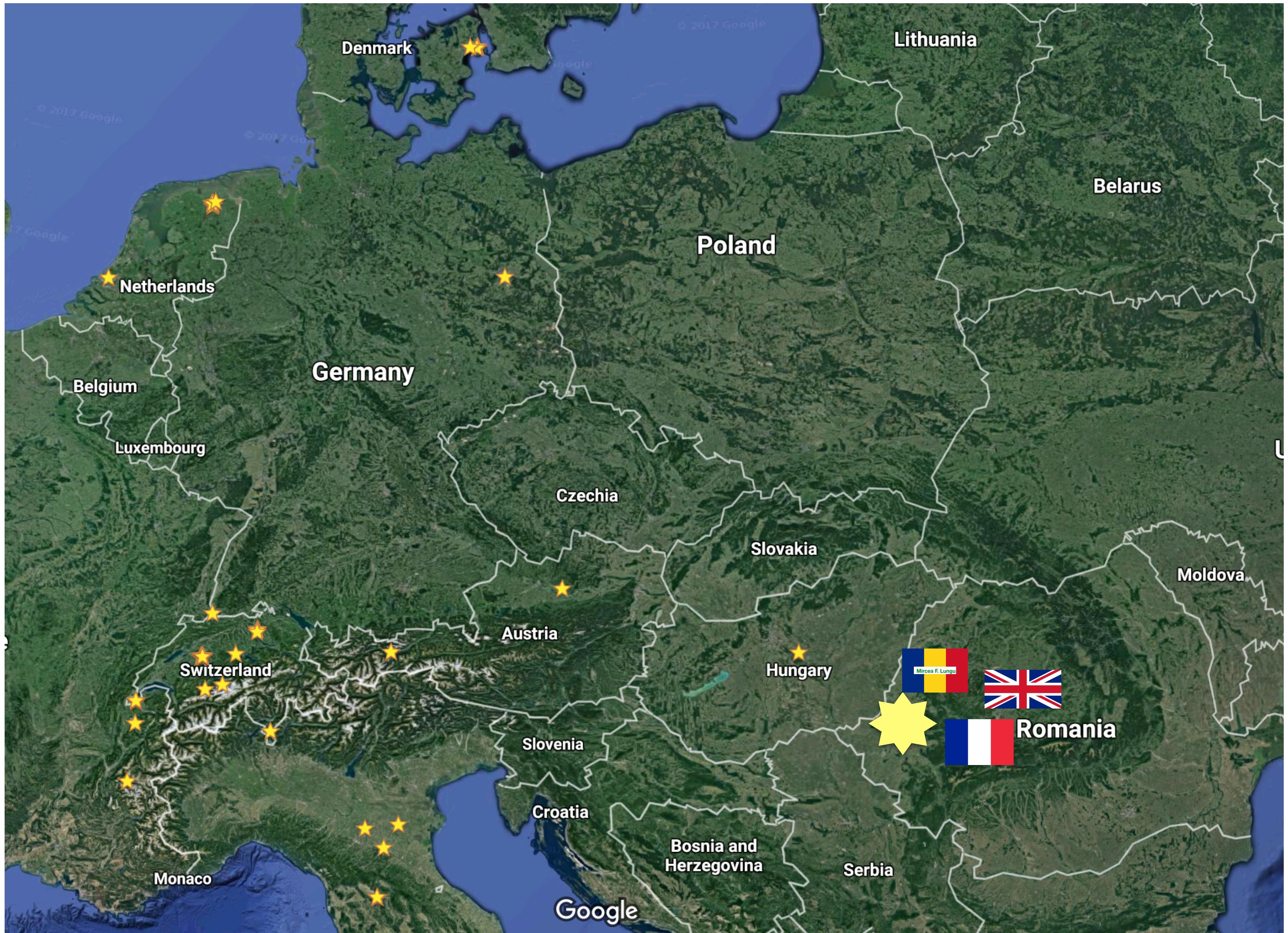


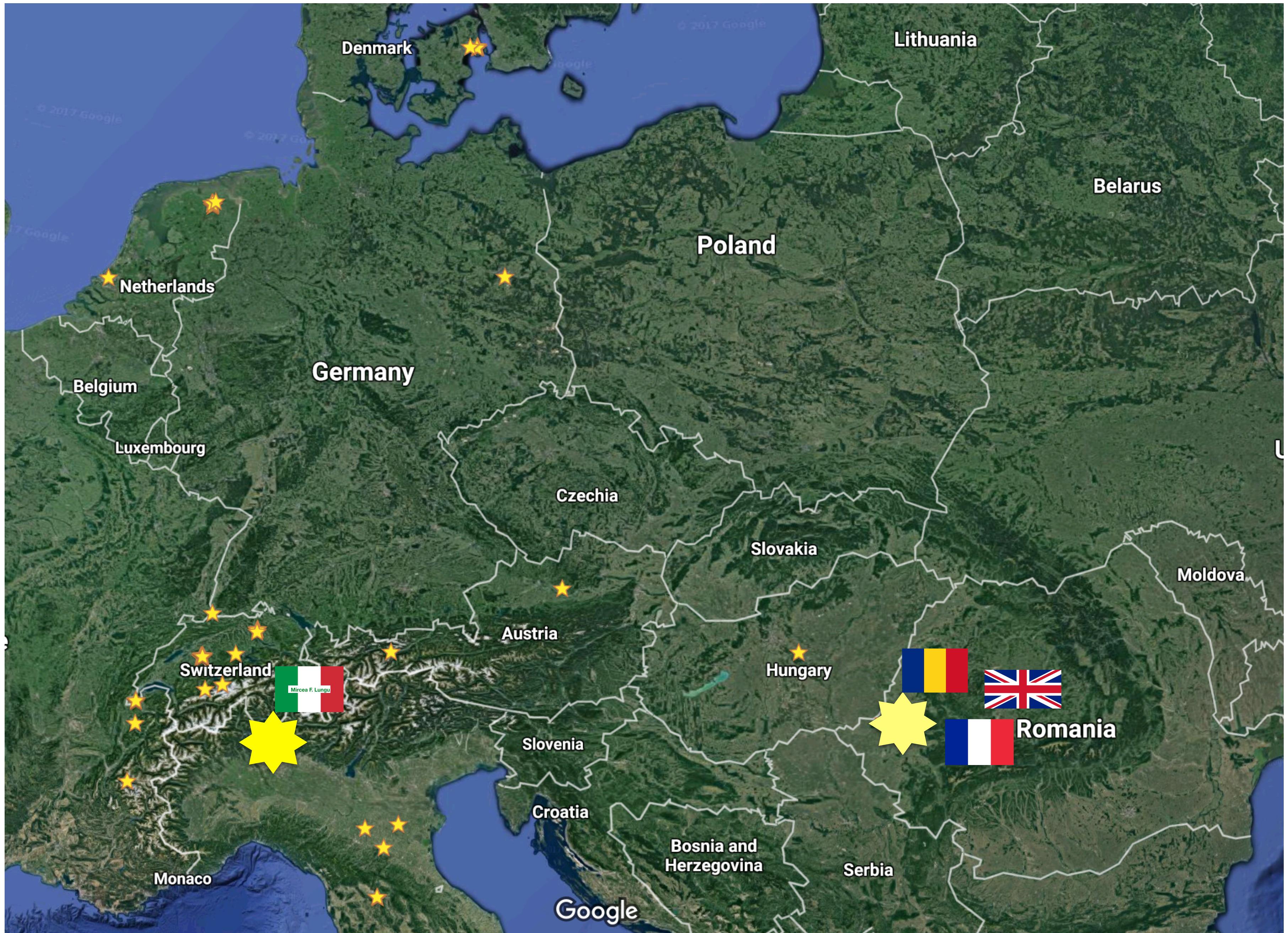
Reading Your Way to Danish Fluency: The Zeeguu Approach

