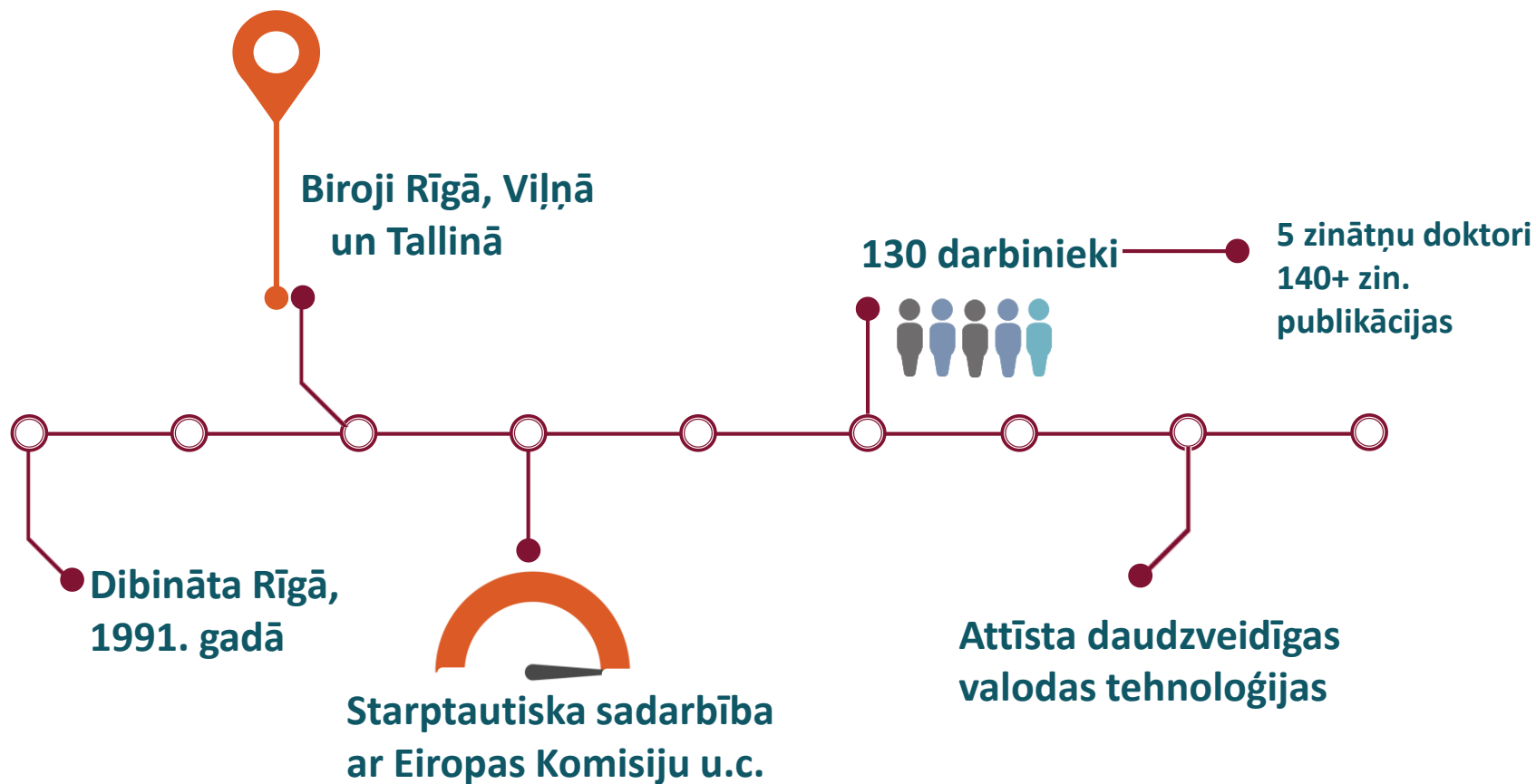


# VALODAS TEHNOLOĢIJAS UN TULKOŠANA

Raivis SKADIŅŠ, Tilde

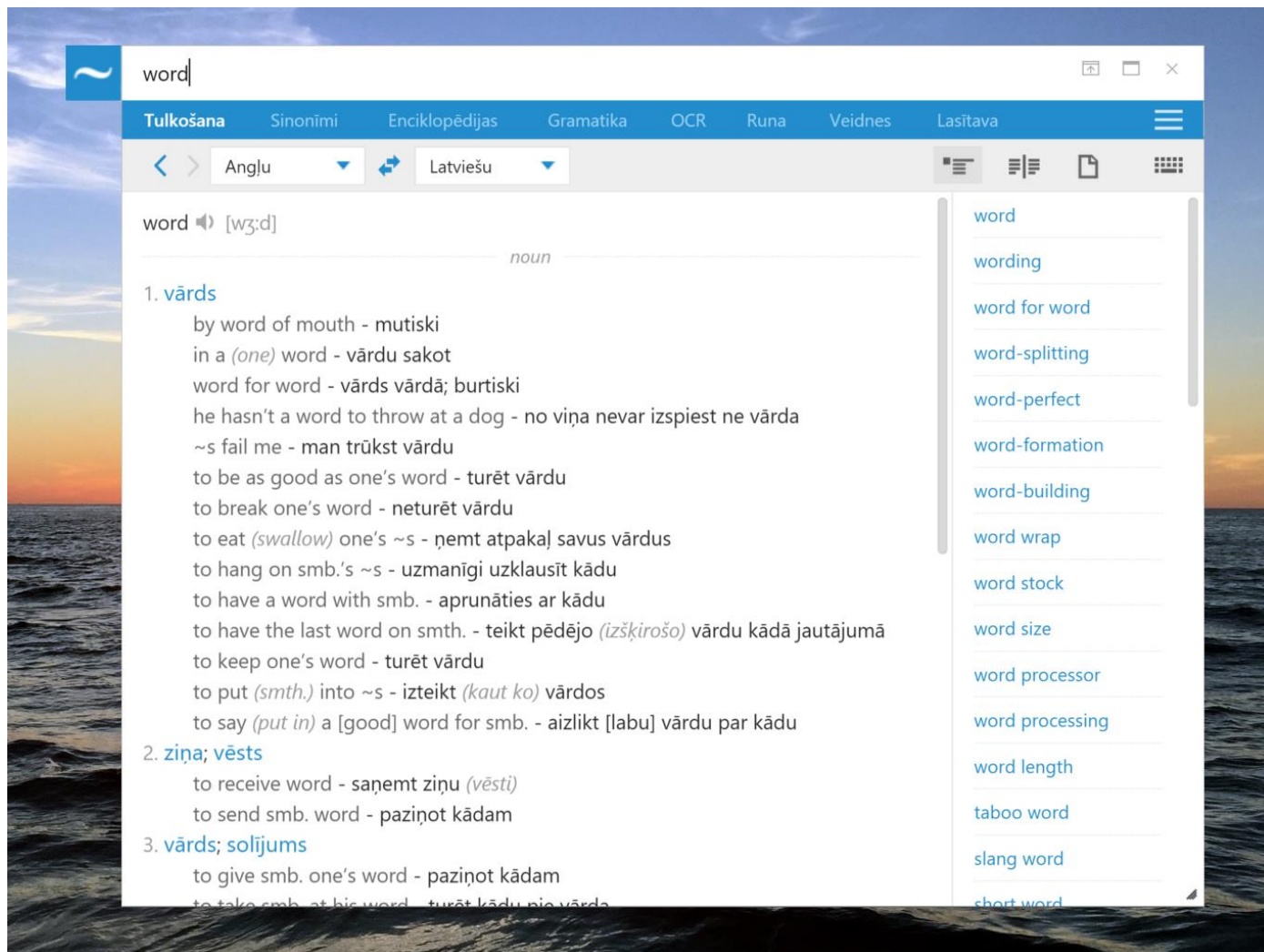
*Latviešu valoda ES – tulkošanas nozares resursi ES iestādēs un Latvijā*

28.10.2016.



- mērķis:  
tehnoloģiju atbalsts latviešu valodai –  
tāds pats kā lielajām valodām
- fonti
- pareizrakstība
- vārdnīcas
- gramatikas pārbaude
- terminu vārdnīcas, [eurotermbank.com](http://eurotermbank.com)





The screenshot displays the Tildes Birojs web application interface. At the top, there is a search bar with the word "word" entered. Below the search bar, a navigation menu includes options like "Tulkošana", "Sinonīmi", "Enciklopēdijas", "Gramatika", "OCR", "Runa", "Veidnes", and "Lasītava". The interface is set to translate from "Angļu" (English) to "Latviešu" (Latvian). The main content area shows the dictionary entry for "word" with its phonetic transcription [wɜːd] and the part of speech "noun". The entry is organized into three numbered sections: "1. vārds" (word), "2. ziņa; vēsts" (message; news), and "3. vārds; solījums" (word; promise). Each section contains several example phrases and their Latvian translations. A right-hand sidebar lists related terms such as "word", "wording", "word for word", "word-splitting", "word-perfect", "word-formation", "word-building", "word wrap", "word stock", "word size", "word processor", "word processing", "word length", "taboo word", "slang word", and "short word".

word |

Tulkošana   Sinonīmi   Enciklopēdijas   Gramatika   OCR   Runa   Veidnes   Lasītava

< >   Angļu   ↕   Latviešu

word [wɜːd]

noun

1. vārds

- by word of mouth - mutiski
- in a (one) word - vārdu sakot
- word for word - vārds vārdā; burtiski
- he hasn't a word to throw at a dog - no viņa nevar izspiest ne vārda
- ~s fail me - man trūkst vārdu
- to be as good as one's word - turēt vārdu
- to break one's word - neturēt vārdu
