

Writing assistants: from word lists to NLP and artificial intelligence

S. Verlinde, L. De Wachter, A. Laffut, K. Blanpain, G. Peeters, K. Sevenants, M. D'Hertefelt
with (a little) help from F. Jacobs

KU Leuven – Leuven Language Institute

Writing assistants?

online dictionaries

LEAD
Netspeak
Just-the-Word
Writeful
...

AWSuM
ColloCaid

Grammarly
...

writing assistants

= an intuitive [...] resource

accessed from within digital writing environments

(Frankenberg-Garcia e.a. 2019)

Text review

Vocabulary resources

LANGUAGE ISSUES

- ▶ Spellchecker UK
- ▶ Spellchecker US
- ▶ Language errors
- ▶ Tense use
- ▶ Adverb Position

COHERENCE AND COHESION

- ▶ Readability
- ▶ Sentence length
- ▶ Paragraphing
- ▶ Reference
- ▶ Linking words and phrases
- ▶ Recurring words and patterns

STYLE

- ▶ Personal language
- ▶ Informal language

Schrijfhulp Academisch Nederlands
Gerealiseerd door het Instituut voor Levende Talen

Tekst controleren

Tekst verrijken

STRUCTUUR EN SAMENHANG

- ▶ Tekstanalyse
- ▶ Verwijswoorden
- ▶ Signaal- en structuurwoorden
- ▶ Frequente inhoudswoorden
- ▶ Vaak terugkerende patronen
- ▶ Zinslengte
- ▶ Alinealengte

STIJL

- ▶ Formele en archaische woorden
- ▶ Passieven
- ▶ Omslachtig taalgebruik
- ▶ Naamwoordstijl
- ▶ Vage en algemene woorden
- ▶ Persoonlijk taalgebruik
- ▶ Informele, spreektaalige en gekleurde woorden

SPELLING

- ▶ Spelfouten
- ▶ Aaneenschrijven of niet?
- ▶ Afkortingen

Skrifhulp Afrikaans
Maandeliks gemaak deur die Instituut voor Levende Talen



Teks hersiening

Teks verryking

STRUCTUUR EN SAMEHANG

- ▶ Teksanalise
- ▶ Skakelwoorde
- ▶ Diskoersmerkers
- ▶ Veel voorkomende woorde
- ▶ Veel voorkomende frases
- ▶ Sinslengte
- ▶ Paragraaflengte

STYL

- ▶ Formele en argaise woorde
- ▶ Passiewe konstruksies
- ▶ Persoonlike taalgebruik
- ▶ Informele woorde en woorde wat kan aanstoot gee/Onvanpaste woorde/Nie-akademiese woorde
- ▶ Omslagtige taalgebruik
- ▶ Naamwoordstyl
- ▶ Vae en algemene woorde

SPELLING

- ▶ Spelfoute
- ▶ Afkortings
- ▶ Aaneengeskrif of nie?

vrt taal

Schrijfassistent Nederlands

Wachten op input

Plaats je tekst hier...

CONTROLEER DE SPELLING

- > Spelling
- > DT en co.

vrt taal

Schrijfassistent Nederlands voor anderstaligen

Plaats je tekst hier...

CORRECT GESCHREVEN?

1. Spelling
2. Aaneengescreven of niet?

CORRECT GEBRUIKT?

3. Zin
4. Substantief
5. Werkwoord
6. Signaal- en structuurwoorden

TEKST VERRIJKEN

7. met verwante woorden
8. met adjectieven
9. met werkwoorden
10. met preposities

TAALTIPS

Uitspraak

Instituut v



Text
review

Vocabulary
resources

Welcome to the Academic Writing Assistant, a writing tool designed to help students and researchers write academic texts. The tool comprises two components.

- **Text review**

This component offers advice on common language issues, coherence and cohesion, and stylistic matters.

For each of these aspects, the writing tool will highlight **possible problems** in your text. As comments are generated automatically, not all the elements which are highlighted are necessarily wrong or relevant to your text. Conversely, the tool only focuses on a limited number of aspects, so some mistakes may not be indicated.

- **Vocabulary resources**

This component provides information on specific words and word combinations (e.g., to carry out / conduct research) as well as on the contexts in which they appear.

