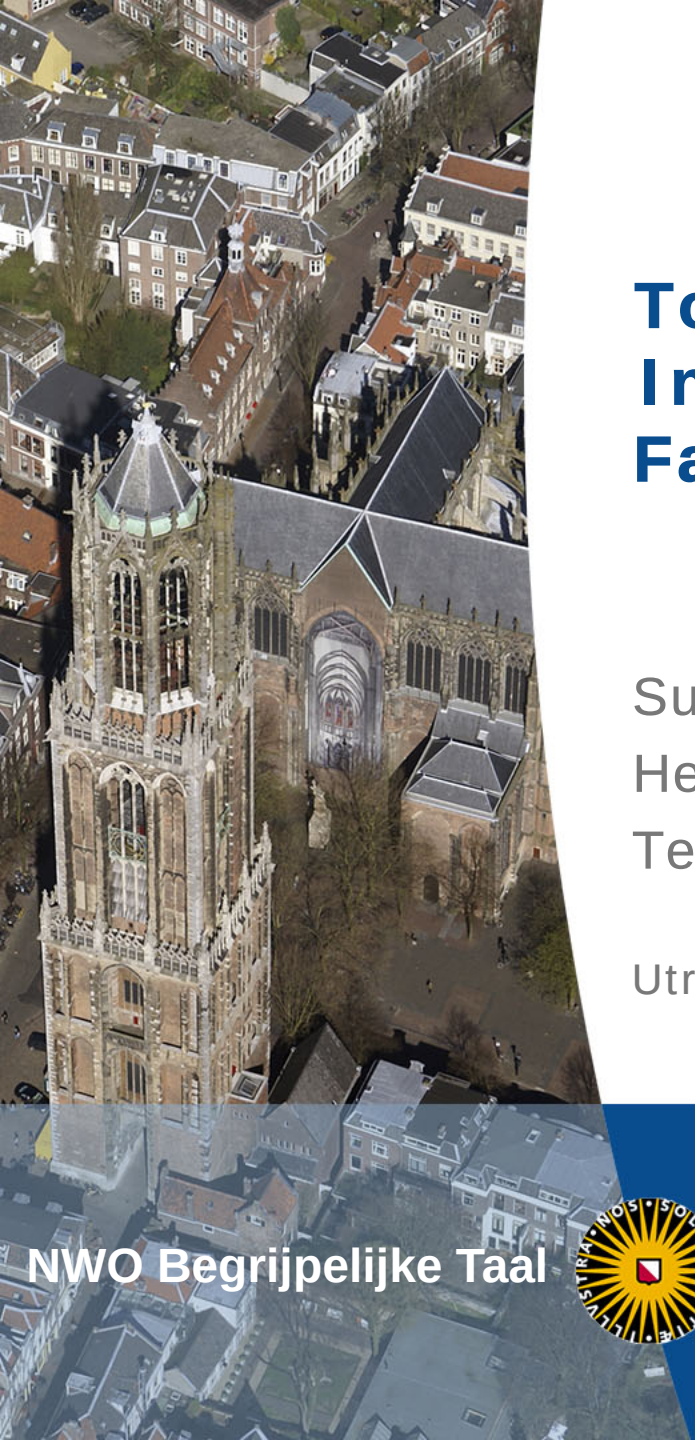


Towards a Validated Readability Index for Dutch: Fact(or)s and Figures

Suzanne Kleijn (s.kleijn1@uu.nl)
Henk Pander Maat
Ted Sanders

Utrecht Institute of Linguistics OTS – Utrecht University

NWO Begrijpelijke Taal

Universiteit Utrecht

Faculty of Humanities

Readability Index for Dutch (LIN)

- LIN: LeesbaarheidsIndex voor het Nederlands
- Project partners: Utrecht University, Radboud University, CITO and Nederlandse Taalunie
- Goal: to build a new and improved readability formula
  - automated (online tool)
  - based on real comprehension and processing data
  - provides an interpretable readability level prediction of a text (i.e., 'Your text is suited for readers of level X')

Readability research

Using objective, quantitative measures to predict the difficulty level of text (e.g., word length)

Two different areas of application:

1. *Readability prediction*: assessing whether a text is appropriate for a certain target group.

2. *Readability improvement*: diagnosing (potential) problems for a certain target group for improvement purposes.

Some issues in traditional readability research

- Predictors are not causally relevant to comprehension
- Predominately use expert judgments to index texts
- Content (message) and style (manner) are confounded
- Higher-level text features like coherence are ignored
- Reader-text interactions are ignored
- Effects on on-line processing are ignored

Our approach to readability

- ‘Causally inspired’ predictors (incl. high-level text features)
- Real comprehension *and* processing data of target group
- Multiple text versions to separate effects of content and style
- Regard for reader-text interactions by including readers with different skills

Steps to building the index

Step 1: Build a tool to automatically extract features from texts (T-Scan)

Step 2: Relate these features to comprehension and processing data

Step 3: Analyze data and build readability index

T-Scan

- T-Scan is a tool which automatically extracts 400 text features from Dutch text.
- It currently provides features describing *lexical complexity*, *sentence complexity*, *referential and relational coherence*, *concreteness*, *person-oriented writing* and *word prediction*.
- During the break: T-Scan demo by Henk Pander Maat

From T-Scan to data collection

- T-Scan only tells us the value of features and not its relation to readability.