- to eat (swallow) one's ~s - ņemt atpakaļ savus vārdus
- to hang on smb.'s ~s - uzmanīgi uzklaustīt kādu
- to have a word with smb. - aprunāties ar kādu
- to have the last word on smth. - teikt pēdējo (izšķirošo) vārdu kādā jautājumā
- to keep one's word - turēt vārdu
- to put (smth.) into ~s - izteikt (kaut ko) vārdos
- to say (put in) a [good] word for smb. - aizlikt [labu] vārdu par kādu

2. ziņa; vēsts

- to receive word - saņemt ziņu (vēsti)
- to send smb. word - paziņot kādam

3. vārds; solījums

- to give smb. one's word - paziņot kādam
- to take smb. at his word - turēt kādu pie vārda

word

wording

word for word

word-splitting

word-perfect

word-formation

word-building

word wrap

word stock

word size

word processor

word processing

word length

taboo word

slang word

short word



letonika.lv. Vārdnīcas

letonika.lv/groups/default.aspx?g=2

Lietotājs: TILDES MĀJA Iziet  
Piekļuve aktīva līdz: 04.01.2017 plkst. 23:59

Enciklopēdijas Vārdnīcas Lasītava Multivide Valoda Ceļvedis Pasakas Kultūras kanons Mācībām ABONĒ

## Vārdnīcas

- Latviešu—angļu
- Angļu—latviešu
- Latviešu—krievu
- Krievu—latviešu
- Latviešu—vācu
- Vācu—latviešu
- Latviešu—lietuviešu
- Lietuviešu—latviešu
- Latviešu—latīņu
- Latīņu—latviešu
- Igauņu—latviešu
- Latviešu—igauņu
- Citvalodu personvārdu atveides vārdnīca

Meklēt: Vārdnīcas

Sadaļā **Vārdnīcas** atrodas visas datu bāzes *letonika.lv* elektroniskās vārdnīcas. Angļu, krievu, vācu un lietuviešu valodai ir atvēlētas savas pamatvārdnīcas, kas papildinātas ar ļoti plašu dažādu nozaru terminoloģijas klāstu. Specifiskie latīņu valodas vārdi un to tulkojumi ir apkopoti atsevišķā vārdnīcā. Meklējot no sadaļas **Vārdnīcas** pirmās lapas, ar vienu meklēšanas pieprasījumu var iegūt jebkura vārda un frāzes tulkojumus. Meklēšanā tiek apkopoti rezultāti no vairāk nekā pusmiljona dažādu valodu vārdu un izteicienu tulkojumiem.