Double-click a word in your text to copy it automatically to the search field.

LANGUAGE ISSUES

▶ [Spellchecker UK](#)

▶ [Spellchecker US](#)

▶ [Language errors](#)

▶ [Tense use](#)

▶ [Adverb Position](#)



COHERENCE AND
COHESION

▶ [Readability](#)

▶ [Sentence length](#)

▶ [Paragraphing](#)

▶ [Reference](#)

▶ [Linking words and phrases](#)

▶ [Recurring words and
patterns](#)

STYLE

▶ [Personal language](#)

▶ [Informal language](#)

Enter your text here...



Tense use

This part of the tool focuses on the aspects of the English tense system which are most relevant within the context of formal writing.

- The past and present perfect are frequently confused. [Click here](#) for a short overview of how to use these tenses, and to see correspond

Past versus present perfect

The past tense (**highlighted** in your text below) is used to refer to something that happened at a specific moment in the (distant or recent) past and is used.

- The reformers claimed a small victory in 2005.
- Last year, they extended the experiment to the entire country.
- The company saw its shares fall by nearly 5% yesterday.

The **present perfect** is used to refer to events or activities that took place in the past but are still connected with the present. Indications of duration should be used.

- They have been working together since the late 1980s.
- Much of what has been published so far is ambiguous and imprecise.
- In the last few years, we have set forth many initiatives.

Note that the system may not highlight all relevant forms.

- Progressive forms are often overused. [Click here](#) for a short overview of how to use these tenses, and to see corresponding verb forms h

✓ Processed

The reformers **have claimed** a small victory in 2005.

Language errors

The highlighted words possibly contain a language error. Read the information provided by the tool and, if necessary, correct the mistake.

