



Easy Language in Belgium

Prospects for automatic simplification

Vincent Vandeghinste

/instituut voor de
Nederlandse taal/

Easy Language in Belgium

Easy language in Belgium

- Easy language
- Plain vs. easy language
- Background: linguistic knowledge in Belgium
- Motivation
- Societal and legal context
- Stakeholders
- Practical outcomes
- Recent call for proposals

Research on Readability and Simplification

- Readability research
 - old formulas
 - contemporary approaches
- Resource: Wablieft corpus
 - comparison with *regular* text
- Other Dutch *easy* corpora
- NT2Lex
- Automated Simplification
 - Knowledge-based
 - Data-driven
 - Building a comparable corpus
 - Unsupervised simplification
- Alternative augmentative communication

Easy Language

Lindholm & Vanhatalo (2021):

"...modified forms of standard languages that aim to facilitate reading and language comprehension."

"...adapted in terms of content, vocabulary, and structure to make them more readable and comprehensible."

"...intended for people who have difficulties understanding standard language."

EASY - PLAIN - ACCESSIBLE



Handbook of
Easy Languages in Europe

Camilla Lindholm and Ulla Vanhatalo (eds.)

Frank & Timme

/instituut voor de
Nederlandse taal/

Easy Language in Belgium

Vincent Vandeghinste, Adéline Muller, Thomas François and Orphée Declercq (2021). Easy Language in Belgium.
In: Camilla Lindholm and Ulla Vanhatalo (eds.): *Handbook of Easy Languages in Europe*. Frank & Timme, Berlin.
pp. 53 - 90.

EASY - PLAIN - ACCESSIBLE



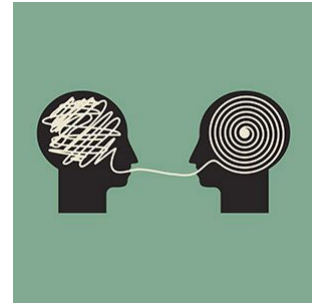
Handbook of Easy Languages in Europe

Camilla Lindholm and Ulla Vanhatalo (eds.)

Frank & Timme

/instituut voor de
Nederlandse taal/

Plain Language -- Easy Language



Plain language

- communication between government / public institution and **general public**
- "A communication is in plain language if its wording, structure, and design are so clear that the intended audience can easily find what they need, understand what they find, and use that information."

International Plain Language Federation

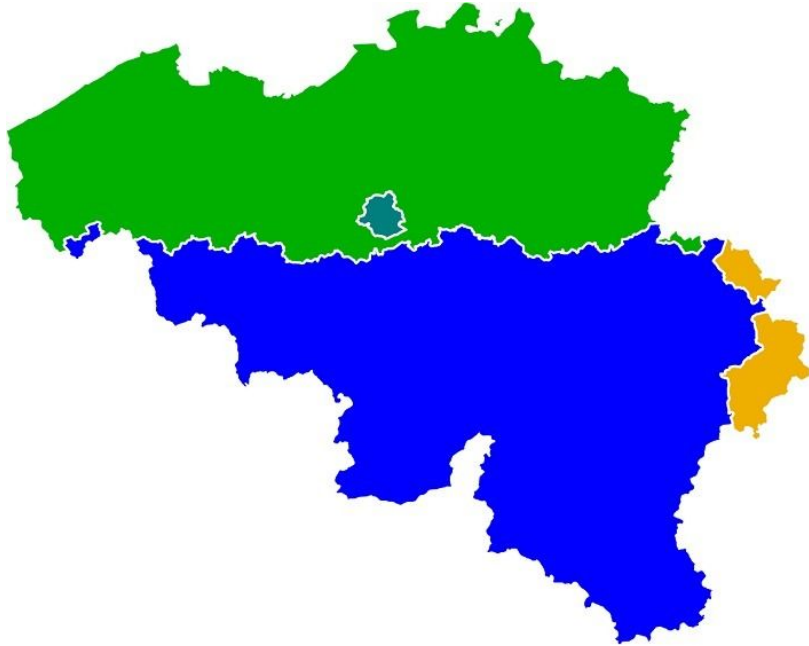
Easy language

- targets **people with low literacy or impairments**
- "No difficult words, jargon, or figurative speech, but articles that everyone can understand"

Wablieft (2021)



Background: Linguistic knowledge in Belgium



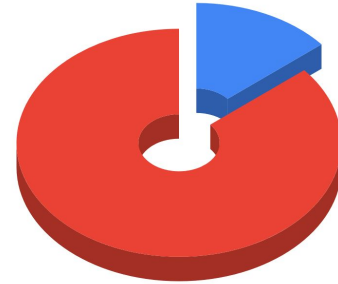
60% of population in Flemish Community,
40% in Francophone Community.

