

XI чтения памяти С. А. Старостина

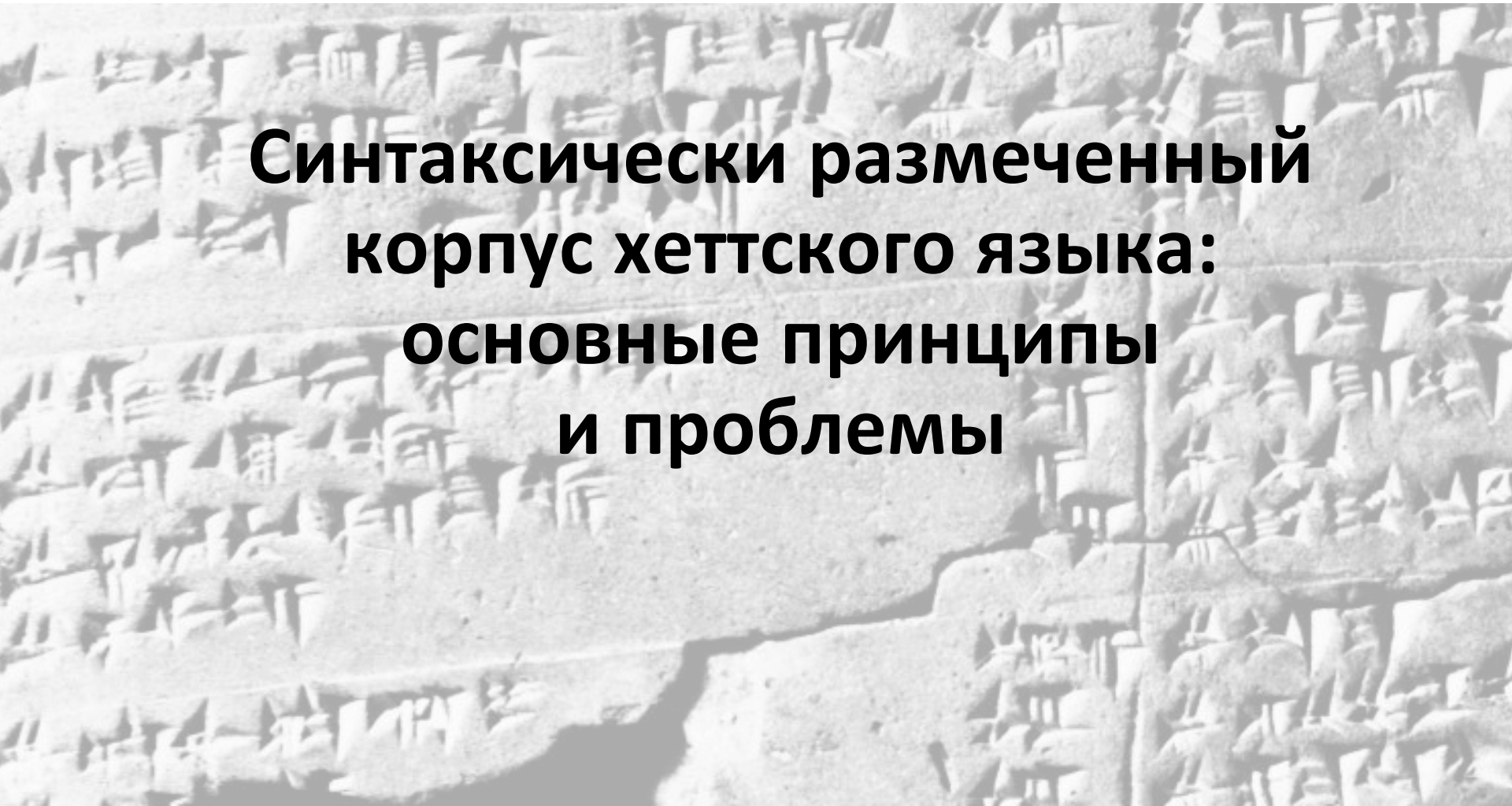
24–25 марта 2016 г.

Российский государственный гуманитарный университет

Мария Молина

Институт языкознания РАН

[maria.molina@me.com](mailto:maria.molina@me.com)



**Синтаксически размеченный  
корпус хеттского языка:  
основные принципы  
и проблемы**

# Хеттский язык

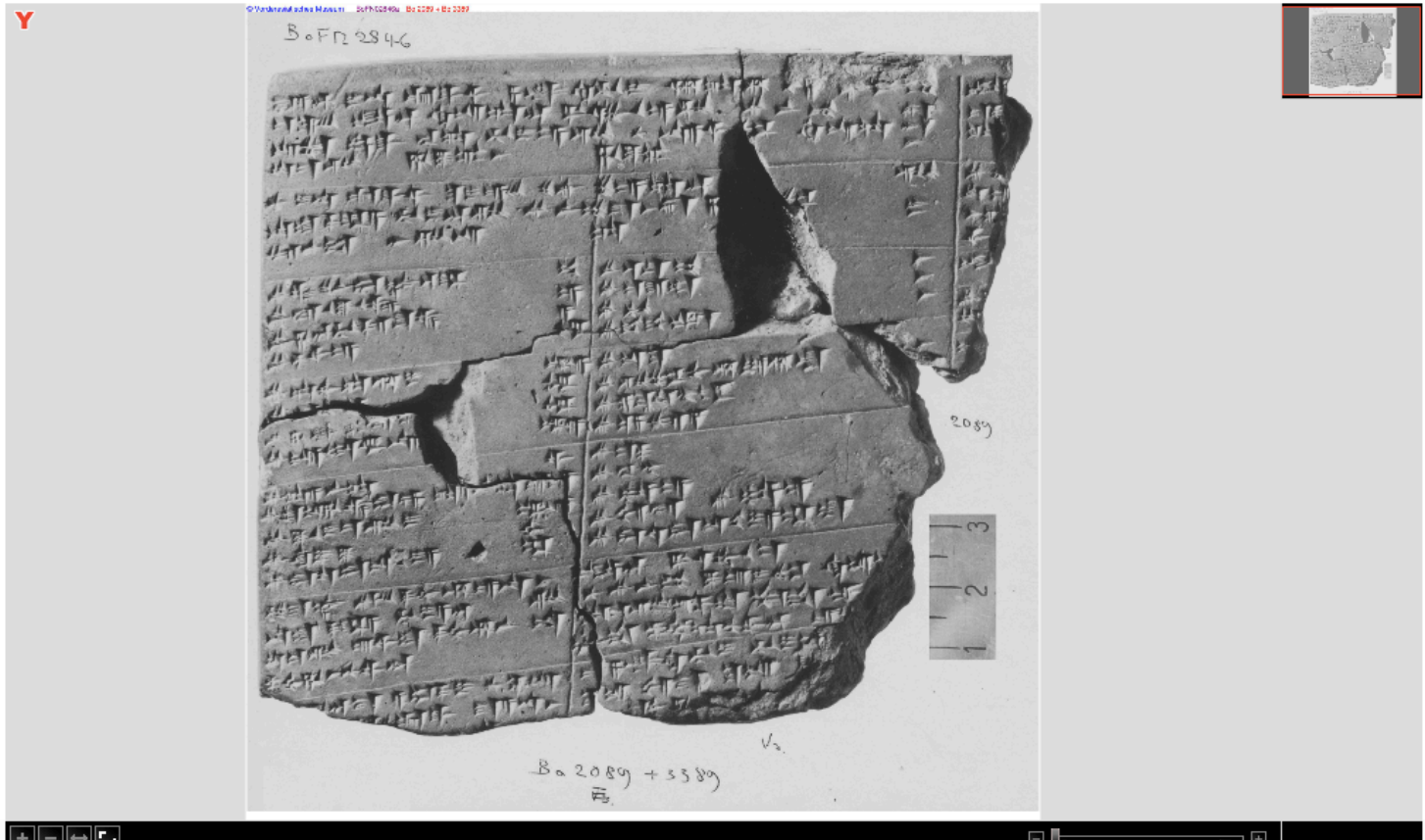
- древнейший зафиксированный в письменности индоевропейский язык анатолийской группы
- вид письменности: клинопись
- время: XVIII–XII вв. до н. э.
- территория: Центральная Анатолия
- нейтральный порядок слов: SOV
- язык с развитой системой сентенциальных клитик