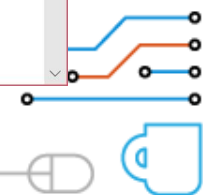
Šos rezultātus papildina daudznozaru daudzvalodu terminoloģijas portāls **EuroTermBank**, kura saturs tagad pieejams ērtā formā, izmantojot *letonika.lv* saskarni un meklēšanas iespējas. Šī publiski pieejamā bezmaksas resursa krājumā ir iekļauti vairāk kā 1,5 miljoni terminu, tai skaitā vairāk kā 200 tūkstoši dažādu nozaru terminu latviešu valodā. EuroTermBank terminu bankas pamatu veido latviešu, lietuviešu, igauņu, ungāru un poļu valodas termini un tiem atbilstošie angļu, vācu, kā arī franču, spāņu, krievu un citu valodu termini. Ērtais un viegli pieejamais izziņas avots paredzēts tulkjiem, redaktoriem, valodniekiem, terminologiem un citiem valodas nozares darbiniekiem, kā arī studentiem, izglītības darbiniekiem, zinātniekiem un citiem interesentiem.

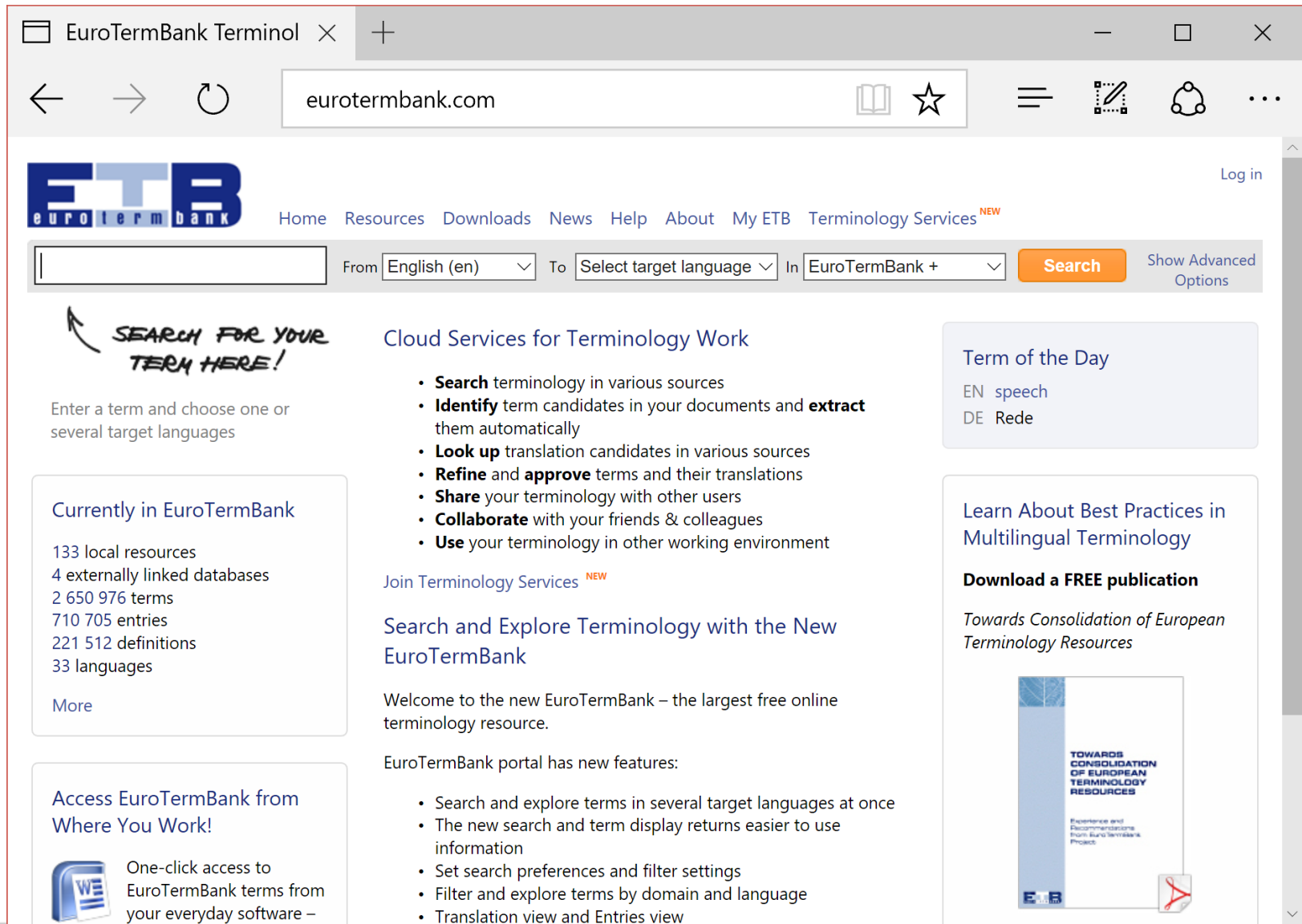
### LATVIEŠU—ANĢĻU UN ANĢĻU—LATVIEŠU VĀRDNĪCAS

Publicēto vārdnīcu pamatā ir profesora A. Veisberga 2001. gadā izdotā "Jaunā latviešu-angļu vārdnīca" un autoru kolektīva — Z. Belzējas, I. Birzvalkas, L. Jurkas, R. Mozeres, J. Raškeviča, A. Treilona — sagatavotā **"Angļu-latviešu vārdnīca"**. Vārdnīcas ir papildinātas ar terminoloģiju un "Wordnet" krātuves skaidrojumiem angļu valodā. "Latviešu-angļu vārdnīcā" ir iekļauti gandrīz 80 000 šķirķļu ar latviešu valodas vārdiem un izteicieniem, savukārt **"Angļu-latviešu vārdnīcā"** ir iekļauti vairāk nekā 70 000 šķirķļu.