Based on Ginsburgh & Weber (2007) -- data
from 2000

- 59% of people in Flanders know French, 19% of people in Wallonia know Dutch
- In Brussels, 96% know French and 59% know Dutch
- English is known by 52% in Flanders, by 42% in Brussels and by 17% in Wallonia

Easy Language in Belgium: Motivation

International Adult Literacy Survey (IALS 1996) found that **14% of adults suffer from low literacy** in Flanders



People may not be sufficiently **functionally literate** to perform their daily and professional activities without problems

Low literacy is a bigger social problem than previously assumed

Easy Language in Belgium: Motivation

In Flanders (2012):

- 14% of men and 15% of women show low literacy
- 26% of people over 55, 9% of people between 16--24
- 41% of 1st generation migrants, 18% of 2nd generation
- 35% of non-native Dutch speakers
- 12% of people who work, 16% of people who are looking for work, 23% of people who do not work and are not looking for work

In French-speaking Belgium:

- no data -- did not take part in IALS 1996 nor in OECD 2016
- similar numbers are expected

Easy Language in Belgium: Societal and legal context

Plain language

- official status
- federal: Charter of the User of Public Services (1993)
 - Dutch + French
 - examples
 - Legislative drafting manual of the Council of State
 - Code on Economic Law
- Flanders:
 - Decree on Open Government (2004)
 - Decree on on Flemish government communication (2016)

Easy language

- only "indirect" legal status through United Nations Convention on the Rights of Persons with Disabilities (CRDP)

people with disabilities are no longer expected to adapt to their environment. It is up to the policymakers to address the barriers to their participation in economic, social and cultural life

- federal manual on how to apply CRPD led to the *Handstreaming* plan (2016)

includes "*Improving the accessibility of the means of communication with the general public, which should be available in a form that is accessible to all technologies adapted to different types of disability*"

Easy Language in Belgium: Stakeholders

Plain language

- *Droits Quotidiens (Helder Recht)* helps people understand legalese
- *Juristenkrant* has a column focusing on plain legal language
- *eWBS* -- Wallonia Brussels Simplification offers awareness-raising activities



Easy language

- Inclusion Europe (located in Brussels)
 - Easy-to-Read guidelines (1998)
 - Easy-to-Read logo to make accessible website recognizable
- Wabliefert
 - newspaper
 - Easy Language services
- WAI-NOT
 - safe web-environment for people with cognitive disabilities
- Inter-Actions
 - *Visa pour le net* promoted simple structure websites
- ASBL Inclusion
 - involved in FALC
- Stan - Trefpunt Verstandelijke Handicap

/instituut voor de
Nederlandse taal/

Het centrum voor duidelijke taal

Wablieft wil de wereld begrijpelijk maken voor iedereen. Met nieuws dat eenvoudig en duidelijk is. Prachtige boeken die vlot lezen. En door organisaties te helpen om hun communicatie te vereenvoudigen. Ieder jaar bekronen we ook een persoon of organisatie die uitblinkt in duidelijke taal. Dat doen we met onze Wablieft-prijs.

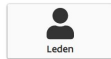


Online lezen? Ga naar de online krant →

Info over Corona



WAI-NOT

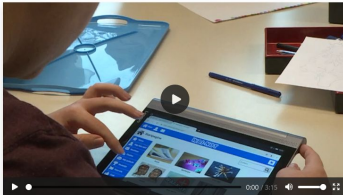


Leden



Bezoekers

[Gratis registreren](#)



Media and Literature

Nouveau roman : *L'herbe dorée*, de Jean-Marc Ceci, paru en décembre 2020

Plus d'infos...