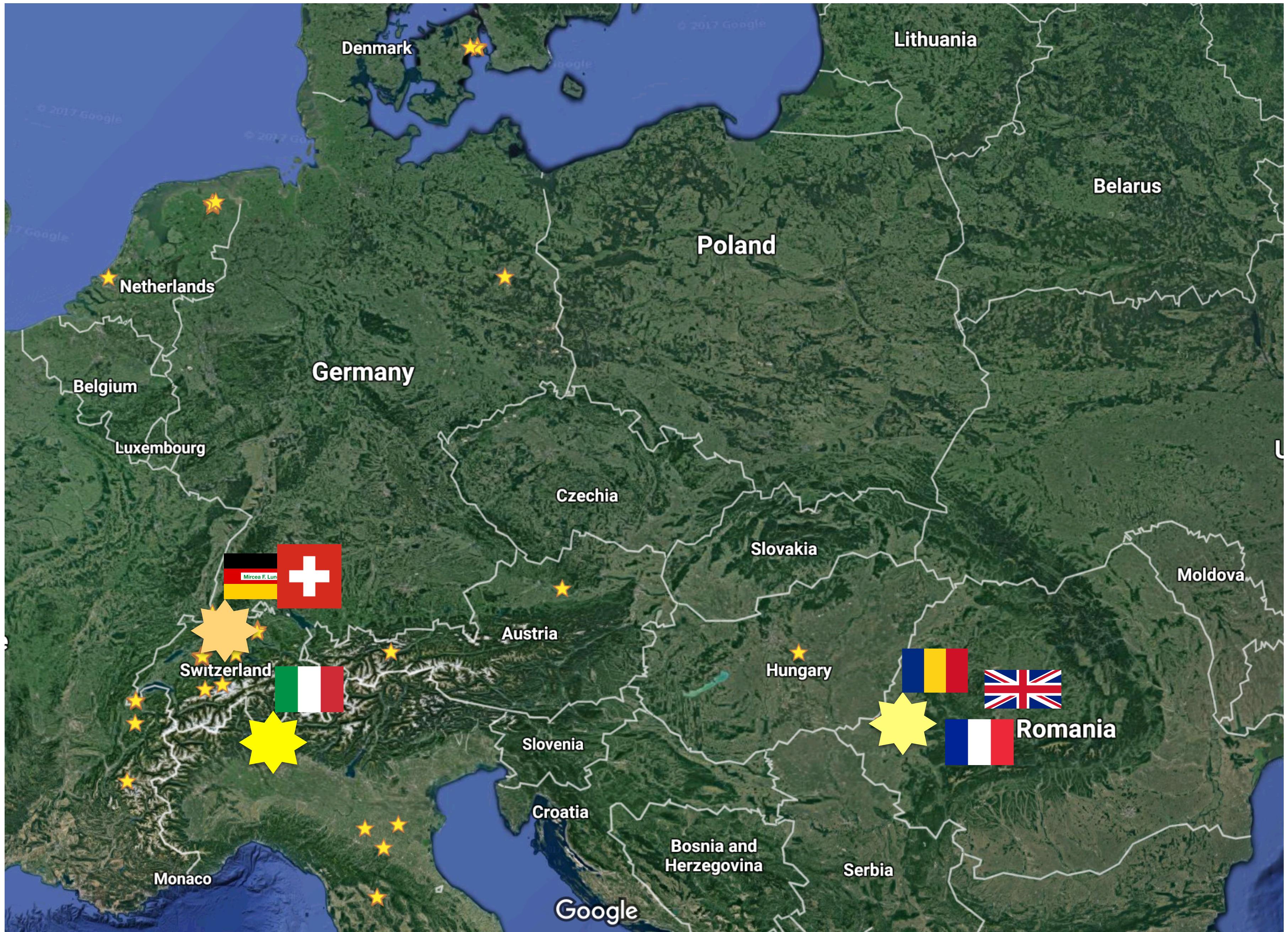
Mircea Lungu

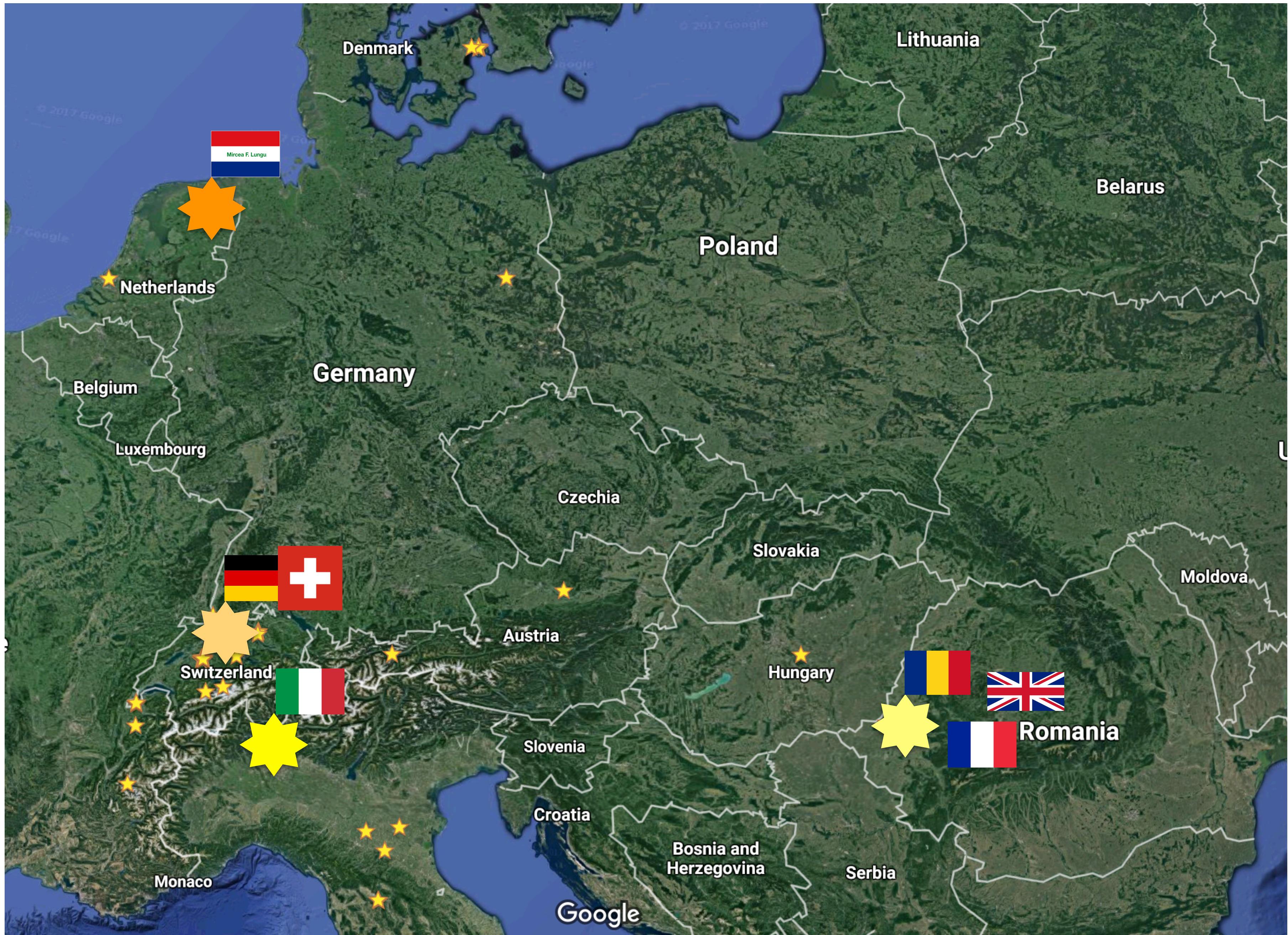
Nov 2025, Copenhagen













Blauzungekrankenheit!



After a while textbooks and apps like Duolingo become boring

Most language materials are designed for the *average learner*

No Freedom =>
Demotivation

**There is no
such thing as
an *average* user**

When U.S. air force discovered the flaw of averages

TR

By Todd Rose

▲ Sat., Jan. 16, 2016 | 11 min. read



<https://www.thestar.com/news/insight/2016/01/16/when-us-air-force-discovered-the-flaw-of-averages.html>

NZZ – Neue Zürcher Zeitung | +

Issues - it-university | NZZ – Neue Zürcher | Der Bund | Ihre Tradi | +

<https://www.nzz.ch>

<https://www.issues.it-univ.ch>

<https://www.derbund.ch>

The screenshot shows the homepage of the NZZ (Neue Zürcher Zeitung) website. At the top, the logo "Neue Zürcher Zeitung" is displayed next to a user icon. Below the logo is a large image of a classical building with columns and statues, with the Latin inscription "LEX JUSTITIA PAX" visible on the pediment. A blue banner at the bottom left reads "KOMMENTAR". The main headline is "Intrigen, Selbstherrlichkeit und Absurditäten: Das Parlament wird von der Schweizer Justiz hinters Licht geführt". A brief summary below the headline states: "Ein umstrittener Bundesgerichtspräsident, skurrile Geschichten am Bundesstrafgericht und Fehden am Bundesverwaltungsgericht – die Schweizer Justiz auf Abwegen." The author's name, Katharina Fontana, and the date, 12.12.2022, are at the bottom left. A bookmark icon is at the bottom right.

The screenshot shows the homepage of the Der Bund website. The logo "Der Bund" is at the top, followed by a user icon and a search icon. Below the logo are navigation links: Bern, Aus den Regionen, Schweiz, and Ukraine-Krieg. A large image in the center shows a man in a white lab coat standing in a cheese factory, leaning against a stack of large cheese wheels. A blue banner at the bottom left reads "Abo Besuch in Gerbers Fondué-Fabrik". The main headline is "Heimat, Natur und Eigenständigkeit – vereint in 400 Gramm Käse". A summary below the headline asks: "Warum ist das Schweizer Nationalgericht so beliebt? Und wie gut ist die Mischung vom Fliessband wirklich? Besuch".

