WG1 hands-on workshop

Gothenburg, 5-7 December 2018

Learning materials through crowdsourcing: teachers, perspectives & scenarios

Gothenburg, 5-7 December 2018

L2 exercises

Vocabulary Multiple Choice

B ▾ | Change level

Click to generate!

4 | Den som accepterar Kristus som uppenbarare av Guds väsen har livsavgörande _____ att hämta i bibeln . | halvsyster ▾ | ➡

> ✓ halvsyster
> fastighet
> feber
> telefonsvarare
> vetande
> parlament

3 | POLITIK : För bara några veckor sedan avböjde Sverige att öva tillsammans med Israels _____ . | ✓

2 | Politikerna i landets _____ röstade mot henne i en omröstning i onsdags . | ✓

1 | Uttryck som _____ och be och bocka och buga tillhör en bild av underkastelse . | böna ▾ | ✓

Context-free, understandable, level-appropriate

Appropriate

https://spraakbanken.gu.se/larkalabb/infl-mc

LᴧRK 🜂

Language Acquisition Reusing Korp

LiWriX – Listen and Write Exercise

Mode: 1 Level: A1 **Change settings (new game)**

Appropriate

Correct: 0/1   **Score:** 0

▶   krona   **Check**   **Hide sentences**

– Vi måste vara väldigt försiktiga med hur vi använder varenda **k____** .
Varje **k____** från skattepengarna ska användas till att göra vården bättre .
Lars Ocklind i Åre är en av dem som har höjt priset på mjölk med en **k____** .
– De ska inte få en **k____** till av mig .
Panten för flaskorna blir en **k____** .
Där höjs skatten med en **k____** för varje hundralapp .
Den som pantar en burk får då 1 **k____** i stället för 50 öre .
I annonser står det ofta att mobil-telefoner säljs för en **k____** .
Han får köpa det för en **k____** .
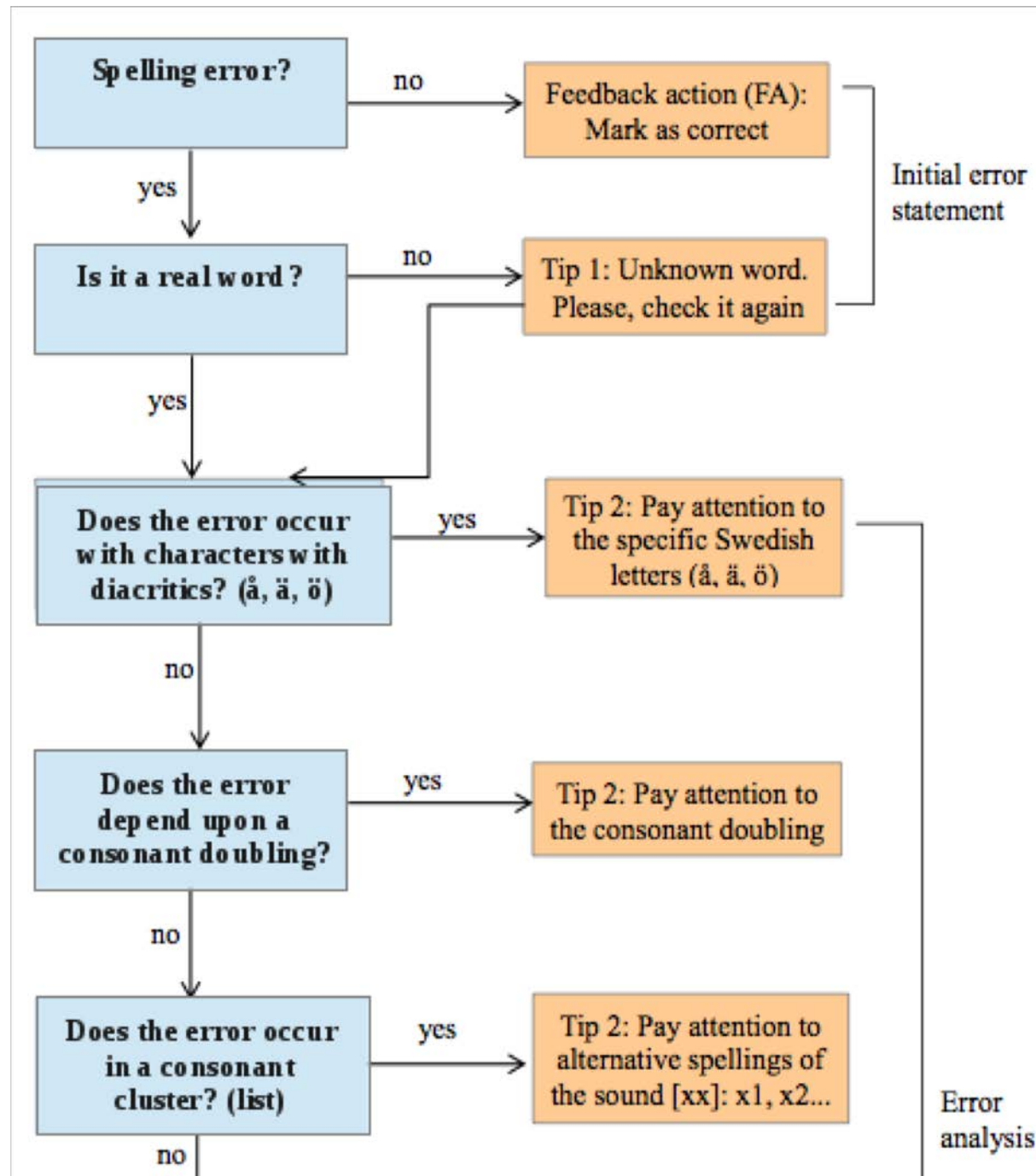Alla de dömda säger att de är oskyldiga och att de inte har tjänat en **k____** på Pirate Bay ,

