueness of accusative and dative clitics and the ban on multiply filled slots are respected. To do so we are going to employ the chi-square test to compare the differences in observed data distribution in Table 16 and expected data distribution in Table 17 under the null hypothesis that the variables are independent.

Table 17. Expected frequencies

	slot 2 filled	slot 2 empty	Total
slot 4 filled	15 (14% of 107)	533 (14% of 3829)	548 (14%)
slot 4 empty	92 (86% of 107)	3296 (86% of 3829)	3388 (86%)
Total	107 (2,7%)	3829 (97,3%)	3936

The chi-square test shows that the deviation of the observed frequencies from the expected ones is too large to be accidental (p-value=0,0000229). Therefore, the constraint on the appearance of 1-2p accusative clitics in the presence of a dative clitic should be included into the list of grammatical rules for Hittite.

The traditional Hittitology attempts to deal with this constraint in the template framework, see already Laroche (1958: 161); Friedrich (1960: §288); Hoffner (1986: 93-94).

However, the rule that slots 2 and 4 are mutually exclusive is rather a descriptive generalization than an explanation. Besides, this is a unique constraint on the simultaneous occupancy of slots for Hittite; on the contrary, clitics of the same functional class are allowed to cooccur, provided that they belong to different slots. In (15), for example, we observe the combination of the dative argument clitic -*šmaš* 'to you' (functioning as possessive "your") in slot 2 and reflexive clitic -*za* in slot 5. Importantly, the reflexive clitic is argumental here and corresponds to the Recipient dative ('for themselves').

(15) *kuiēš*=(*š*)*maš*=*za* $^{LÚ.MEŠ}$APIN.LÁ $^{LÚ.MEŠ}$NU.GIŠ.KIRI$_6$.GEŠTIN

　　which.NOM.PL.C=you. DAT.PL= REFL plowmen   vinedressers

$^{LÚ.MEŠ}$NU.^{GIŠ}M[(Ú.SAR   MUNUS^{MEŠ} ^{N})]^{A4}ARA$_5$ *danna*   *šanḫiškanzi*

gardeners   women   mill   take.INF seek.IPF.3PL.PRS

'Others wish to capture your plowmen, vinedressers, gardeners and grinding-women' (NH (CTH 377.B) rev. 9'-10').

Therefore, the mutual exclusiveness of slots 2 and 4 looks like a stipulation (i.e. an ad hoc assumption), and it is desirable to find another explanation for this unexpected gap in the data. The explanation might come from the hypothesis that the restrictions on argument clitics cooccurrence are syntactic rather than phonological, i.e. related to the formal templatic structure of the clitic complex.

In Lyutikova and Sideltsev (2020), we suggest that the peculiarities of clitics cooccurrence in Hittite are a realization of a well-known constraint on combining the person feature and the syntactic position of phonologically weak elements — the so-called Person Case Constraint (PCC) (Bonet 1991, 1994; Anagnostopoulou 2003, 2005, 2017; Béjar 2003; Béjar, Rezac 2003; Nevins 2007; Adger and Harbour 2007; Rezac 2007, 2011; among many others). In many languages, including Romance languages, Greek, Czech, Basque, and Georgian (see Haspelmath 2004 for a (non-exhaustive) list), certain person combinations are restricted when two phonologically "weak" arguments (clitics, weak pronouns, agreement markers) occupy the same domain. Thus, in French,

in combinations of a direct and indirect object, both of which are phonologically weak, the direct object may not be 1p or 2p, cf. (16):

(16) a. *Paul*   ***le***   ***lui***   *a*   *présenté.*
       Paul   3.SG.M.ACC   3.SG.DAT has   introduced
       'Paul introduced him to her.'
     b.   *Paul   ***me*** / ***te***   ***lui***   *a*   *présenté.*
          Paul   1.SG.ACC / 2.SG.ACC   3.SG.DAT has   introduced
     c.   *Paul   ***lui***   ***m'*** / ***t'***   *a*   *présenté.*
          Paul   3.SG.DAT 1.SG.ACC / 2.SG.ACC   has   introduced
   int. : 'Paul introduced me / you to her.'

It is easy to see that in Hittite, exactly the same generalization holds: in a combination of two argument clitics, the accusative / nominative clitic can only be 3p.