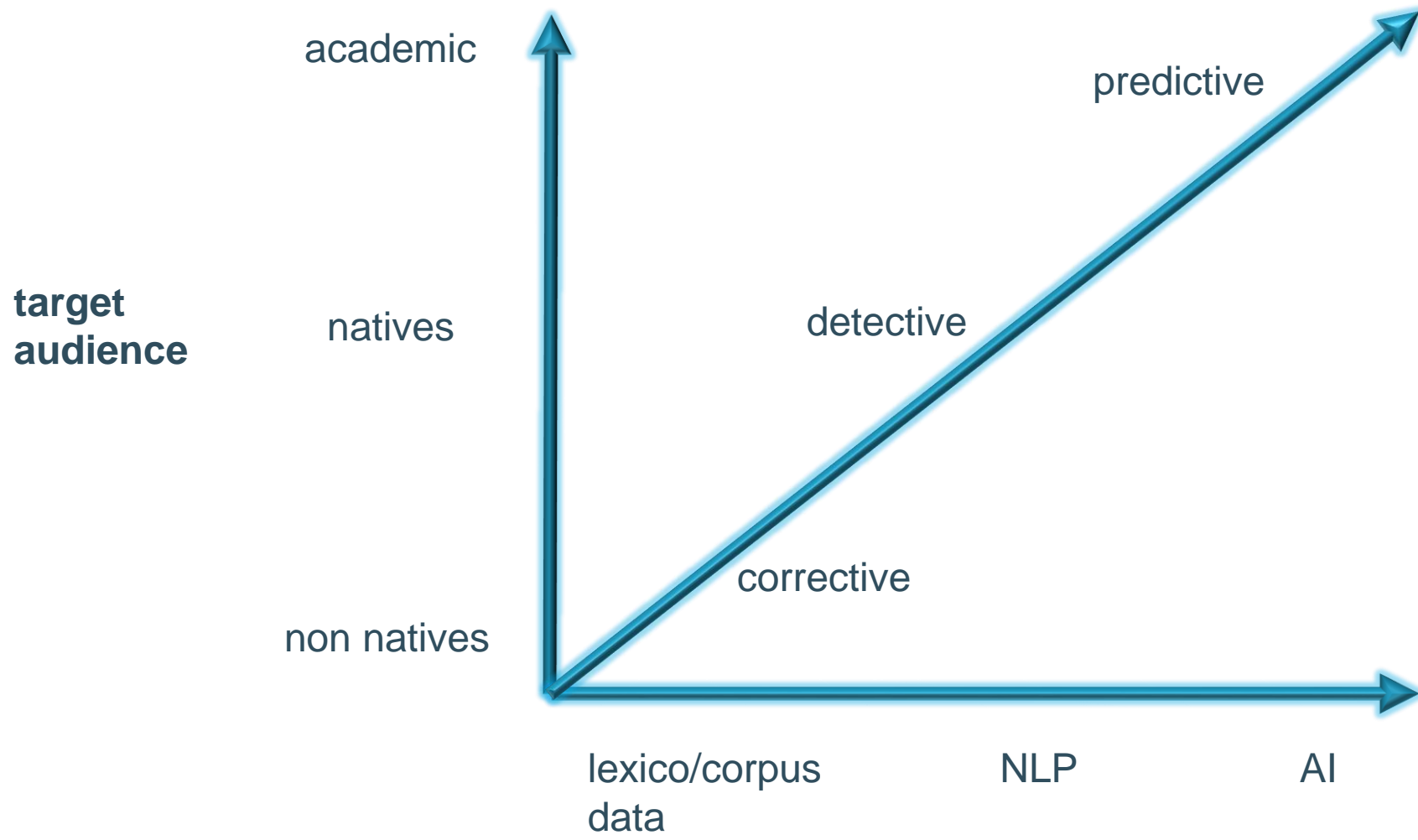
✓ Processed

Let's **discuss about** this topic.

Discuss is not followed by *about*.

✗ The final section discusses about the types of policy responses needed to address these health issues.

✓ The final section discusses the types of policy responses needed to address these health issues.



functions
 cf. Tarp e.a. 2017,
 Ziyuan 2012

**data/
 technology**

Spell checker

- Rationale:
compounds: Dutch, German vs English
- Goals:
avoiding false positives
identifying potential compound errors + suggesting correct form
(interfix -s-!)
- Workflow:
breaking down existing compounds
 - > list of potential modifiers and heads
 - > set of restriction rules

bewegingskunst
bewegingskunstenaar
bewegingsas
bewegingsensor
bewegingslaag
bewegingsles
bewegingsleer
bewegingsleerkracht

Aaneenschrijven of niet?

- Samenstellingen worden in het Nederlands in de regel **aan elkaar** geschreven. Er worden voornamelijk fouten gemaakt bij langere woorden (*computerprogramma*) en bij werkwoorden (*gebruikmaken, tekortkomen*). Daarbij moet wel rekening gehouden worden met de **klinkerbotsing** (*programma-administratie*). Voor de **tussen-s** zijn er geen officiële regels. Het gebruik primeert.
Het algoritme dat mogelijke samenstellingen herkent, maakt fouten. Denk dus goed na voor je iets verandert aan je tekst.
- Het **liggend streepje** wordt gebruikt bij klinkerbotsing, maar ook in een aantal specifieke gevallen, zoals bij prefixen (*oud-leerling*) of leenwoorden (*rendez-vous*).