La Traversée

Une initiative de Lire & Ecrire Luxembourg
avec la participation des éditions Weyrich

Easy Language in Belgium: Practical Outcomes

Media and literature

- Wablieft weekly newspaper
- Wablieft Start
- WAI-NOT website (news)
- Literature
 - French publishers Hachette and Clé International
 - Original books in Easy French: La Traversée

Informative texts: COVID-19

- governmental information website
 - contains various posters and fact sheets in Easy language (Dutch and French)
 - is not easy to use
 - links to Wablieft for further information
- Wablieft and Inclusion asbl

Other projects

- Week van de Geletterdheid
 - annual campaign drawing attention to low literacy issues
- Klare Taal Rendeert
 - guidelines for easy language in adult education
- Diversiteitspraktijk
 - entire section on accessible communication
- Nedbox.be
 - exercises at different difficulty levels / from different domains
- International Literacy Day (8th of September)
 - awareness campaign on e.g. digital literacy

/instituut voor de
Nederlandse taal/

Recent call for Easy / Plain Language projects

Based on citizen's requests: Accessible documents

Call launched on Sept. 15.

An application that automatically converts difficult to understand texts into easier language. Official documents and forms are often in difficult language with plenty of jargon. An automatic "translator" could convert these texts to simpler language use, leading to easier to understand texts, without losing important nuances or details.

The screenshot shows the 'am! ai!' website interface. At the top, there is a purple navigation bar with 'Registreer' and 'Aanmelden' buttons. Below the navigation bar, there are dropdown menus for 'Doe mee', 'Over het traject', and 'Lees de AI verhalen'. A search icon and social media icons for Facebook and Twitter are also present. The main content area features a heading 'Ontdek de 17 gekozen AI-concepten' and a sub-heading 'Toegankelijke documenten'. A call to action reads: 'Als je al een idee hebt om je vraag op te lossen met de hulp van AI, beschrijf je idee dan kort.' Below this, there is a paragraph explaining the application: 'Een toepassing die automatisch moeilijk begrijpbare teksten kan omzetten in eenvoudiger taalgebruik. Officiële documenten en formulieren zijn vaak opgesteld in moeilijke taal en bevatten veel jargon. Een automatische 'hertaler' kan deze teksten omzetten naar een eenvoudiger taalgebruik, zodat ze gemakkelijker te begrijpen zijn, zonder dat belangrijke nuances of details daarbij verdwijnen.' To the right of the main text, there are social media sharing options for Facebook and Twitter, both showing 0 shares. Below the main text, there are three more social media sharing options: a trophy icon, Facebook, and Twitter, all showing 0 shares. At the bottom right, there is a section titled 'Zwerfvuil herkennen en aanpakken' with a brief description of a tool for identifying litter and a 'lees meer' button.

Readability Research: Old formulas = superficial factors

- Flesch (1948) = $206.83 - (1.015 \times w/sen) - (84.6 \times syl/w)$
- Flesch Douma (1960) = $206.83 - (0.93 \times w/sen) - (77 \times syl/w)$
- CLIB (1994) = $46 - (6.603 \times c/w) + (0.474 \times fwr) - (36.5 \times ttr) + (1.425 \times w/sen)$
- CILT (1994) = $114.49 + (0.28 \times fwr) - (12.33 \times c/w)$ (used for AVI levels)
- Leesindex A (1963) = $195 - (2 \times w/sen) - (66.7 \times syl/w)$ (old AVI levels)

Only superficial factors

- w/sen = words / sentence
- syl/w = syllables / word
- c/w = characters / word
- fwr = frequent words ratio (77% most frequent words (types) in reference corpus)
- ttr = type token ratio = nr of different words in a text / nr of words in a text

Readability Research: Old formulas applied

Example sentence: *"De hond blaft naar de postbode."*

words / sentence: 6 words / 1 sentence = 6

syllables / word: 8 syllables / 6 words = 1.33

- Flesch: $206.83 - (1.015 \times w/\text{sen}) - (84.6 \times \text{syl}/w) = 88.22$
- Flesch Douma: $206.83 - (0.93 \times w/\text{sen}) - (77 \times \text{syl}/w) = 98.84$

Readability Research: Contemporary approaches

- More complex text characteristics
 - Morpho-syntactic metrics (e.g. part of speech ratios)
 - Syntactic metrics (e.g. tree depth, dependents/head)
 - Readability metrics
- Machine learning techniques
 - Readability prediction system for Dutch (De Clercq and Hoste, 2016)
 - Automatic classification of text difficulty (Vandeghinste & Bulté, 2019)
 - Wablieft vs. De Standaard

