

SENIE

latviešu valodas seno tekstu korpuss
korpus früher lettischer texte
the corpus of early written latvian texts

The Corpus of Early Written Latvian (XVI – XVIII c.)



Everita Andronova
*Institute of Mathematics and Computer Science,
University of Latvia*
everita.andronova@lumii.lv

- Latvian diachronic corpus among others
- The Corpus of Early Written Latvian
(XVI – XVIII c.)
- Lessons we have learned
- Historical Dictionary of Latvian
(XVI – XVII c.)
- Future plans

Old German Reference Corpus (750–1050)

Reference Corpus Middle High German (1050–1350)

ReN - Referenzkorpus Mittelniederdeutsch/Niederrheinisch (1200-1650)

Reference Corpus Early Modern New High German (1350–1650)

GerManC (1650-1800)

Korpus tekstów staropolskich do roku 1500

Korpus polszczyzny XVI wieku

Elektroniczny korpus tekstów polskich z XVII i XVIII w. (do 1772 r.)

KorBa (KORpus BARokowy) – "The Electronic Corpus of the 17th and 18th century Polish (Korpus tekstów polskich z XVII i XVIII w.)"

IMPACT GT corpus (Korpus GT projektu IMPACT)

Church-Slavonic

Middle Russian

Old Russian

Birchbark corpora

The Regensburg Russian Diachronic Corpus (RRuDi)

The Corpus of the Velikie Minei Čet'i

The "Manuscript" corpus

The Corpus of Northern Russian Hagiographical Texts

Diakorp

Old Hungarian Corpus (XII–XVII c.)

Corpus of Old Literary Finnish (1543–1810)

Corpus of Old Written Estonian

<http://seniejirastai.lki.lt/>

The Old Lithuanian Corpus SLIEKKAS

HIPKON

XV century New Testament translations

Corpus of Old and Middle Hungarian court records and private correspondence

Saarbrücken Cookbook Corpora (1569—1800)

Main trends

- There are languages with more than 4 diachronic corpora available (Russian, Polish, German); more common practice – there is just one diachronic corpus
- Diachronic corpus is a sub-corpus of the National corpus (Czech, Russian)
- Corpora may consist of sample selections (GerManC, HIPKON) or the whole texts
- The bulk of corpora comprises transliterated texts, although some corpora are ‘diplomatic transcribed’ (ReN), partly normalised to modern spelling (Old Hungarian) or transcribed (ReF, Diakorp)
- Apart from (rich) structural mark-up many corpora are POS tagged, some of them are annotated using modern language processing tools (Diakorp, Old Estonian)
- Representativeness or balance is not a criteria applied for design of diachronic corpora (apart from GerManC and new version of Diakorp to come)
- Although a diachronic corpus itself refers to a special one, there are special corpora among them as well (Saarbrücken Cookbook Corpora (1569—1800), Corpus of Old and Middle Hungarian court records and private correspondence, HIPKON (sermons in Upper German (Bavarian-Alemannic))

Prehistory of Latvian Corpus

- .. *the earliest texts were “written incorrectly (by Germans!)” and the language in the texts is “full of mistakes”* (Endzelīns, 1951: 22, 20)
- 1930s Jānis Endzelīns: ***Thesaurus linguae letticae*** — all the Latvian words should be collected both from spoken language and written texts (Endzelīns 1933: 818)
- K. Mühlenbachs *Lettisch-deutsches Wörterbuch* (1923–32; revised and supplemented by J. Endzelīns and E. Hauzenberga 1934–1946)
- V. Rūķe-Draviņa: proposal to develop ***A dictionary of the early written texts*** (Rūķe 1961)
- idea to start work on ***The historical dictionary of the Latvian language*** (Baldunčiks 1994)
- **1995–96** — database of early written Latvian texts was initiated and several early printed Latvian sources (XVII c.) were digitalised at IMCS, UL
- “FrakturaSpecial” fonts (Latvian software company Tilde, **1996**)
- **2001** prof. Trevor Fennell invited and encouraged scholars of Latvia to start work on the *Dictionary of Old Latvian*.

The Corpus of Early Written Latvian (www.korpuss.lv/senie)

- launched in 2003 (collaboration between IMCS, UL, Institute of the Latvian Language, UL, Faculty of Humanities, UL)
- covers written sources (both printed books and manuscript transcripts) from the 16th–18th c., texts included *in toto*
- word lists, frequency lists, inverse vocabularies and a concordancer available
- navigation within corpus
- structural marking only
- no POS tagging, no lemmatization
- new sources added constantly

Selection of texts

“Gesamtkatalog Die Älteren Drucke in Lettischer Sprache (1525–1855)”

Task — to cover the texts of the 16th–18th century

Decision in 2002 — to include the first editions only

F(ar f)uture plans — all printed sources should be presented

Challenge — search for the (only) copy; close cooperation with libraries

Size

Decision — to include source *in toto*

Observations — influence of one particular author’s language in general corpus; text fragments commuting from source to source

Representativity — should only printed texts be included?

Decision — to add some of available manuscript dictionary transcripts and documents

Annotation

Rich structural mark-up (to exclude words of foreign origin from word indices)

The Corpus of Early Written Latvian (www.korpuss.lv/senie)

Earliest texts – Lord's Prayers of the 16th c. (study of numerous copies in books printed in Latin or German), the 1st printed book only in 1585

Mostly **religious** texts (catechism, pericopes, hymnals, Bible translation, sermons..)

Lexicographical sources («Lettus» (1638), manuscript transcripts (Fuerecker's dictionary) from mid-17th.c.), Elger's dictionary (1683) is in progress

Limited number of **secular** texts (statutes of the linen-weavers, court-martial laws, oaths, dedication poetry)

The screenshot shows a web browser window displaying the SENIE website. The URL is senie.korpuss.lv/source.jsp?codificator=VLH1685_Syr. The page title is "SENIE latviešu valodas seno tekstu korpuss". Below the title, there is a search bar and a list of search results for "Das Haus=, Zucht= vnd Lehrbuch Jesus Syrachs (VLH1685_Syr)". The results include:

- bibliogrāfija
- statisks indekss un teksts
- vārdformu biežuma saraksts (CS): [top 1000] [pilns]
- vārdformu biežuma saraksts (LC): [top 1000] [pilns]
- vārdformu indekss (CS): [pilns]
- vārdformu indekss (LC): [pilns]
- oriģināla faksimils
- Unicode

At the bottom of the page, there is a footer with the following text:

LU Matemātikas un informātikas institūta Mākslīgā intelekta laboratorija
 LU Humanitāro zinātņu fakultātes Baltu valodu katedra
 LU Latviešu valodas institūts
 2002-2018

On the right side of the footer, there is a Windows watermark: "Activate Windows Go to Settings to activate Windows."