- In een aantal gevallen geeft de schrijfwijze ook aanleiding tot een **betekenisverschil**: *ten slotte* en *tenslotte*.

✓ tekst verwerkt!

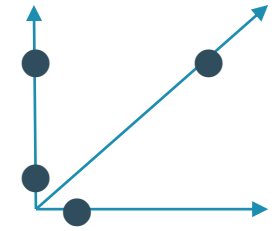
Dit toestel is een beweging trainer.

Als dit een samenstelling zou zijn, schrijf dan

bewegingstrainer (met tussen-s in meer dan 75% van gelijkaardige gevallen)

bewegingtrainer (zonder tussen-s in minder dan 25% van gelijkaardige gevallen)

Word finder: word combinations



- Rationale:
receptive vocabulary > productive vocabulary (Laufer 1998)
word combinations!
- Goal:
adding relevant adjectives to nouns
- Workflow:
 - v. Dutch as a foreign language:
isolating [^adjective] + noun patterns by parsing
n-grams (Google ngrams, Brants 2006)
suggesting semantically ordered, frequent word combinations
 - v. Academic English:
linking to external data: Flax database (univ. of Waikato)

Kwaliteit	
	zeer goed (bv. "een uitstekend boek")
	goed (bv. "een goed boek")
	slecht (bv. "een saai boek")
	zeer slecht (bv. "een vreselijk boek")
Kwantiteit	
	zeer groot (bv. "een reusachtig boek")
	groot (bv. "veel boeken")
	klein (bv. "weinig boeken")
	zeer klein (bv. "een miniscuul boek")

Tijd	
	tijd (bv. "een tijdelijk boek")
Type	
	type (bv. "een modern boek")
Sequentie	
	sequentie (bv. "het eerste boek")
Andere eigenschap/kenmerk	
	andere eigenschap/kenmerk (bv. "een nieuw boek")

✓ Tekst verwerkt!

