Various factors of the evaluation of text simplification

Natalia Grabar[1] & Rémi Cardon[2]

(1) UMR8163 STL, CNRS, Université de Lille, France
(2) CENTAL, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

SimpleText@CLEF: 22/09/2021

Context

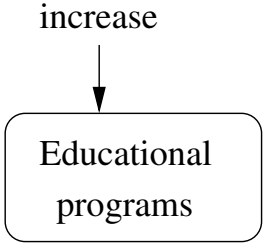
- Automatic text simplification
  - adapt the content of a text
  - so that it becomes easily understandable
  - for a given type of population:
    - children
    - foreigners
    - lay people
    - people with neurodegenerative disorders
    - ...

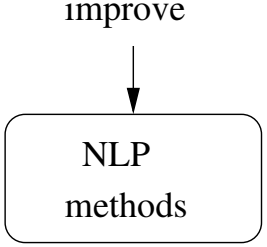
Context: literacy and readability

Literacy
diagnose and increase

Readability
diagnose and improve

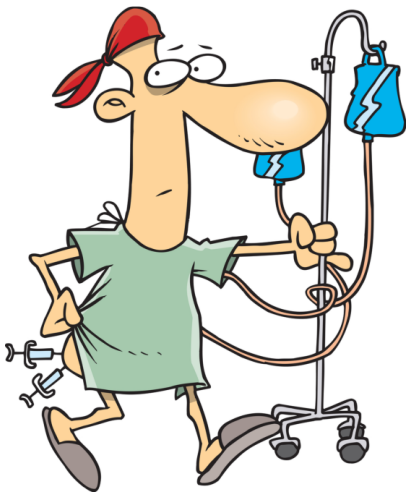
Provide
educational material

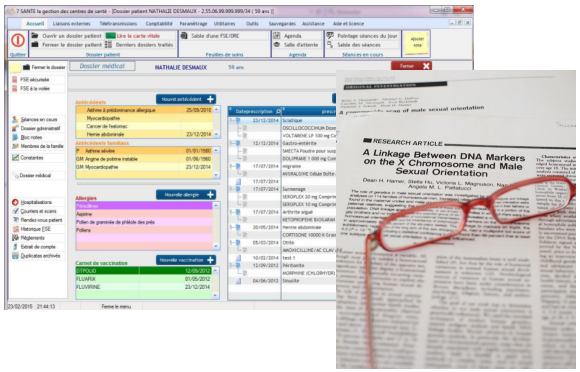




- on the example of children and adults

Context: literacy and readability





Literacy
diagnose and increase

↓

Educational programs

Readability
diagnose and improve

↓

NLP methods

Create
reliable documents

- on the example of lay people and experts

Context

- Automatic text simplification
  - recent research area
- General methodological principles well defined:
  - resources, steps, language levels of simplification...
- Evaluation: not well defined

Context

General evaluation principle

- reference data
- output from automatic systems
- comparison between the two datasets
- computing of standard evaluation measures

is difficult to apply

Context

- Contrary to other NLP tasks
  - information retrieval and extraction, discursive relations, terminology structuring, question-answering...
- ATS: less factual
- Everyone has an opinion on simplification
  - we are all affected

Factors in evaluation of simplification

1. End user
2. Reference data
3. Source document contents
4. Approaches used for the simplification
5. Evaluation measures

End user

- Different types of target population
  - children, foreigners, lay people, people with neurodegenerative disorders...
- Different needs:
  - precise situations calling for scenarii
  - types of documents and information
- Inside a given target population:
  - people with different literacy levels

Literacy levels in adults (Bernèche & Perron, 2006; OECD, 2019)

0. read brief texts on familiar topics, locate a single piece of information, know basic vocabulary
1. read short texts, locate synonymous information, recognise basic vocabulary, determine the meaning of sentences
2. make matches between the text and information, paraphrase, make low-level inferences
3. read and navigate in dense, lengthy or complex texts
4. integrate, interpret information from complex texts, identify and understand non-central ideas, interpret or evaluate subtle evidence-claim or persuasive discourse relationships
5. search for, and integrate, information across multiple texts, construct syntheses of similar and contrasting ideas, evaluate evidence based arguments, understand subtle cues, make high-level inferences, use specialised background knowledge