▶   ☹   barnprogram

Context-free, understandable, level-appropriate

Error data

Error types	Nr,%	Example (correct → *incorrect*)
Competence-based errors	**55**	
Consonant doubling	28	sto**pp**a → *sto**p**a*
Diacritics (å, ä, ö)	23	h**ö**gre → *h**o**gre*
Phonetic errors (e.g. voiced vs voiceless)	25	relevan**s** --> *relevan**z***
Consonant clusters (phoneme-grapheme mappings, incl. cases of homonyms)	20	**sk**ön → ***sj**ön*
Other (unclassifiable)	4	**Israel** → ***visträv***
Performance-based errors	**17**	
Typos (neighbouring keys, addition, deletion, insertion, replacement)	17	förb**ä**ttra → *förb'ttra*
Across one word (phrases & sentences)	**28**	**se en** bild → ***sen** bild*

Error data

Spelling error?
— no → Feedback action (FA): Mark as correct
— yes ↓

Is it a real word?
— no → Tip 1: Unknown word. Please, check it again
— yes ↓

Does the error occur with characters with diacritics? (å, ä, ö)
— yes → Tip 2: Pay attention to the specific Swedish letters (å, ä, ö)
— no ↓

Does the error depend upon a consonant doubling?
— yes → Tip 2: Pay attention to the consonant doubling
— no ↓

Does the error occur in a consonant cluster? (list)
— yes → Tip 2: Pay attention to alternative spellings of the sound [xx]: x1, x2...
— no ↓

Error analysis

6

En hundraårig kvinna har avlidit på sjukhus efter ett brutalt rån i Derby , Storbritannien . Det var i måndags som kvinnan blev påhoppad bakifrån , omkullslagen och rånad på sin handväska . I samband med det fick hon flera skador – bland annat en bruten nacke . Efter två dagars vård har hon avlidit , meddelar polisen . " Det här brottet är alldeles fruktansvärt " , säger polisens utredare Darren De ' ath i ett pressmeddelande . " Det finns folk i Derby som känner personen som bär ansvaret för det här brottet , och jag uppmanar dem att träda fram med all information som kan bidra till att de ansvariga ställs inför rätta . " En 39-årig man frihetsberövades inom ramen för utredningen , men släpptes senare . Polisen söker nu information om en personbil som sågs i området vid tidpunkten för rånet . Kvinnans väska återfanns i ett annat kvarter . Hennes kropp kommer att genomgå en rättsmedicinsk undersökning för att slå fast dödsorsaken .

Resources

https://spraakbanken.gu.se/larkalabb/texteval

What do you want to assess? ❓

Learner essay | **Text readability**

Show all words of the following CEFR level(s) ❓

☑ A1 ▢ ▢
☑ A2 ▢ ▢
☑ B1 ▢ ▢
☑ B2 ▢ ▢
☑ C1 ▢ ▢

Additional options ❓

☑ Mark all potentially incorrect words
☐ Use Spellchecker

Edit text | **Reset**

Evaluation

Suggested overall level: C1

Given the limited amount of underlying data, this CEFR level should be considered as a suggestion and its use as a basis for decisions in high-stakes assessment is discouraged.