# Клинописные таблички на хеттском



Mainzer Photoarchiv -flash-

Статус: [hethiter.net/](http://hethiter.net/): PhotArch BoFN02846a





# Публикации

СТН 263

(Протокол для дворцовых привратников)

(A iv 9') *ma-aḥ-ḥa-an-ma ḥa-li-in-du-wa ta-ra-an-zi nu* <sup>LÚ</sup>ME-ŠE-DI A-N[A  
...] (10') *ḥal-za-a-i ka-a-az-zu-e nu ša-aš-ta-an ša-ra-a da-a-i* (11') *na-an-  
kán kat-ta pé-e-da-i nu* <sub>ḥa</sub> <sub>li-in-du-wa-aš</sub> *ma-aḥ-ḥa-a[n]* (12') UD-at  
UD-at *ak-ku-<sub>uš-ke</sub> -ez-zi na-aš QA-TAM-MA e-ku-z[i]*

(A iv 13'–17'; B iv 1'–2') (13') *n[a-aš]-ta DUMU É.GAL šu-uḥ-<sub>ḥa</sub> ša-ra-a  
pa-iz-zi pé-ra-an-ma-aš-ši* [<sup>LÚ</sup>Ú.ḤÚB]<sup>79</sup> (14') *ḥu-ia-an-za nu* <sup>LÚ</sup>Ú.[Ḥ]ÚB  
<sup>GIŠ</sup>AB<sup>HL.A</sup> *an-da iš-ta-a-pí nu DUMU [É.GAL]* (15') <sup>GIŠ</sup>*za-ak-<sub>ki-in</sub> [p]é-eš-*

# Слитная транслитерация

## Protocol for the Palace Gatekeeper (CTH 263)

All duplicates are OH-MH/NS

A: KBo 50.275 (1344/v), **KBo 5.11(+)**Bo 2089 (A i 1–20, ii 1–12), KUB 26.23(+)**Bo 3389**

B: KUB 26.28 (Bo 3256); C: KBo 50.270 (260/v)

- iv 9'            *maḥḥan=ma ḥalinduwa taranzi*  
But when they say “to the palace complex” (*ḥalinduwa*)
- iv 9'–10'      *nu <sup>LÚ</sup>MEŠEDI AN[A ...] ḥalzāi kāzzue*  
the royal bodyguard <sup>(10')</sup>calls out *kāzzue* to [...]
- iv 10'          *nu šaštan šarā dāi*  
and he picks up the bed
- iv 11'          *n=an=kan katta pēdai*  
and he carries it down
- iv 11'–12'     *nu ḥalinduwaš maḥḥa[n] UD-at UD-at akkuškezzi*  
And just like he drinks in the palace complex day to day
- iv 12'          *n=aš QATAMMA ekuz[i]*  
so he drinks

# Размеченные корпуса ХЕТТСКИХ ТЕКСТОВ

Hethitologie Portal Mainz

<http://www.hethport.uni-wuerzburg.de>

<http://hethiter.net>

TITUS (Thesaurus Indogermanischer Text- und Sprachmaterialien), Hittite corpus

<http://titus.uni-frankfurt.de/indexe.htm>

- тексты разделены на клаузы
- выполнена слитная транслитерация
- сделана морфологическая разметка
- нет синтаксической разметки и нет разметки информационной структуры
- жанры: договора, законы, мифы и ритуалы

# Синтаксически размеченный корпус хеттского языка

<http://hittitecorpus.ru>

- **Цель:** синтаксическая разметка и разметка информационной структуры
- **Базовый элемент:** клауза
- **Жанры:** письма, инструкции, молитвы
- **ПО:** Реляционная база данных MsSQL с онлайн-интерфейсом поиска
- **Связанные таблицы:** тексты — клаузы — словоформы
- **Метаданные для текста:** публикация, дубликат, период создания
- **Метаданные для клаузы:** параграф и строки в табличке, уровень битости текста
- **Клауза:** силлабическая транслитерация, слитная транслитерация, перевод на английский
- **Словоформа:** слоги, слитная транслитерация, глоссы
- **Синтаксическая разметка:** порядок слов, структура составляющих (клауза), разметка части речи — PoS (словоформа)
- **Разметка информационной структуры:** фокус — топик



## New Hittite Letters

<a href="#">KUB 19.5 + KBo 19.79</a>	To King Muwattalli II from Manapa-Tarḫunta of the Šeḫa River Land
<a href="#">KUB 14.3</a>	From King Ḫattušili III to the King of Ahḫiyawa ("Tawagalawa Letter")
<a href="#">KUB 19.55 + KUB 48.90</a>	From King Tudḫaliya IV to Tarkašnawa of Mira ("Milawata Letter")
<a href="#">KBo 18.15</a>	To King Muršili II of Ḫatti from Mašḫuiluwa of Mira-Kuwaliya
<a href="#">KUB 23.102</a>	From King Muwattalli II or Muršili III (Urḫi-Teššub) to King Adad-nirāri I of Assyria
<a href="#">KUB 23.103 (Dupl. A)</a>	From King Tudḫaliya IV of Hatti to an Assyrian Nobleman Bāba-ah-iddina
<a href="#">KBo 18.2</a>	From the King to the Queen Mother; Piggyback Letter from the King to Several Persons
<a href="#">Güterbock 1979, 142–44</a>	To the King from the Queen; Piggyback Letter from mNÍG.BA-dU (= Ari-Teššub or Piyama-Tarḫunta) to [PN]
<a href="#">KBo 18.4</a>	From the King of Išūwa to his Father, the Chief of the Charioteers
<a href="#">KBo 18.48</a>	From the King to Prince Ḫešni
<a href="#">KBo 13.62</a>	School Tablet Based Upon a Real Letter from [...] to [the Queen?]
<a href="#">KBo 18.54</a>	To the King from Kaššū; Piggyback Letter to the King from Zarna-ziti
<a href="#">KBo 18.29</a>	To the King from UR.MAḪ-[...]
<a href="#">KUB 57.123</a>	To the King from Taki-Šarruma

**Text**

KUB 19.5 + KBo 19.79 Duplicate Time All sources All brokenness

**Clauses**

Paragraph Lines Syllables Normalized spelling Translation

**Word form**

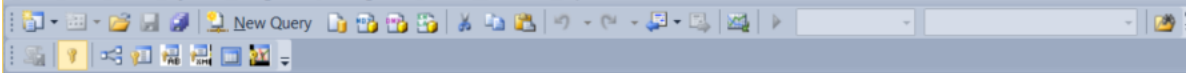
Syllables Normalized spelling Translation

Match phrase Search

1 texts

[Link to the query results](#)

KUB 19.5 + KBo 19.79 NH/NS <a href="#">edit</a>	1 obv. 1 <input type="button" value="morphology"/>	[A-NA <sup>d</sup> UTU-ŠI EN-Y]A? QÍ-BI-MA ANA <sup>d</sup> UTU-ŠI EN-YA QÍBI-MA To His Majesty, my lord, speak!
	1 obv. 1 <input type="button" value="morphology"/>	U[M-M]A <sup>m</sup> Ma-na-pa- <sup>d</sup> U ARAD-KA-MA UMMA <sup>m</sup> Manapa- <sup>d</sup> U ARAD-KA-MA Thus says Manapa-Tarhunta, your servant!
	2 obv. 2 <input type="button" value="morphology"/>	[ka-a-ša-kán ŠÀ KUR-]TI ḫu-u-ma-an SIG <sub>5</sub> -in kāša=kan ŠÀ KUR-TI ḫūman SIG <sub>5</sub> -in At present all is well in the country
	3 obv. 3 <input type="button" value="morphology"/>	[m o o o o o ] ú-it