Literacy levels

Inner differences within population types:

1. children
   - depends on their age
2. neurodegenerative disorders
   - depends on the stage of the disease
3. foreigners
   - depens on the distance between languages
     and on the literacy of foreigners
4. lay people
   - depends on their expertise
     and on duration of exposure to specialized knowledge

Literacy levels

Difference between population types:

- Literacy levels
  - differ between the target populations
  - are not comparable
- Not a continuum
- Intricated scales
- *Specific needs of each population type*
- Possible to link with the 6 standard literacy levels
- Lambda users: levels 2-3

Factors in evaluation of simplification

1. End user
2. Reference data
3. Source document contents
4. Approaches used for the simplification
5. Evaluation measures

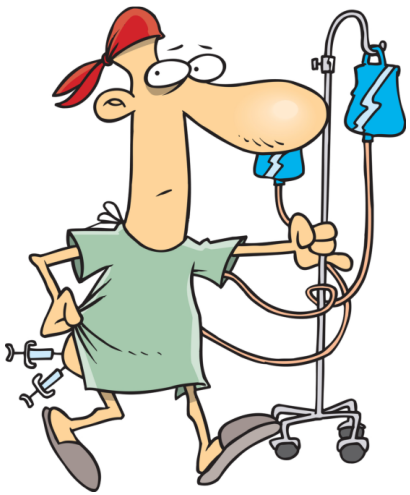
Reference data

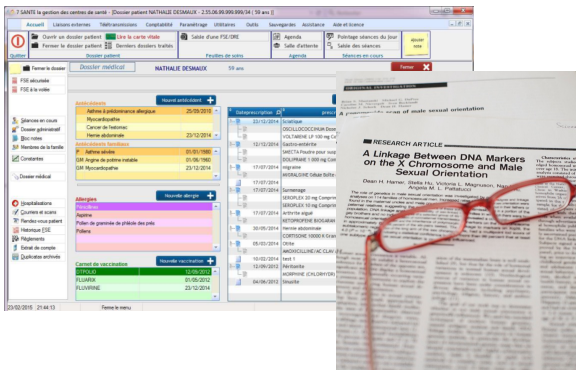
- *expert judgment*: an idea on target population's needs (Clercq *et al.*, 2014)
  - difficulties to know the real needs
- *textbooks*: created for a population according to their readability levels, such as school books (François & Fairon, 2013; Gala *et al.*, 2013)
  - usually created by experts using theoretical observations
- *crowdsourcing*: involves large population (Clercq *et al.*, 2014; Xu *et al.*, 2016; Alva-Manchego *et al.*, 2020b)
  - population involved: uncontrolled and unknown
- *eye-tracking*: fine-grained analysis of reading difficulties (Yaneva *et al.*, 2015; Grabar *et al.*, 2018)
  - use of short text spans
- *manual annotation* by humans (Grabar & Hamon, 2016)
  - large variability across the annotators
  - face-saving strategies, inconsistency

Reference data

- Different approaches for the creation of the reference data
- Inherent limitations
- Bias with each approach
- Difficulty to generalize data and models generated
  - data from experts are difficult to generalize over the population (Clercq *et al.*, 2014)
- Target population often missing in the process
- Reference data differ:
  - approach
  - human expertise
  - simplification operations (Vásquez-Rodríguez *et al.*, 2021)
- *Varying content of the reference data*

Reference data





Literacy
diagnose and increase

Educational programs

Readability
diagnose and improve

NLP methods

Create
reliable documents

- Ideally: literacy and readability should match

Reference data

- Simplification should fit a given literacy level for a given target population
- *Native simplified-language speaker does not exist* (Siddharthan, 2014)
- Yet, simplification levels are vague and subjective
  - complicated to define simplification rules
  - complicated to respect these rules systematically
    - in manual and automatic approaches
- Need for simplification guidelines

Reference data

Simplification guidelines (Ruel *et al.*, 2011; OCDE, 2015; UNAPEI, 2019) :