Detailed evaluation

Number of sentences	10
Number of tokens	169
Non-lemmatized forms	4
Average sentence length	16.9
Average token length	4.7
Average dependency length	2.27
LIX score	42 (normal)
Nominal ratio	1.64
Pronoun-to-noun ratio	0.31

Linguistic insights based on data modeling

Text Inspector

textinspector.com Text Inspector
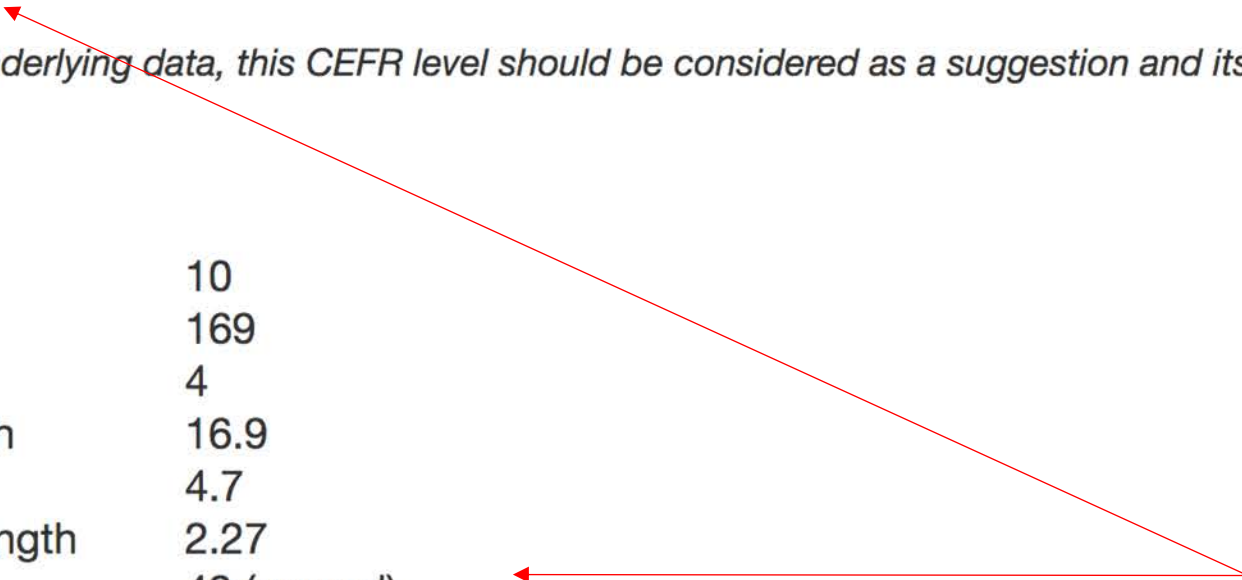
Word List	Types	Tokens
A1	27 (46.55%)	47 (58.02%)
A2	5 (8.62%)	5 (6.17%)
B1	8 (13.79%)	9 (11.11%)
B2	7 (12.07%)	8 (9.88%)
C1	3 (5.17%)	3 (3.70%)
Unlisted	8 (13.79%)	9 (11.11%)
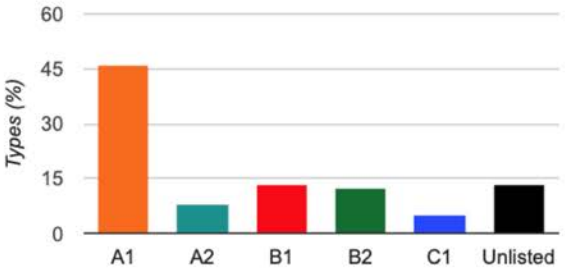
Statistics and details (Toggle all words)

On the main site you can process 10,000 words, and export yo...
(Subscription required).

A1	27 types / 47 tokens 46.55% / 58.02% (Show words)
A2	5 types / 5 tokens 8.62% / 6.17% (Show words)
B2	7 types / 8 tokens 12.07% / 9.88%
C1	3 types / 3 tokens 5.17% / 3.70%
Unlisted	8 types / 9 tokens 13.79% / 11.11%