Whatever the reason for PCC might be (both formal and functional analyses have been proposed in the literature, see Anagnostopoulou (2017) for more detail and Lyutikova, Sideltsev (2020) specifically for Hittite), it is clearly a more general constraint than the language-specific rules of arranging clitics in the clitic complex. Therefore, we might hypothesize that the constraint on multiply filled slots and the constraint on the simultaneous filling of slots 2 and 4 is a descriptive model only. Indeed, they are fully covered by the assumption that only one dative and only one nominative/accusative clitic can be licensed in Hittite clause, plus the PCC constraint. We believe that this analysis is superior to the traditional one, since it relies on cross-linguistically attested constraints rather than on language-specific rules.

To sum up, we have demonstrated that applying statistical techniques to data collection allows us to establish significant generalizations about the independence or association of various parameters, as well as identify relevant gaps in logically possible combinations which cannot be accidental. In this way, we obtain a sounder basis for linguistic theorizing.

6. *Conclusions*

In this methodologically oriented study we demonstrated on the basis of the Hittite system of enclitic personal pronouns and the cooccurrence of different enclitic pronouns within the enclitic chain that the application of modern digital methods to data collection and statistical evaluation of its resources enables us to establish the substantive basis for developing and supporting theoretical hypotheses even on the material of a dead language with its fairly limited general number of texts, on the one hand, and lack of on-line digital corpora with search tools, on the other hand.

Namely, we showed how the eternal problem whether absence of a certain form or meaning or a combination of forms and meanings is significant or simply incidental in a dead language can be solved by a representative database and statistical evaluation of its resources even in the absence of a full searchable corpus.

In this way the data of dead languages which is undeservedly ignored by modern theoretical linguistics due to insufficiently reliable generalisations, including lack of negative data, obtains the chance to enter the dataset of any linguistic theory.

7. Bibliography

Adger, David & Daniel Harbour. 2007. "Syntax and Syncretisms of the Person Case Constraint", *Syntax* 10: 2–37.

Alexandrov, Boris. 2019. "Posing a Question in Akkadian: *Wh*- and Other Interrogatives in Old Babylonian Letters", in Grégory Chambon, Michael Guichard & Anne-Isabelle Langlois (eds.) *De l'argile au numérique. Mélanges assyriologiques en l'honneur de Dominique Charpin* (Publications de l'Institut de Proche-Orient Ancien du Collège de France 3). Leuven, Paris, Bristol: Peeters: 13–46.

Anagnostopoulou, Elena. 2003. *The Syntax of Ditransitives. Evidence from Clitics*. Berlin, New York: Mouton de Gruyter.

Anagnostopoulou, Elena. 2005. "Strong and Weak Person Restrictions: A feature checking analysis", in Lorie Heggie & Francisco Ordoñez (eds.), *Clitic and Affix Combinations*, 199–235. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Anagnostopoulou, Elena. 2017. "The Person Case Constraint", in Martin Everaert & Henk van Riemsdijk (eds.), *The Wiley Blackwell companion to syntax*, 2nd edition, Oxford: Blackwell, 1–47. https://doi.org/10.1002/9781118358733.wbsyncom101.

Arkhipov Ilya & Sergei Loesov. 2020. "The *t*-Form of the Sargonic Verb" , in Ilya Arkhipov, Leonid Kogan & Natalia Koslova (eds.), *The Third Millennium Studies in Early Mesopotamia and Syria in Honor of Walter Sommerfeld and Manfred Krebernik* (CM 50), 35–53. Leiden, Boston: Brill.

Arkhipov, Ilya & Sergey Loesov. 2019. "Two genitive constructions of Old Babylonian", *BSOAS* 82/3: 395–403

Béjar, Susana. 2003. *Phi-Syntax. A Theory of Agreement.* Ph.D. Dissertation. University of Toronto.

Béjar, Susana & Milan Rezac. 2003. "Person licensing and the derivation of PCC effects", in Ana Teresa Perez-Leroux & Yves Roberge (eds.), *Romance linguistics: theory and acquisition*, 49–62. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Bickel, Balthasar & Johanna Nichols. 2007. "Inflectional morphology", in Timothy Shopen (ed.), *Language typology and syntactic description*, 169–240. Cambridge: CUP.