Object Explorer

Connect

- dbo.HittiteClause
  - Columns
    - Id (PK, int, not null)
    - EnglishTranslation (nvarchar(MAX), not null)
    - NormalizedSpelling (nvarchar(MAX), not null)
    - Paragraph (int, not null)
    - Syllables (nvarchar(MAX), not null)
    - TextId (FK, int, not null)
    - Order (int, not null)
    - FormattedNormalizedSpelling (nvarchar(MAX), not null)
    - FormattedSyllables (nvarchar(MAX), not null)
    - EndLine (int, not null)
    - LinePrefix (nvarchar(MAX), not null)
    - StartLine (int, not null)
    - ParagraphPostfix (nvarchar(MAX), not null)
    - Brokenness (int, not null)
  - Keys
  - Constraints
  - Triggers
  - Indexes
  - Statistics
- dbo.HittiteText
  - Columns
    - Id (PK, int, not null)
    - Duplicate (nvarchar(MAX), not null)
    - Time (nvarchar(MAX), not null)
    - Order (int, not null)
    - Source (nvarchar(MAX), not null)
    - Publication (nvarchar(MAX), not null)
  - Keys
  - Constraints
  - Triggers
  - Indexes
  - Statistics
- dbo.HittiteWordForm
  - Columns
    - Id (PK, int, not null)
    - ClauseId (FK, int, not null)
    - NormalizedSpelling (nvarchar(MAX), not null)
    - Syllables (nvarchar(MAX), not null)
    - Translation (nvarchar(MAX), not null)
    - Order (int, not null)
    - FormattedNormalizedSpelling (nvarchar(MAX), not null)
    - FormattedSyllables (nvarchar(MAX), not null)
  - Keys
  - Constraints
  - Triggers
  - Indexes
  - Statistics
- Views

HOSTING.HittiteCo...o.HittiteWordForm

Column Name	Data Type	Allow Nulls
Id	int	<input type="checkbox"/>
ClauseId	int	<input type="checkbox"/>
NormalizedSpelling	nvarchar(MAX)	<input checked="" type="checkbox"/>
Syllables	nvarchar(MAX)	<input checked="" type="checkbox"/>
Translation	nvarchar(MAX)	<input checked="" type="checkbox"/>
[Order]	int	<input type="checkbox"/>
FormattedNormalizedSpelling	nvarchar(MAX)	<input checked="" type="checkbox"/>
FormattedSyllables	nvarchar(MAX)	<input checked="" type="checkbox"/>

Column Properties

HOSTING.HittiteCor... - dbo.HittiteClause

Column Name	Data Type	Allow Nulls
Id	int	<input type="checkbox"/>
EnglishTranslation	nvarchar(MAX)	<input checked="" type="checkbox"/>
NormalizedSpelling	nvarchar(MAX)	<input checked="" type="checkbox"/>
Paragraph	int	<input checked="" type="checkbox"/>
Syllables	nvarchar(MAX)	<input checked="" type="checkbox"/>
TextId	int	<input type="checkbox"/>
[Order]	int	<input type="checkbox"/>
FormattedNormalizedSpelling	nvarchar(MAX)	<input checked="" type="checkbox"/>
FormattedSyllables	nvarchar(MAX)	<input checked="" type="checkbox"/>
EndLine	int	<input type="checkbox"/>
LinePrefix	nvarchar(MAX)	<input type="checkbox"/>
StartLine	int	<input type="checkbox"/>
ParagraphPostfix	nvarchar(MAX)	<input checked="" type="checkbox"/>
Brokenness	int	<input checked="" type="checkbox"/>

Column Properties

HOSTING.HittiteCorpus - dbo.HittiteText

Column Name	Data Type	Allow Nulls
Id	int	<input type="checkbox"/>
Duplicate	nvarchar(MAX)	<input checked="" type="checkbox"/>
Time	nvarchar(MAX)	<input checked="" type="checkbox"/>
[Order]	int	<input type="checkbox"/>
Source	nvarchar(MAX)	<input checked="" type="checkbox"/>
Publication	nvarchar(MAX)	<input checked="" type="checkbox"/>

Column Properties

Object Explorer

- Connect
- dbo.HittiteClause
  - Columns
    - Id (PK, int, not null)
    - EnglishTranslation (r)
    - NormalizedSpelling
    - Paragraph (int, null)
    - Syllables (nvarchar(r
    - TextId (FK, int, not n
    - Order (int, not null)
    - FormattedNormaliz
    - FormattedSyllables (
    - EndLine (int, not null
    - LinePrefix (nvarchar(
    - StartLine (int, not nu
    - ParagraphPostfix (nv
    - Brokenness (int, null)
  - Keys
  - Constraints
  - Triggers
  - Indexes
  - Statistics
- dbo.HittiteText
  - Columns
    - Id (PK, int, not null)
    - Duplicate (nvarchar(
    - Time (nvarchar(max)
    - Order (int, not null)
    - Source (nvarchar(ma
    - Publication (nvarchara
  - Keys
  - Constraints
  - Triggers
  - Indexes
  - Statistics
- dbo.HittiteWordForm
  - Columns
    - Id (PK, int, not null)
    - ClauseId (FK, int, not
    - NormalizedSpelling
    - Syllables (nvarchar(r
    - Translation (nvarchara
    - Order (int, not null)
    - FormattedNormaliz
    - FormattedSyllables (
  - Keys
  - Constraints
  - Triggers
  - Indexes
  - Statistics
- Views

```
SQLQuery2.sql - 192...iteCorpus (sa (56)) x  
/***** Script for SelectTopNRows command from SSMS *****/  
SELECT TOP 1000 [Id]  
    ,[Duplicate]  
    ,[Time]  
    ,[Order]  
    ,[Source]  
    ,[Publication]  
FROM [HittiteCorpus].[dbo].[HittiteText]
```

100 %

Results Messages

	Id	Duplic...	Time	Or...	Sour...	Publication
1	271	NULL	NH/NS	1	NULL	KUB 19.5 + KBo 19.79
2	272	NULL	NH/NS	2	NULL	KUB 14.3
3	273	NULL	NH/NS	3	NULL	KUB 19.55 + KUB 48.90
4	274	NULL	NH/NS	4	NULL	KBo 18.15
5	275	NULL	NH/NS	5	NULL	KUB 23.102
6	276	A	NH/NS	6	NULL	KUB 23.103
7	277	NULL	NH/NS	7	NULL	KBo 18.2
8	278	NULL	NH/NS	8	NULL	Güterbock 1979, 142-44
9	279	NULL	NH/NS	9	NULL	KBo 18.4
10	280	NULL	NH/NS	10	NULL	KBo 18.48
11	281	NULL	NH/NS	11	NULL	KBo 13.62
12	282	NULL	MH?/NS	12	NULL	KBo 18.54
13	283	NULL	NH/NS	13	NULL	KBo 18.29
14	284	NULL	NH/LNS	14	NULL	KUB 57.123
15	285	NULL	NH/NS	15	NULL	KUB 19.23
16	286	NULL	NH/LNS	16	NULL	KBo 9.82
17	287	NULL	NH/NS	17	NULL	KBo 2.11
18	288	NULL	NH/NS	18	NULL	KBo 18.79
19	289	NULL	NH/NS	19	NULL	KUB 57.1
20	290	NULL	NH/NS	20	NULL	KUB 40.1
21	291	NULL	NH/NS	21	NULL	Bo 2810
22	292	NULL	NH/NS	22	NULL	KUB 23.85
23	293	NULL	NH/NS	23	NULL	VS 28.129
24	294	NULL	NH/NS	24	NULL	SMEA 45-T 1
25	295	NULL	NH/NS	25	NULL	BLMJ 1143
26	296	NULL	NH/NS	26	NULL	AT 125
27	297	NULL	NH/NS	27	NULL	ATT 35