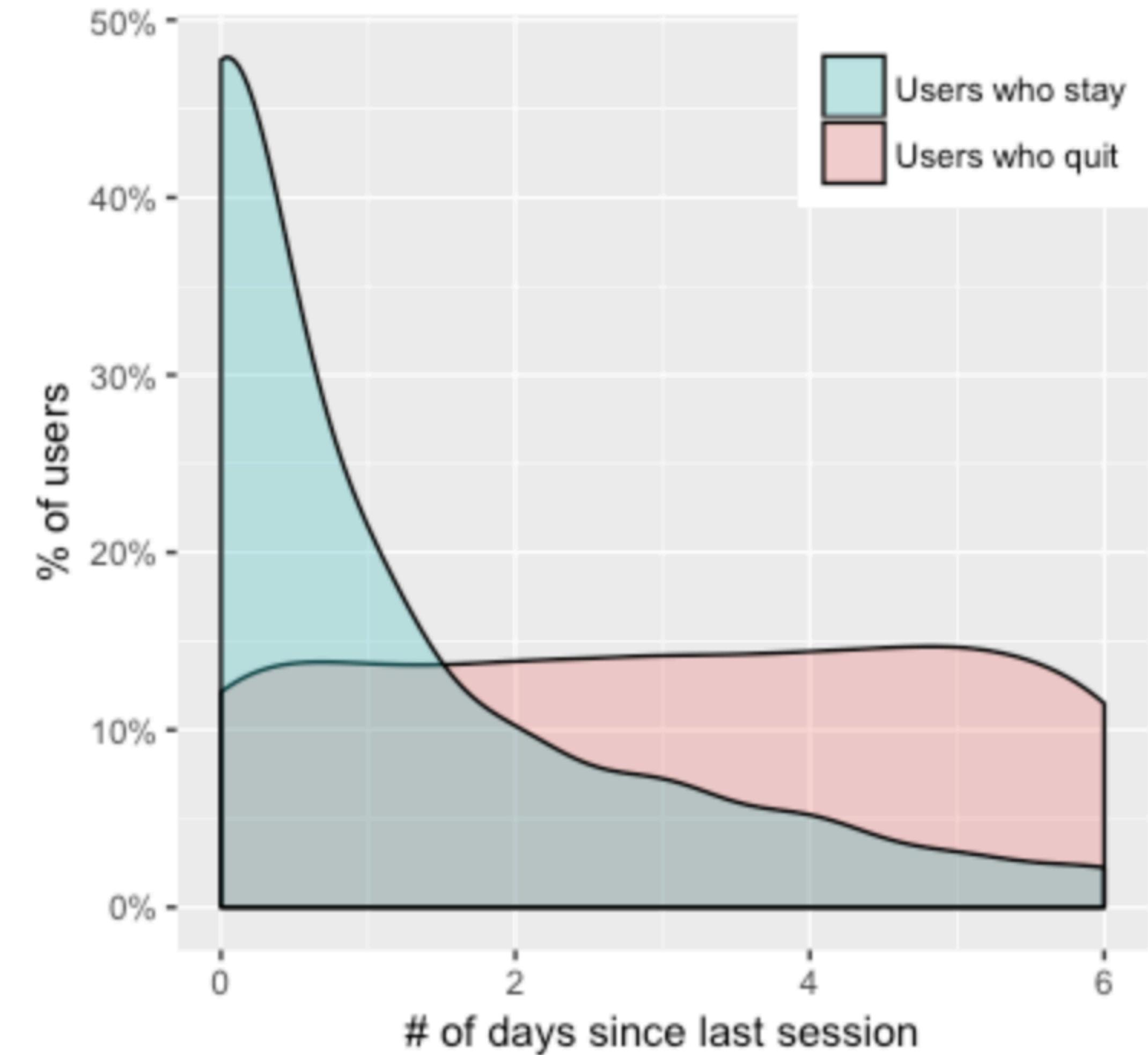
No Competence =>
Demoralization

Best predictor of language learning? Motivation.

Suzanne Hidi and Judith M Harackiewicz.

Motivating the academically unmotivated: A critical issue for the 21st century. *Review of educational research* 70, 2 (2000), 151–179.

William G Brozo, Gerry Shiel, and Keith Topping. **Engagement in reading: Lessons learned from three PISA countries.** *Journal of Adolescent & Adult Literacy*



Vision: Hyper-Personalizing the Internet for Learners

Using personalization *for good*

1. Personalized article recommendations based on **difficulty and topic**
2. Minimal-effort **interactive translations** to compensate for the imperfection of recommendations
3. **Optimal vocabulary consolidation** with exercises based on past interactions in readings

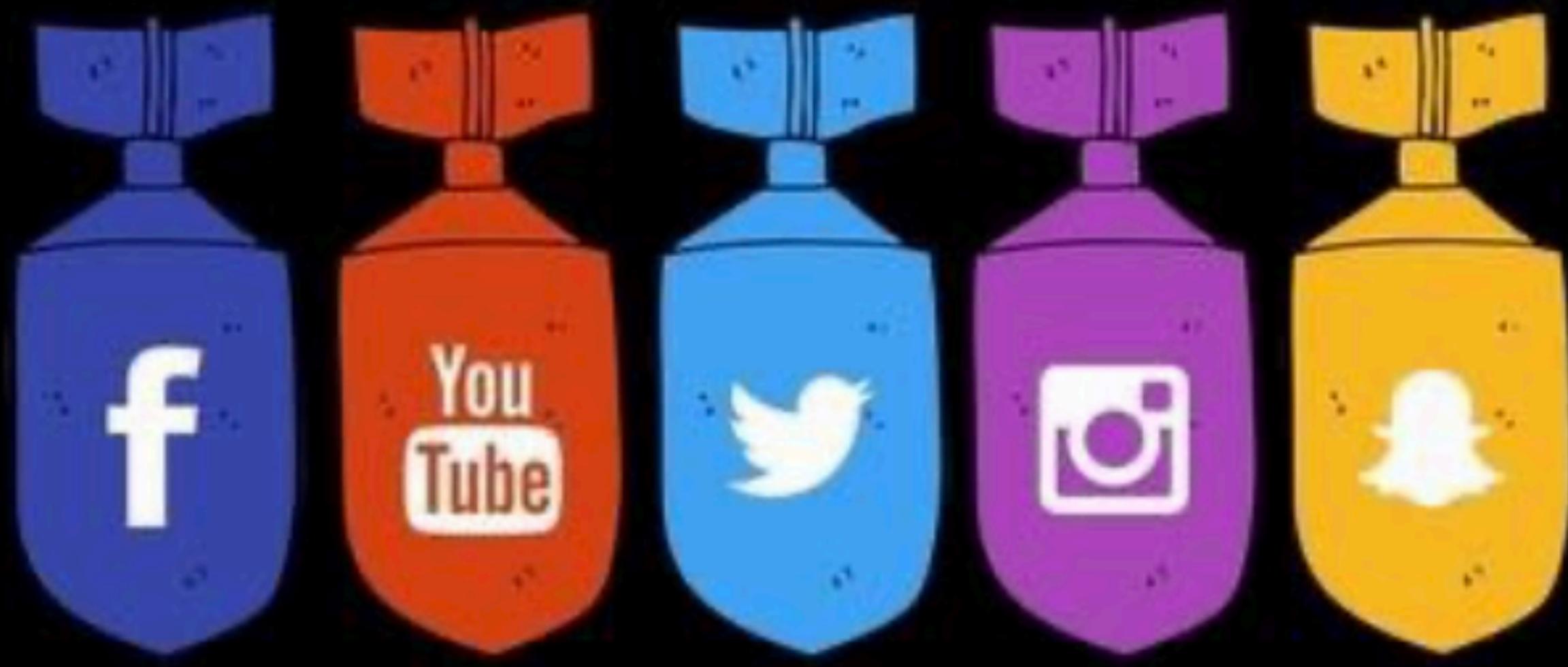
as opposed to...



Apropos of personalization...

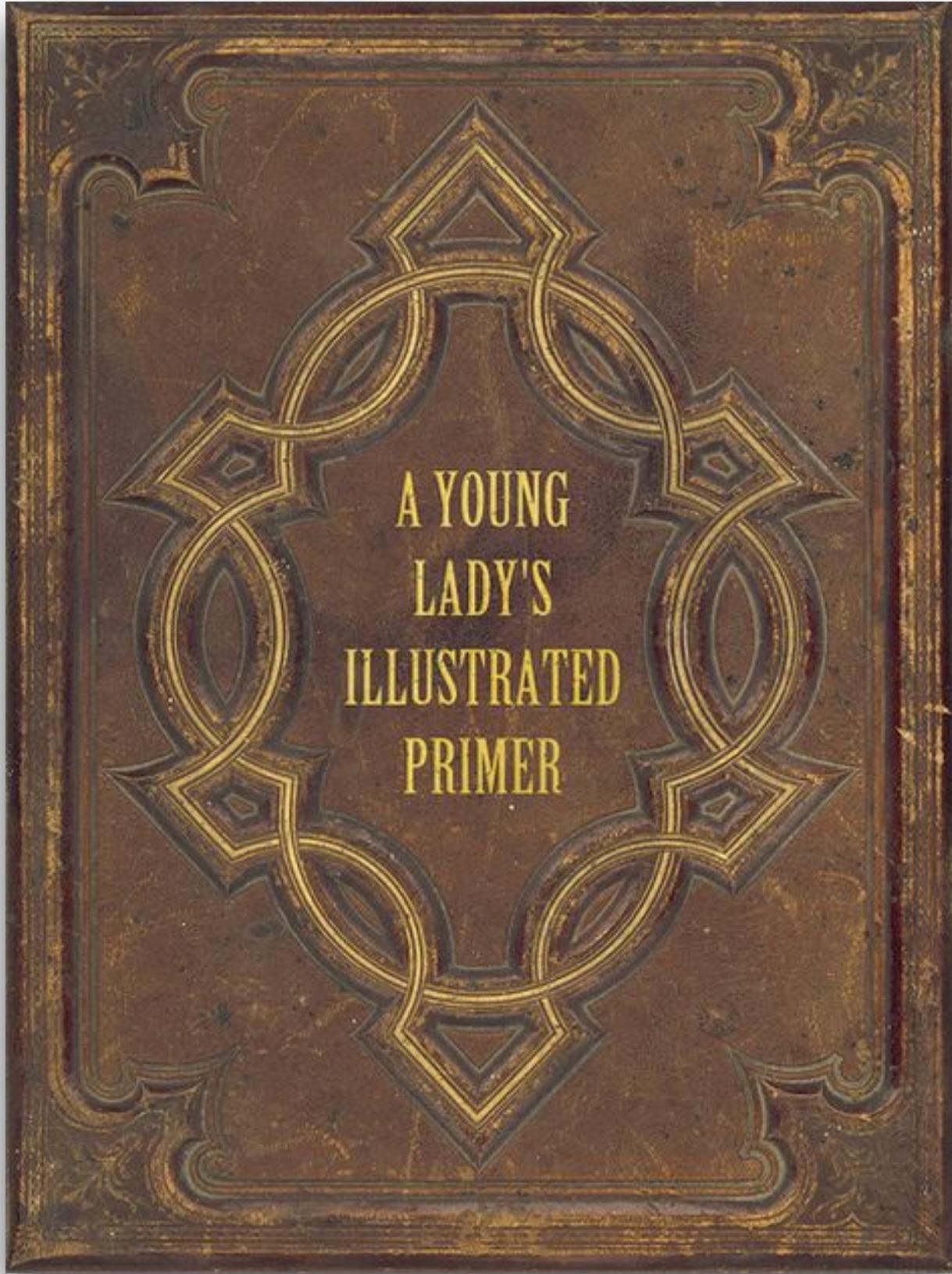
Who else uses it and why?

WEAPONS OF



MASS DISTRACTION

Use personalization
to motivate us to
stay online



... fully interactive, able to teach reading, science, history and martial arts (among other things) while adapting to the current need of the reader.