The Windows taskbar at the bottom shows the time as 4:43 PM on 7/30/2018.

The Corpus of Early Written Latvian (www.korpuss.lv/senie)

Corpus (in 2016):

Century	Sources	Occurrences
16th	11	57 476
17th	45	935 510
18th	18	95 500
Total	74	1 088 486

1) textological studies are required

- only the first editions of the books are included; but: nothing should be left
- search for copies (G.Elger's hymnals in Vilnius, G.Dressel's Catechism in Halle, Agenda Parva in Poland)
- the border line in time – when does the period of Old Latvian end? – mid- 19th c. would be more precise

2) wide (spelling, morphological) variation

- variant detector would be nice (then we could apply Modern Latvian tools for language processing created at IMCS, UL)

māja (142) *s. f.* debbes=**mahju** (2), mahja (3), mahjahm (6), mahjahs (4), mahjahß (28), mahjams (1), mahjas (5), majahm (1), mahja^ (2), mahja^s (20), mahju (8), **mahya** (1), mahyahß (6), mahyahs (3), mahyu (1), **maias** (2), maie (1), **majahs** (3), majahß (11), maja` (1), maja^ (1), **mayas** (18), mayaas (2), mayahs (4), mayahß (4), mayäs (1), mayo (2)

The Corpus of Early Written Latvian (www.korpuss.lv/senie)

3) transcription of Gothic script

- plain text with combined symbols for some special characters, not very user-friendly (users hate ξ used for f)
- switch to Unicode is in progress (started in 2017, Old Testament will be converted in 2018), converted files are offered in a separate entry to use for publications

**1: Pinnekļo[^]s/ in tawu Kaklu
wiņņas**

2: Kakla=dξelī[^]s.

**3: 26. Lohki tawus Kammeēšchus/
in**

**4: neśš to/ in ne turrees pretti
wiņņas**

5: Šaitehm.

**6: 27. Turrees pee tahs no
wiśša[^]s**

**Pinnekļô^s/ in tawu Kaklu
wiņņas**

Kakla=dfelfîs.

**26. Lohki tawus
Kammeefchus/ in**

**nefs to/ in ne turrees pretti
wiņņas**

Šaitehm.

**27. Turrees pee tahs no
wiffâs**

The Corpus of Early Written Latvian (www.korpuss.lv/senie)

4) transcription of manuscripts

- often quite chaotic usage of small and capital letters

5) involving students in text processing

- experience at UL, at St. Petersburg University
- individual research

Special thanks to E. Kazakēnaitē – VLH1685

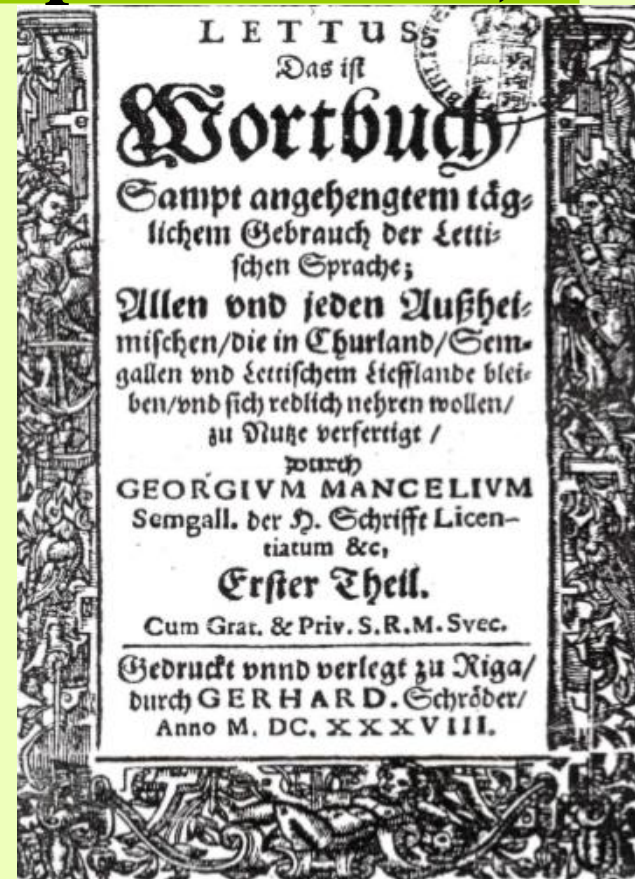
6) facsimiles

- co-operation with LNB, digital copies
- users appreciate them



7) corpus management system

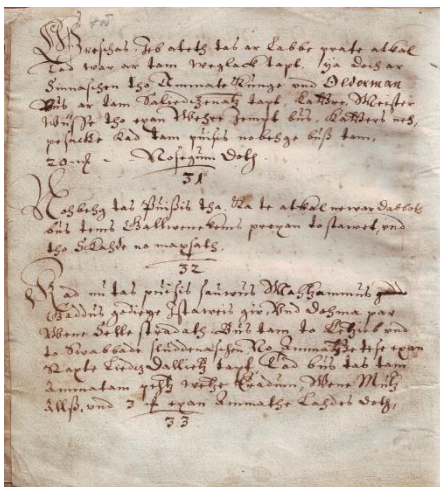
- N. Grūzītis, IMCS, 2002
- crucial step – modernisation (L. Pretkalniņa, IMCS, 2019 (?))



SENIE

latviešu valodas seno tekstu korpuss
korpus früher lettischer texte
the corpus of early written latvian texts

The Corpus of Early Written Latvian (www.korpuss.lv/senie)



Latviešu valodas vēsturis | www.korpuss.lv/senie/sta | Linaudēju šrāga #13

← → C www.korpuss.lv/senie/static/LS1625.html

Linaudēju šrāga

[LS1625](#) [SĀKUMLAPA](#)

Vārdformu skaits: 1562
Vārdlietojumu skaits: 5181

abbe
1 | 1 | [LS1625 13v.lpp. 19.r.](#)
1

abbems
1 | 1 | [LS1625 12v.lpp. 6.r.](#)
1 | 1 | [LS1625 17v.lpp. 11.r.](#)
2

abbeššchen
1 | 1 | [LS1625 21v.lpp. 1.r.](#)
1

abdohmašchen
1 | 1 | [LS1625 16v.lpp. 11.r.](#)
1

abdomath
1 | 1 | [LS1625 17r.lpp. 1.r.](#)
1

abdšerth
1 | 1 | [LS1625 11v.lpp. 18.r.](#)
1

22: 30.