- Connect
- dbo.HittiteClause
  - Columns
    - Id (PK, int, not null)
    - EnglishTranslation (r)
    - NormalizedSpelling
    - Paragraph (int, null)
    - Syllables (nvarchar(r))
    - TextId (FK, int, not null)
    - Order (int, not null)
    - FormattedNormalizedSpelling
    - FormattedSyllables
    - EndLine (int, not null)
    - LinePrefix (nvarchar(r))
    - StartLine (int, not null)
    - ParagraphPostfix (nvarchar(r))
    - Brokenness (int, null)
- Keys
- Constraints
- Triggers
- Indexes
- Statistics
- dbo.HittiteText
  - Columns
    - Id (PK, int, not null)
    - Duplicate (nvarchar(max))
    - Time (nvarchar(max))
    - Order (int, not null)
    - Source (nvarchar(max))
    - Publication (nvarchar(max))
  - Keys
  - Constraints
  - Triggers
  - Indexes
  - Statistics
- dbo.HittiteWordForm
  - Columns
    - Id (PK, int, not null)
    - Classed (FK, int, not null)
    - NormalizedSpelling
    - Syllables (nvarchar(r))
    - Translation (nvarchar(r))
    - Order (int, not null)
    - FormattedNormalizedSpelling
    - FormattedSyllables
  - Keys
  - Constraints
  - Triggers
  - Indexes
  - Statistics

```

/***** Script for SelectTopNRows command from SSMS *****/
SELECT TOP 1000 [Id]
, [EnglishTranslation]
, [NormalizedSpelling]
, [Paragraph]
, [Syllables]
, [TextId]
, [Order]
, [FormattedNormalizedSpelling]
, [FormattedSyllables]
, [EndLine]
, [LinePrefix]
, [StartLine]
, [ParagraphPostfix]
, [Brokenness]
FROM [HittiteCorpus].[dbo].[HittiteClause]