### LATVIEŠU-LATĪŅU UN LATĪŅU-LATVIEŠU VĀRDNĪCA

Šajos tulkošanas resursos ir apkopoti vairāk nekā 2000 dažādu nozaru vārdu, izteicienu un terminu latīņu valodā un to tulkojumi latviešu un citās valodās. **"Latīņu-latviešu vārdnīcā"** publicētie termini ļauj precīzi identificēt un izmantot botānikas, Latvijas putnu, augu ģinšu taksonu, agronomijas, arheoloģijas, dabas un citu nozaru informāciju. Terminu tulkojumu valodas katrai nozarei var atšķirties. **"Latviešu-latīņu vārdnīcā"** var meklēt tulkojumus latviešu valodas vārdiem.





The screenshot shows the EuroTermBank website in a browser window. The browser's address bar displays "eurotermbank.com". The website header includes the ETB logo, navigation links (Home, Resources, Downloads, News, Help, About, My ETB, Terminology Services <sup>NEW</sup>), and a "Log in" link. A search bar is prominently featured with dropdown menus for "From" (English (en)), "To" (Select target language), and "In" (EuroTermBank +), along with a "Search" button and a "Show Advanced Options" link. The main content area is divided into several sections: a search prompt with a handwritten note "SEARCH FOR YOUR TERM HERE!", a "Cloud Services for Terminology Work" section with a list of features, a "Term of the Day" box, a "Learn About Best Practices in Multilingual Terminology" section with a "Download a FREE publication" link, and an "Access EuroTermBank from Where You Work!" section with a Microsoft Word icon. A statistics box on the left lists: 133 local resources, 4 externally linked databases, 2 650 976 terms, 710 705 entries, 221 512 definitions, and 33 languages. A welcome message states: "Welcome to the new EuroTermBank – the largest free online terminology resource. EuroTermBank portal has new features:" followed by a list of features. A document icon is visible in the bottom right corner of the page content.

EuroTermBank Terminol × +

eurotermbank.com

ETB euro term bank

Home Resources Downloads News Help About My ETB Terminology Services <sup>NEW</sup>

Log in

From English (en) To Select target language In EuroTermBank + Search Show Advanced Options

SEARCH FOR YOUR TERM HERE!

Enter a term and choose one or several target languages

### Cloud Services for Terminology Work

- **Search** terminology in various sources
- **Identify** term candidates in your documents and **extract** them automatically
- **Look up** translation candidates in various sources
- **Refine** and **approve** terms and their translations
- **Share** your terminology with other users
- **Collaborate** with your friends & colleagues
- **Use** your terminology in other working environment

Join Terminology Services <sup>NEW</sup>

### Search and Explore Terminology with the New EuroTermBank

Welcome to the new EuroTermBank – the largest free online terminology resource.

EuroTermBank portal has new features:

- Search and explore terms in several target languages at once
- The new search and term display returns easier to use information
- Set search preferences and filter settings
- Filter and explore terms by domain and language
- Translation view and Entries view

### Term of the Day

EN speech  
DE Rede

### Learn About Best Practices in Multilingual Terminology

**Download a FREE publication**

*Towards Consolidation of European Terminology Resources*

133 local resources  
4 externally linked databases  
2 650 976 terms  
710 705 entries  
221 512 definitions  
33 languages

More

### Access EuroTermBank from Where You Work!

One-click access to EuroTermBank terms from your everyday software –

TOWARDS CONSOLIDATION OF EUROPEAN TERMINOLOGY RESOURCES

Experience and Recommendations from EuroTermBank Project

# Tilde Terminology



The screenshot shows a web browser window with the URL `term.tilde.com/lookup`. The page has a navigation bar with three main sections: **MACHINE TRANSLATION**, **TERMINOLOGY** (which is underlined), and **LOCALIZATION**. A **TILDE** logo is in the top right corner. Below the navigation bar, there is a **Log in / Sign up** link. The main content area features a large search input field with the placeholder text **Type here to begin your term search**. Below the search field, it says **ADVANCED SEARCH** on the left and **Powered by TaaS** on the right. A dark red navigation bar contains the following links: **HOME**, **LOOKUP**, **RESOURCES**, **TECHNOLOGY**, **PRICING**, and **INTEGRATION**.

## Quick Look up

Paste a short text below. Hover over highlighted terms to see translations. Max 2000 characters for unregistered users. [Register here.](#)

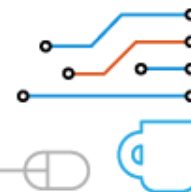
A computer is a general purpose device that can be programmed to carry out a set of arithmetic or logical operations automatically. Since a sequence of operations can be readily changed, the computer can solve more than one kind of problem. Conventionally, a computer consists of at least one processing element, typically a central processing unit (CPU), and some form of memory. The processing element carries out arithmetic and logic operations, and a sequencing and control unit can change the order of operations in response to stored information. Peripheral devices allow information to be retrieved from an external source, and the result of operations saved and retrieved.