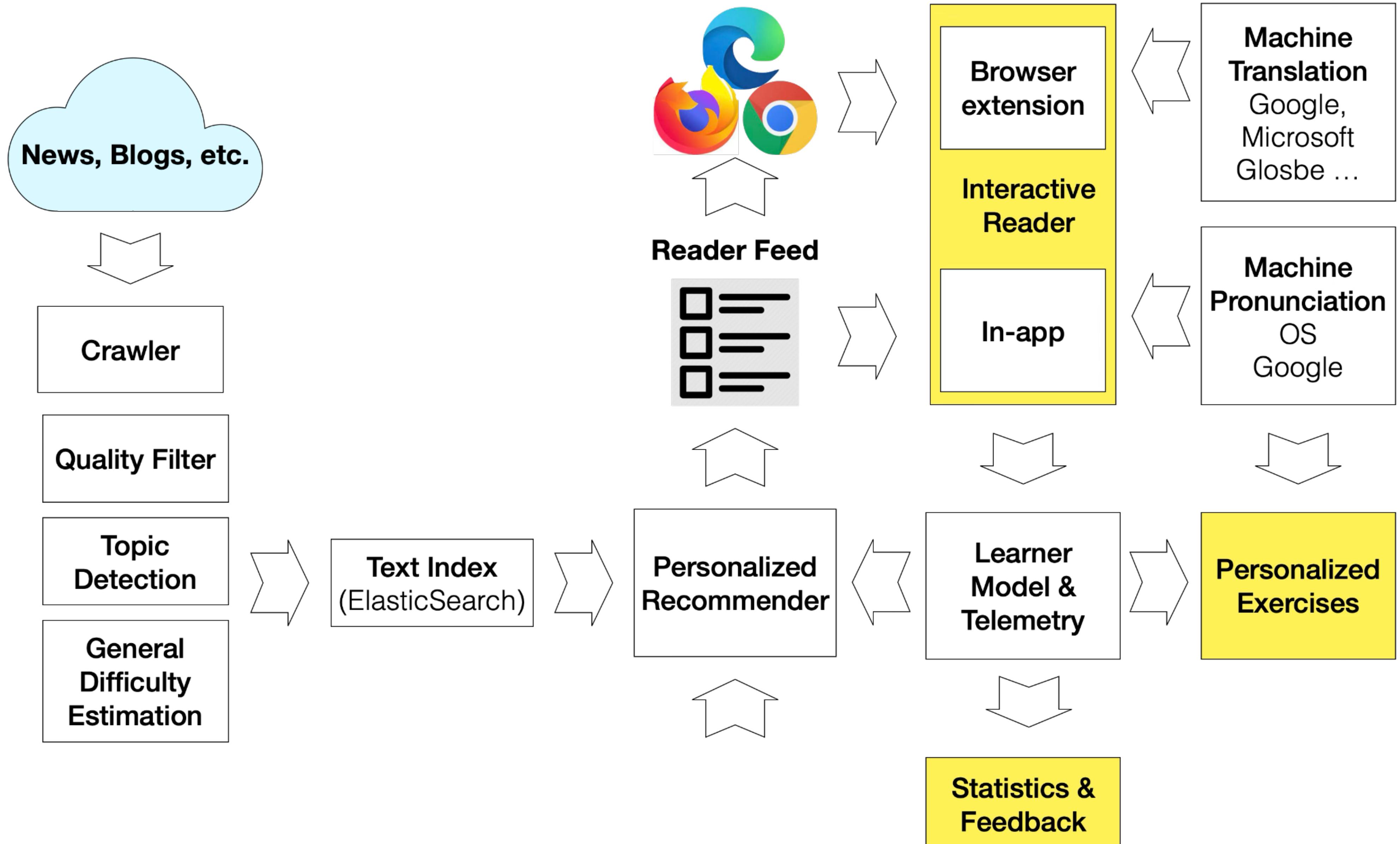
Neal Stephenson, *The Diamond Age or A Young Lady's Illustrated Primer*

Interactive demo reading



How do we recommend articles?

we seed this
with sources we trust



Difficulty Estimation - Traditional with FK

- Calculates three metrics
 - Words per sentence,
 - Syllables per word*
 - Total sentences
- Uses language-specific formulas
 - e.g., English: $206.835 - 1.015 \times WPS - 84.6 \times SPW$)
- Converts to grade between 0-100 (Inverted FK score)
- Usability feedback
 - Users don't care about 67 vs. 68
 - Switched to CEFR levels – A1 (>83) to C2 (<15)

* Using *pyphen* library for accurate hyphenation

Difficulty Estimation - LLM-based

- Since two months we are experimenting with an LLM for estimating difficulty
- We still need to evaluate it

This Danish text is at **CEFR level C1 (Advanced)**.

Here's why this text qualifies as C1 level:

Vocabulary complexity:

- Highly specialized scientific terminology (kosmologi, galakser, lysår, udvidelse, tomrum)
- Abstract concepts requiring advanced comprehension
- Technical phrases like "kosmologiske model" and "Hubble Tension"

Grammatical structures:

- Complex sentence structures with multiple subordinate clauses
- Passive constructions and advanced verb forms
- Sophisticated conjunctions and linking words

Text characteristics:

- Academic/scientific discourse with formal register
- Dense information requiring sustained attention
- Abstract reasoning about cosmological theories
- References to scientific literature and research

Cognitive demands:

- Understanding of complex cause-and-effect relationships

Topic Assignment? Three Types of Topics

1. Website-Specific Topics

- a. Predefined article feeds: The Onion, Lercio, Der Postillon -> "satire"

2. URL Keyword-Based Topics

- a. Some urls we know for sure the topic (e.g. `/politics/article` → "politics")

3. Content-based Inference

- a. Uses semantic similarity via Elasticsearch
- b. Finds $k=9$ similar articles using text embeddings
- c. - Collects topics from these similar articles
- d. - If the most common topic appears in $\geq 50\%$ of neighbors, assigns it

Catering to more specific interests

Searches

- Keyword-based
- Language specific (e.g. København for Danish)
- **Saved** searches are automatically included in recommendations

Exclusion filters

- Keyword based
- Language specific (e.g. død for Danish)
- Disturbing news

Student: “I do not want to read about accidents and deaths!”

Translations

- **Machine translation from multiple APIs**
 - Learner can choose alternatives
 - Learner can upload own
 - Needs to understand enough of the context to choose the right one
- **(Word, translation, context) saved in DB**
 - Context is tokenized with [Stanza](#) (sentence end for context extraction)
 - Translation saved together with token index to DB (in case word occurs twice in text)