Bornkessel-Schlesewsky, Ina, Andrej Malchukov & Marc Richards (eds.). 2014. *Scales and Hierarchies: A Cross-Disciplinary Perspective*. (Trends in Linguistics. Studies and Monographs.; Vol. 277). Berlin: Mouton de Gruyter.

Bonet, Eulàlia, 1991. *Morphology after Syntax: Pronominal Clitics in Romance Languages.* PhD. Dissertation, MIT.

Bonet, Eulàlia, 1994. "The Person-Case Constraint: A Morphological Approach", in *MIT Working Papers in Linguistics 22. The Morphology-Syntax Connection*: 33–52.

Daniel, Michael. 2005. "Understanding inclusives", in Elena Filimonova (ed.). *Clusivity: Typology and case studies of the inclusive exclusive distinction*, 3–48. Amsterdam, Philadelphia: John Benjamins.

Charpin, Dominique. 2014a. "The Assyriologist and the Computer: The "Archibab" Project", *Hebrew Bible and Ancient Israel* 3: 137–153.

Charpin, Dominique. 2014b. "Ressources assyriologique sur internet", *Bibliotheca orientalis* LXXI 3-4: 331–358.

CHD – Güterbock, Hans, Harry Hoffner, & Theo van den Hout (eds.) 1989–. *The Hittite Dictionary of the Oriental Institute of the University of Chicago*, Chicago: The Oriental Institute of the University of Chicago.

Cohen, Eran. 2005. *The Modal System of Old Babylonian* (HSS 56). Winona Lake: Eisenbrauns.

Cohen, Eran. 2012. *Conditional Structures in Mesopotamian Old Babylonian* (LANE 4). Winona Lake: Eisenbrauns.

Croft, William. 1990. *Typology and universals*. Cambridge University Press.

van Gelderen, Elly. 2018. *Analyzing Syntax through Texts: Old, Middle and Early Modern English*. Edinburgh University Press.

Corbett, Greville G. 2000. "Number", in Martin Haspelmath (ed.), *Language Typology and Linguistic Universals: An International Handbook*, 816–831. Berlin: de Gruyter.

Déchaine, Rose-Marie & Martina Wiltschko, 2002. "Decomposing pronouns", *Linguistic Inquiry* 33/3: 409–442.

Garrett, Andrew, 1990. *The syntax of Anatolian pronominal clitics*. PhD dissertation, Harvard University.

Garrett, Andrew, 1996. "Wackernagel's Law and Unaccusativity in Hittit", in Aaron L. Halpern & Arnold M. Zwicky (eds.), *Approaching Second Position Clitics and Related Phenomena* (CSLI Lecture Notes Number 61), 85–134. Stanford, California: CSLI Publications Center for the Study of Language and Information.

Goedegebuure, Petra, 2014. *The Hittite Demonstratives. Studies in Deixis, Topics and Focus* (StBoT 55), Wiesbaden: Harrassowitz.

Goedegebuure, Petra. 2018. "The Packagers -ant- and -a-, and the Origin of Split-Ergativity in Hittite (and Lycian)", in David M. Goldstein, Stephanie W. Jamison, and Brent Vine (eds.), *Proceedings of the 29th Annual UCLA Indo-European Conference*, 77–115. Bremen: Hempen.

Friedrich, Johannes. 1960. *Hethitisches Elementarbuch. 1. Teil: Kurzgefasste Grammatik*. Heidelberg: Winter.

Harley, Heidi & Elizabeth Ritter. 2002. "Person and number in pronouns: A feature-geometric analysis", *Language* 78: 482–526.

Haspelmath, Martin. 2004. "Explaining the Ditransitive Person-Role Constraint: a usage-based account", *Constructions* 2/2004.

Hernáiz, Rodrigo. 2019. *Studies in linguistic and orthographic variation in Old Babylonian letters* (AOAT 465). Müster: Ugarit-Verlag.

Hoffner, Harry A., Jr. 1986. "Studies in Hittite Grammar", in Harry A. Hoffner, Jr. and Gary M. Beckman (eds.), *Kaniššuwar. A tribute to Hans G. Güterbock on his seventy-fifth birthday* (AS 23), 83–94. Chicago: The Oriental Institute of the University of Chicago.

Hoffner, Harry A., Jr. 2009. *Letters from the Hittite Kingdom* (SBL WAW 15), Atlanta.