# Mašīntulkošana



- vispārēja lietojuma mašīntulkošana
  - Google, Microsoft
  - [translate.tilde.com](https://translate.tilde.com), Tildes Birojs
- pielāgota mašīntulkošana
  - MT@EC
  - HUGO.LV
  - Tilde MT / LetsMT.eu





Tulkot tekstu

hugo.lv

LATVIJA TULKO AR **HUGO**.LV

LV RU EN

Tulkot tekstu Tulkot dokumentu Tulkot tīmekļa vietni

Angļu - latviešu (vispārēja) Tulkot

Mašīntulkošanas rezultāti ļauj saprast teksta nozīmi, bet nevar aizstāt cilvēka radītu tulkojumu.

<https://hugo.lv/lv/Translate/Website> centrs | E-pasts: [help@kis.gov.lv](mailto:help@kis.gov.lv) Par hugo.lv

ERAF IEGULDĪJUMS TAVĀ NĀKOTNĒ

EUROPS SAVIENĪBA



# Tilde MT



The screenshot shows a web browser window displaying the Tilde MT website. The browser's address bar shows the URL `tilde.com/mt/machine-translation`. The website's navigation bar includes the following items:

- MACHINE TRANSLATION** (highlighted with a red underline)
- TERMINOLOGY**
- LOCALIZATION**
- Tilde MT**
- Features**
- News**
- Case Studies**
- Data Library**
- Contacts**
- Pricing**
- Tools** (with a dropdown arrow)
- Log-in** (in the top right corner)

The main banner features a large image of a cable-stayed bridge over water. The text on the banner reads:

## Cross language barriers with machine translation

**FREE TRIAL**

At the bottom of the page, there are four categories listed in a red bar:

- TRANSLATION PROFESSIONALS**
- ENTERPRISE SOLUTIONS**
- GOVERNMENT INSTITUTIONS**
- ACADEMIC PARTNERSHIPS**



# Vai MT palīdz tulkošanā?

- kā to lietot?
  - pēcredīgēšana?
  - integrēšana CAT rīkos?
  - paredzošā rakstīšana?
  
- vai MT tiešām palīdz celt tulkotāja produktivitāti?



# 1. eksperiments

- IT jomai pielāgota MT sistēma
- SDL Trados Studio
- 6 valodas
  - latviešu, lietuviešu, igauņu, poļu, čehu, ungāru
- mašīntulkojumi tiek doti tikai kā ieteikumi

Skadiņš, R., Pinnis, M., Vasiljevs, A., Skadiņa, I., & Hudik, T. (2014). Application of Machine Translation in Localization into Low-Resourced Languages. In M. Tadić, P. Koehn, J. Roturier, & A. Way (Eds.), *Proceedings of the 17th Annual Conference of the European Association for Machine Translation EAMT2014* (pp. 209–216). Dubrovnik: European Association for Machine Translation. Retrieved from [http://hnk.ffzg.hr/eamt2014/EAMT2014\\_proceedings.pdf](http://hnk.ffzg.hr/eamt2014/EAMT2014_proceedings.pdf)



# 1. eksperiments

The screenshot displays two windows from the LetsMT! Machine Translation Provider. The top window shows a translation of a document with the following content:

Source	Quality	Match	Target
1 Click the button to see the demo.	82%	90%	Noklikšķiniet uz pogas, lai skatītu demonstrāciju.
2 Click another button to see the demo again.	AT	100%	Noklikšķiniet uz citas pogas, lai skatītu demonstrāciju.

The bottom window shows a different translation of the same document:

Source	Quality	Match	Target
1 Click the button to see the demo.	AT	100%	Noklikšķiniet uz pogas, lai skatītu demonstrāciju.
2 Click another button to see the demo again.			
3 Click the button to see the demo again.			

Arrows labeled 1, 2, and 3 point to the following elements:

- 1: Points to the source text "Click the button to see the demo." in both windows.
- 2: Points to the target text "Noklikšķiniet uz pogas, lai skatītu demonstrāciju." in the top window, and the target text "Noklikšķiniet uz pogas, lai skatītu demonstrāciju." in the bottom window.
- 3: Points to the target text "Noklikšķiniet uz citas pogas, lai skatītu demonstrāciju." in the top window, and the target text "Noklikšķiniet uz pogas, lai skatītu demonstrāciju." in the bottom window.



# 1. eksperiments

- produktivitātes pieaugums visām valodām
- kvalitāte pasliktinās, bet paliek tajā pašā līmenī

<b>MT System</b>	<b>Scenario 1 (1)</b>	<b>Scenario 2 (2)</b>	<b>Increase (3)</b>
EN-LV	550	731	32.9 %
EN-PL v1	305	392	28.5 %
EN-PL v2	294	357	21.5 %
EN-CS v1	315	394	25.1 %
EN-CS v2	291	351	20.8 %
EN-HU	287	339	18.0 %



## 2. eksperiments

- pēcredīgēšana
- medicīnas teksti – receptes, instrukcijas u.tml.
- angļu-latviešu
- ļoti precīzi mērījumi
  - milisekundes, klikšķi, taustiņu nospiedieni u.t.t
- 350 000 vārdu
- katram teikumam ieguvām
  - salabotu teikumu un mērījumus
  - vērtējumu – nav jāmaina, gandrīz perfekts, labs, slikts, ļoti slikts
- ne visai ērts CAT rīks ☹️