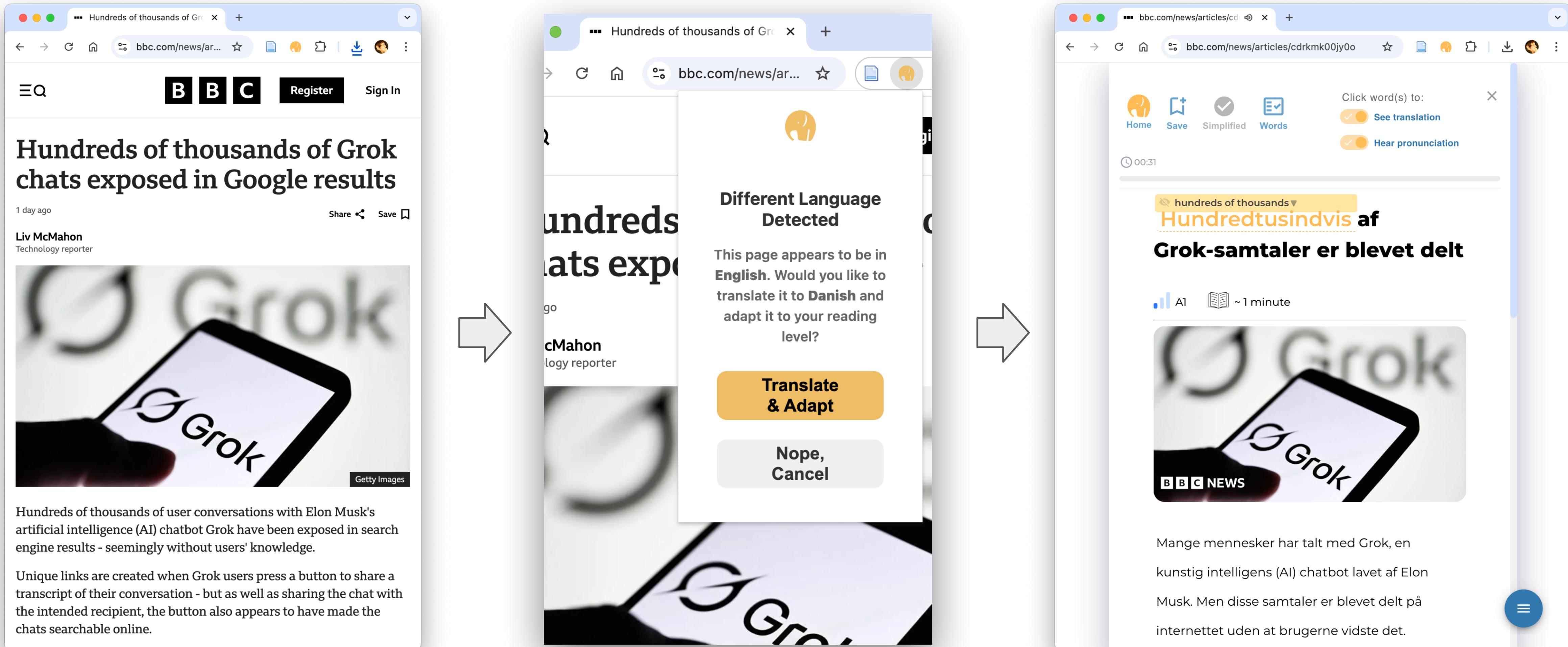
Approx. 1/7 translations => dropdown

The teacher who liked the imperfection

Simplification

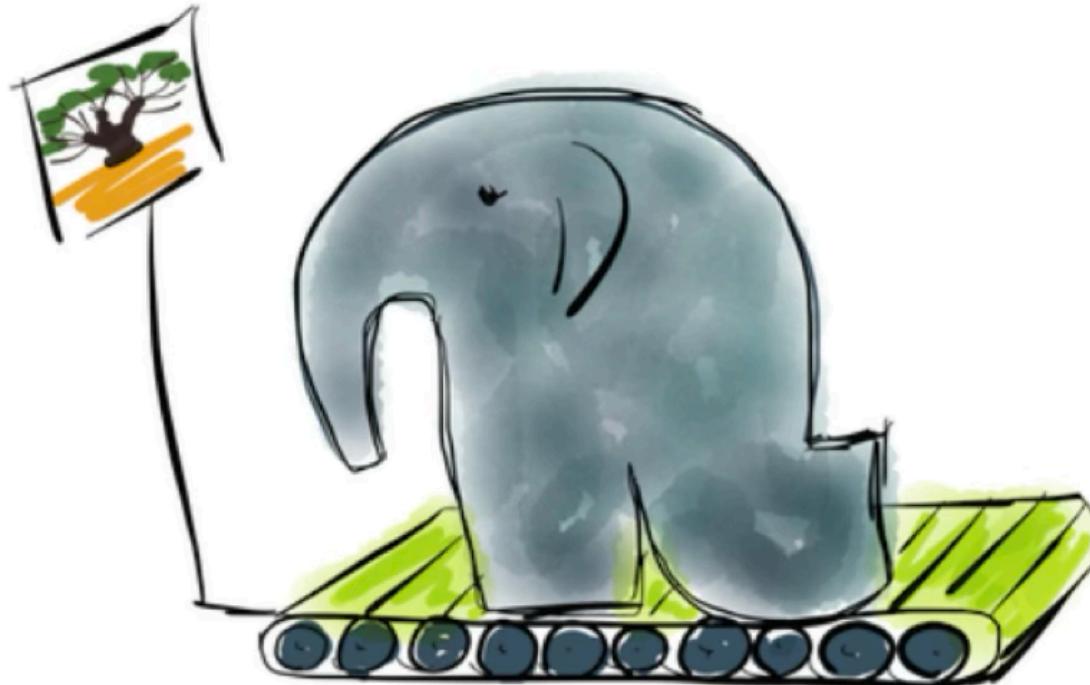
- Problem: **very few articles on the net are A1 or A2**
- Idea: what if we have the LLM automatically simplify existing articles?
 - Same LLM as difficulty estimation
 - For every article we save the simplified versions in the DB
 - When the article would be recommended based on the interests, we swap it in the recommendation feed instead of the original
- Limitations
 - Not evaluated yet
 - Not perfect (machine job)
 - A1, A2 – more related to personal life - we aim to do the simplification for any topic!

Translation + Simplification = Brand new idea!



Towards a lens over the whole Internet?

Interactive demo vocabulary practice



10 Exercise types in 4 levels

Based on two dimensions

- i. **Cognitive Demand:** Recognition (selecting from options) vs. Recall (producing from memory)
- ii. **Skill Type:** Receptive (understanding L2) vs. Productive (generating L2)

Level 1 – Recognition/Receptive (Easiest)

Understanding L2 with multiple choice support:

- **MultipleChoiceL2toL1** – See L2 word, select L1 translation
- **MultipleChoiceAudio** – Hear L2, select from options
- **Match** – Match L2 words to L1 translations
- **ClickWordInContext** – Click on target word in L2 text
- **FindWordInContext** – Identify target word in L2 text

Level 2 – Recognition/Receptive (Context-based)

Understanding L2 in context with options:

- **MultipleChoice** – Read L2 context, select fitting L2 word
- **MultipleChoiceContext** – See L2 word, select fitting L2 context

Level 3 – Recall/Receptive

Understanding L2 and producing L1 (no options):

- **TranslateL2toL1** – See L2, type L1 translation
- **TranslateWhatYouHear** – Hear L2, type L1 translation

Level 4 – Recall/Productive (Hardest)

Generating L2 from memory:

- **SpellWhatYouHear** – Hear L2, type L2 spelling
- **FindWordInContextCloze** – Fill blank with L2 word

Exercise Scheduling

- Spaced repetition is more challenging with multiple exercise types
 - You progress the word every time it's correct
 - A word has to go through four levels before is learned
 - Each level has two different exercises (1 day and 2 days cooling interval)
 - With cooling intervals: and then you move
 - The fastest to learn a word: 8 times correctly, about two weeks
- Maximum number of words the user can have In learning at a given point
- As words are learned, new words are introduced based on
 - Quality checks (expressions not more than 3 words)
 - Word rank (lower ranks have priority)

Han spiser is, **selv om** det er meget koldt udenfor.

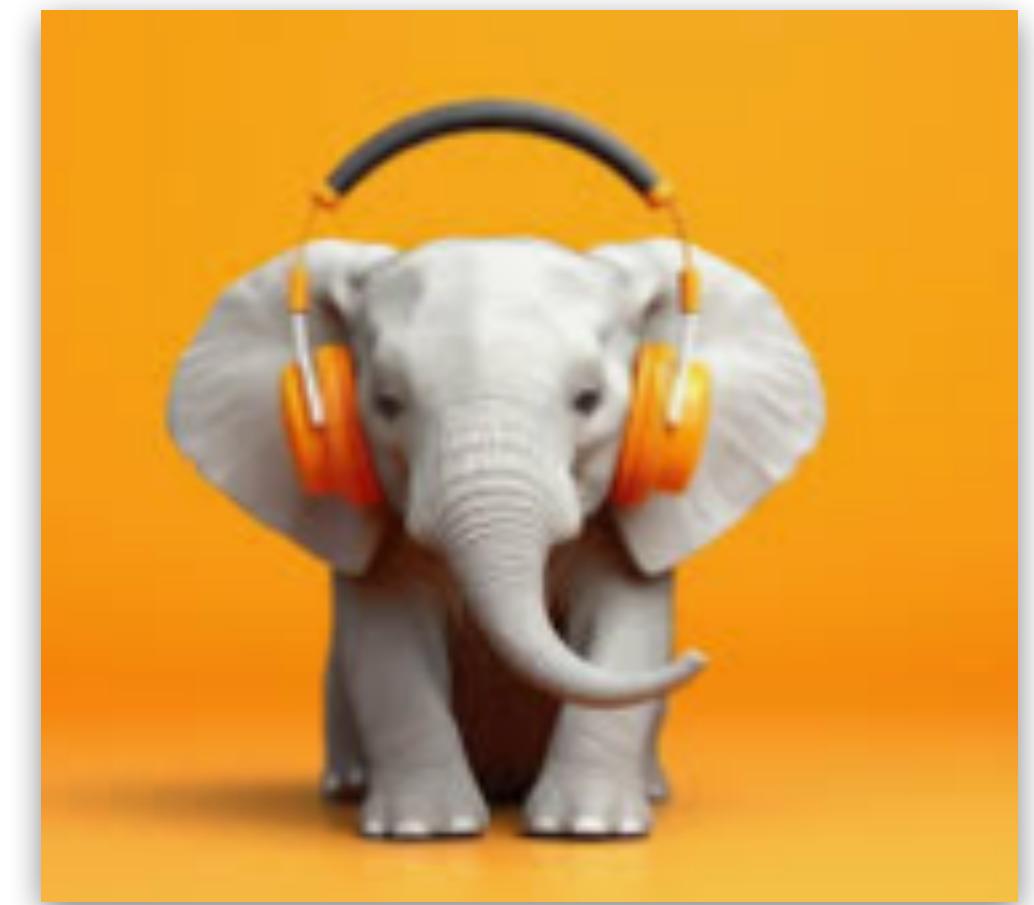
[Change example](#)

although



Audio lessons - useful for beginners

- Every day automatically chose three main words to focus on
- Generate lesson script with LLMs
 - Short dialogue
 - “*Repeat after me*” kind of challenges
 - Sentences on the dialogue
 - A few more sentences
- Based on a traditional approach
 - No speech recognition - too difficult with non-natives
 - Comprehensive **input** is our goal
- Generate audio with TTS technology



Alternative Examples In Exercises

- Problem: **The original context is not always the best**
 - Some sentences do not make sense on their own
 - Some sentences are too long and make exercises unpleasant
- Solution
 - Create a **DB of alternative examples** for a given word/translation pair
 - **Allow the learner to change the example that is used in exercises**
 - Preliminary, subjective, & ongoing evaluation
 - Good because otherwise one memorizes the solution based on context
 - Useful to see alternative uses

Select example for **selv om/although** ×

Det er et stort projekt, som på sigt kan komme til at omfatte 14.000 satellitter, **selv om** der i første omgang kun er planlagt omkring 1.000.
From: Kina i hælene på SpaceX og Starlink: Vil opsende 14.000 internetsatellitter [past encounter](#)

Han spiser is, **selv om** det er meget koldt

Jeg går en tur, **selv om** det regner udenfor

Jeg går ud og leger, **selv om** det regner i dag.

Vi går til stranden, **selv om** klokken er mange.