- use short words
- use frequent and non-ambiguous words
- avoid abbreviations
- limit the variability of the vocabulary used
- make syntactically simple sentences
- avoid sentences in passive or negative voice
- use personal style
- explain difficult concepts
- use pictures

Simplification principles remain vague
must be re-interpreted

Reference data

(Allen, 2009)'s principles:

- structural approach:
  - use of wordlists and lists of structures graded to different levels of complexity
  - constrains the author of simplified materials into conformity with the boundaries set by the guidelines
- intuitive approach:
  - rely on intuition to guide the process of simplification
  - dominates in simplification
  - is what learners of English are most likely to come across in the classroom
- Few studies on the effects of modification upon the linguistic features of simplified texts

Reference data

Some examples of the reference data in English:

- Wikipedia and Simple Wikipedia: 2 levels
  (Zhu *et al.*, 2010; Biran *et al.*, 2011; Coster & Kauchak, 2011)
- revision history of articles from Simple Wikipedia: several versions (Yatskar *et al.*, 2010)
- simplified versions of scientific articles[1]: 2 levels
- simplified versions of novels[2]: 2 levels
- *Newsela* (Xu *et al.*, 2015) : 5 levels
- *Split & Rephrase* (Narayan *et al.*, 2017) : 2 levels

[1] `http://www.reutershealth.com`
[2] `www.onestopenglish.com`

Reference data

Some examples of the reference data in other languages:

- Basque: CBST (Gonzalez-Dios *et al.*, 2018)
- Danish: DSim (Klerke & Søgaard, 2012)
- French: (Brouwers *et al.*, 2012) , CLEAR (Grabar & Cardon, 2018) , ALector (Gala *et al.*, 2020)
- German: (Klaper *et al.*, 2013; Säuberli *et al.*, 2020)
- Italian: PaCCSS–IT (Brunato *et al.*, 2014; Brunato *et al.*, 2015; Brunato *et al.*, 2016) , SIMPITIKI (Tonelli *et al.*, 2016)
- Japanese: (Goto *et al.*, 2015)
- Portuguese: (Aluisio *et al.*, 2008; Caseli *et al.*, 2009)
- Russian: (Dmitrieva & Tiedemann, 2018)
- Spanish: (Collados, 2013; Bott *et al.*, 2014)

Factors in evaluation of simplification

1. End user
2. Reference data
3. Source document contents
4. Approaches used for the simplification
5. Evaluation measures

Document content

- General or specialized language
- Specialized languages:
  - lexicon: rich in specific terminology
  - syntax: specific syntactic structures

Document content

- Intensive lexical transformations
  - increase the distance with the source text

  Medication inhibiting the peristalsis are counter-indicated in this situation.
  In this case, do not take medication for stopping or decreasing the intestinal transit.

- Similarly for:
  - texts for children: structural transformation
  - texts for diseased people: lexical, syntactic transformations
  - ...

Document content

- Simplification $\sim$ generalization
- Lesser association with the meaning preservation
- No association with form preservation:
  - more transformations cause lesser similarity

Factors in evaluation of simplification

1. End user
2. Reference data
3. Source document contents
4. Approaches used for the simplification
5. Evaluation measures

Approaches for simplification

In addition to the creation of the reference data

- Manual, automatic and hybrid systems
  - need for post-edition (Cardon, 2021)
- Levels of simplification
  - lexical, syntactic, structural...
- Evaluation results differ on the same dataset

Factors in evaluation of simplification

1. End user
2. Reference data
3. Source document contents
4. Approaches used for the simplification
5. Evaluation measures

Evaluation measures

- *Precision, accuracy* (Horn *et al.*, 2014)
  - the higher the better
- Textual similarity (Levenshtein, 1966; Vásquez-Rodríguez *et al.*, 2021) :
  - EditNTS (Dong *et al.*, 2019) :
    - detects and predicts: ADD, DELETE, KEEP
  - SeqLabel (Alva-Manchego *et al.*, 2020a) :
    - automatic identification of operations in the original parallel corpus
    - creation of new annotated corpus