A1 by all\_files editing... partial: 54s revisions: 0 total: 0s

7/2000  
0 saved

- aktīvā viela ChondroCelect sastāv no ārstēšanas devu dzīvotspējīgu autoloģās cilvēka skrimšļa šūnas flakonos, kas satur 4 miljonus šūnas 0.4 ml šūnu suspensijā, kas atbilst koncentrācijai ①	
Do not use the pre-filled syringe if it has been dropped on a hard surface.	Nelietojiet pilnšļirci, ja tā ir nokritusi uz cietas virsmas. ②
In clinical studies with ChondroCelect a periosteal flap was used as a biological membrane.	Klīniskajos pētījumos ar ChondroCelect kaulplēves audus lietoja kā bioloģisko membrānu.
In studies in goats, mild signs of synovitis were observed in the majority of the animals, including controls at 10 weeks post surgery.	Pētījumos ar kazām, viegli pazīmes sinovīts lielākai daļai pacientu tika novērota arī kontroles grupas dzīvnieku, pēc 10 nedēļām pēc operācijas.
- The active substance of ChondroCelect consists of a treatment dose of viable autologous human cartilage cells in vials containing 4 million cells in 0.4 ml, corresponding to a concentration of ③	- aktīvā viela ChondroCelect sastāv no ārstēšanas devu dzīvotspējīgu autoloģās cilvēka skrimšļa šūnas flakonos, kas satur 4 miljonus šūnas 0.4 ml šūnu suspensijā, kas atbilst koncentrācijai ④
After any dose or schedule adjustment the haemoglobin should be monitored every one or two weeks.	Vai shēma pēc jebkuras devas korekcijas hemoglobīna līmenis ir jāpārbauda vienu reizi nedēļā vai reizi divās nedēļās. ⑤
The additional sugar residues are molecularly indistinct from those on the endogenous hormone.	Papildu cukura atliekas no tām ir molecularly indistinct endogēnajam hormonam.
Aranesp 10 syringe	Aranesp 10 šļirce
substance (9): vielas (6) cartilage (9): skrimšļa (8) concentration (13): koncentrācija (13) ⑥ cartilage cells (15): skrimšļa šūnas (14)	