Input

purpose B1 what A1 is A1 the A1 rationale behind A1 the A1 task B2 learners B2 of A1 english have A1

great A1 difficulty B1 grasping C1 phrasal verbs B1 especially A2 for A1 languages A1 that A1 do A1 not A1

have A1 prepositions B1 like A1 turkish what A1 is A1 its A1 main B1 objective B2 produce B1

equivalents C1 for A1 english phrasal verbs B1 in A1 multiple C1 languages A1 why A1 is A1 it A1

suitable B1 for A1 crowdsourcing it A1 is A1 a A1 well A1 defined B2 task B2 it A1 can A1 be A1

subdivided in A1 ways A2 that A1 do A1 not A1 force B2 everybody A2 to A1 answer A1 everything A2

the A1 knowledge B1 is A1 spread B2 widely B2 in A1 the A1 crowd A2 some A1 people A1 will A1

know A1 polish some A1 portuguese etc

Text Inspector shows the lowest value of each item by default. For more accuracy, click on an item and choose the correct use from the list. Then click on the UPDATE button below to update your statistics.

UPDATE Back

Fill-In-The-Blank-Tests (Cloze tests)

house
wife
car
job
hand
keys
glasses
wallet
soul
phone

"He lost his marbles *yesterday"*

confidence
hope
will
future
degree
friend
head
book
pen
rubber
hat
ring

Tobias Horsmann

Types of Gap-fill Exercises

The students have to ____ the test.

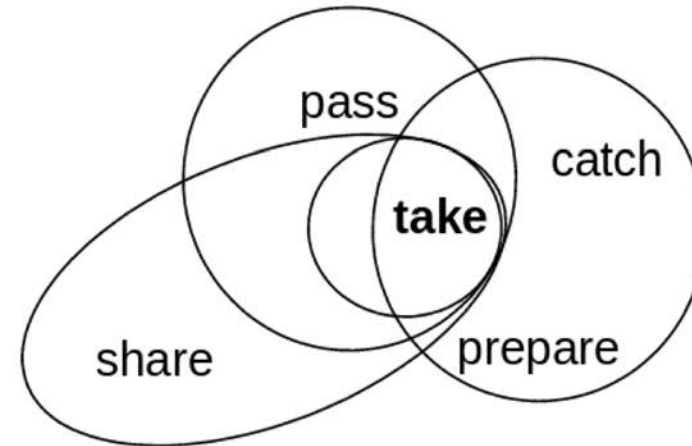The students have to ____ the test.

a) **take**
b) fold
c) entertain
d) fry

- high ambiguity
- production task

- low/moderate ambiguity
- recognition task

Michael Wojatzki

Bundled Gap-fill [Wojatzki et a. 2016]

The students have to ___ the test.

Their cook will ___ three salmons.

All passengers should ___ their seats.

Both authors ___ credit for this.

- low ambiguity
- production task

Example: VIEW

- Corpora creation costs both in time and money, but:

- Well-documented, representative, reliably annotated and available corpora are used far beyond their initial research purpose
  - Penn TreeBank (Marcus et al., 1993; cited 6813 times), is still used for research (e.g. Pawar, A., & Mago, V., 2018)
  - ICLE (Granger, 1998; cited 358 times) → modern research (e.g. Möller, 2017)

- Whereas tools trained on corpora get outdated as research makes progress
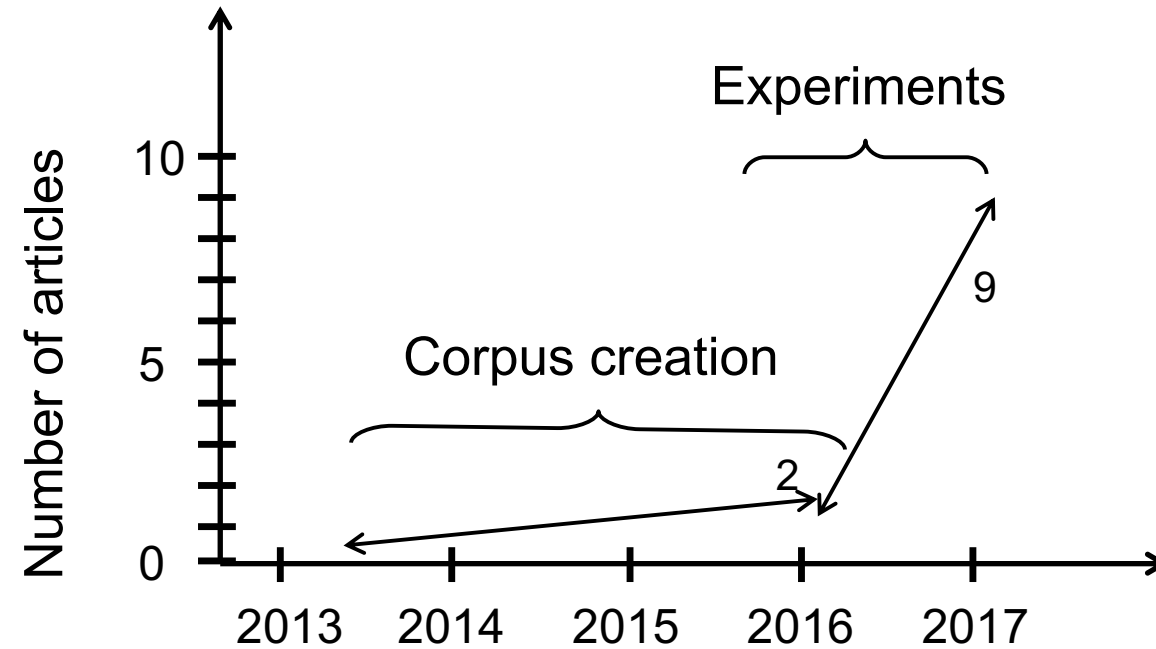
→