**Try it on your own
(or in groups of two)**

Go to: <https://zeeguu.org/>
Click on **Start Learning**
Use invite code: **Meetup25**

Plenary Discussion

- Was there something that was not clear?
- What do you think about the app?
- Can you see yourself using the app?
- Do you have ideas of how to improve it?
-

Research Studies

Current Users

- Dutch highschool students - French learners
 - Multiple highschoools
 - Using the platform for many years now
 - Task is - free reading at home
 - Reward is - 0.5 point for 1/2h
- University of Groningen - Dutch learners
 - Using the platform for multiple years
 - Retention is low - one in 50 learners stays with it for the long term

Several small studies with learners

- Participants
 - appreciate the browser extension
 - consider that audio exercises make the system more fun
- Inconclusive study on the impact of progress visualization across learning levels for a word
- Study on augmenting video the way we do articles - participants are positive

<https://github.com/zeeguu/studies>

Calls to action

- It's an open source project - join to help with coding :)
- Just use the app for a while and let me know what could make it better for you
- **Commit to participating in a one-week study (send mail to mlun@itu.dk)**



<https://forms.gle/YiTsThRA6m1yfFV08>



A little language practice every day builds your vocabulary all the way



Android



iOS

1. Tap the Menu icon (:) while viewing this page on your mobile device
2. Select "Add to Home Screen" (or "Install App")
3. Confirm by tapping "Add"



Android



iOS

1. Tap the Share icon (↑) while viewing this page on your mobile device
2. Select "Add to Home Screen"
3. Confirm by tapping "Add"

Backup Slides

“From my grandfathers’s father, I learned to dispense with attendance at public schools and to recognize that on such things money should be eagerly spent”

Marcus Aurelius

From translations to meanings

- Initially every translation would be scheduled for learning
- But then, if a learner translated the same word twice, then it would be scheduled twice
- To avoid this, we model in our DB distinctly the **translation** and the **meaning** it points to
 - Challenge: mapping syntactic variation on the same meaning
 - *Oh, this is quite an **august** entourage.*
 - ***August** entourage is what we have here, I'm telling you.*
 - Challenge: disambiguate between meanings
 - *Let's invite **August** to tag along!*
 - *I proposed we do our vacation in **August**.*

Richer semantic modelling?

But if we know that the learner already knows the word **hurtig** (i.e., fast) then maybe we don't even have to schedule the word **hurtigere** (faster)?

Then again... learner did click on it to get a translation... so maybe they should learn it.

Also, if it's too easy they can remove it from exercises

ClickWordInContext

00:26

faster



Et tildækket sår heller **hurtigere**.

[Other contexts](#)



Next

[Remove word from exercises](#)

Auto-pronounce solution

[Feedback](#) [Share exercise](#)

Vocabulary Progress?



Congratulations! You have learned **149** words so far.

A word is learned when done correctly in exercises in all the four levels. Or when it is marked as "too easy" by you after you have done several exercises with it.

August 2025 (36 words, median rank: 1311)

utrolig 1639 - incredibly	ved 32 - at	løfte 1682 - promise	desuden 1671 - also
undersøge 1619 - investigate	overstået 1251 - ended	virker 351 - seems	gjorde 128 - did
som 36 - like	ødelægger 1384 - ruins	væk 132 - away	målet 1550 - the target
forsvundet 1424 - disappeared	gemmer 1373 - hide	hemmelighed 1344 - secret	
næsten 362 - almost	behøver 250 - need	intet 173 - nothing	bud 1662 - bid
angrebet 1634 - attacked	skilt 1614 - divorced	sager 1463 - things	skød 728 - shoots
ligner 409 - look like	Disse 301 - These	ældre 1293 - older	har tit 1224 - have often
vender 960 - turn over	halvdelen 1461 - half	overleve 1331 - survive	forestille 1328 - imagine
dårlige 1247 - bad	overhovedet 806 - even	forventer 1431 - expects	støtte 1409 - support
kunne 70 - could			

Learning with realistic content... ... not always smarter than DuoLingo!



lungs ▾

Hvis ens **lunger** af en eller anden grund havde taget skade, ville man så kunne

to breathe through your anus ▾

trække vejret gennem anus ?

Det lyder som en joke, men ideen vandt i 2024 Ig Nobel-prisen – Nobelprisens sære tvilling, der hvert år uddeles til spøjs, overraskende og forunderlig forskning.

Da forskerne blev tildelt en

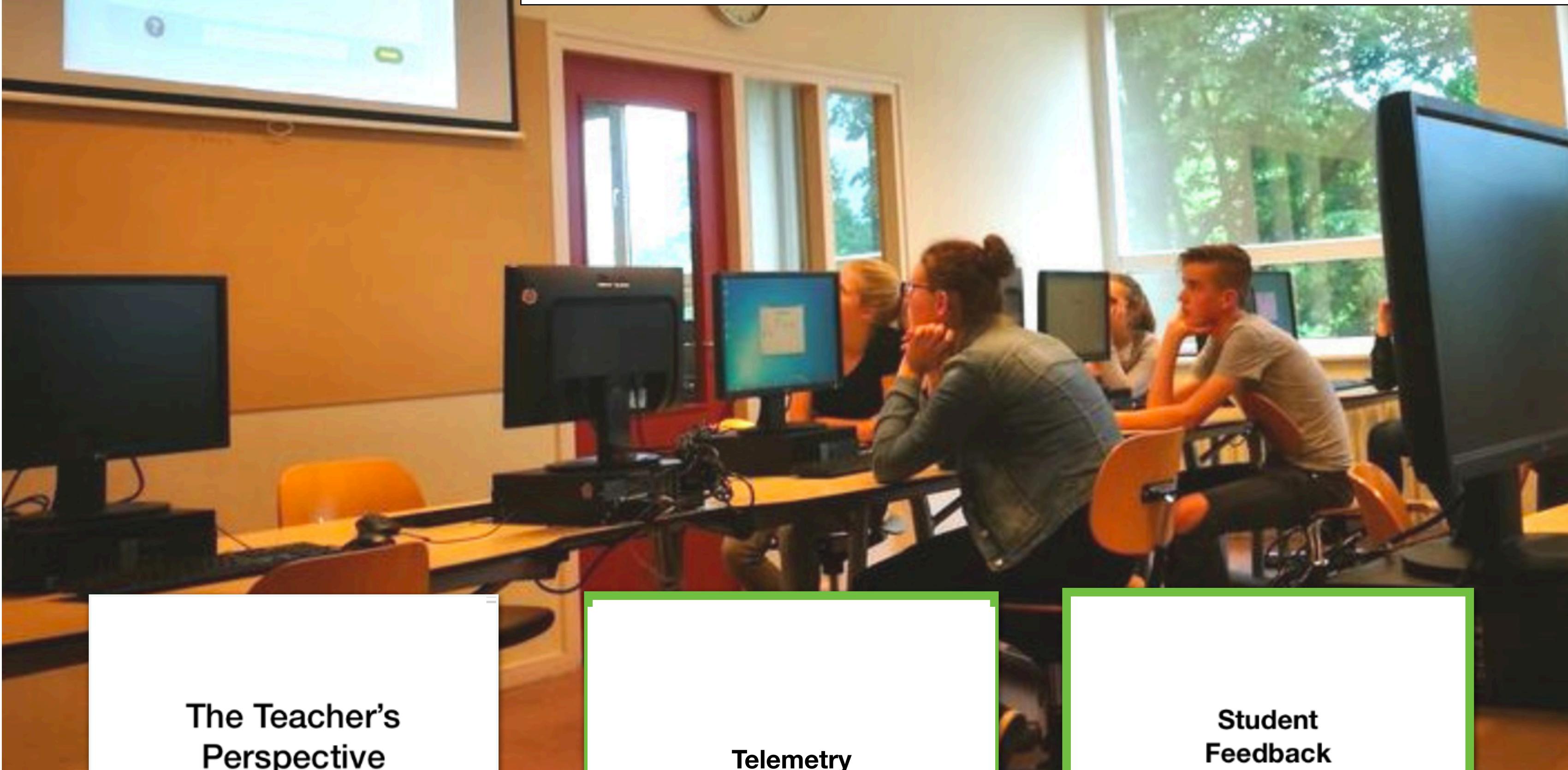
Open

Save

Hide

3 Weeks & 60 French Students

(bilingual Dutch-English)



The Teacher's Perspective

Presentation by Wim Gombert

Telemetry

How do learners use the system?

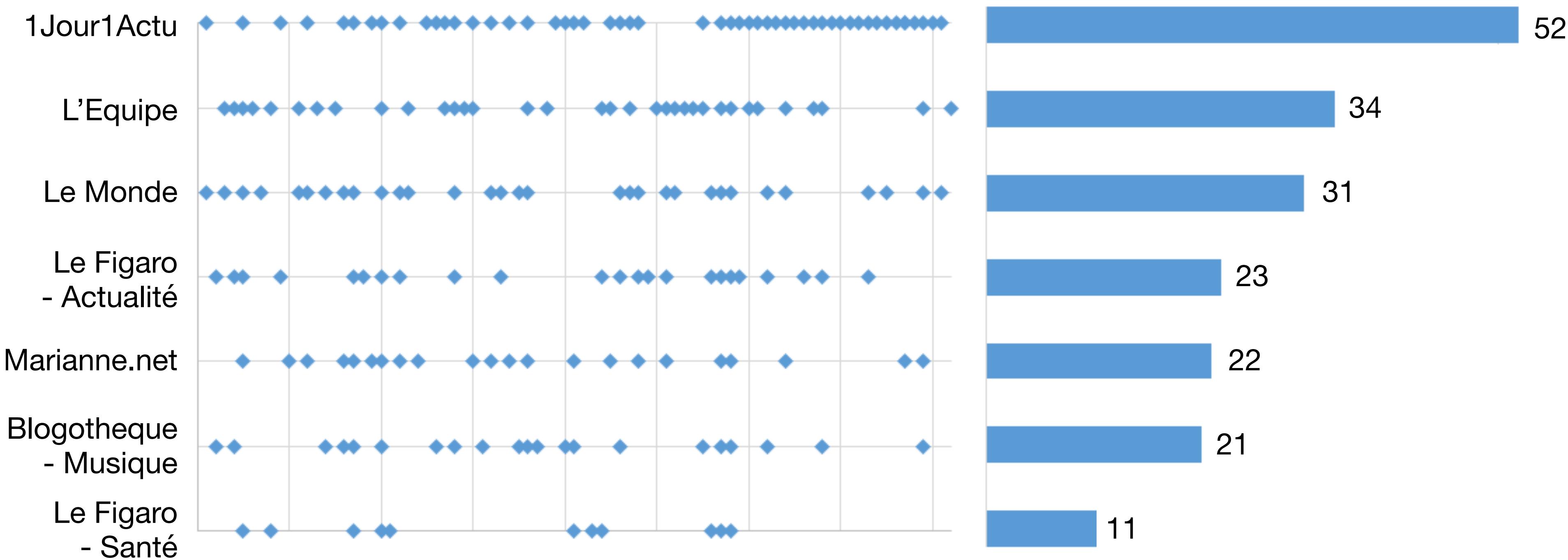
Student Feedback

My vocabulary truly is improving, but you do have to use it more than a few times. A very nice website, easy to use and with nice topics