Resource: Wabliefte corpus

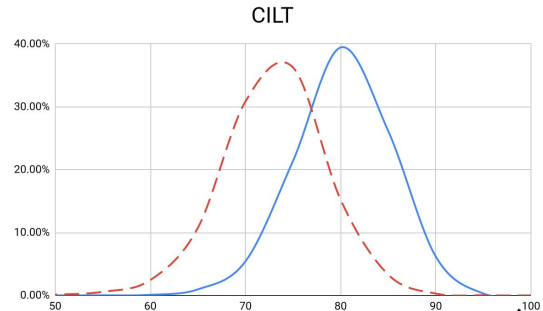
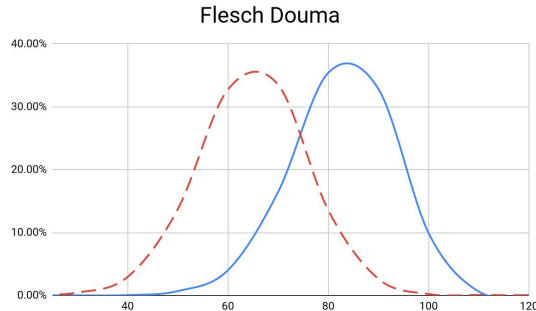
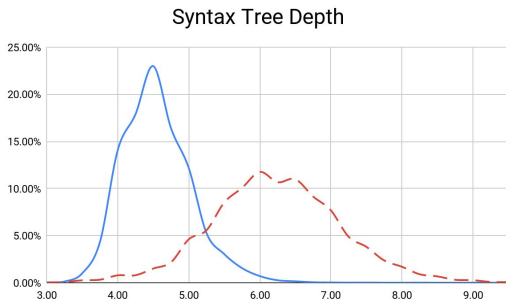
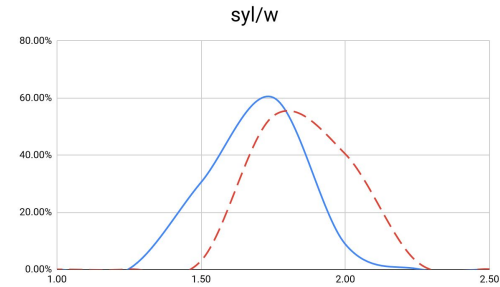
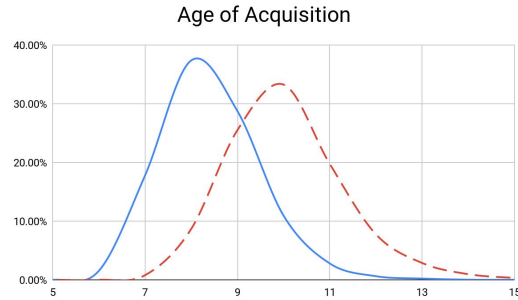
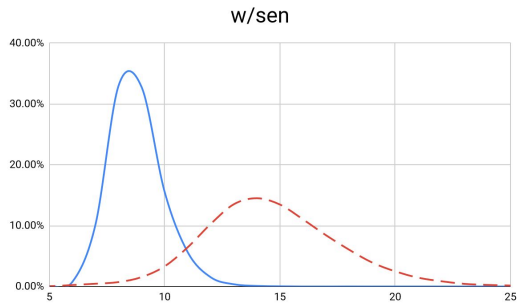


- Archive of Wabliefte newspaper (2011-2017)
- about 2 million words
- available for non-commercial purposes
- <http://hdl.handle.net/10032/tm-a2-q6>

Vincent Vandeghinste, Bram Bulté & Liesbeth Augustinus (2019). [Wabliefte: An Easy-to-Read Newspaper corpus for Dutch](#). In *CLARIN Annual Conference 2019 Proceedings*. pp.188-191. Leipzig, Germany.

Resource: Wabliedt corpus

Vincent Vandeghinste and Bram Bulté (2019). [Linguistic Proxies of Readability: Comparing Easy-to-Read and regular newspaper Dutch](#). *Computational Linguistics in the Netherlands Journal*, vol. 9. p. 81-100.