Data

Curios "time & effort" fact:
Data vs experiments

Essay corpus, SweLL-corpus, creation and SweLL-based publications

Lifetime of corpora vs tools

- Corpora creation costs both in time and money, but:

- Well-documented, representative, reliably annotated and available corpora are used far beyond their initial research purpose
    - Penn TreeBank (Marcus et al., 1993; cited 6813 times), is still used for research (e.g. Pawar, A., & Mago, V., 2018)
    - ICLE (Granger, 1998; cited 358 times) ➔ modern research (e.g. Möller, 2017)

- Whereas tools trained on corpora get outdated as research makes progress

➔

Tools decay, data stay

Annotation of data

Annotation…

- …is now the place where linguistics hides in NLP (Fort, 2016)
  - Parts of speech
  - Base forms of the words (lemmas)
  - Syntactic and semantic information
  - …

Karën Fort. 2016. Collaborative Annotation for Reliable Natural Langugage Processing. Wiley.

Annotation…

- …can "hide" other disciplines than linguistics
    - *(e.g. so called)* Error annotation
    - Target skills
    - Receptive vs productive skills
    - Level of proficiency in a (second/foreign) language
    - Text genres
    - …

Annotation…

- Manually by experts
- Automatically (based on previously manually annotated data)
- Through crowdsourcing, i.e. by non-experts

Reliability of tools & algorithms
teacher attitudes

Common "beginner" pitfalls
when developing ICALL applications

Lesson 1

- Do not underestimate the time it takes to collect and prepare data

Time-effect ratio consequences

- Researchers skip compiling own data
  → use what is available
  → in the end often targeting English

Lesson 2

- Take time to study legal regulations, not to waste previously collected data

    → There are loopholes, but not without information loss

Different worlds?

L2 teachers vs NLP researchers

or do we really work interdisciplinary?

- Terminology (e.g. corpora & annotation vs genre pedagogy & processability theory)
- Example: normalization

Two views

- Teachers
  → want control

- NLP researchers
  → want to automatize

Two views

- Teachers
  → want control

  → keep to fixed practices

  → want 100% correctness

  → sceptic about automatic
  solutions

- NLP researchers
  → want to automatize

  → want to revolutionize

  → work within reasonable
  margins

  → enthusiastic about
  automatic solutions

Two views

- Teachers

  → want control

  → keep to fixed practices

  → want 100% correctness

  → sceptic about automatic solutions

- NLP researchers

  → want to automatize

  → want to revolutionize

  → work within reasonable margins

  → enthusiastic about automatic solutions

Meeting half-way?

- Educational Testing Service (ETS)

    → from Criterion/e-rater to Language Muse (20 years)

    → http://languagemuse.10clouds.com/

- WERTI/VIEW
    → no automatic pre-selection of exercise text

Lesson 3

- Take time to study what makes L2 researchers & L2 teachers "tick", and vice versa

- Be ready to compromise → on both sides

Application life cycle

Recent example

Some add-ons have been disabled

The following add-ons have not been verified for use in Firefox. You can find replacements or ask the developer to get them verified.
Learn more about our efforts to help keep you safe online.
Developers interested in getting their add-ons verified can continue by reading our manual.