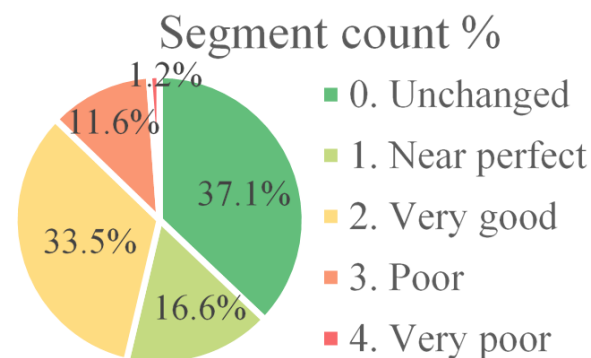




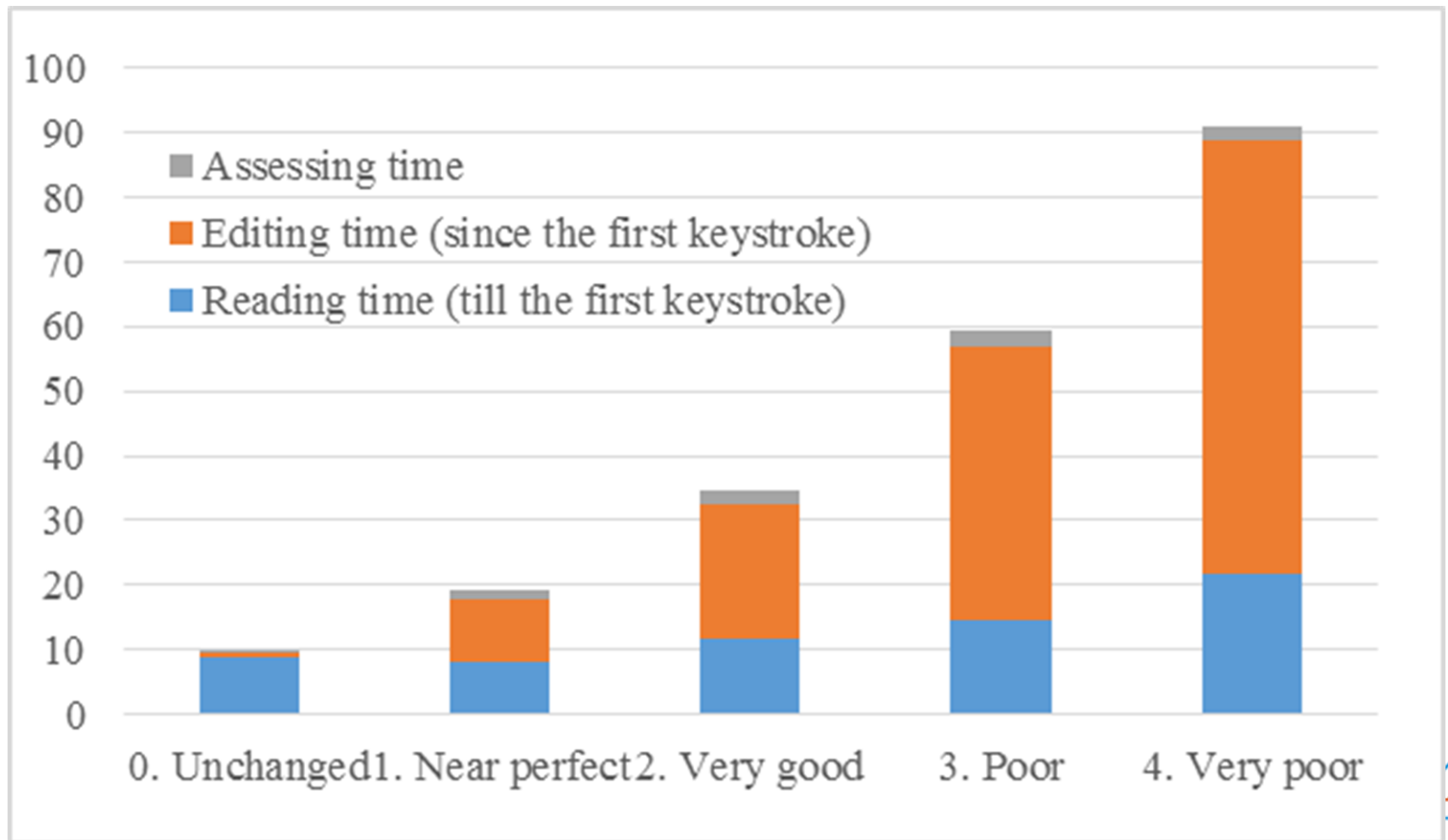
# 2. eksperiments



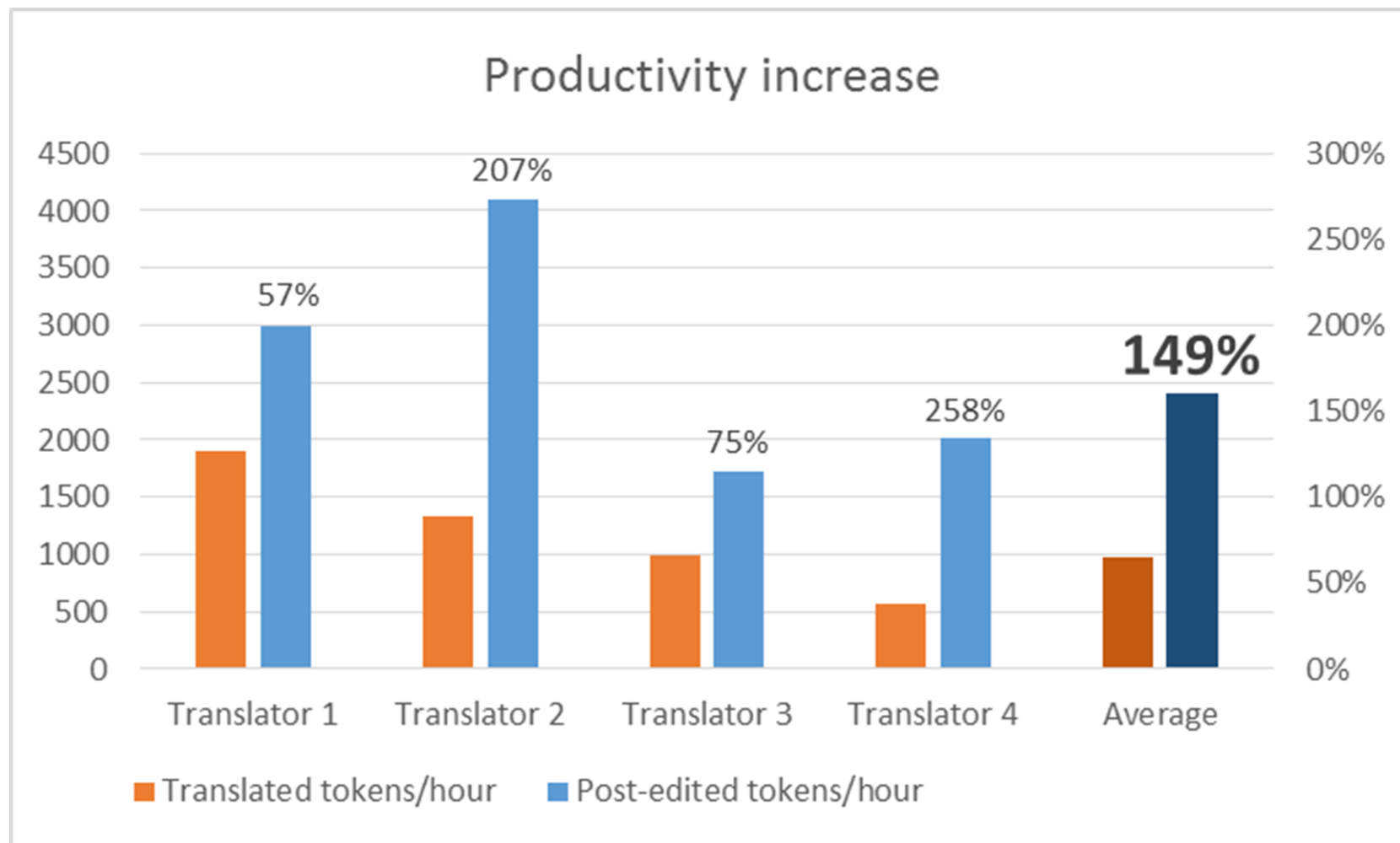
MT suggestion assessment	Total editing time	Total source length in tokens	Segment count	Productivity (tokens post-edited in one hour)
0. Unchanged	14:15:43	83,661	5,488	5,865
1. Near perfect	12:12:01	46,108	2,458	3,779
2. Very good	44:11:31	102,309	4,962	2,315
3. Poor	26:40:50	37,956	1,717	1,422
4. Very poor	04:13:39	3,582	175	847
Grand Total	101:33:46	273,616	14,800	2,694