7v. lpp.
1: Greščas jeb ateth tas ar labbe prate atkal
2: tad war ar tam weglack tapt, ya doch ar
3: šimnašchen tho ammatte kunge vnd olderman
4: bus ar tam šaliedczenatz tapt, katre meišter
5: wušše tho exan wehre jempt bus, katters neh
6: pešacke kad tam pušis nobehge buš tam,
7: 20 mk nošegum doth.
8: 31
9: Nohbehg tas pušis tha ka te atkal newar dabboth
10: bus tems gallwenekems prexan to štawet, vnd
11: tho škahde no maxšath.
12: 32
13: Kad nu tas pušis šauvus mahtzammus
14: gaddus godiege jštaweis gir, vnd dolhma par
15: wene šelle štradath, bus tam to letzieb vnd
16: to šwabade šluddenašchen no ammatte teše exan
17: raxte liedczdallietz tapt, tad bus tas tam
18: ammatam pelitz wetze eradum, wene mutz
19: allš, vnd 3 mk exan ammatte lahdes doth.
20: 33.

8r. lpp.
1: Grib tas wehl jlgack par allgepuše py šauwe
2: meišter jeb meišterenes štradath, gir tems walle

Linaudeju_sraga_13.jpg ^

Rādīt visu X

Handwritten statutes of linen-weavers of 1625

Applications:

- 1) Studies in the history of the Latvian language (P.Vanags, A.Trumpa, R.Siliņa-Piņķe, A.Frīdenberga, E.Andronova)
- 2) Ph.D. thesis (A.Trumpa (Adjektīvu semantiskā diferenciacija latviešu un lietuviešu valodā (Rīga, 2005)), G.Elksnīte (Nominālās vārdkopas Georga Manceļa tekstos (Liepāja, 2011)), A.Frīdenberga (Nominal Word-Formation in the Works of G.Mancelius' (Rīga, 2016)), on-going studies: J.Wiśnioch (“Verbs E.Glika Bībeles tulkojumā”), E. Kazakēnaitē)
- 3) Use of corpus in teaching process (University of Latvia, Stockholm University, St.Petersburg State University, Liepāja University)
- 4) Aim – to encourage non-linguistics studies (M.Grudule on the history of Latvian hymnals)
- 5) Historical Dictionary of Latvian (2012-2016) (E.Andronova, A.Frīdenberga, R.Siliņa-Piņķe, A.Trumpa, P.Vanags)

The Corpus of Early Written Latvian : feedback

Google results (July, 2018):

Latviešu valodas seno tekstu korpus – ~ 35,000 results

Corpus of Early Latvian – ~ 2 m. results

Seno tekstu korpus SENIE – 50,400 results

Šis promocijas darbs ir pirmais darbs, kurā salīdzināmi vēsturiskā aspektā sistemātiski ir pētīta formāli identisku latviešu un lietuviešu valodas adjektīvu semantiskā diferenciacija (Trumpa 2005:7).

..pētot latviešu valodas vārdu semantisko attīstību, ir izmantots **salīdzinoši vairāk** (..) **seno tekstu** materiālu (galvenokārt balstoties uz **Latviešu valodas seno tekstu korpusā** [LVSTK] ietverto materiālu) (..) (Trumpa 2005: 42)

Promocijas darba mērķis ir apzināt, izpētīt un noskaidrot, kā latviešu valodas sistēmā iekļaujas G. Manceļa tekstos lietotās nominālās vārdkopas. (Elksnīte 2011: 3)

Promocijas darbā izmantotais materiāls iegūts, frontāli ekscerpējot G. Manceļa vārdnīcas un tās pielikumu tekstus. Piemēri no sprediķu grāmatas atlasīti, izmantojot Latvijas Universitātes Matemātikas un informātikas institūta Mākslīgā intelekta laboratorijas (<http://www.ailab.lv>) izstrādāto **projektu** Latviešu valodas seno tekstu korpus. Piemēru atlasei izmantots lietojumrīks konkordance, ar kuras palīdzību, ierakstot meklētājā vārdformu vai tās daļu, atrastas nominālās vārdkopas vārdlietojumu kontekstā (ibid: 8).

The Corpus of Early Written Latvian : feedback

- Pētījuma avots ir visi minētie G.Manceļa darbi. Šobrīd tie visi ir pieejami latviešu valodas seno tekstu **korpusā** (<http://www.korpuss.lv/senie>), kas dod iespēju ekscerpēt un analizēt apjomīgu materiāla daudzumu (Frīdenberga 2016: 5).
- Vārddarināšanas sistēma latviešu rakstu valodas agrīnajā periodā līdz šim nav pietiekami pētīta. Šādu pētījumu **dara iespējamu arī izveidotais seno tekstu korpuss un tā izmantošana**. (ibid:6)
- **Pirmo reizi pētījumā ir aplūkots viss Georga Manceļa darbu korpuss**, konstatēti tajā izmantotie vārddarināšanas paņēmieni un līdzekļi. (ibid:6)
- Balstoties uz šī materiāla, **izdarīt secinājumus par latviešu valodas vārddarināšanas sistēmu 17. gadsimta pirmajā pusē**. (ibid:6)

The Corpus of Early Written Latvian : feedback

- A. Frīdenberga (2016:192) compares the spread of nominal suffixes in G.Mancelius texts and in the Corpus of Modern Latvian (miljons 2.0m) (3.5 m.tokens)

3. tabula

Atvasinājumu lietojuma salīdzinājums G. Manceļa darbos un līdzsvarotā mūsdienu latviešu valodas tekstu korpusā “miljons 2.0 m”