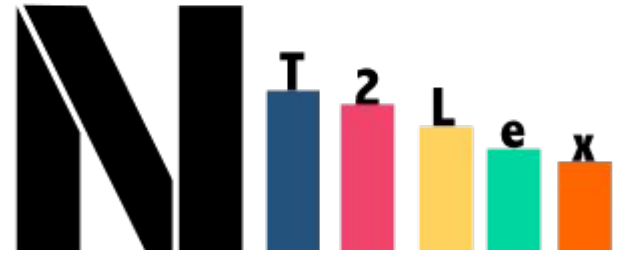
Other Dutch easy corpora (work in progress)

- the Netherlands
 - Basilex: Dutch language directed to children (NL)
 - <http://hdl.handle.net/10032/tm-a2-n4>
 - commercial + non-commercial
 - Oefenen.nl
 - trying to get a license
 - Startkrant
 - negotiations are starting
- Flanders
 - Youth journal subtitles *Karrewiet*
 - trying to get a license
 - WAI-NOT very easy language -- just released last week
 - <http://hdl.handle.net/10032/tm-a2-t9>
 - non-commercial use

/instituut voor
de Nederlandse
taal/taalmaterialen

/instituut voor de
Nederlandse taal/

NT2Lex



- a lexical database for Dutch as a foreign language (NT2)
 - includes frequency distributions of words observed in texts graded along the six-level scale of the Common European Framework of Reference for Languages
- a receptive graded lexicon, with word frequencies observed in
 - textbook reading activities and
 - simplified readers targeting learners of Dutch.

Tack, A., François, T., Desmet, P. & Fairon, C. (2018). [NT2Lex: A CEFR-Graded Lexical Resource for Dutch as a Foreign Language Linked to Open Dutch WordNet](#). In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 137-146).

Automated Simplification: Knowledge-based



Lexical simplification + wikification

- replace low frequency words with higher frequency synonyms
- synonym detection using WordNet
- link words / entities to Wikipedia

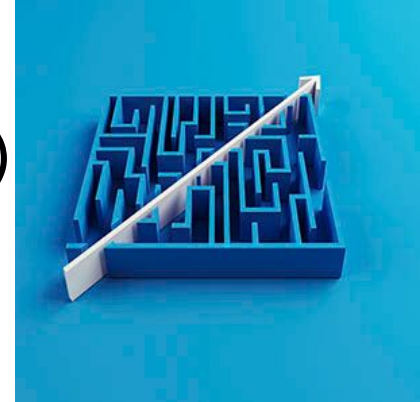
Bram Bulté, Leen Sevens and Vincent Vandeghinste (2018). [Automating lexical simplification in Dutch](#). *Computational Linguistics in the Netherlands Journal*, vol. 8. pp. 24-48.

Syntactic simplification

- detection of complex syntactic phenomena
 - subordinate and relative clauses
 - passive constructions
- replace by *easier* constructions
- used in the context of text-to-pictograph translation

Leen Sevens, Vincent Vandeghinste, Ineke Schuurman and Frank Van Eynde (2018). [Less is More: A Rule-Based Syntactic Simplification Module for Improved Text-to-Pictograph Translation](#). In: *Data and Knowledge Engineering*. Volume 117, pp. 264-289. Elsevier.

Automated Simplification: Data-driven (ongoing)

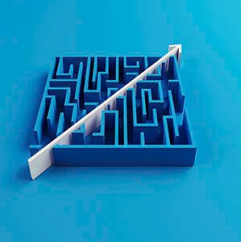


Building comparable corpus *regular news* \leftrightarrow *easy news*

- De Standaard newspaper and Wablieft newspaper
 - Match articles about same event (article alignment)
 - Different degrees of comparability
 - entirely different, about same topic (sports), about same topic (football), about same event
 - Quality of alignment?
- Train a machine translation engine on this parallel / comparable corpus
 - Article alignment \neq Sentence alignment
 - Quality of translation?