Evaluation measures

- Evaluation measures from MT (Vu *et al.*, 2014) :
  - *BLEU* (*bilingual evaluation understudy*) (Papineni *et al.*, 2002)
    - adaptation of Precision + word order (*n-grams*)
    - the higher the better
    - correlation with grammaticality (Wubben *et al.*, 2012; Martin *et al.*, 2018)
    - correlation with semantics (Martin *et al.*, 2018)
    - at the level of corpus: unsuitable for sentences
  - *TERp* (*Translation Edit Rate plus*) (Snover *et al.*, 2009)
    - number of edition operations: insertion, suppression, substitution, inversion
    - the lower the better: less transformation required to fit the reference
  - *OOV* (*out of vocabulary*): rate of words missing from the reference vocabulary (Vu *et al.*, 2014)
    - *Basic English 850 Words list*
    - the lower the better

Evaluation measures

- Evaluation measures from readability:
  - *classical readability scores* (Flesch, 1948; Gunning, 1973; Björnsson & Härd af Segerstad, 1979)
    - values depend on scores
    - syntactic simplicity (Vu *et al.*, 2014)
      simplification outputs longer sentences (Wubben *et al.*, 2012)
    - not correlated with real difficulty (Woodsend & Lapata, 2011; Wubben *et al.*, 2012; Zheng & Yu, 2017)
    - not a simplification measure (Tanprasert & Kauchak, 2021)

Evaluation measures

- Native evaluation measures
  - *Changed* (Horn *et al.*, 2014) : percentage of the test examples where the system suggested some change, correct or not:
    - the higher the better
  - *Potential* (Paetzold & Specia, 2016) : rate of instances among which at least one proposed candidate is in the reference data
    - the higher the better
  - *SARI* (Xu *et al.*, 2016)
    - comparison with reference and source data
    - the higher the better
    - more reliable if several reference datasets are available (Alva-Manchego *et al.*, 2020c; Zhang & Lapata, 2017)
    - no correlation with simplicity (Alva-Manchego *et al.*, 2020b)

Evaluation measures

- Evaluation suite of the simplification outputs:
  - EASSE (Alva-Manchego *et al.*, 2019)
  - comparison with the reference dataset(s) - BLEU, SARI
  - BLEU, SARI, FKGL (Flesch Kincaid Grade-Level)
  - no correlation with simplicity (Woodsend & Lapata, 2011; Wubben *et al.*, 2012; Zheng & Yu, 2017; Alva-Manchego *et al.*, 2020b; Tanprasert & Kauchak, 2021)
  - for the evaluation of English simplification

Evaluation measures

Examples :

- **1:5** – *Source*: Claustrophobia is an important and irrational fear of small enclosed spaces.
  – *Reference*: Claustrophobia is the fear of enclosed spaces, closed places, and small rooms.
  – *Simplification*: Communication is a malign complex product (then that stimulated bone production)
- **1:75** – *Source* : Arterial hypotension can be observed in case of an intraveinous administration performed too quickly, in less than 60 minutes (see rubric 4.2).
  – *Reference* : Arterial hypotension can be observed in case of an intraveinous administration performed too quickly, in less than 60 minutes
  – *Simplification* : A decrase of arterial tension can be observed in case of an intraveinous administration performed too quickly, in less than 60 minutes.

Evaluation measures

	WikiLarge FR			*CLEAR*		
Model	*BLEU*	*SARI*	*Kandel*	*BLEU*	*SARI*	*Kandel*
Indentity	60.02	25.05	81.15	55.00	23.73	76.67
CLEAR	0.15	20.52	94.32	21.59	22.07	84.15
1:5	23.98	33.68	95.56	39.07	**40.94**	87.36
1:10	30.94	34.05	94.61	38.17	36.38	86.72
1:25	**37.29**	34.74	91.40	42.92	39.14	88.22
1:50	32.68	**36.73**	**98.81**	**49.72**	37.52	90.60
1:75	34.20	36.47	89.05	40.16	38.58	**92.35**

(Cardon, 2021)

Evaluation measures

Three criteria for human judgement about the simplification:

- Semantics (*adequacy*):
  - is the meaning preserved?
- Grammaticality (*fluency*):
  - is the simplified text grammatical and understandable?
- Simplicity:
  - is the content simpler than the source text?