⚠ **VIEW could not be verified for use in Firefox and has been disabled.** **More Information**

VIEW (disabled)
VIEW is an intelligent computer-assisted language learning (ICALL) system designed to provide supplementary lan… More   Remove

⚠ **WERTi could not be verified for use in Firefox and has been disabled.** **More Information**

WERTi (disabled)
WERTi is an intelligent computer-assisted language learning (ICALL) system designed to provide supplementary la… More   Remove

Lesson 4
more of an insight

- (Most?) ICALL research remains within ICALLers' "comfortable zone", i.e. on their desks; at the best goes into a prototype

- Researchers can at most develop prototypes as a "proof-of-concept", but cannot maintain full-scale applications

- There is a need for a new type of funding, the one that would bring research findings to end-users

MWEs and crowdsourcing workshop

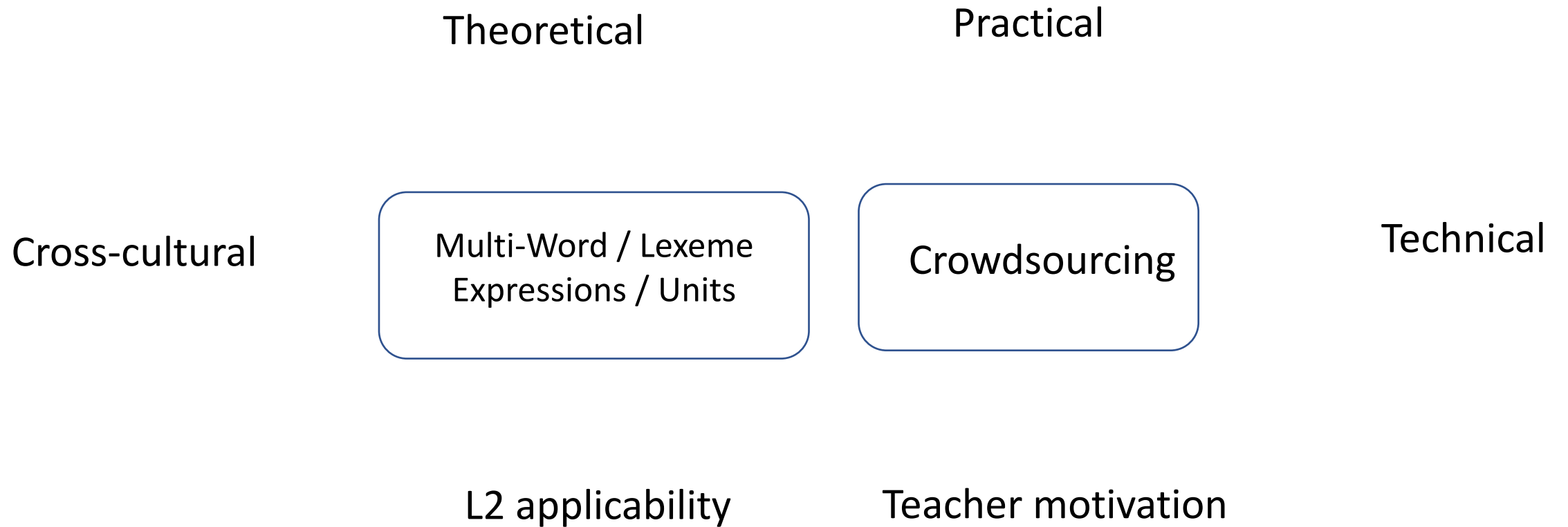- https://spraakbanken.gu.se/eng/wg1-dec18-gbg

  - Analysis of the experiment results
  - Assessment – can crowdsourcing be used for language learning resources and materials?
  - Who should be the crowd?
  - How many are a crowd?
  - Is it possible to set up a multi-lingual experiment?
  - Etc-etc

Lesson X...

...to be continued

- ...
- ...

During the workshop…

Theoretical

Practical

Cross-cultural

Multi-Word / Lexeme
Expressions / Units

Crowdsourcing

Technical

L2 applicability

Teacher motivation

Motivation for this hands-on meeting

- Crowdsourcing for language learning
  - Assessing the quality of text questions (*Chinkina & Meurers, 2017*)
  - Feedbook (*Ramon Ziai, Bjoern Rudzewitz, Kordula De Kuthy, Florian Nuxoll & Detmar Meurers, 2018*)
  - DuoLingo (*Settles, Brust, Gustafson, Hagiwara & Madnani, 2018*)
  - …
  - but --> not much to go, in fact, to validate crowdsourcing for LL

- We need to have a proof-of-concept that
  - crowdsourcing is valid for annotating/creating language learning data

Conclusions about MWE experiment

- Together during the WG1 workshop – analysis & conclusions

- Likely to see clashes between teachers, SLA and NLP researchers