Nr.	Manceļa darbos	Vārdlietojums	Nr.	Mūsdienu korpusā	Vārdlietojums
1.	-īb-	5652	1.	-ums	83 322
2.	-šan-	3612	2.	-ība	72 368
3.	-tāj-	1415	3.	-šana/-šanās	44 307
4.	-nieks/-niece	1294	4.	-tājs/-tāja	16 552
5.	-um-	1266	5.	-nieks/-niece	13 510
6.	-kl-	912	6.	-kl-	11 368
7.	-iņ-	789	7.	-ējs/-ēja	5955
8.	-īt-	186	8.	-iņš/-iņa	5141
9.	-ēj-	177	9.	-ītis/-īte	2341
10.	-āj-	35	10.	-ājs	396

Technical problems

- Collapse of the old server, reconstruction was / is **1000000** long (special thanks to M.Šimkūnas for back-up!)
- Change of web address (senie.korpuss.lv, www.korpuss.lv/senie)

Maintenance

- Lack of English interface

Advertising for a wider public

- Introducing data for non-linguists
- Complains about difficult reading of texts

Only state support

- No international projects so far
- Good practise: co-operation with the Institute of the Estonian Language, a book ‘Common Roots of the Latvian and Estonian Literary Languages’ in 2008)

Aim – to compile a representative pilot of a modern electronic dictionary of the 16th–17th century Latvian.

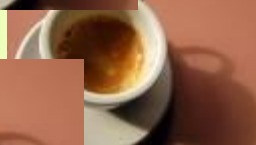
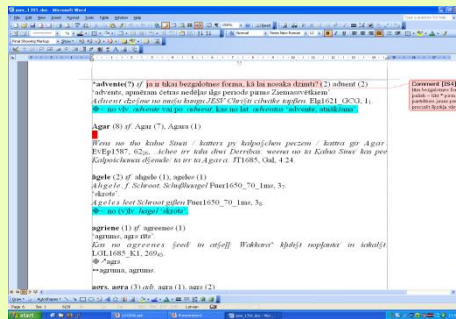
- ca. 1200 entries covering all parts-of speech, different types of meaning explanations
- 40% are proper names
- the oldest and the newest attested example in the period under investigation is given for each sense of a word
- citations from the earliest dictionaries are included as separate field in the entry
- editorial guidelines have been prepared
- *TLex 2013* as a lexicographer's work-bench
- Andronova E., Siliņa-Piņķe R., Trumpa A., Vanags P. The Electronic Historical Latvian Dictionary Based on the Corpus of Early Written Latvian Texts. *Acta-Baltico Slavica 40. Pogranicze bałtycko-słowiańskie w aspekcie leksykalnym i leksykograficznym*. Warszawa: Instytut Slawistyki Polskiej Akademii Nauk, 2016, pp.1–37.



Pre-electronic dictionary (2006-2010)

Lexicographer

- Corpus – word-form index, frequency list, ‘address’ of word-form, examples
- Draft of dictionary entry
- Discussion at the seminar
- Entry in .doc format, published on the website



Entry schema

- helps to stick to the structure of the entry
- automatizes various actions
- supports the export in XML format which in turn was imported in our developed electronic version
- facilitates coordination of team work (5 persons involved)

Lemma: abējāds LemmaSign=abējāds, Frequency=6, Grammar=īpašības vārds

- Utterances
 - WordForm: WordForm=abbeide, Frequency=1
 - WordForm: WordForm=abbejads, Frequency=1
 - WordForm: WordForm=abbejadu, Frequency=1
 - WordForm: WordForm=abbejahds, Frequency=1
 - WordForm: WordForm=abbeyadu, Frequency=2
- Dictionaries
 - Citation
 - Text: Text=%uAbbejads%u der es mit bey"den halt, neuter
 - Source: Source=Fuer1650_70_1ms, Page=1, Line=23
 - Citation
 - Text: Text=%uAbbejahds%u, der es mit beiden hält, neutral
 - Source: Source=Fuer1650_70_2ms, Page=1, Line=20
- Sense: 1 SenseNumber=1
 - Definition: Definition=(tikai daudzskaitlī) abi
 - Examples
 - Example
 - Text: Text=Dod mums tho Thewe patteše adšidt / matce mums arri
 - Source: Source=Ps1615, Page=76, Line=2
- Sense: 2 SenseNumber=2
 - Definition: Definition=abējāds, divējāds
 - Examples
 - Example
 - Text: Text=%uAbbejads%u der es mit bey"den halt, neuter
 - Source: Source=Fuer1650_70_1ms, Page=1, Line=23
 - Example
 - Text: Text=Tōw %uabbejadu%u Gharru teešcham šaprohtam.
 - Source: Source=LGL1685_V5, Page=87, Line=27
- Etymology
 - BasicWord
 - WordRef: Lang=latviešu, Word=abējs
- References
 - References: ↑ norāde uz pamatvārdu abējs, abēja
 - References: ↓ vārd darināšanas norāde abējādi

Entry structure

The headword is followed by the total frequency of the lemma in the corpus, grammatical information, and all the attested word forms with corresponding token frequency in the corpus. Evidence from the lexicographical sources of the period. Explanation of meaning, first and last citation in the corpus, idioms, collocations, etymology, elaborated system of cross-references.

		↓ nagliņi , ↔ nagla
23	nags	<p>nags (53) s. m. nagga=mällums (2), naggeem (13), naggeems (3), nagghus (1), naggi (3), naggis (4), naggohß (4), naggs (10), naggu (2), naggus (4), nags (2), putna=nags (2), saggja=naggeem (1), sag</p> <p><i>Nagel an Fingern oder Zäen/ Naggs.</i> Manc1638_L, 129A₂₁. <i>Finger=Nagel/ Naggs.</i> Manc1638_L, 60A₃. <i>Klaw der Vogel vnd Thiere/ Naggs.</i> Manc1638_L, 102B₂₃. <i>Huff oder Solen an Pferden/ Naggs.</i> Manc1638_L, 94A₃. <i>Nags ein Nagel an fingern.</i> Fuer1650_70_1ms, 163₄. <i>Nags ein Nagel an fingern.</i> Fuer1650_70_2ms, 237₅.</p> <p>1. nags (civēka, dzīvnieka, velna u. c.) <i>Finger=Nagel/ Naggs.</i> Manc1638_L, 60A₃. <i>Laid tas Welns muhs ne šagrahb/ Dohd Spehka pretti kautees/ No viņņa Naggeem ņautees.</i> LGL1685_K1, (</p> <p>◇ kā naga mellums (2) ka nagga=mällums (2) ļoti maz <i>To nu atkal taß Kungs JEšus nhe by pellnijis / aišto tohß Leezineekus dširrdādami / nhe warrejam me winjam tick dauds Ghräki / ka Nagga=Mällums / ka Atzy bahšt / warrātu parahditi kļuht.</i> Manc1654_I</p> <p>◇ putna nags (2) putna=nags (2) <i>Putna=nags Ein karger-filtz.</i> Fuer1650_70_1ms, 177₇.</p>