Evaluation measures

- Examples of simplicity scales:

Score	Interpretation
+2	much simpler
+1	somewhat simpler
0	equally difficult
-1	somewhat more difficult
-2	much more difficult

(Nisioi *et al.*, 2017)

Score	Interpretation
5	more than one good simplification operation
4	one good simplification operation
3	no notable change
2	one phenomenon that makes the sentence more difficult
1	much more difficult

(Cardon, 2021)

Evaluation measures

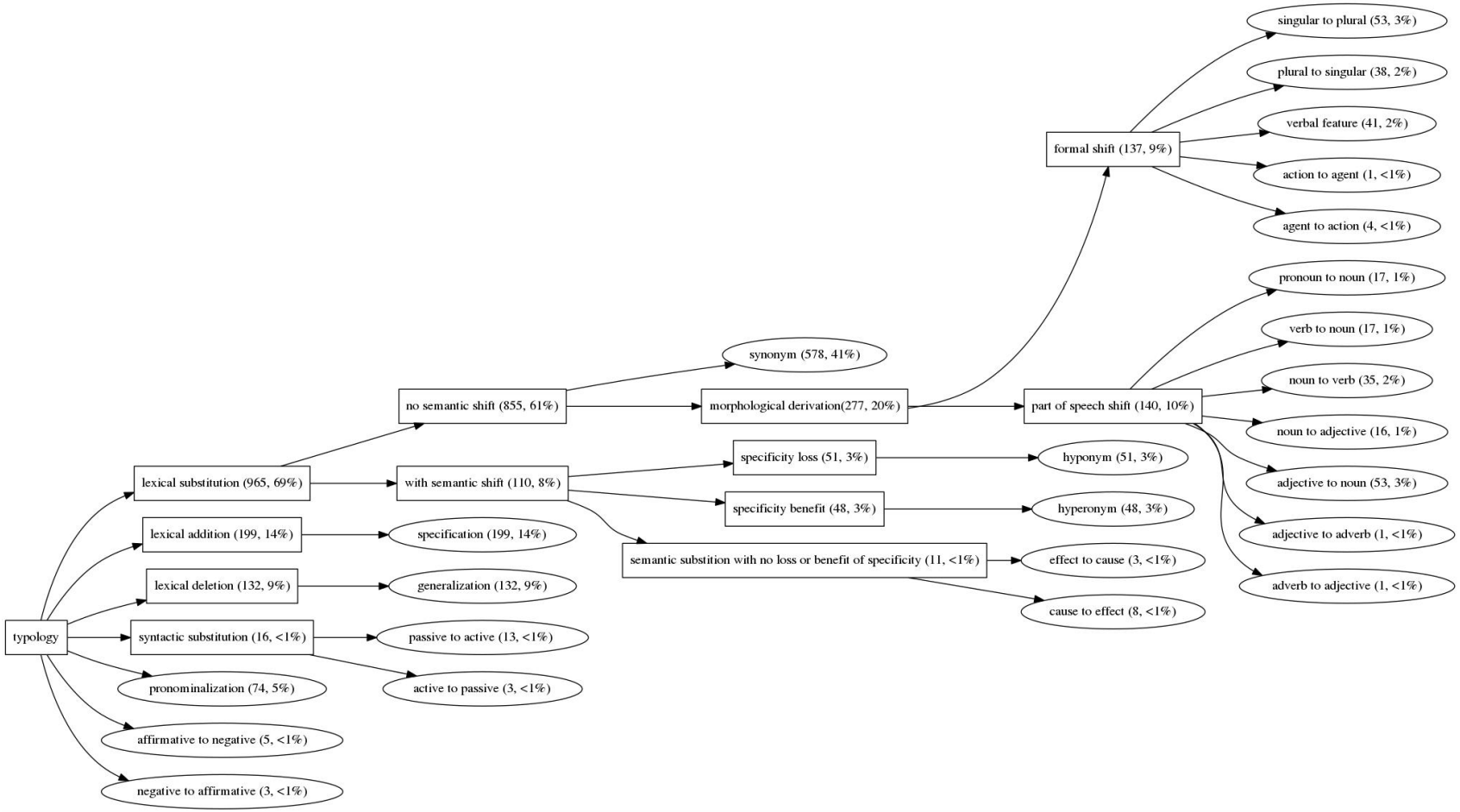
Difficult to implement the criteria:

- guidelines are vague (Stodden, 2021)
- background of annotators
- intuition of annotators
- low reproducibility

Evaluation measures

- The level of evaluation
  - lexical, syntactic...
- The granularity of evaluation
  - general categories of transformations:
    - insert, delete, rephrase...
    - edition distance with the source document
  - number of transformations
  - precise categories: typology of transformations (Brunato *et al.*, 2014; Koptient *et al.*, 2019)
  - ...

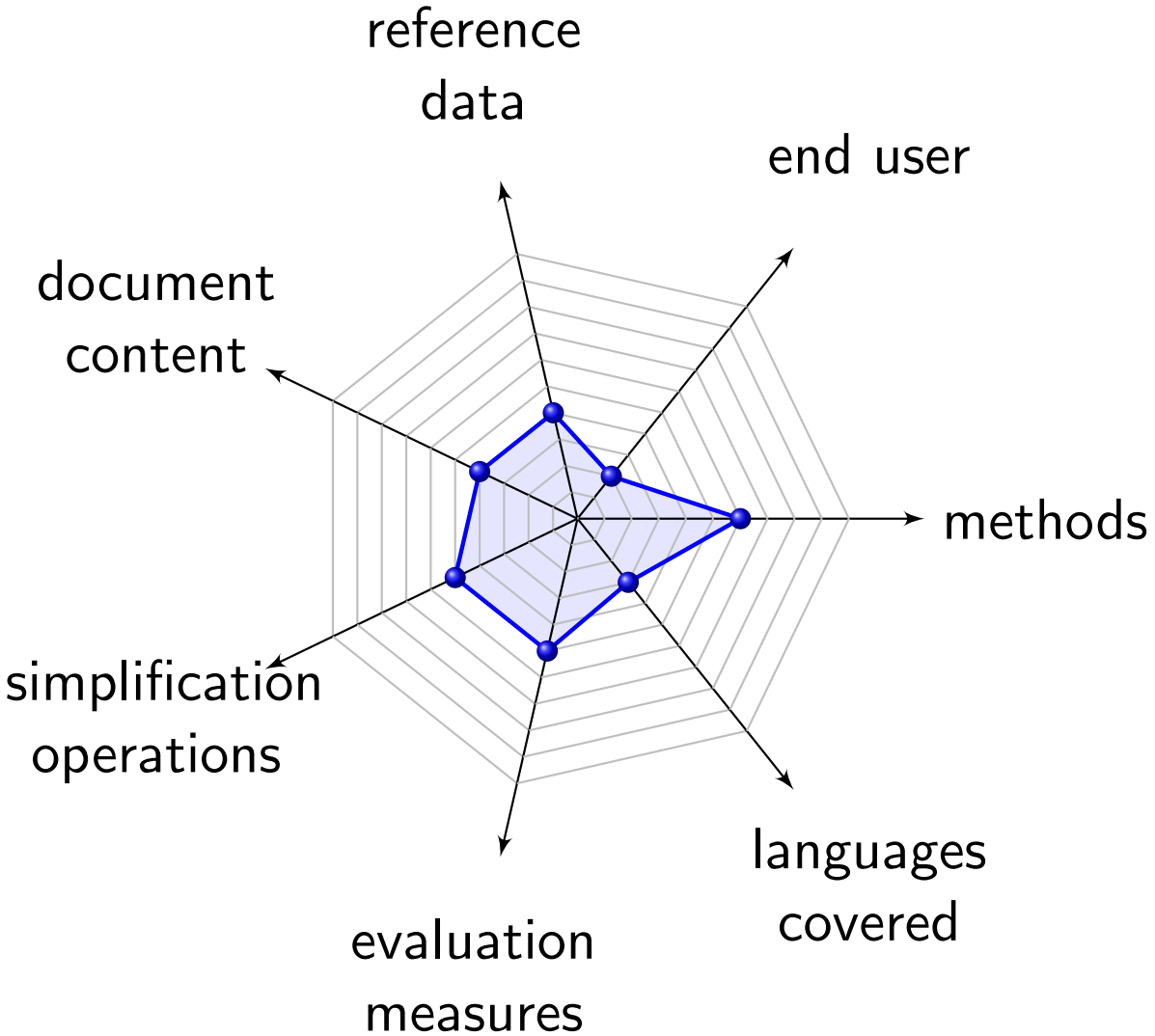
Evaluation measures



Conclusion

- New research area
  - different from other areas (less factual)
- Target population not involved
  - precise needs must be defined
- Reference data missing
- Evaluation measures:
  - fuzzy for human judgement
  - automatic measures not suitable
  - quality estimation: evaluation without reference (Saggion, 2017)
- Mostly oriented on English language

Conclusion


reference data
end user
methods
languages covered
evaluation measures
simplification operations
document content

Allen, D. (2009).
A study of the role of relative clauses in the simplification of news texts for learners of English.
System, **37**(4), 585–599.

Aluisio, S., Specia, L., Pardo, T., Maziero, E. & de Mattos Fortes, R. (2008).
Towards Brazilian Portoguese automatic text simplification systems.
In *Proc of the eighth ACM symp on Document engineering*, pp. 240–248.

Alva-Manchego, F., Bingel, J., Paetzold, G., Scarton, C. & Specia, L. (2020a).
Learning how to simplify from explicit labeling of complex-simplified text pairs.
In *Proc of the Eighth Inter Joint Conf on Natural Language Processing*, pp. 295–305, Taipei, Taiwan.

Alva-Manchego, F., Martin, L., Scarton, A. B. C., Sagot, B. & Specia, L. (2020b).