- Empirical data
  - Comprehension → cloze tests
  - On-line processing → eye-tracking

Data collection

Cloze study:

- 2850 Dutch 8th-10th grade students
  - Enrolled in different levels of Dutch secondary education (vmbo-b/k/g; havo; vwo)
- 60 texts
  - 2 text versions to separate effects of content and style
- 4 cloze texts per person
- 30 – 40 items per cloze text

Cloze fragment

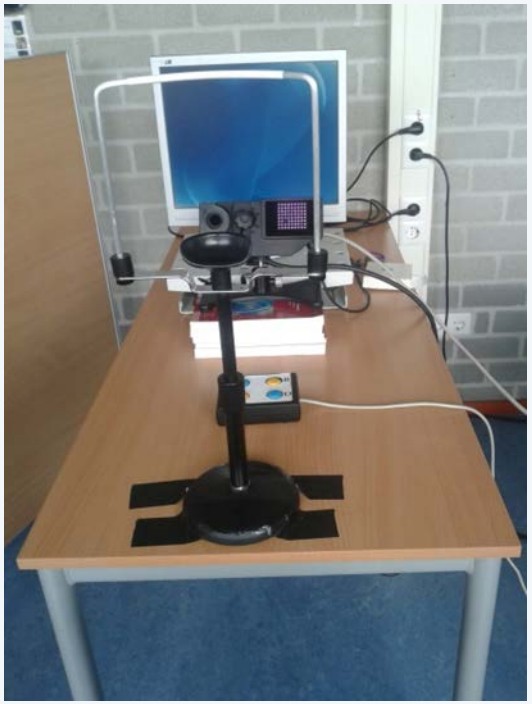
Bankrekeningfraude

Steeds meer jongeren worden slachtoffer van criminelen die via de bankrekeningen van deze jongeren geld witwassen. Met een _____ beloning in het vooruitzicht geven de scholieren toestemming grote bedragen tijdelijk op hun rekening te zetten. Zodra de bank _____ merkt, moet de jongere het vaak alweer _____ geld terugbetalen. Met een hoge _____ tot gevolg. Ook hangt de jongere een zware _____ boven het hoofd voor fraude. De politie en banken waarschuwen jongeren hier dan ook _____ voor. Criminele organisaties spreken _____ vaak aan met een smoes of ze geld op hun bankrekening mogen zetten. De _____ doen dit om illegaal verkregen geld wit te wassen. Stemmen de _____ in? Dan _____ ze een groot geldbedrag op hun bankrekening gestort. De jongeren krijgen de opdracht het _____ op te nemen en af te geven aan de ronselaar. Als beloning voor het _____ van de rekening stellen de criminele organisaties een leuk geldbedrag in het vooruitzicht. Dat wordt meestal _____ uitbetaald, omdat de ronselaar _____ al met de buit is verdwenen. De

Data collection (2)

Eye-tracking study:

- 181 Dutch 9th grade students
  - Enrolled in different levels of Dutch secondary education (vmbo-k; havo; vwo)
- 8 texts taken from the cloze study
- Multiple choice questions after each text



Screen presentation

Griep en verkoudheid

In de herfst en winter lopen veel mensen een verkoudheid of griep op. Griep- en verkoudheidsvirussen bevinden zich in druppeltjes snot, slijm en speeksel. Ze worden door praten, hoesten of niezen verspreid. De kans op besmetting is groot in ruimten waar mensen dicht bij elkaar zitten en waar onvoldoende geventileerd wordt, bijvoorbeeld in een trein of bus, een school of kinderdagverblijf. Virussen worden ook overgedragen via handen als iemand je bijvoorbeeld de hand schudt en via voorwerpen zoals een deurknop. Besmetting is nooit volledig te voorkomen, maar goede hygiëne kan verspreiding en besmetting beperken.

Separating content from style

- All 60 texts were manipulated to create 2 text versions:
  - 20 texts on lexical complexity
  - 20 texts on syntactic complexity
  - 20 texts on relational complexity

Our lexical manipulation

- Text were manipulated to create a lexically easy and a lexically difficult version.
  - 20% of content words were replaced by a more frequent or less frequent synonym.
  - Manipulated words in ‘easy’ text version are on average 14 times more frequent than in the ‘difficult’ text version (using Subtlex NL)
  - Natural language: no stilted or archaic language
  - Text content was left intact
- While minimizing (systematic) confounds:
  - Content, word length, syntactic structure, argument overlap and type-token ratio were kept constant between text versions.

Lexical manipulation

Higher frequency text version
Lower frequency text version

Examples:

1. Rabies is een infectieziekte die de hersenen beschadigt/aantast.
   “Rabies is an infectious disease that damages/impairs the brain.”

2. Iemand heeft genoeg geld/voldoende middelen om er een tijdje tussenuit te kunnen.
   “Someone has enough money/sufficient means to take a break for a while.”

Combined results of cloze and eye movements

- Increasing word frequency has a positive effect on comprehension and on-line processing:
  - Higher cloze scores
  - Higher multiple choice score
  - Shorter reading times

- Clear main effects of Educational level and Grade in the expected directions

Conclusion lexical manipulation

- Word frequency affects readability, but the effects are relatively small.

- Word frequency is only one measure of lexical complexity → a combination of features may prove far more successful

Where are we now?

- We have T-Scan to automatically extract text features.
- We have empirical data which show which features influence comprehension and text processing.
- We know how they affect different level readers.

- All that is left is to actually build the index (LIN)

Questions?