Entry structure in TLex2013

The screenshot displays the TLex software interface for a new document object model. The main window shows the entry structure for the lemma 'dūša'. The structure is as follows:

- Lemma: dūša LemmaSign=dūša, Frequency=10, Grammar=lietvārds sieviešu dzimte, Createc...
- Utterances
 - WordForm: WordForm=duhscha, Frequency=1
 - WordForm: WordForm=duhšcha, Frequency=4
 - WordForm: WordForm=duhšchas, Frequency=2
 - WordForm: WordForm=duhšchu, Frequency=1
 - WordForm: WordForm=duscha, Frequency=1
 - WordForm: WordForm=dušchu, Frequency=1
- Dictionaries
 - Citation
 - Text: Text=%uDuhšcha%u, die Gemüths bewegung Lith.
 - Source: Source=Fuer1650_70_1ms,Page=62,Line=1
 - Citation
 - Text: Text=%uDuhšcha%u die Gemüths=bewegung. Lith.
 - Source: Source=Fuer1650_70_2ms,Page=104,Line=17
 - Sense: 1 SenseNumber=1

The interface also shows a list of words on the left side, including 'dūša', 'dvešingieris', 'dverģis', 'dvērgis', 'Dvīnis', 'dvīniši', 'dvīņi', 'Dvīņis', 'dzegus', 'dzeguse', 'dzeguiziesietav', 'dzeinis', 'dzeldēt', 'dzeļdēt', 'dzeldināšana', 'dzeldināt', 'dzeldišana', 'dzelkne', 'dzelme', 'dzelna', 'dzelne', 'dūša', 'dvešingieris', 'dilles', 'Daugava'. The main window displays the entry for 'dūša' with its historical and linguistic information, including citations and senses.

dūša (10) s. f. duhscha (1), duhšcha (4), duhšchas (2), duhšchu (1), duscha (1), dušchu (1)
Duhšcha, die Gemüths bewegung Lith. Fuer1650_70_1ms, 62₁.
Duhšcha die Gemüths=bewegung. Lith. Fuer1650_70_2ms, 104₁₇.

1. **fiziskā pašsajūta**
 ar Širdī man Šlimma duscha mir ist nicht wohl. Fuer1650_70_2ms, 104₁₈.

2. **noskaņojums, garastāvoklis**
 man newa labbas duhšchas us wiņņu ich habe kein gut gemüht zu ihm. Fuer1650_70_2ms, 104₂₀.
 Tad manna Širds gribb pušcham šprahgt / Šmags Akmins Pohl uhs dušchu. LGL1685_V5, 359₂₉.

♦ **akmins pol ūz dūšu** (2) akmins pohl uhs dušchu (1), akmins pohl us duhšchu (1)
 sākas nepatikami pārdzīvojumi
 Tad manna Širds gribb pušcham šprahgt / Šmags Akmins Pohl uhs dušchu. LGL1685_V5, 359₂₉.

♦ : la. *dvest, dvēst* 'elpot, dvašot'
 [peteris] [anta]

dveņģeris (1) s. m. dwengheris (1)
Zwinger/ Dwengheris. Manc1638_PhL, 390₁₉.

♦ **cietums (?)**; eja starp pilsētas nocietinājumiem (?)
Zwinger/ Dwengheris. Manc1638_PhL, 390.
 ♦ < vlv. *dwenger* 'cietums'
 [everita] [anta]

dverģis, dvērgis (3) s. m. dwehrghis (1), dwerģis (2)
Zwerg/ Dwehrghis. Manc1638_L, 222A₃₀.
Dwergis, Ein Zwerg. Germ. Fuer1650_70_1ms, 57₁₄.
Dwergis, ein Zwerg. Germ. Fuer1650_70_2ms, 106₅.

punduris

LVVV

Historical Dictionary of Latvian (<http://tezaurs.lv/lvvv/>)

- A custom electronic dictionary version was developed to tailor it for our specific requirements (I.Poikāns, UL)

The screenshot shows the website 'tezaurs.lv/lvvv/' with a search interface on the left and a results table on the right.

Search Filters (Left Panel):

- Izvērstā meklēšana [?]**
- Vārds:** [Empty input field]
- Meklēt nozīmēs
- Meklēt visā tekstā
- Vārdšķira:** lietvārds
- Vārdu savienojums:** *
- Cilme:** aizguvums, viduslejasvācu
- Meklēt** button
- Vienkāršā meklēšana

Navigation Menu: Par vārdnīcu, Šķirkļa uzbūve, Meklēšana, Avoti, Saīsinājumi, Autori, Publikācijas, Kontakti

Results Table:

#	Lemma	Šķirklis
1	Brencis	Brencis (1) s. m. <i>npers.</i> Brentzis (1) Brencis <u>Brentzis/</u> <i>Jahka. Manc1638_Run, 460B₄.</i> ◊ < vlv. <i>Lawrenz</i>
2	cābele	cābele (2) s. f. zahbele (2) <u>Zahbele.</u> <i>ein Zobel, melna Zauna. Fuer1650_70_1ms, 311₂.</i> <u>Zahbele.</u> <i>eine Zobel, melna Zauna. Fuer1650_70_2ms, 538₂.</i> sabulis <u>Zahbele.</u> <i>ein Zobel, melna Zauna. Fuer1650_70_1ms, 311₂.</i> <u>Zahbele.</u> <i>eine Zobel, melna Zauna. Fuer1650_70_2ms, 538₂.</i> ◊ < vlv. <i>zabel 'sabulis'</i>
3	dīķis	dīķis (7) s. m. dieki (1), diekis (1), dihkis (2), dihķi (2), dihķi[^] (1) <i>Teich/ Diekis. Manc1638_L, 182A₂₄.</i> <i>Weier/ Dihkis. Manc1638_L, 204B₁₈.</i> dīķis <i>Teich/ Diekis. Manc1638_L, 182A₂₄.</i> Bet tur Jeruhšaleme [^] irr pee teem Avju (Wahrtim) <u>Dihķis/</u> ar Pawahrda Ebrei~šķi Betešda/ tam bija peežas Jštabas. JT1685 Jn 5:5. ◊ < vlv. <i>dīķis</i>

Word-forms in the entry are linked with concordance lines of subset of the Corpus (only the 16-17th c.)