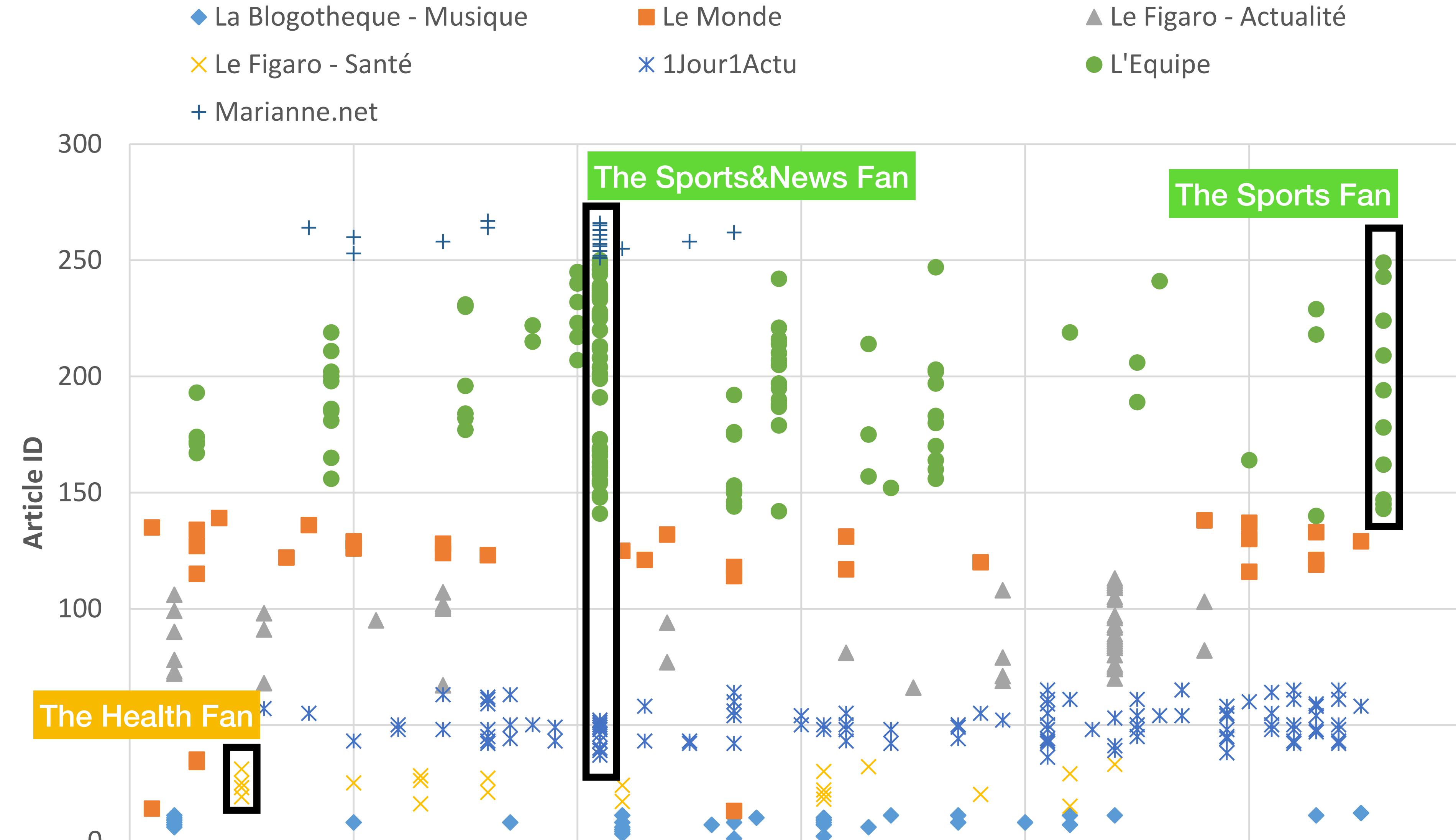
Telemetry

**what are we learning by studying the
interactions?**

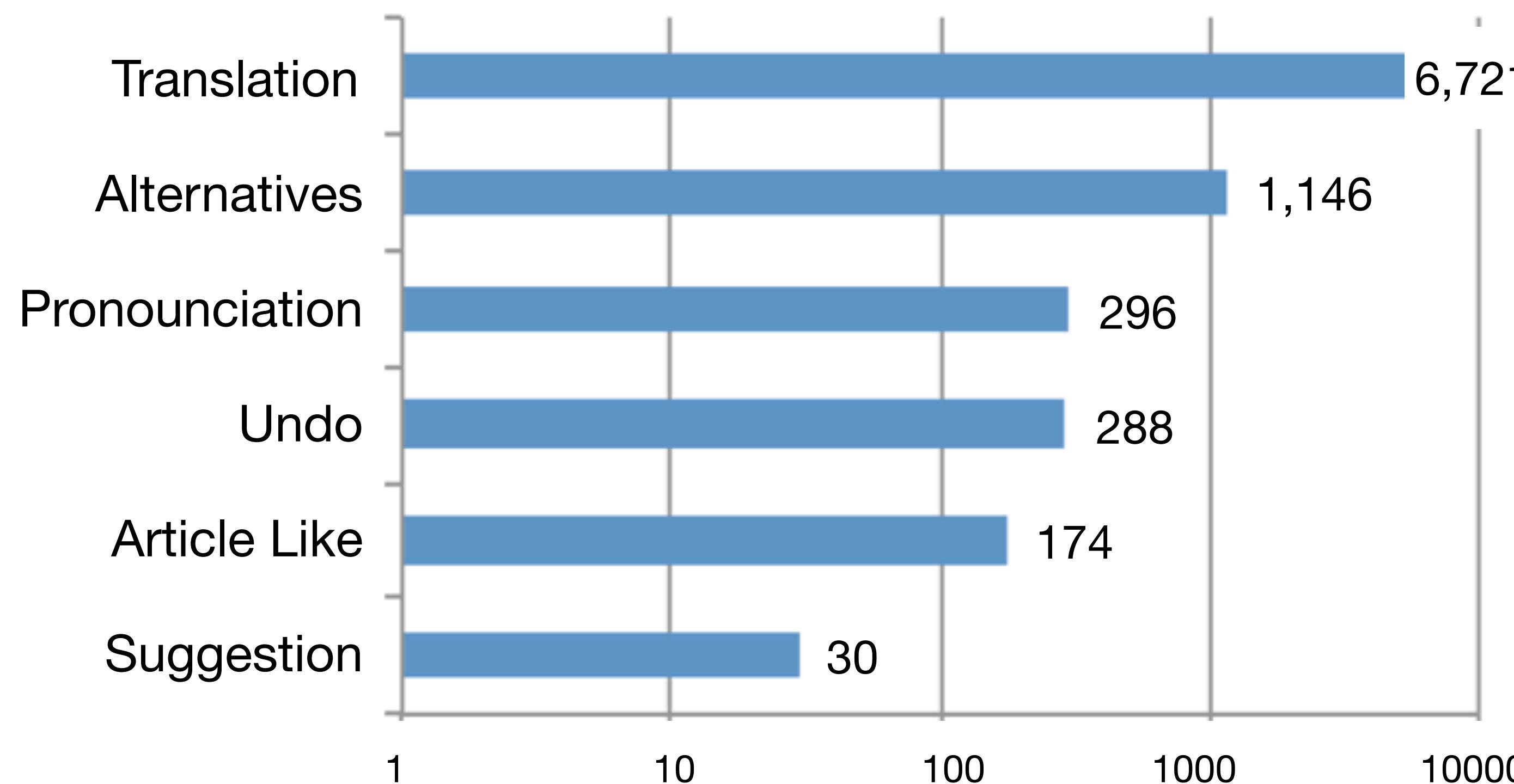
Source Subscriptions



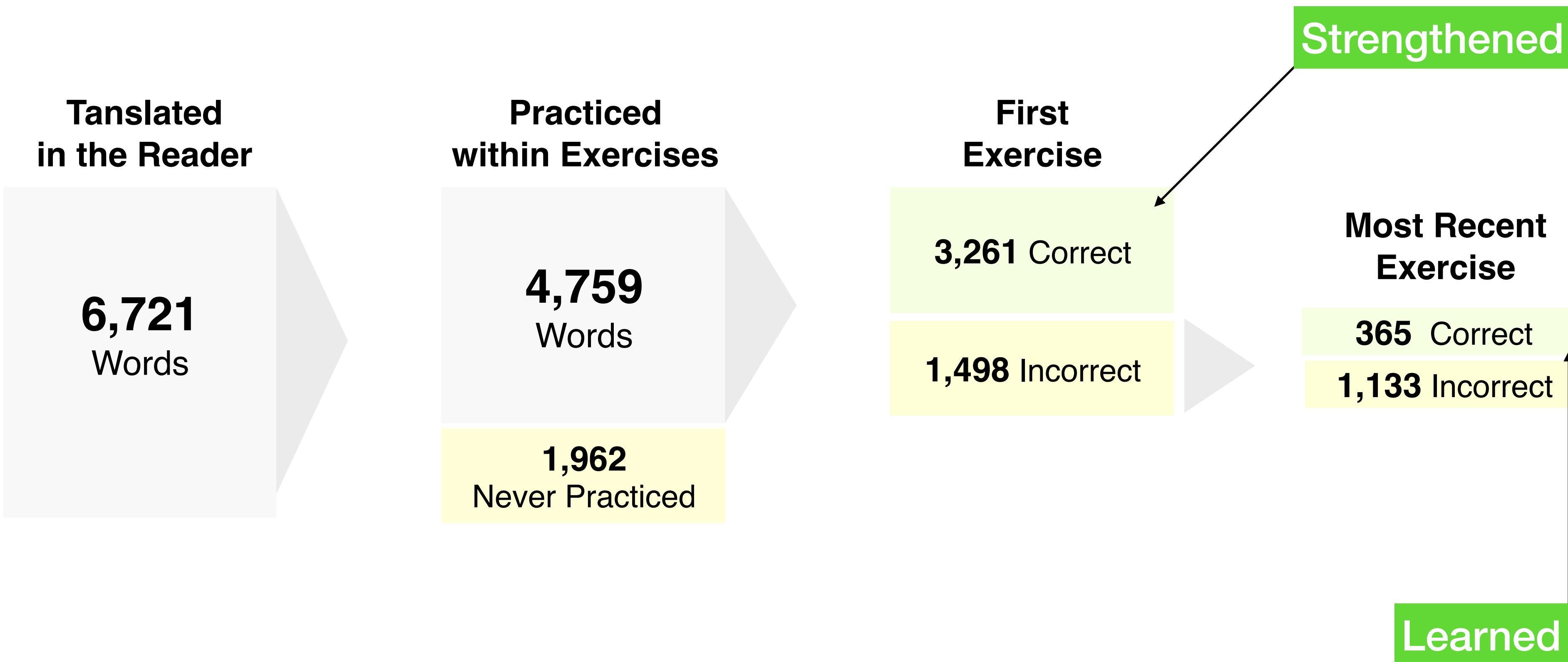
Article Interactions



Usage of Reader Features



Word Interactions In the Exercises



Student Feedback

what do the kids say?