Ik woon in een huis.

+ adjectief



| een prachtig huis | een heerlijk huis | een geweldig huis | een schitterend huis | een fantastisch huis | het prachtige huis | het ideale huis | het grootste huis



| een mooi huis | een leuk huis | het comfortabele huis | een goed huis | een gezellig huis | het veilige huis | het juiste huis | een schoon huis | het mooiste huis | een fijn huis | het mooie huis | een bijzonder huis | een comfortabel huis | een veilig huis | het goede huis | het gezelligste huis | een net huis | een prettig huis | een gaaf huis | een fatsoenlijk huis



| de week huis | het verkeerde huis | een vreemd huis | een passief huis | een vies huis | het vreemde huis | het passief huis | een lelijk huis | een slecht huis | een verkeerd huis | een eng huis | het passieve huis | het zieke huis | het sombere huis | een vuil huis



| een kapitaal huis | een smerig huis



| een heel huis | het ganse huis | een compleet huis | het volledige huis | een gigantisch huis | de meeste huis | een

adjective + model

🍒 five-factor model	36	🍒 theoretical model	21
🍒 exemplar model	21	🍒 integrative model	20
🍒 full interaction model	12	🍒 computational model	11
🍒 final model	10	🍒 interactive activation model	10
🍒 statistical model	9	🍒 general linear model	9

>>> more

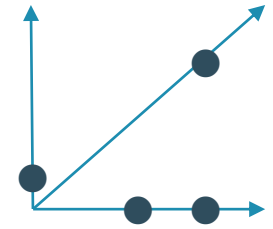
adjective + model

🍒 standard model	155	🍒 kinetic model	148
🍒 simple model	114	🍒 theoretical model	107
🍒 mathematical model	98	🍒 new model	76
🍒 numerical model	68	🍒 final model	68
🍒 structural model	60	🍒 statistical model	57

>>> more

- Interactions between frontal cortex and basal ganglia in working memory: A **computational model**.
- Interactions between frontal cortex and basal ganglia in working memory: A **computational model**.
- On the relationship between repetition priming and recognition memory: Insights from a **computational model**.
- The **computational model** accommodates the amount of blending with a separate quantitative variable specifying the amount of blending allowed.

Word finder: related words



- Rationale:
passive vocabulary > active vocabulary (Laufer 1998)
- Goal:
suggesting related words to extend the scope of the text
- Workflow:
detecting nouns, verbs and adjectives by parsing
suggesting semantically related words using word embeddings
(fastText word vectors for 157 languages, Mikolov e.a. 2018)

✓ Tekst verwerkt!

Ik woon in een huis.

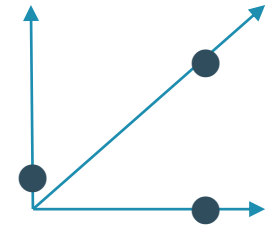
woning | thuis | tuin | appartement | pand | gezin | woonkamer | buurt | voordeur | kantoor | dorp | schuur | keuken | buiten | kamer
| boerderij | villa | huishouden | deur |

✓ Tekst verwerkt!

Ik woon in een mooi huis.

prachtig | schitterend | geweldig | fantastisch | leuk | fraai | knap | goed | gaaf | indrukwekkend | fijn | bijzonder | super | schattig |
heel | heerlijk | tof | erg | gezellig | verrassend |

Use of prepositions



- Rationale:
 - 13.5% of all errors in Cambridge Learner Corpus (excl. spelling)
(Leacock e.a. 2014)
- Goal:
 - making reliable suggestions
- Workflow:
 - detecting prepositions (with spaCy)
 - submitting context (4 words L/R) to language model
 - fast.ai pre-trained Wikitext model
 - + fine-tuned using training data
 - + classifier

- original sentence:
*It was closed **of** some reason I still don't know.