The screenshot shows a web browser window with the URL tezaurs.lv/vvv/. The page title is "Latviešu valodas vēsturiskā vārdnīca". The entry for "nasta" is displayed, including its etymology and a list of concordance lines from the 16-17th century corpus.

apķesejščas **nahšis**, einbeschmutzte nister. Fuer1650_70_1ms, 94₄.
 Iskapti **nahši** mehdsinaht . die Sense an der Nasen prufen. Fuer1650_70_2ms, 222₆.
 ◆ : liet. *nósis* 'deguns', pr. *nozy* 't. p.'

33 nasta

nasta (162) s. f. *deenas=naštu* (1), *ghrāko=našta* (1), *ghrāko=naštu* (1), *ghrāko=našta* (2), *grehku=naštahm* (1), *grehku=naštu* (1), *grehkunaštu* (1), *kruhta=naštu* (1), *nasta* (2), *našta* (4), *našta`* (2), *naštu* (6), *našt* (3), *našt'* (1), *našta* (11), *našta=nāššātais* (1), *našta=neššētajā* (1), *naštahm* (2), *naštahs* (1), *naštams* (1), *naštas* (5), *našte* (14), *našt*h (1), *našt*he (3), *naštu* (35), *naštu=neššājas* (1), *našta* (9), *našt*he (1), *naštu* (42), *naštunāššētajas* (1), *naštunāššētajas* (2), *nāšt* (3), *nāšt*he (1)

Būrde/ Našta. Manc1638_L, 43B₂.
Lašt/ Našta. Manc1638_L, 112A₈.
eine Būrde/ Našta. Manc1638_L, 187A₉.
Nasta eine last, *būrde* a` nest. Fuer1650_70_1ms, 164₉.
Nasta eine last, a` nest. it. *būrde*. Fuer1650_70_2ms, 238₁₄.

1. *nasta*, *smags nesamais*; *sainis*
 Rouge / tows Köninx nake py thöw leenpratyx / vnd yaye vs wene Ešsel / vnd vs wene kummelle tās **našte** neššētajes Ešsellemates. EvEp1587, 4₁₃.
 Jo manni Grehki pahr mannu Galwu augušchi/ ka` kahda gruhla **Našta** irr tee man wissai šmaggi tappušchi. LGL1685_K1, 265₃₃.

2. *pārn*. *apgrūtinājums*; *rūpes*, *raizes*; *kaut kas ļoti nepatīkams*
 Aešto yckwens Bouwe **Našte** neBys. EvEp1587, 175₂₁.
 US kurreen buhs man behgt? Kad Širds tik gruhla irr/ No leelahm Grehku **Naštahm**. Kur buhs man Padohm' atrašt? LGL1685_K1, 94₁₈.

- (at)**ņemt** / (no)**ņemt** (**nost**) **nastu**; **nastu atņemt** (7) atņemm.. *naštu* (1), *jemm*.. *nohšt*.. *našt'* (1), *naštu atņehmis* (1), *nojemt*.. *naštu* (1), *ņemm nohšt*.. *naštu* (2), *ņemm*.. *nohšt*.. *našt* (1)

- **grēku nasta** (43) (*ghrāko*) *naštu* (1), *ghrāko(ghrāto)* *naštu* (1), *ghrāko našta* (1), *ghrāko našta* (1), *ghrāko naštu* (1), *ghrāko naštu* (15), *ghrāko=našta* (1), *ghrāko=naštu* (1), *ghrāko=našta* (2), *ghrāku naštu* (1), *ghreko naštu* (1), *grāhko naštu* (1), *grāko naštu* (1), *greeke našte* (1), *greeke nāšt*he (1), *grehku naštahm* (1), *grehku naštas* (1), *grehku naštas* (1), *grehku naštu* (6), *grehku=naštahm* (1), *grehku=naštu* (1),

Vārdformas šablons: **naštu**

Avots: LVVV

Statistika: vārdformas - 1, vārdlietojumi - 39

leelu gruhtu [Naštu](#) Jo tu pats par to zeetis ešš
Nā warr nojemt šcho [Naštu](#) Kaut tetzeys tahļ
štahw Und pallieds to [Naštu](#) pannešt Kad neds Laime
To [Naštu](#) ko tu mums usleez
Šcho [Naštu](#) wiņšch uš šewim jemdams lokahs
Deews usleek mums weenu [Naštu](#) bett wings
to Ghrāku [Naštu](#) und tahß Elles Mohkahms tad by tawam
[Naštu](#) vslickt
nešše us šawu Mugguru to smaggu ghruhtu [Naštu](#) wiššo Laušcho
Tad tohpu eš peštieti no Ghrāko [Naštu](#) Kungs turri mann štippre
arri šawu [Naštu](#) ghann ghrute neššis gir nhe top scheitan
in Baggatu zittadi tu leelu [Naštu](#)
mohzihts in wiņņas [Naštu](#) tam ne
ka taß to J E S V packal [naštu](#) Bett leels pullx Lauššcho und Šeewas ghaja
ka ir tu ne tohpi kahrдинаhts Neššah zits zitta [Naštu](#) tad juhs Krištus Bauslibu peepildiiseet Bet ja kas
hds in ne pee zitta Jo ikweens neššihs šawu pašchu [Naštu](#) Bet kas

- The main task was to understand the procedure of dictionary compilation and coordination of team-work and to find out our needs in future stages.
- The result is not only the dictionary itself and its platform, but also a set of guidelines.
- The link to concordance lines of the wordforms is a good chance for users to verify our conclusions with corpus data.
- Even a limited-size dictionary can give some insight into the grammatical and semantic phenomena and processes of the language, as well as the general use of written Latvian during the 16th and 17th centuries.
- ‘So-far only two dictionaries have been compiled in Latvia applying corpus data: *Latviešu valodas vēsturiskā vārdnīca* ‘Historical Dictionary of the Latvian Language’ that is based on the one-million-word Corpus of Early Written Latvian and *Franču-latviešu vārdnīca* ‘French-Latvian Dictionary’ (..)’ (Karpinska 2015:37).