```

100 %

Results Messages

Id	EnglishTranslation	NormalizedSpelling	Paragra...	Syllables	TextId	Or...	FormattedNormalizedSpelling	FormattedSyllables
1	16772 To His Majesty, my lord, speak!	ANNA dUTU-ŠI EN-YA QIBI-MA	1	[A-NA dUTU-ŠI EN-YA? QI-BI-MA	271	1	<A>ANA</> <sup>d</sup><sup>UTU</sup><sup>ŠI</sup><sup>EN</sup><sup>YA</sup><sup>QI</sup><sup>BI</sup><sup>MA</sup>	<A>NA</> <sup>d</sup><sup>UTU</sup><sup>ŠI</sup><sup>EN</sup><sup>YA</sup><sup>QI</sup><sup>BI</sup><sup>MA</sup>
2	16773 Thus says Manapa-Tarhunta, your servant!	UMA mManapa-du ARAD-KA-MA	1	[UM-MJA mMa-na-pa-du ARAD-KA-MA	271	2	<U>MA</> <sup>m</sup><sup>Ma</sup><sup>na</sup><sup>pa</sup><sup>du</sup> <sup>ARAD</sup><sup>KA</sup><sup>MA</sup>	<U>M</> <sup>m</sup><sup>Ma</sup><sup>na</sup><sup>pa</sup><sup>du</sup> <sup>ARAD</sup><sup>KA</sup><sup>MA</sup>
3	16774 At present all is well in the country	kāša-kan ŠA KUR-TI hūman SIG5-in	2	[ka-a-ša-kan ŠA KUR-TI hū-ma-an SIG5-in	271	3	<ka>a</><sup>ša</sup><sup>kan</sup> <sup>ŠA</sup> <sup>KUR</sup><sup>TI</sup> <sup>hū</sup><sup>ma</sup><sup>an</sup> <sup>SIG5</sup><sup>in</sup>	<ka>a</><sup>ša</sup><sup>kan</sup> <sup>ŠA</sup> <sup>KUR</sup><sup>TI</sup> <sup>hū</sup><sup>ma</sup><sup>an</sup> <sup>SIG5</sup><sup>in</sup>
4	16775 [...] came [...]	m[...] uit	3	[mo o o o o] ú-it	271	4	<sup>m</sup><sup>[...]</sup> <sup>uit</sup>	<sup>m</sup><sup>[...]</sup> <sup>uit</sup>
5	16776 And they [...] went back to the country of Wiluša ...	n=at [...] an EGIR-pa KURWiluša GUL-uwanzi pāer	3	[na-at o o] x=an EGIR-pa KURW5-lu-ša GUL-u-wa...	271	6	<n>a</><sup>t</sup> <sup>[...]</sup> <sup>an</sup> <sup>EGIR</sup><sup>pa</sup> <sup>KUR</sup><sup>W5</sup><sup>lu</sup><sup>ša</sup> <sup>GUL</sup><sup>u</sup><sup>wa</sup>...	<n>a</><sup>t</sup> <sup>[...]</sup> <sup>an</sup> <sup>EGIR</sup><sup>pa</sup> <sup>KUR</sup><sup>W5</sup><sup>lu</sup><sup>ša</sup> <sup>GUL</sup><sup>u</sup><sup>wa</sup>...
6	16777 And he brought Hittite troops with him	ÉRIN.MEŠ KURHittite-ya uwaiet	3	ÉRIN.MEŠ KURHittite-ya ú-wa-te-et	271	5	&#201;.RIN.MEŠ <sup>KUR</sup><sup>Hittite</sup><sup>ya</sup> <sup>u</sup><sup>wa</sup><sup>te</sup><sup>et</sup>	&#201;.RIN.MEŠ <sup>KUR</sup><sup>Hittite</sup><sup>ya</sup> <sup>u</sup><sup>wa</sup><sup>te</sup><sup>et</sup>
7	16778 I, however, am ill	amuk=ma ištarkezi	3	[am-mu-uk-m]a iš-tar-ek-zi	271	7	<am</><sup>mu</sup><sup>uk</sup><sup>m</sup> <sup>a</sup> <sup>i</sup><sup>š</sup><sup>t</sup><sup>ar</sup><sup>ek</sup><sup>zi</sup>	<am</><sup>mu</sup><sup>uk</sup><sup>m</sup> <sup>a</sup> <sup>i</sup><sup>š</sup><sup>t</sup><sup>ar</sup><sup>ek</sup><sup>zi</sup>
8	16779 I am seriously ill	GIG-zi=ma mu HUL-lu	3	GIG-zi=ma mu HUL-lu	271	8	GIG<sup>zi</sup>=ma mu<sup>mu</sup> <sup>HUL</sup><sup>lu</sup>	GIG<sup>zi</sup>=ma mu<sup>mu</sup> <sup>HUL</sup><sup>lu</sup>
9	16780 Illness has prostrated me!	GIG-aš-mu mekki tamaššan harzi</>	3	GIG-aš-mu [me?ek?ki?] ta-ma-aš-ša-an har-zi	271	9	GIG<sup>a</sup>š<sup>mu</sup> <sup>me</sup><sup>ki</sup><sup>ki</sup> <sup>ta</sup><sup>ma</sup><sup>a</sup>š<sup>ša</sup><sup>an</sup> <sup>har</sup><sup>zi</sup>	GIG<sup>a</sup>š<sup>mu</sup> <sup>me</sup><sup>ki</sup><sup>ki</sup> <sup>ta</sup><sup>ma</sup><sup>a</sup>š<sup>ša</sup><sup>an</sup> <sup>har</sup><sup>zi</sup>
10	16781 When Piyama-radu had humiliated me	mPiyamaraduš=ma=mu GIM-an luriyāta	4	[mPi-ya-m]a-ra-du-uš-ma-mu GIM-an lu-ri-ya-āh-ta	271	10	<sup>m</sup><sup>Piyamaraduš</sup>=ma=mu <sup>GIM</sup><sup>an</sup> <sup>luri</sup><sup>yā</sup>ta	<sup>m</sup><sup>Piyamaraduš</sup>=ma=mu <sup>GIM</sup><sup>an</sup> <sup>luri</sup><sup>yā</sup>ta
11	16782 Set up Atpā over? me	nu=mu=kan maTāpn [...] UGU tittanut	4	nu-mu-kán maT-pa-a-an [o o o] UGU ti-it-ta-nu-ut	271	11	<nu</><sup>mu</sup>=kan <sup>ma</sup><sup>Tāpn</sup> <sup>[...]</sup> <sup>UGU</sup> <sup>ti</sup><sup>t</sup><sup>an</sup><sup>ut</sup>	<nu</><sup>mu</sup>=kan <sup>ma</sup><sup>Tāpn</sup> <sup>[...]</sup> <sup>UGU</sup> <sup>ti</sup><sup>t</sup><sup>an</sup><sup>ut</sup>
12	16783 And attacked (the country of) Lazpa	nu KURLazpan GUL-aḫta	4	nu KURLa-az-pa-an GUL-aḫ-ta	271	12	<nu</> <sup>KUR</sup><sup>Lazpan</sup> <sup>GUL</sup><sup>a</sup>ḫ<sup>ta</sup>	<nu</> <sup>KUR</sup><sup>Lazpan</sup> <sup>GUL</sup><sup>a</sup>ḫ<sup>ta</sup>
13	16784 All of the purple-dyers who were mine	[...] LÚ.MEŠSARIPUTI kuēš kuēš ammel ešer	4	[o o LÚ] MEŠSA-RI-PU-TI ku-e-eš ku-e-eš am-mi...	271	13	<sup>[...]</sup> <sup>LÚ</sup> <sup>MEŠSA</sup><sup>RI</sup><sup>PU</sup><sup>TI</sup> <sup>ku</sup><sup>e</sup><sup>eš</sup> <sup>ku</sup><sup>e</sup><sup>eš</sup> <sup>am</sup><sup>mi</sup>...	<sup>[...]</sup> <sup>LÚ</sup> <sup>MEŠSA</sup><sup>RI</sup><sup>PU</sup><sup>TI</sup> <sup>ku</sup><sup>e</sup><sup>eš</sup> <sup>ku</sup><sup>e</sup><sup>eš</sup> <sup>am</sup><sup>mi</sup>...
14	16785 Without exception they joined with him	nu=šāi=kan hūmanduš=pat anda ḫandaer	4	[nu-uš-šai-kán] hū-u-ma-an-du-uš-pát an-da ḫa-an-d...	271	14	<nu</>=šāi=kan <sup>hū</sup><sup>manduš</sup>=pat anda ḫandaer<sup>an</sup><sup>d</sup>...	<nu</>=šāi=kan <sup>hū</sup><sup>manduš</sup>=pat anda ḫandaer<sup>an</sup><sup>d</sup>...
15	16786 And all of the purple-dyers of His Majesty	ŠA dUTU-ŠI-ya kuēš kuēš ešer [...] LÚ.MEŠSARIPUTI	4	ŠA dUTU-ŠI-ya ku-e-eš [ku-e-eš e-še-er] [o o LÚ] M...	271	15	<sup>ŠA</sup> <sup>dUTU</sup><sup>ŠI</sup><sup>ya</sup> <sup>ku</sup><sup>e</sup><sup>eš</sup> <sup>[...]</sup> <sup>LÚ</sup> <sup>MEŠSA</sup><sup>RI</sup><sup>PU</sup><sup>TI</sup>	<sup>ŠA</sup> <sup>dUTU</sup><sup>ŠI</sup><sup>ya</sup> <sup>ku</sup><sup>e</sup><sup>eš</sup> <sup>[...]</sup> <sup>LÚ</sup> <sup>MEŠSA</sup><sup>RI</sup><sup>PU</sup><sup>TI</sup>