/instituut voor de
Nederlandse taal/

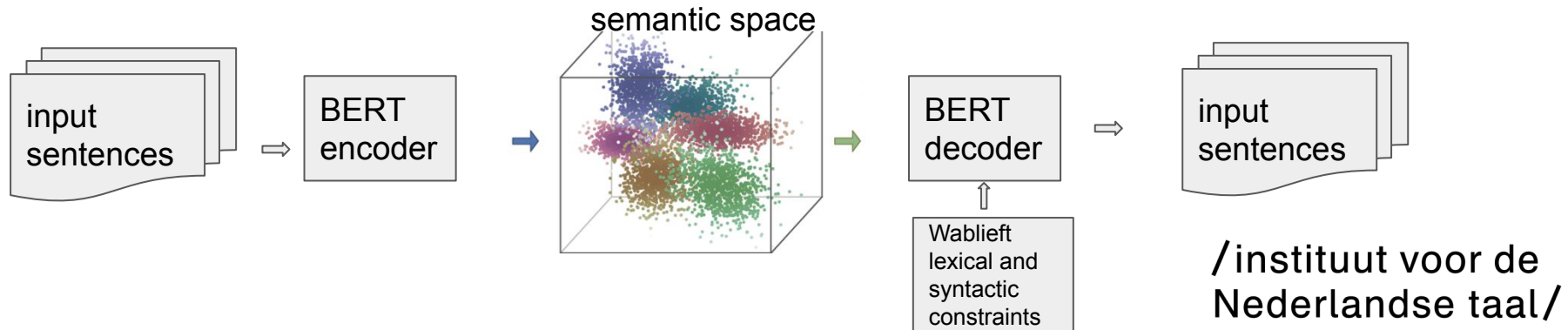


Automated Simplification: Data-driven (plans)

Unsupervised simplification

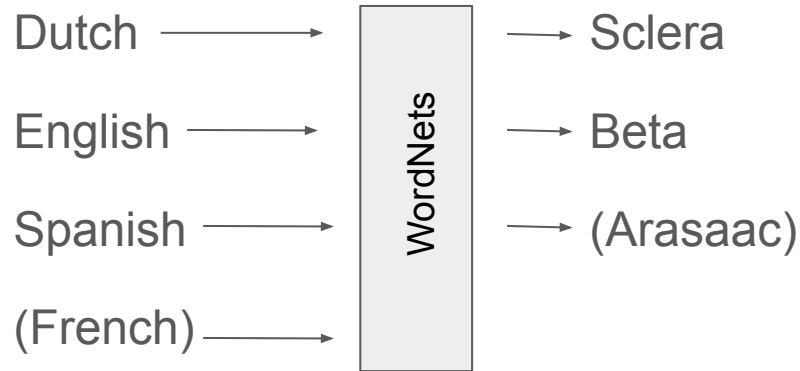
BERT2BERT encoder-decoder model with constraints on the decoder

- projects meaning of regular sentence in the semantic space
- generates new sentence from the point in the semantic space (with constraints)
 - can only generate words from Wablieft corpus
 - can only generate Wablieft-like sentences



Alternative Augmentative Communication (AAC)

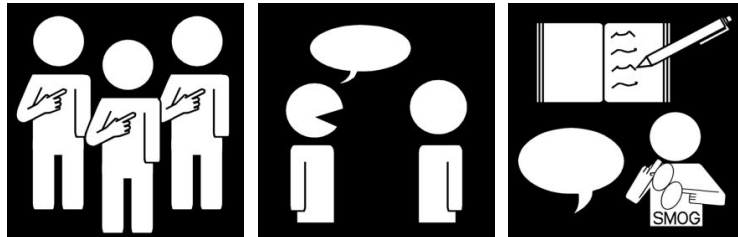
Text-to-pictograph translation (<http://picto.ccl.kuleuven.be>)



Magali Norré, Vincent Vandeghinste, Pierrette Bouillon and Thomas François (2021). [Extending a Text-to-Pictograph System to French and to Arasaac](#). *International Conference Recent Advances in Natural Language Processing (RANLP)*

Leen Sevens (2018). [Words Divide, Pictographs Unite: Pictograph Communication Technologies for People with an Intellectual Disability](#). PhD thesis. LOT.

Vincent Vandeghinste, Ineke Schuurman, Leen Sevens and Frank Van Eynde (2015). Translating Text into Pictographs. *Natural Language Engineering*. Cambridge University Press. Available on CJO 2015 [doi:10.1017/S135132491500039X](https://doi.org/10.1017/S135132491500039X)



"we talk about communication" in Sclera

/instituut voor de
Nederlandse taal/



/instituut voor de
Nederlandse taal/