One of the resources available from www.korpuss.lv

The screenshot displays the website www.korpuss.lv in a browser window. The page is titled "AiLab korpusi" and features a navigation menu with "Kas ir korpuss?" and "Kur izmanto korpusu?". The main content area is divided into several sections, each representing a different corpus resource:

- LVK2018**: 2016–2018, 10 milj. vārdlietojumu. Līdzsvarotais mūsdienu latviešu valodas tekstu korpuss. Includes buttons for "Vairāk informācijas" and "Meklēt korpūsā".
- LVK2013**: 2007–2013, 4,5 milj. vārdlietojumu. Līdzsvarotais mūsdienu latviešu valodas tekstu korpuss. Includes buttons for "Vairāk informācijas" and "Meklēt korpūsā".
- MuLa**: 2011–2013, 1 milj. vārdlietojumu. Mūsdienu latgaliešu tekstu korpuss. Includes buttons for "Vairāk informācijas" and "Meklēt korpūsā".
- Saeima**: 2013–, 21 milj. vārdlietojumu. Latvijas Republikas 5.–12. Saeimas sēžu stenogrammu korpuss. Includes buttons for "Vairāk informācijas" and "Meklēt korpūsā".
- LiLa**: 2011–2013, 8 milj. vārdlietojumu. Lietuvišu-latviešu-lietuvišu paralēlo tekstu korpuss. Includes buttons for "Vairāk informācijas" and "Meklēt korpūsā".
- Senie**: 2002–2015, 902 986 vārdlietojumu. Latviešu valodas seno tekstu korpuss SENIE. Includes buttons for "Vairāk informācijas" and "Meklēt korpūsā".

The browser's address bar shows the URL www.korpuss.lv. The taskbar at the bottom indicates the system time as 12:03 PM on 7/27/2018.

Modernization of the Corpus platform:

- New platform dealing with Unicode files (L.Pretklaniņa, UL)
- Switch to NoSketchEngine (?)

The screenshot shows the NoSketchEngine web interface. The browser address bar displays the URL: `nosketch.korpuss.lv/run.cgi/wordlist_form?corpname=saeima`. The page title is "Word list options".

Word list options:

- Corpus: Saeima
- Subcorpus: None (whole corpus) [info](#) [create new](#)
- Search attribute: word
- use n-grams. Value of n: from 2 to 2
- hide/nest sub-n-grams

Filter options:

- Filter word list by: Regular expression:
- Minimum frequency: 5
- Maximum frequency: 0 (0 = no maximum frequency)
- Whitelist: No file chosen
- Blacklist: No file chosen [format](#)
- Include non-words

Output options:

- Frequency figures: Hit counts Document counts ARF
- Output type: Simple Keywords
- Reference (sub)corpus: Saeima (whole corpus)
- Prefer: rare words common words
- Change output attribute(s)

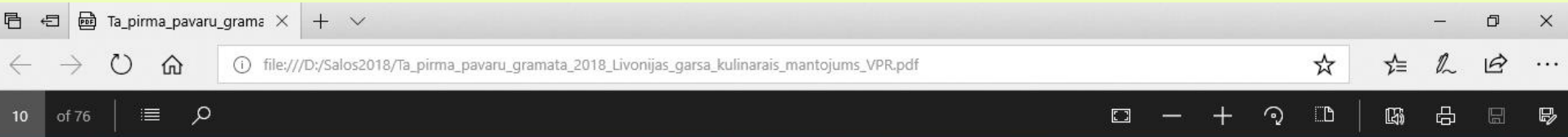
The interface is displayed in a browser window with several tabs open, including "Korpuss | AiLab", "Word list", and "Search corpus". The Windows taskbar at the bottom shows the date and time as 12:05 PM on 7/27/2018.

www.garamantas.lv – 75 405 pages of manuscripts already typed (July 2018)

The screenshot shows the website www.garamantas.lv in a browser window. The page title is "LATVIEŠU FOLKLORAS KRĀTUVES DIGITĀLAIS ARHĪVS". The navigation bar includes "Par projektu", "Resursi", "Karte", and a search bar. The main content is organized into three columns:

- Jaunumi (News):** A post titled "Aktuāli pētniekam: LFK 150" dated 11.07.2018. It features a handwritten manuscript snippet and a small graphic with the text "150 GADI". Below the image, it states: "Latviešu folkloras krātuves pētnieks dr. philol. Aigars Liebārdis aicina folkloras materiālu atšifrētājus pievērsties 150 manuskriptam, lai kopīgi pārrakstītu vairāk nekā 2500 lapaspuses. Atšifrējumi veidos būtisku daļu no latviešu buramvārdu digitālā kataloga, kas tiek izstrādāts buramvārdu pētnieka pēcdoktorantūras projekta "Latviešu buramvārdu..."
- Iepazīsti kolekcijas (Get to know collections):** A post titled "Folkloras kolekcijas no Jelgavas puses" dated 25.06.2018. It describes collections gathered in the Jelgava region. A list of collections follows:
 - [Bb44] Aleksandra Cimzes folkloras vākums,
 - [3] Garozas pamatskolas vākums,
 - [12] Ūziņu pamatskolas vākums,
 - [32] Irmas Mednes folkloras vākums,
 - [61] Marijas Andersones folkloras vākums,
 - [65] M. Tereses folkloras vākums Elejā,
 - [115] Elizas Vidzes folkloras vākums.
- Simtgades burtnieki (Centennial letter carriers):** A large colorful graphic with the text "Latvija 100 GADI". Below it, it says "Dāvini laiku Latvijai! Seko jaunākajām ziņām @garamantas #LV100 #burtnieki". It lists digital releases:
 - Kopā atšifrētas 75405 manuskriptu lappuses!
 - Jaunākie atšifrējumi
 - Pirms 3 minūtēm [Linda_Pavlovska:0150-Buramvardi-14-0136](#)
 - Pirms 6 minūtēm [Aija_Jonina:0150-](#)