16	16787 Without exception joined with him?	n=at=kan hūmanduš=pat anda ḫandaer	4	na-at-kán hū-ma-an-du-uš-pát an-da ḫa-an-da-er	271	16	<n>a</><sup>t</sup><sup>ka</sup>n <sup>hū</sup><sup>manduš</sup>=pat<sup>an</sup><sup>d</sup><sup>a</sup> <sup>ḫa</sup><sup>an</sup><sup>da</sup><sup>er</sup>	<n>a</><sup>t</sup><sup>ka</sup>n <sup>hū</sup><sup>manduš</sup>=pat<sup>an</sup><sup>d</sup><sup>a</sup> <sup>ḫa</sup><sup>an</sup><sup>da</sup><sup>er</sup>
17	16790 And xx-huha, the domestic end table man	[...] hūbaš kuš LÚAMATU LÚ GIŠBANSUR	4	[o o o (-?)] hū-ha-aš ku-š LÚAMA A TU LÚ <sup>GIŠ</sup><sup>B...</sup>	271	17	<sup>[...]</sup> <sup>hū</sup><sup>baš</sup> <sup>kuš</sup> <sup>LÚ</sup><sup>AMATU</sup> <sup>LÚ</sup> <sup>GIŠ</sup><sup>BANSUR</sup>	<sup>[...]</sup> <sup>hū</sup><sup>baš</sup> <sup>kuš</sup> <sup>LÚ</sup><sup>AMATU</sup> <sup>LÚ</sup> <sup>GIŠ</sup><sup>BANSUR</sup>
18	16791 [...] to Kupanta-LAMMA [...]	[...] ANA mKupa-na-ta dLAMMA [...]	4	[... A-N]A mKu-pa-na-ta dLAMMA [...]	271	47	<sup>[...]</sup> <sup>ANA</sup> <sup>m</sup><sup>Ku</sup><sup>pa</sup><sup>na</sup><sup>ta</sup> <sup>dLAMMA</sup> <sup>[...]</sup>	<sup>[...]</sup> <sup>ANA</sup> <sup>m</sup><sup>Ku</sup><sup>pa</sup><sup>na</sup><sup>ta</sup> <sup>dLAMMA</sup> <sup>[...]</sup>
19	16792 [...] he shallid [...] the basket-weavers	[...] LÚAD KID-ta-ra-aš-wa<sup>ra</sup>-aš-kán [...]	4	[...] LÚAD KID-ta-ra-aš-wa<sup>ra</sup>-aš-kán [...]	271	46	<sup>[...]</sup> <sup>LÚ</sup><sup>AD</sup> <sup>KID</sup><sup>ta</sup><sup>ra</sup><sup>a</sup>š<sup>wa</sup><sup>ra</sup><sup>a</sup>š<sup>kán</sup> <sup>[...]</sup>	<sup>[...]</sup> <sup>LÚ</sup><sup>AD</sup> <sup>KID</sup><sup>ta</sup><sup>ra</sup><sup>a</sup>š<sup>wa</sup><sup>ra</sup><sup>a</sup>š<sup>kán</sup> <sup>[...]</sup>
20	16793 "To Atpā I did write about the purple-dyers!"	ANA maT-pa-a-wa [...] AŠPUR	4	A-NA maT-pa-a-wa [...] AŠ-PUJR	271	45	<sup>ANA</sup> <sup>ma</sup><sup>T</sup><sup>pa</sup><sup>a</sup><sup>wa</sup> <sup>[...]</sup> <sup>AŠ</sup><sup>PUR</sup>	<sup>ANA</sup> <sup>ma</sup><sup>T</sup><sup>pa</sup><sup>a</sup><sup>wa</sup> <sup>[...]</sup> <sup>AŠ</sup><sup>PUR</sup>
21	16794 "Write to Atpā about the purple-dyers!"	ANA maT-pa-a-wa [...] ŠUPUR	4	[A-NA maT-pa-a-wa [...] ŠU?PUJR	271	44	<sup>ANA</sup> <sup>ma</sup><sup>T</sup><sup>pa</sup><sup>a</sup><sup>wa</sup> <sup>[...]</sup> <sup>ŠU</sup><sup>PUR</sup>	<sup>ANA</sup> <sup>ma</sup><sup>T</sup><sup>pa</sup><sup>a</sup><sup>wa</sup> <sup>[...]</sup> <sup>ŠU</sup><sup>PUR</sup>
22	16795 "What you said to me"	tuk=wa=mu kuit TÁQBI	4	tu-uk[=wa-mu ku-it TÁQ]BI	271	43	<sup>tuk</sup>=wa=mu <sup>ku</sup>it <sup>TÁQ</sup>BI	<sup>tuk</sup>=wa=mu <sup>ku</sup>it <sup>TÁQ</sup>BI
23	16796 "We did"	nu=wa iyawen	4	[nu-wa i-ya-u-en]	271	42	<sup>nu</sup>=wa <sup>i</sup><sup>ya</sup><sup>u</sup><sup>en</sup>	<sup>nu</sup>=wa <sup>i</sup><sup>ya</sup><sup>u</sup><sup>en</sup>
24	16797 And Kupanta-LAMMA wrote to me as follows	nu=mu mKu-pa-a]n-ta d[LAMMA-aš kiššan iŠPUR	4	nu-mu mKu-pa-a]n-ta d[LAMMA-aš kiš-ša-an iŠ-?P...	271	41	<sup>nu</sup>=mu <sup>[...]</sup> <sup>Kupa</sup><sup>na</sup><sup>t</sup><sup>a</sup> <sup>d[LAMMA</sup><sup>a</sup>š <sup>kiš</sup><sup>šan</sup> <sup>iŠ</sup><sup>PUR</sup>	<sup>nu</sup>=mu <sup>[...]</sup> <sup>Kupa</sup><sup>na</sup><sup>t</sup><sup>a</sup> <sup>d[LAMMA</sup><sup>a</sup>š <sup>kiš</sup><sup>šan</sup> <sup>iŠ</sup><sup>PUR</sup>
25	16798 He (Atpā) let all (of them) without exception go ...	n=aš hūmanduš=pat arḫa tarneš	4	n]a-aš hū-ma-a]n-du-uš-pát ar]ḫa [tar]n-a-aš	271	40	<n>a</><sup>š</sup> <sup>hū</sup><sup>manduš</sup>=pat<sup>ar</sup>ḫa <sup>tar</sup>neš	<n>a</><sup>š</sup> <sup>hū</sup><sup>manduš</sup>=pat<sup>ar</sup>ḫa <sup>tar</sup>neš
26	16799 And the purple-dyers who belong to the gods (...)	nu LÚ.MEŠSARIPUTI ŠA DINGIR.MEŠ kuēš ŠA dU...	4	nu LÚ.MEŠSA-RI-PU-TI ŠA DINGIR.MEŠ ku-e-eš Š...	271	39	<sup>nu</sup> <sup>LÚ</sup> <sup>MEŠSA</sup><sup>RI</sup><sup>PU</sup><sup>TI</sup> <sup>ŠA</sup> <sup>DINGIR</sup>.MEŠ <sup>ku</sup><sup>e</sup><sup>eš</sup> <sup>ŠA</sup> <sup>dU</sup>...	<sup>nu</sup> <sup>LÚ</sup> <sup>MEŠSA</sup><sup>RI</sup><sup>PU</sup><sup>TI</sup> <sup>ŠA</sup> <sup>DINGIR</sup>.MEŠ <sup>ku</sup><sup>e</sup><sup>eš</sup> <sup>ŠA</sup> <sup>dU</sup>...
27	16800 "Let them go home!"	nu=war=aš arḫa tarini	4	[nu-wa-ra-aš ar]ḫa tar-ni	271	38	<sup>nu</sup>=wa<sup>ra</sup>=aš <sup>ar</sup>ḫa <sup>tar</sup>ni	<sup>nu</sup>=wa<sup>ra</sup>=aš <sup>ar</sup>ḫa <sup>tar</sup>ni
28	16801 "The purple-dyers of His Majesty who are there..."	LÚ.MEŠSARIPUTI]H I-A-wa kuēš ŠA dUTU-ŠI a]pi	4	[LÚ.MEŠSA-RI-PU-TI]H I-A-wa ku-e-eš ŠA dUTU-ŠI]...	271	37	<sup>LÚ</sup> <sup>MEŠSA</sup><sup>RI</sup><sup>PU</sup><sup>TI</sup>]H I-A-wa <sup>ku</sup><sup>e</sup><sup>eš</sup> <sup>ŠA</sup> <sup>dUTU</sup><sup>ŠI</sup>]...	<sup>LÚ</sup> <sup>MEŠSA</sup><sup>RI</sup><sup>PU</sup><sup>TI</sup>]H I-A-wa <sup>ku</sup><sup>e</sup><sup>eš</sup> <sup>ŠA</sup> <sup>dUTU</sup><sup>ŠI</sup>]...
29	16802 Kupanta-LAMMA sent a message to Atpā	mKupa-na-ta d[LAMMA-aš-m]e ANA maT-pā iŠPUR	4	mKu-pa-na-ta d[LAMMA-aš-m]e A-NA maT-pa-a iŠ-?P...	271	36	<sup>m</sup><sup>Ku</sup><sup>pa</sup><sup>na</sup><sup>ta</sup> <sup>d[LAMMA</sup><sup>a</sup>š<sup>m</sup>]e <sup>ANA</sup> <sup>ma</sup><sup>T</sup><sup>pa</sup> <sup>iŠ</sup><sup>PUR</sup>	<sup>m</sup><sup>Ku</sup><sup>pa</sup><sup>na</sup><sup>ta</sup> <sup>d[LAMMA</sup><sup>a</sup>š<sup>m</sup>]e <sup>ANA</sup> <sup>ma</sup><sup>T</sup><sup>pa</sup> <sup>iŠ</sup><sup>PUR</sup>