My vocabulary truly is improving but you have to use it more than a few times. Easy to use and with nice topics.

and we did not even pay this one!!!

It's good for improving reading skills. It would be even better if this tool was **available in Dutch**

Good that you can choose what you read. Were it possible to **translate to Dutch, it would be better**

Challenge: Dutch 

Add a choice for **different topics** not only for the sources

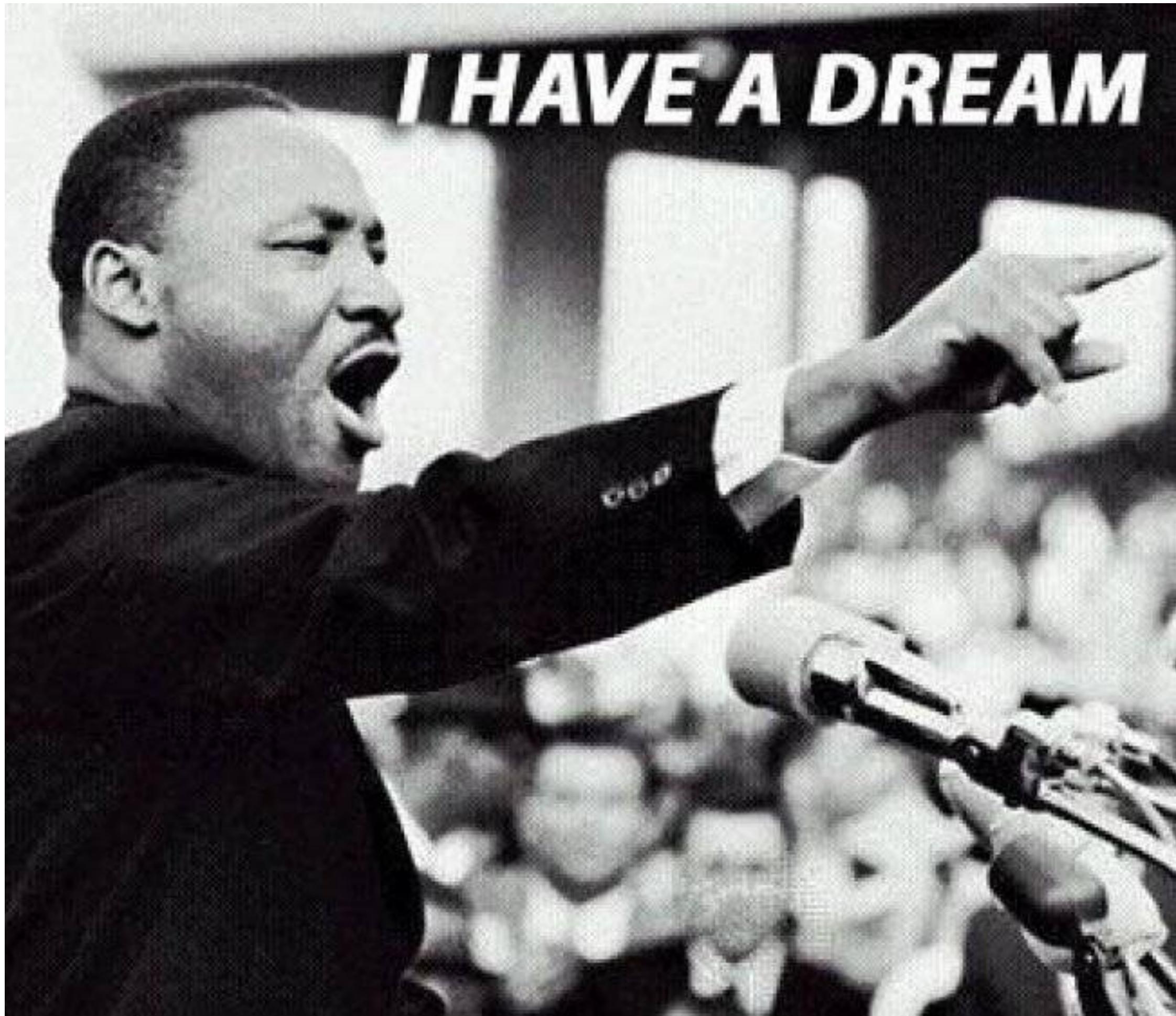
Better display of the articles and **tags** such as
Gaming or **News**

Challenge: Topics

Order articles in different subjects like Animals, Fashion... **not p**

I would like to **avoid articles** with information about **accidents** w
human casualties

Challenge: Filters



**that I don't read
any more news
about the 45th
POTUS...**

Challenge: Filters

Challenge: Topics

Challenge: Filters

almost there...



Languages add

English (x)

French (x)

Spanish (x)

What Do You
Want To Read?

Topics add

Health (x)

Science (x)

Sources add

Le Figaro (x)

New
Scientist (x)

Keywords add

Mexico (x)

Elon Musk (x)

Filters add

Keyword:
Trump (x)

Topic: War (x)

Source: Le
Figaro (x)

As We May Study: Towards the Web as a Personalized Language Textbook

Mircea F. Lungu, Luc van der Brand, Dan Chirtoaca, Martin Avagyan

Johann Bernoulli Institute, University of Groningen
The Netherlands

ABSTRACT

We present a system designed to enable learners of a foreign language to read materials that are personally interesting to them from the web and practice vocabulary with interactive exercises based on their past readings. We report on the results of deploying the system for one month with three classes of Dutch highschool students learning French. The students and their teacher were positive about the system and in particular about the personalization aspects that the system enables.

ACM Classification Keywords

H.5.2. Information Interfaces and Presentation (e.g. HCI): User Interfaces

Author Keywords

language learning; personalization; education

INTRODUCTION

At any given moment, numerous people are learning new languages. English, estimates the British Council, will be learned by two billion people by 2020. Although a plethora of tools and techniques exist for beginners, few exist to support the intermediates and advanced learners, for whom one of the best possible activities is extensive reading [8, 3, 24].

However, when reading in a foreign language, most of the intermediate learners still require language textbooks. Such textbooks are designed by experts who make sure that the texts are simple enough for the desired language level of a broad audience and that exercises that allow the readers to practice newly learned concepts accompany the texts. In spite of becoming more colorful, being sold with complementary audio or video lessons, their main limitation remains unchanged since the last century: by being designed for any average learner they are not exciting for any individual learner [18]. A student interested in sports might not be motivated to read about *Maria who is a babysitter in Spain*. The lack of motivation, a well known problem, especially with young learners [18, 33], might be solved if learners could read materials that are

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI 2018, April 21–26, 2018, Montreal, QC, Canada

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 978-1-4503-5620-6/18/04... \$15.00

DOI: <https://doi.org/10.1145/3173574.3173912>

personally interesting. If they would spend more time reading, their capabilities would increase, and enjoyment would result in further reading, in a virtuous circle [4, 16].

Given the vast amounts of multi-language content available and added daily on the Internet (e.g., blogs, news articles, eBooks) it is likely that every student can find materials that are *personally interesting* for them in almost any language they are learning. This would fit a general trend where old systems designed for the average user are being replaced with personalized attention across domains: in medicine¹, computer security, web design [32], or mathematical education [30].

Although individual components have been proposed before for free reading (e.g. browser extensions) and interactive exercises (e.g. Duolingo), a system that integrates free reading with personalized exercises into a *personalized textbook* has not been evaluated. The first contribution of this paper is describing the architecture and user experience of a system which combines:

1. **Reading comprehension support on both desktop and mobile devices for texts that are interesting to the user.** If ideal comprehension support should work “*without requiring even a single click*” [31], the next best thing is one-click (or one-touch) support. This is combined with a system for estimating text difficulty that allows the learner to choose articles that are within their capabilities².

2. **Integration between vocabulary practice and reading history.** Instead of manually adding words to an external vocabulary practice system, the most relevant unknown words encountered in the readings are automatically scheduled for practice, when possible in the context of the original text, since learning a word in context is more effective [26].

The second contribution is the evaluation of the usefulness and usability of such a system by deploying it with three classes of French learning highschool students for one month. We analyze their usage of the system, their feedback, and the feedback of their teacher to understand when and how such a system fits the modern language classroom.

¹The nascent discipline of *personalized medicine* suggests that analysis of the genetic makeup of an individual may guide health care decisions far more precisely than big group studies do

²This to address the problem of a learner choosing an article randomly from a website and discovering that it is too difficult and giving up.

Read More about the case study in the paper from CHI'18

UX design
data analysis
teacher feedback
student feedback

“Such a system is
critical for language
education in schools,
since the possibility of
choosing their topics of
interest is motivating”

The Teacher