- spaCy output

It was closed of some reason I still don't know.												
Woord:	It	was	closed	of	some	reason	I	still	do	n't	know	.
Lemma:	-PRON-	be	close	of	some	reason	-PRON-	still	do	not	know	.
POS-tag:	PRON	VERB	VERB	ADP	DET	NOUN	PRON	ADV	VERB	ADV	VERB	PUNCT

- context submitted to language model:
it was closed # some reason i still

- Output:

```
- resultaten: {
  as: "0,00060874",
  at: "0,00046935",
  between: "3,036E-5",
  by: "0,0029481",
  for: "99,961",
  from: "0,0034452",
  if: "0,00016193",
  in: "0,0018144",
  into: "0,00013193",
  of: "0,00065962",
  on: "0,0030632",
  to: "0,010829",
  under: "0,00010877",
  with: "0,012762",
  within: "0,0022585",
  totaal%: 100.0002911,
  max: "99,961"
```

- Cut-off point?

- $p \geq 0.975 \rightarrow \geq 95\%$ accuracy $\rightarrow \pm 45\%$ of tested sentences
suggestion for correction + 5 sample sentences taken from a corpus

- $0.975 < p > 0.85 \rightarrow = 70\%$ accuracy $\rightarrow \pm 20\%$ of tested sentences
5 sample sentences taken from corpus for both prepositions
(original & suggested preposition)

- $p \leq 0.85$
no suggestion made

Performance?

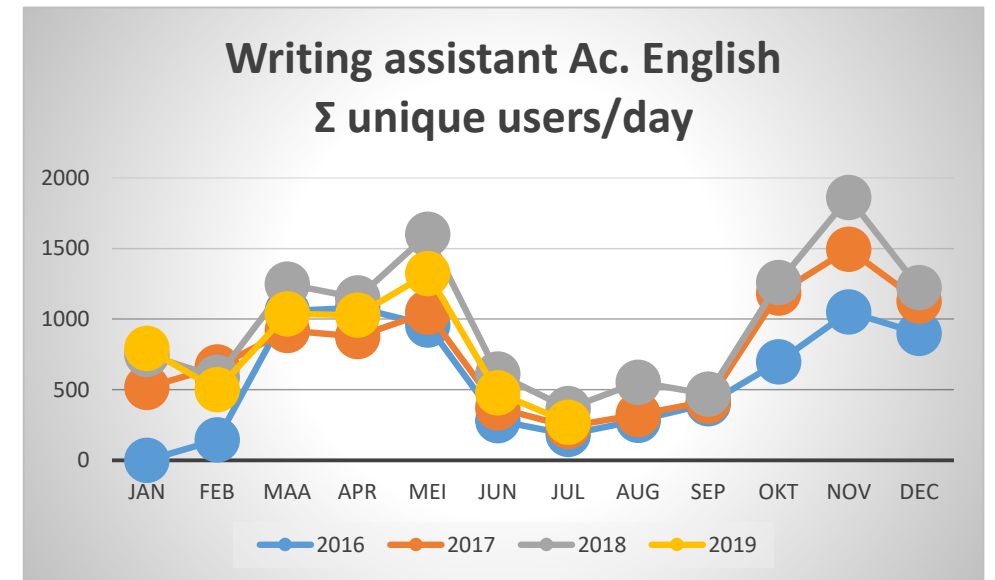
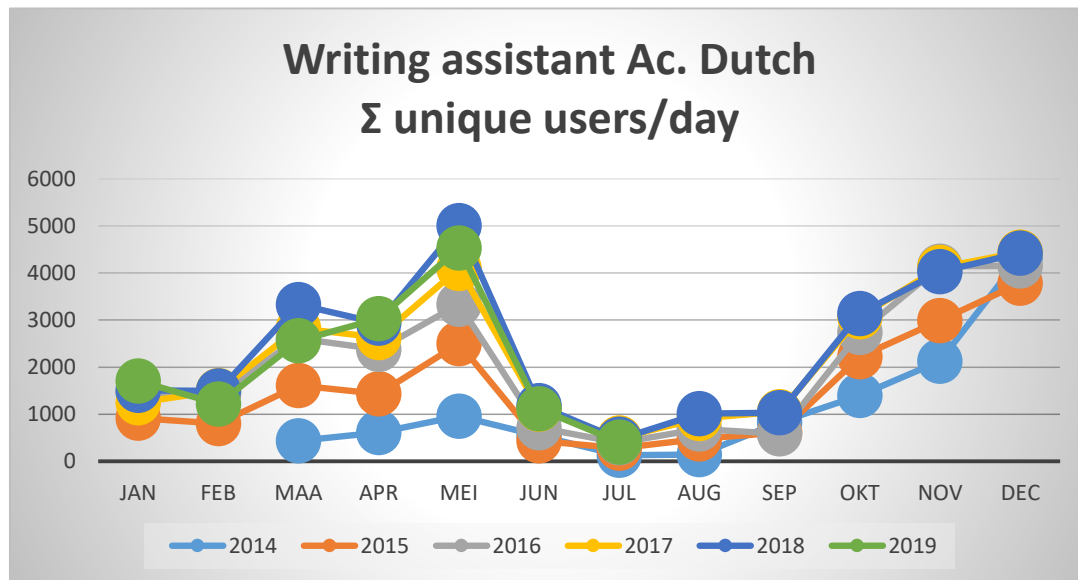
- state-of-the-art for some features:
 - spell checker for Dutch