Query executed successfully.



# Формат xls-файлов для выгрузки данных в MsSQL

Для аннотации и выгрузки информации в базу данных и на сайт используется Excel — просто и доступно

Clause Metadata				Clause Table			Clause annot						
Brokenness	Paragraph	Lines	Clause index	syllables	normalized spelling	English translation	1						
3	1	obv. 1	YY1	[A-NA <sup>d</sup> UTU-ŠI EN-Y]A <sup>7</sup> QI-BI-MA	!ANA <sup>d</sup> UTU-ŠI EN-YA QI-BI-MA	To His Majesty, my lord, speak!	V						
2	1	obv. 1	YY2	!U[M-M]A <sup>7</sup> Ma-na-pa- <sup>d</sup> U ARAD-KA-MA	!UMMA <sup>7</sup> Manapa- <sup>d</sup> U ARAD-KA-MA	Thus says Manapa-Tarhunta, your servant!	SV						
2	2	obv. 2	YY3	[ka-a-ša-kán ŠĀ KUR-]TI ħu-u-ma-an SIG <sub>3</sub> -in	!kāša=kan ŠĀ KUR-TI ħūman SIG <sub>3</sub> -in	At present all is well in the country	SV						
4	3	obv. 3	YY4	[ <sup>m</sup> o o o o o o] ū-it	! <sup>m</sup> [...] uit	[...] came [...]	V						
1	3	obv. 3	YY5	ÉRIN.MEŠ <sup>KUR</sup> Ĥat-ti-ya ū-wa-te-et	ÉRIN.MEŠ <sup>KUR</sup> Ĥatti-ya uwatet	And he brought Hittite troops with him	OV						
3	3	obv. 4	YY6	[na-at o o o]x-an EGIR-pa <sup>KUR</sup> Wi-lu-ša GUL-u-wa-an-zi pa-a-er	!n=at [...]an EGIR-pa <sup>KUR</sup> Wiluša GUL-uwanzi pāer	And they [...] went back to the country of Wiluša in order to attack (it)	V						
3	3	obv. 5	YY7	[am-mu-uk-m]a iš-tar-ak-zi	!ammuk=ma ištarkzi	I, however, am ill	SV						
1	3	obv. 5	YY8	GIG-zi-ma-mu ĤUL-lu	GIG-zi=ma=mu ĤUL-lu	I am seriously ill	V						
3	3	obv. 5-6	YY9	GIG-aš-mu [me <sup>3</sup> -ek <sup>3</sup> -ki <sup>3</sup> ] ta-ma-aš-ša-an ĥar-zi	GIG-aš=mu mekki tamaššan ĥarzi	Illness has prostrated me!	SV						
2	4	oi	A	B	C	D	E	F	G	H	I	J	
3	4	oi	Metadata				Word form				Syntax		
1	4	oi	2	Text index	Line	Clause index	Word Form ind	Syllables	Normalized spelling	Translation	PoS	Constituent	Synt.Role
			3	XX1	obv. 1	YY1	ZZ1	[A-NA	!ANA	to	P	PP	IO
			4	XX1	obv. 1	YY1	ZZ2	! <sup>d</sup> UTU-ŠI	! <sup>d</sup> UTU-ŠI	My Majesty	N	NP	
			5	XX1	obv. 1	YY1	ZZ3	EN-Y]A <sup>7</sup>	EN-YA	My Lord	N		
			6	XX1	obv. 1	YY1	ZZ4	!QI-BI-MA	!QI-BI-MA	speak	V	VP	V
			7	XX1	obv. 1	YY2	ZZ1	!U[M-M]A	!UMMA	thus	Adv	AdvP	Adv
			8	XX1	obv. 1	YY2	ZZ2	! <sup>m</sup> Manapa- <sup>d</sup> U	! <sup>m</sup> Manapa- <sup>d</sup> U	Manapa-Tarhunta	N	NP	S
			9	XX1	obv. 1	YY2	ZZ3	ARAD-KA-MA	ARAD-KA-MA	your servant	N		
			10	XX1	obv. 2	YY3	ZZ1	[ka-a-ša	!kāša	behold	Adv		
			11	XX1	obv. 2	YY3	ZZ2	-kán	!=kan	LOC.ENCL	Part	CliticP	
			12	XX1	obv. 2	YY3	ZZ3	ŠĀ	ŠĀ	in	P	PP	IO
			13	XX1	obv. 2	YY3	ZZ4	KUR-]TI	KUR-TI	country	N	NP	
			14	XX1	obv. 2	YY3	ZZ5	ħu-u-ma-an	!ħūman	all	Adj	NP	S
			15	XX1	obv. 2	YY3	ZZ6	SIG <sub>3</sub> -in	SIG <sub>3</sub> -in	well	Adv	VP	V
			16	XX1	obv. 3	YY4	ZZ1	[ <sup>m</sup> o o o o o o]	! <sup>m</sup> [...]	[...]	N	NP	S
			17	XX1	obv. 3	YY4	ZZ2	ū-it	!uit	come.3SG.PRT	V	VP	V
			18	XX1	obv. 3	YY5	ZZ1	ÉRIN.MEŠ	ÉRIN.MEŠ	troops	N	NP	O
			19	XX1	obv. 3	YY5	ZZ2	! <sup>KUR</sup> Ĥat-ti	! <sup>KUR</sup> Ĥatti	Hittite	N		
			20	XX1	obv. 3	YY5	ZZ3	-ya	!=ya	and.CONJ.ENCL	Conj		
			21	XX1	obv. 3	YY5	ZZ4	ū-wa-te-et	!uwatet	bring.3SG.PRT	V	VP	V

# Автоматическая токенизация

Все клаузы при выгрузке в MsSQL проходят процедуру автоматического разделения на слова (клитки выделяются как отдельные токены).

The screenshot shows the HittiteCorpus website interface. A popup window titled "Morfology" is open, displaying a table with the following data:

[ka-a-ša	<i>kāša</i>	behold
-kán	= <i>kan</i>	LOC.ENCL
ŠÀ	ŠÀ	in
KUR-]T/	<i>KUR-T/</i>	country
ḫu-u-ma-an	<i>ḫūman</i>	all
SIG <sub>5</sub> -in	<i>SIG<sub>5</sub>-in</i>	well

The background interface shows the "HittiteCorpus Database search" page. It includes search options for "KUB 19.5 + KBo 19.79" and "Paragraph" vs "Lines". A "Search" button is visible. Below the search options, there is a list of query results for "1 texts". The first result is "KUB 19.5 + KBo 19.79 NH/NS edit". The second result is "1 obv. 1 morfology" with the text: "U[M-M]A ṹMa-na-pa-ṹU ARAD-KA-MA UMMA ṹManapa-ṹU ARAD-KA-MA Thus says Manapa-Tarhunta, your servant!". The third result is "2 obv. 2 morfology" with the text: "[ka-a-ša-kán ŠÀ KUR-]T/ ḫu-u-ma-an SIG<sub>5</sub>-in *kāša=kan ŠÀ KUR-T/ ḫūman SIG<sub>5</sub>-in* At present all is well in the country". The fourth result is "3 obv. 3 morfology" with the text: "[ṹo o o o o ] ú-it ṹ[... ] *uit* [...] came [...]".



# Проблема #1: транслитерация

Хеттский текст может содержать:

- слова и фонетические компоненты на хеттском языке;
- аккадограммы (слоговая запись на аккадском языке);
- шумерограммы (логограммы на шумерском языке).
- они могут комбинироваться в одном слове, напр.:  
аккадограмма + фонетический компонент на хеттском;  
шумерограмма + фонетический компонент на хеттском;  
шумерограмма + аккадограмма + фонетический компонент на хеттском, и т. п.
- составляющие слова, записанные на разных языках, разделяются дефисом:  
EN-YA, Si×SÁ-at.

# Проблема #1: транслитерация

Принципы транслитерации:

- хеттские слова и комплементы: **курсив, строчные** — *ḫūtan*
- аккадограммы: **курсив, прописные** — *ANA*
- шумерограммы: **прописные** — *ÉRIN.MEŠ*
- детерминативы: **верхний регистр** — *<sup>m</sup>Manara-dU*
- индексы шумерских и аккадских клинописных знаков, необходимые для различения омофонов: **нижний регистр** — e.g. *SIG<sub>5</sub>*
- также в транслитерации для различения омофонов используются **диакритики**; некоторые диакритические знаки традиционно используются для аккадской и хеттской транскрипции: *ŠÀ, Ḫ, Ṣ̌*.

# Принципы поддержки транслитерации в корпусе

- Все форматирование текста делается в программах Word и Excel, используемых для обработки сырого текста — встроенные инструменты позволяют сделать это без применения html-разметки.
- Специальный скрипт в базе данных форматирует тексты при загрузке в MsSQL — автоматически добавляя в нужных местах html-теги (экономим время на html-разметке).
- Результаты помещаются в отдельные поля (<FormattedSyllables>; <FormattedNormalizedSpelling>)
- Поисковый механизм использует поля без форматирования (поиск облегчается благодаря отсутствию тегов); текст публикуется на сайте с форматированием.