## 2. eksperiments

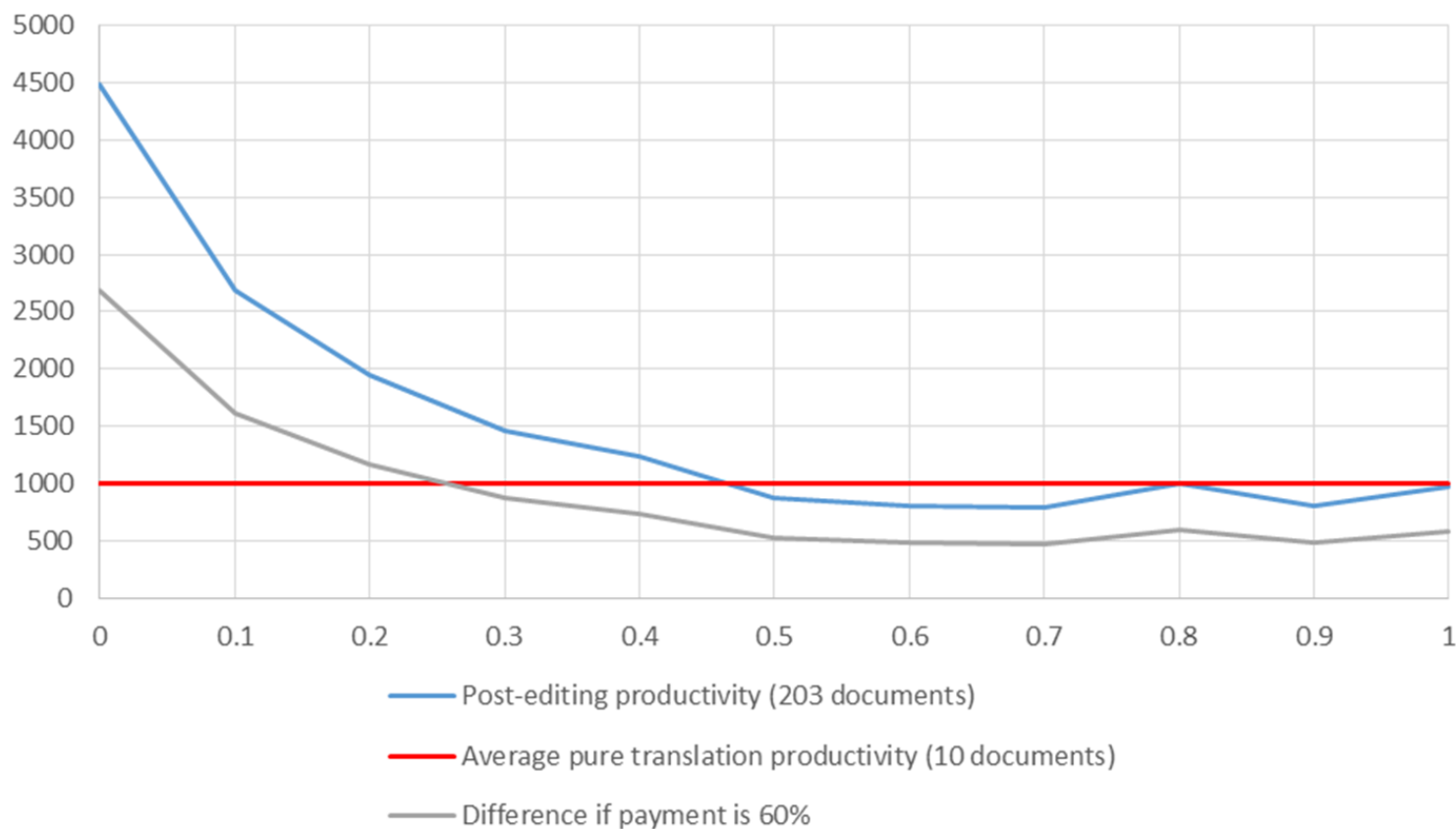


## 2. eksperiments



## 2. eksperiments

Average productivity at different HTER levels for 203 documents and 5 translators



# Kas tālāk?

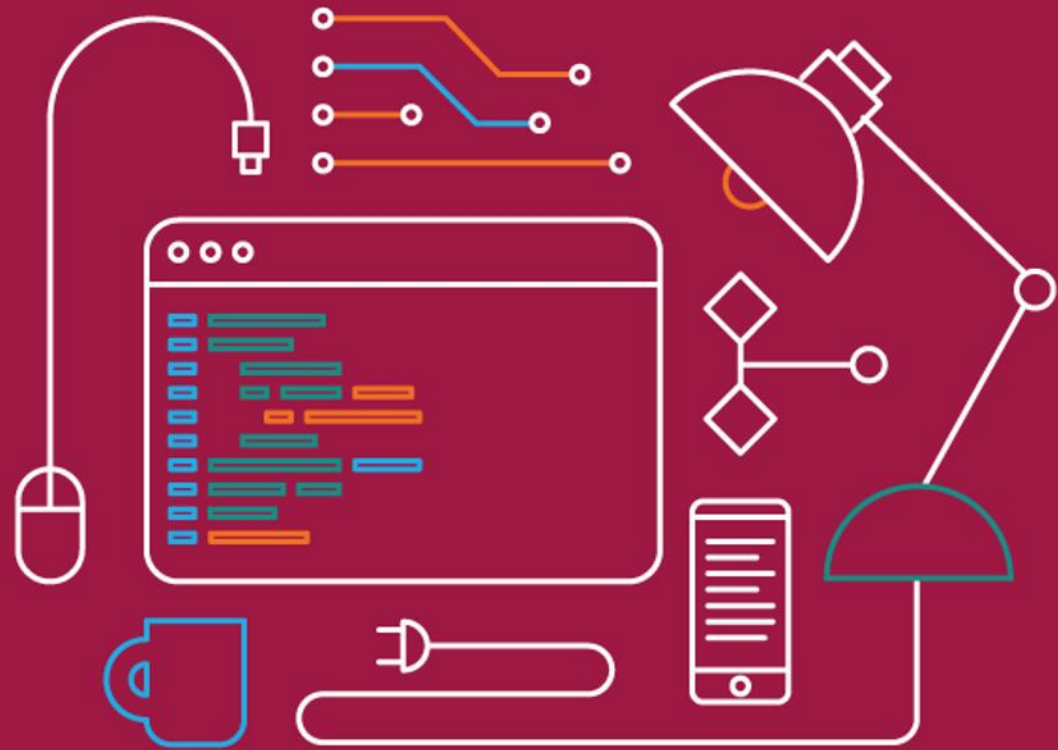
- HUGO.LV 2.0
- Connecting Europe Facility, Automated Translation (CEF.AT)
- Runas atpazīšana
- Paradigmas maiņa mašīntulkošanā – mākslīgo neironu tīklu mašīntulkošana



# kāpēc MT kopiena mīl tulkotājus?

- modernā mašīntulkošana mācās kā tulkot no piemēriem
- Jūs esat tie, kas rada tulkojumus, no kuriem mūsu tehnoloģija mācās
- vienīgi Jūs varat mums pateikt, cik labi mums izdodas
- mēs gribam radīt instrumentu, kas Jums patiešām noder





?