(2017)	non-word errors /100
Valkuil.net	31
Klinkende Taal	88
Word	91
Spelcheck.nl	97
Google Chrome	97
Writing assistant Dutch	100
LanguageTool.org	100
= Spelling.nu	

- specific language errors for English by native Dutch language speakers
- close to state-of-the-art for other features
- correction + detection + prediction

Does it improve writing?

A system that improves documents is not necessarily a system that improves writers. (Leacock e.a. 2014)



Conclusion

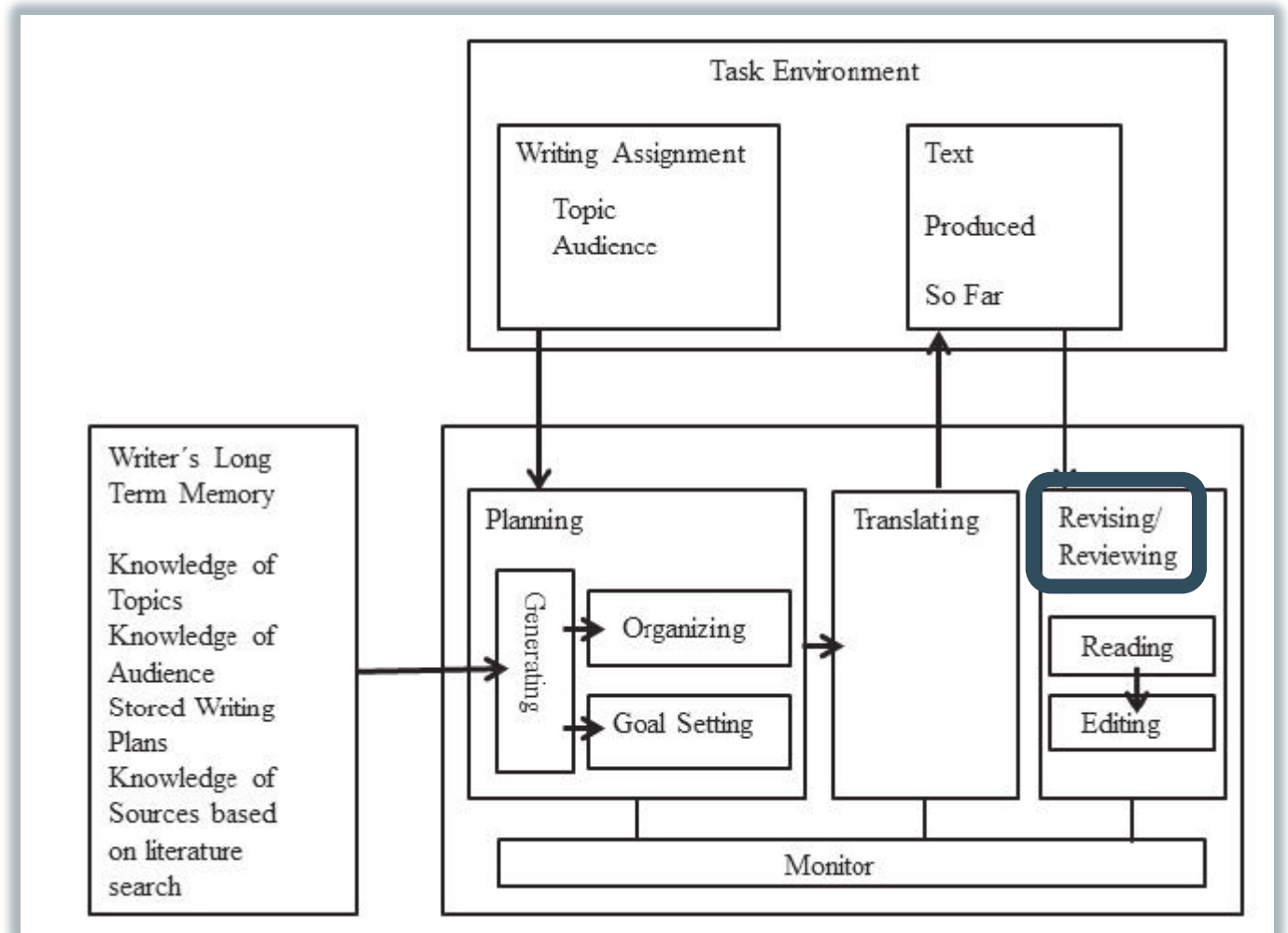
- Writing = a complex process

Hayes-Flower writing model
(1980)

Planning?

Argumentative structure?

'Moves'?



Conclusion

- Corrective function
 - Ac. English: add errors
- Detective function
 - Dutch as a foreign language: word order check
 - Ac. English: increase accuracy for preposition errors
- Predictive function
 - Dutch as a foreign language: add translation to suggested words/
word combinations



thank you!

serge.verlinde@kuleuven.be

References

Brants, Thorsten, and Alex Franz. 2006. *Web 1T 5-gram Version 1 LDC2006T13*. DVD. Philadelphia: Linguistic Data Consortium.

Flower, L., Hayes, J.R. 1980. The dynamics of composing. In L.W. Gregg & E.R. Steinberg (eds.) *Cognitive Processes in Writing*. Hillsdale, NJ: Lawrence Erlbaum Assoc., Pub.

Laufer, B. 1998. The Development of Passive and Active Vocabulary in a Second Language: Same or Different. *Applied Linguistics* 19.2: 255-271.

Leacock, C., Chodorow, M., Gamon, M. and Tetreault, J. 2014. *Automated grammatical error detection for language learners*. Morgan.& Claypool.

Mikolov, T., Grave, E., Bojanowski, P., Puhersch, C., Joulin, A. 2018. Advances in pre-training distributed word representations. *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*.

Strobl, C., Ailhaud, E., Benetos, K., Devitt, A., Kruse, O., Proske, A. and Rapp, C. 2019. Digital support for academic writing: A review of technologies and pedagogies. *Computers & Education*, 131: 33-48.

Tarp, S., Fisker, K. and Sepstrup, P. 2017. L2 writing assistants and context-aware dictionaries: New challenges to lexicography. *Lexikos*, 27: 494–521.

Ziyuan, Y. 2012. *Breaking the language barrier : a game-changing approach*.

(<https://sites.google.com/site/yaoziyuan/publications/books/breaking-the-language-barrier-a-gamechanging-approach>)

Dutch writing assistant: <https://schrijfassistent.be>. The other writing assistants are licenced products.

AWSuM: <http://langtest.jp/awsum/>; *ColloCaid*: <https://www.collocaid.uk/>; *fast.ai*: <https://www.fast.ai/>; *fastText*: <https://fasttext.cc/>; *Flax*:

<http://flax.nzdl.org/greenstone3/flax>; *Just-the-word*: <http://www.just-the-word.com/>; *LEAD*: <https://leaddico.uclouvain.be/login>; *Netspeak*: <http://www.netspeak.org/>;

spaCy: <https://spacy.io/>; *Writeful*: <https://writefullapp.com/>