Interest in old printings - transcribed first Cooking book



Tās pavāru grāmatas

I. nodaļa

Dažādas zupas

1. Zupe ar baltas mandeļu kuljies^{1*}

Ņemi baltas maizītes, vai daudz, vai maz, pēc zupas daudzuma. Nomizo it plāni to brūno garozi, sagriezi tās maizītes gabalos un meti podā. Piemeti arī kādas pētersilju, moru² un burkānu saknes, kas krustiski ir pāršķeltas. Tad uzlej it labu un spēcīgu gaļas zupi³ un vāri to, kamēr tā maize savirusi un tā zupe drusciņ tumīga paliek. Tad sagrūdi tās mandeles it smalki kā vienu miklu. Ņemi arī drusciņ vai teļa, vai cāļu cepeta klāt un kādu pāri cieti novāritu pautu⁴ dzeltenumu, to sakapā un saberzē, kamēr arī paliek kā viena mikla. To visu iemeti tai zupā un samaisi labi. Kad jau mandeles klāt, tad lai šī zupe vairs neverd, jo kad nu vēl dabū virt, tad tā saiet kopā un paliek nejauka; bet lai tāpat vien stāv pret uguni, ka allaž karsta paliek. Spiedi to nu caur astru drēbi jeb sietu, cik vien tur iet caur, tā ka tā zupe tauka, balta un tumīga top. Iemeti sāli, cik vajaga, un, kad gribi, arī drusciņ mušatenblūm⁵. Tā lai nu bez viršanas karsti vien stāv, kamēr uz galdu padod. Kad ir uzdodama⁶, tad *samieķšē vēl maizi un dodī to zupī virsū. Šinī zupē vēl vari ielikt, ko pats gribi, vai cāli, vai teļa gaļu, vai tauka kaula jeb markses kaulu⁷, vai vienu teļa kaulu, ar vilzeli⁸ apsistu⁹.

To varzi jeb vilzeli, jeb pildekli,

ko ap kādu kaulu var sist, to taisi šādā vizē¹⁰; ņemi it tiru teļa gaļu no gūžam, kur visas ādas un cipslas tīri ir izgrieztas, un sagriezi to kantiski¹¹. Cik tev nu ir gaļas, tā ņemi pus tik daudz nieru tauku; un, kad drusciņ vērsa markses¹² jeb kaula tauku ir

kapā vēl, kamēr no tās rīrejes vairāk neko neredzi kā vienu dzeltanu miklu vien. Ja šī varze vai vilzeli pēc tava prāta vēl būtu biežāks, tad to allažīņ šķīstāku vari darīt, krējumu vai pienu pieliedams. Šo vilzeli apsiiti ap vienu teļa kaulu vai iztāisi no tā vienu kluču jeb kukuli, kā patik, nolīdzini to virspusi labi glumi un izrobi atkal ar nažu¹⁷ šā tā, cepi to iekš taršu pannas¹⁸, lieci bļodā un stellē¹⁹ uz galdu.

* Kulije²⁰ ir tāda caurkāsta un caurspiesta spēka zupe. Kad brīžam kādas cepeta atliekas vai šķiņķa gaļu, vai citu ko smalku sakapā un miezeri sagrūdi ar visiem kauliem, tad pašu pēdīgu spēku tur izvāri un caurkāsi, ka tas biežums vien paliek pakaļā²¹; tad tāda izkāsta un caurspiesta zupe ir tā kulije.

* Tā maizes mieķšēšana tā notiek: iemeti zupas bļodā vai restētas maizes šķēles, vai aprīvētas maizes garozes, vai arī veselu vienu aprīvētu maizīti, uzlej labu spēcīgu un klāru²² zupi it verdošu virsū, aplāj ar citu bļodu un celi uz oglēm, ka tā labi velkas iekš to maizi iekšā. Bet kad jau veselu vienu maizīti tā gribi izmieķšēt, tad vajaga dibenī caurumiņu izgriezt un drusciņ mikstuma izņemt, ka tā zupe caur visu maizi labi var ievilkties.

2. Vēžu zupe



Thank you!
Any questions?

The screenshot shows a web browser window with the URL senie.korpuss.lv/kwic.jsp?wordform=pald%25&source=SENIE&context=50&limit=50. The page displays a list of Latvian folk songs, each with a number, a title, and a short excerpt of lyrics. The lyrics are in Latvian, and the word "Paldees" is highlighted in green in the original image. The browser's taskbar at the bottom shows various application icons and the system clock indicating 12:08 PM on 7/31/2018.

13 [Paldees](#) par meelošchanu
Tam labbam šaimneekam [Paldees](#) par šwehtišchanu
2 [Paldees](#) šchim Tehwam ka' eekšch mums
4 [Paldees](#) jums Tehiht Mahmiņa
5 [Paldees](#) par to Mantiņū
[Paldees](#) jums Tehiht Mahmiņa
To wišši prohtam arr [Paldees](#) miħjam grahamatahm
To wišši prohtam arr [Paldees](#) miħjahm grahamatahm
To wišši prohtam arr [Paldees](#) miħjahm grahamatahm
Nu irr ta šelta=meitiņa 69 [Paldees](#) jums tehtiht mahmiņa 77
Maušch [Paldees](#) nu dšerr wehl ohtru
Behrt [Paldees](#) Kungi warr wišš darriht
Wiņšch nahk muš wiššu labba` [Paldeews](#) ak šaldajs Ješuliņšch Ka
13 Ta` šazzidams Deews augštakajs [Paldeews](#) mans miħlajs Radditajs
j nu preezatees Kad mehs no ta dširdim Deewam doht [Paldeews](#) Deews
14 [Paldeews](#) ak miħļa Brahļa=Širds [Paldeews](#) ak Ješus tawa Pirt
14 [Paldeews](#) ak miħļa Brahļa=Širds [Paldeews](#) ak Ješus tawa Pirts
šti mannu Wainu Es ešmu mehšts mans Parads dšehšts [Paldeews](#) par tahdu Maiņū
16 [Paldeews](#) ak šaldajs Eņģeliht Gohds Deewam dohts Mehs
Tad by mums wiššeems pašušt [Paldeews](#) šaldais JEšuliņš
Zitta Lughšchana [Paldeews](#) tam augštam šchehligam Deewam
293 Ka gribbi tapehz Deewu peeluhgt [Paldeews](#) ak šaldais Ješuliņšch Ka tu mums
šwehtkeem [Paldeews](#) mans Ješus Tu šalihdšina man Deewam
4 To Wahrdu buhs teem likt ar Meera In ne [paldeews](#) par to dabuht
TAs [paldeews](#) Deewam nodšihwohts Pezh Dahrbeem kļušt šwehts
TAs [paldeews](#) Deewam nodšiwots