Hoffner, Harry A., Jr. & H. Craig Melchert. 2008. *A Grammar of the Hittite Language*, Part 1: Reference Grammar, Winona Lake, Indiana: Eisenbrauns.

Inglese, Guglielmo. 2017. "A synchronic and diachronic study of Hittite reciprocal constructions, *Studies in Language* 41/4: 956–1006.

Inglese, Guglielmo, Giuseppe Rizzo & Miriam Pflugmacher. 2019. "Definite referential null subjects in Old Hittite: a corpus study", *Indogermanische Forschungen* 134: 137–170.

Jagersma, Bram. 2010. *A Descriptive Grammar of Sumerian*. PhD Dissertation, University of Leiden.

Kalinin, Maxim & Sergei Loesov. 2017. "Lexical Sondergut of Neo-Assyrian", *State Archives of Assyria Bulletin* XXIII: 1–46.

Kayne, Richard. 1975. *French Syntax.* Cambridge MA: MIT Press.

Kühne, Cord & Heinrich Otten. 1971. *Der Šaušgamuwa-Vertrag* (StBoT 16). Wiesbaden: Harrassowitz.

Kouwenberg, Norbertus Johannes Cornelius. 2017. *A Grammar of Old Assyrian* (HdO 118). Leiden, Boston: Brill.

Laroche, Emmanuel. 1958. "Comparison du louvite et du lycien", *BSL* 53: 159–197.

Loesov, Sergei. 2004. "T-Perfect in Old Babylonian: The Debate and a Thesis", *Babel und Bibel* 1: 83–181.

Loesov, Sergei. 2005. "Akkadian Sentences about the Present Time, Part One", *Babel und Bibel* 2: 101–148.

Loesov, Sergei. 2006. "Akkadian Sentences about the Present Time (II/1)", *Babel und Bibel* 3: 133–148.

Lyutikova, Ekaterina & Andrei Sideltsev. 2020. "On the Syntax of Argument Clitics in Hittite", *Transactions of the Philological Society* 118/1: 29–78.

Miller, Jared. 2013. *Royal Hittite Instructions* (SBL WAW 31), Atlanta.

Mora, Clelia & Mauro Giorgieri. 2004. *Le lettere tra i re ittiti e i re assiri ritrovate a Ḫattuša* (HANEM 7), Padova: S.a.r.g.o.n editrice.

Nevins, Andrew. 2007. "The Representation of Third Person and its Consequences for Person-Case effects", *Natural Language and Linguistic Theory* 25: 273–313.

Postal, Paul. 1966. "On so-called pronouns in English", in David Reibel & Sanford Schane (eds.), *Modern studies in English*, 201–223. Englewood Cliffs, N.J.: Prentice-Hall.

Postal, Paul. 1990. "French indirect object demotion", in Paul M. Postal & Brian D. Joseph (eds.), *Studies in Relational Grammar* 3, 104–200. Chicago: University of Chicago Press.

Rezac, Milan. 2007. "Escaping the Person Case Constraint: Referential computation in the phi-system", *Linguistic Variation Yearbook* 6: 97–138.

Rezac, Milan. 2011. *Phi-features and the modular architecture of language.* Berlin: Springer.

de Ridder, Jakob Jan. 2018. *Descriptive Grammar of Middle Assyrian* (Leipzig Altorientalische Studien 8). Wiesbaden: Harrassowitz.

Silverstein, Michael. 1986. "Hierarchy of features and ergativity", in Pieter Muysken & Henk van Riemsdijk (eds.), *Features and Projections*, 163–232. Dordrecht: Foris. Originally published in Robert M.W. Dixon (ed.), 1976, *Grammatical Categories in Australian Languages*, Canberra: Australian Institutes of Aboriginal Studies.

Streck, Michael P. 2010. "Grosses Fach Altorientalistik: Der Umfang des keilschriftlichen Textkorpus", *MDOG* 142: 35–58.

Walkden, George & Kristian A. Rusten. 2016. "Null subjects in Middle English", *Journal of Germanic Linguistics* 28/1: 31–78.

Walkden, George & Donald Morrison. 2017. "Regional variation in Jespersen's Cycle in Early Modern English", *Studia Anglica Posnaniensia* 52/2: 173–201.

Woods, Christopher. 2008. *The Grammar of Perspective. The Sumerian Conjugation Prefixes as a System of Voice* (CM 32). Leiden; Boston.