# Проблема #2: лакуны

- клауза как базовый элемент корпуса
- синтаксическая разметка битого текста: вводим уровни битости, 1–5
- разметка «субъект— объект — глагол» (до уровня битости 4):  
V, OV, SV, SOV, OSV
- разметка структуры составляющих (до уровня 3):  
битые фрагменты = [...];  
ЕСЛИ **ясно из контекста**, что [...] содержит глагол, связку, субъект, объект, ТО фрагмент отмечается соответственно;  
ЕСЛИ [...] явно содержит косв. дополнение или наречие, ТО фрагмент отмечается как зависимый от глагола.

# Уровни битости текста

- 1. полностью целый или восстанавливаемый по дубликатам** (квадратные + круглые скобки):  
e.g. KUB 14.3 iv 49  
SAG.DU-*an ku-ra-an-du* (Hoffner 2009:312)  
Let them cut off his head!
- 2. битый, но восстанавливаемый по контексту:**  
e.g. KUB 14.3 iv 50–51  
[SAG.DU-*an-m*]*a ku-in ku-ra-an-zi* (Hoffner 2009:312)  
And the head that they cut off...
- 3. битый с сохранением ключевых составляющих, структура фразы ясна, процент невозможного текста меньше половины фразы:**  
e.g. KUB 19.55 obv. 3–4  
[*A-BU-KA-ma* ...] [Z]AG.MEŠ-*YA i-la-liš-ke-[et ...]* (Hoffner 2009:317)  
But your father [...] was coveting my border territories (had always desired my border territories)...
- 4. сильно битый текст, порядок слов нельзя восстановить с уверенностью, но значение фразы восстанавливаемо по контексту:**  
e.g. KUB 19.55 obv. 22  
[x-x]x-mu-za le?-e? i[-la?-li?-ya?-ši?...] (Hoffner 2009:317)  
You shall not desire? my (land...)
- 5. сильно битый, предложение можно использовать только для фиксации отдельных слов, офрографических и статистических целей (напр., “какова частотность клауз с nu”):**  
e.g. KUB 19.55 obv. 24  
[...] *A-BU-KA ku-w[a-pí ...]*  
(...) your father whe(n...)

# Проблема #3: парсинг текста

маркеры границ клаузы:

- фразовый коннектор *ni*

ЕСЛИ *ni* ( $ni=$ ,  $n=a..$ ,  $n=e$ ) ТО отмечаем <next clause>

ПЕРЕД *ni*

- цепи ваккернагелевских клитик

ЕСЛИ CliticP (список всех возможных вариаций цепочек клитик) ТО отмечаем <next clause> ПЕРЕД CliticP

- Глагол в конце клаузы

ЕСЛИ Verb ТО ?отмечаем <next clause> ПОСЛЕ Verb

# Проблема #4:

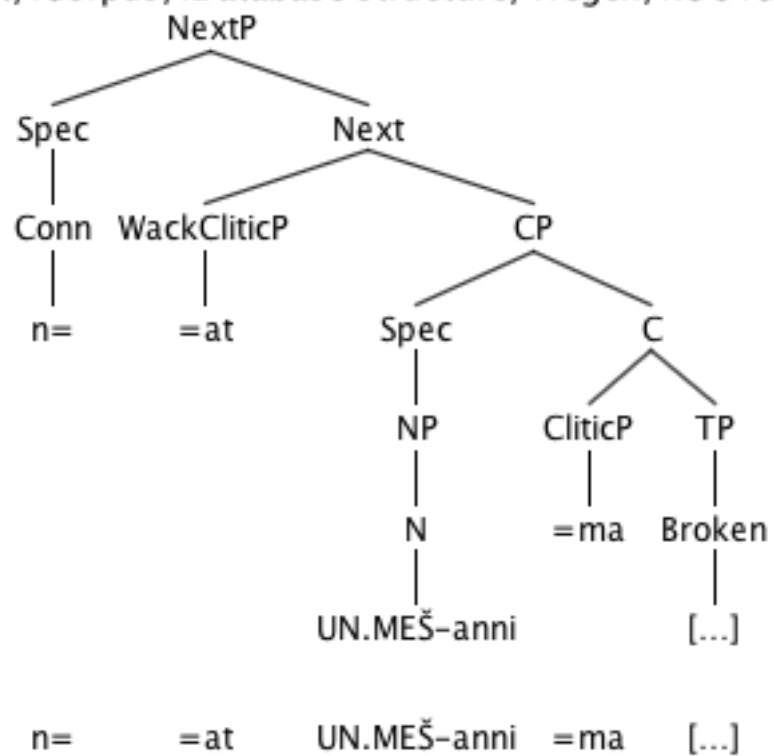
## синтаксическая разметка

- Структура фразы: дерево составляющих
- Теги: набор тегов стандартен (NP, VP, TP, CP, PP, AdvP) с добавлением вспомогательных тегов, необходимых для работы с анатолийскими языками и для разметки информационной структуры: CliticP, ForceP–TopP–FocP–FinP, SpeechActP
- ПО для визуализации синтаксических деревьев: open source Stanford Tregex
- В процессе разработки: встроенный генератор деревьев на сайте





an/!Current/!Corpus/!Database structure/Tregex/ROOT5.txt



# Проблема #5: разметка информационной структуры

- Фокус отмечается как **bold**
- Семантические нюансы фокусов в корпусе не отражены, только общее деление фокус–топик
- *Focus skeleton* — метод определения фокуса

# Focus skeleton

- Каждый контекст сравнивается с предыдущим и последующим текстом в поиске **набора альтернатив (закрытого или открытого)** и **новой / старой** информации.
- Дается перечень **всех возможных вариантов** для данной пропозиции.

E.g. НКМ 17 obv. 12

*nu=ttā*

*kinun=pat ištamašš-er*

CONN=PRON.2SG now=PTCL hear-3PL.PST

*'It is now that they have heard of you'*

**Предыдущий контекст:** “Когда я отправил тебя, Хулла, (прошлой) зимой, (враги) тогда о тебе не слышали”.

**Focus skeleton:** враги слышали о тебе [зимой, весной, летом, осенью, тогда, сейчас] — закрытый ряд альтернатив, предполагающий контраст.

Поскольку присутствует упоминание о том, что «тогда не слышали», focus skeleton = “теперь, а не тогда”, — **контрастивный замещающий фокус** на *kinun=pat*.

# Annotated Corpus of Hittite clauses (Letters, Instructions and Prayers)

Dr. Andrej V. Sideltsev  
Maria Molina  
Contact us: maria.molina@me.com  
Russian Academy of Sciences  
Institute of linguistics  
Department of Anatolian and Celtic Studies

The aim of the project is to develop an online syntactically annotated corpus of Hittite, a dead Indo-European language of 18–12 cc. BC attested in cuneiform script on clay tablets.

Nowadays new electronic and online corpora for different languages emerge every new day. Hittite remains practically the only major Indo-European language with a significant corpus of texts that does not have any electronic syntactically annotated corpus. As the oldest attested Indo-European language, Hittite proves to be more and more interesting for the researchers. A lot of papers on its linguistics, including syntax and morphology, have been published during the last decade, so the need of an online Hittite corpus is more and more compelling.

The project of a syntactically annotated Hittite corpus started in Moscow in May 2014 at the Institute of linguistics (RAS, Moscow). It is based on two publications, Hittite letters by [Hoffner 2009] and Hittite instructions [Miller 2013]. Hittite prayers are planned to be included as well. All letters and instructions have been analyzed and digitalized according to the project guidelines, and a database has been built using FileMakerPro 13 shell. Starting from January 2016, a MySQL relational database system is being developed for the corpus, with an online search interface and limited public access to the database materials.

The basic element of the database is a clause (a simple sentence). By May 2016 the on-going work on clause database is to be finished, the amount of clauses aiming at around 5000 entries. All linguistic material then is planned to be parsed, tokenized, and lemmatized. By the start of 2017 all clauses available online should be annotated for focus-comment and word order (SOV). A treebank is planned to be built according to Penn Treebank standards.