ASSET: A dataset for tuning and evaluation of sentence simplification models with multiple rewriting transformations.
In *Proc of the 58th Annual Meeting of the Assoc for Comp Linguistics*, pp. 4668–4679.

Alva-Manchego, F., Martin, L., Scarton, C. & Specia, L. (2019).
EASSE: Easier automatic sentence simplification evaluation.
In *Proc of 2019 Conf on Empirical Methods in Natural Language Processing and 9th Inter Joint Conf on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations*, pp. 49–54, Hong Kong, China.

Alva-Manchego, F., Scarton, C. & Specia, L. (2020c).
Data-driven sentence simplification: Survey and benchmark.
Computational Linguistics, pp. 1–87.

Bernèche, F. & Perron, B. (2006).
Développer nos compétences en littératie: un défi porteur d'avenir. Enquête internationale sur l'alphabétisation et les compétences des adultes.
Technical report, Institut de la statistique du Québec, Canada.


Biran, O., Brody, S. & Elhadad, N. (2011).
Putting it simply: a context-aware approach to lexical simplification.
In *Annual Meeting of the Association for Computational Linguistics*.


Björnsson, H. & Härd af Segerstad, B. (1979).
Lix pä franska och tio andra språk.
Stockholm: Pedagogiskt centrum, Stockholms skolförvaltning.


Bott, S., Saggion, H. & Mille, S. (2014).
Text simplification tools for Spanish.
In *LREC 2014*, pp. 1–7.


Brouwers, L., Bernhard, D., Ligozat, A.-L. & François, T. (2012).
Simplification syntaxique de phrases pour le français.
In *Traitement Automatique des Langues Naturelles (TALN)*, pp. 211–224.


Brunato, D., Cimino, A., Dell'Orletta, F. & Venturi, G. (2016).
PaCCSS–IT: A parallel corpus of complex–simple sentences for automatic text simplification.
In *Proc of Conf on Empirical Methods in Natural Language Processing (EMNLP 2016)*, pp. 351–361, Austin, Texas, USA.


Brunato, D., Dell'Orletta, F., Venturi, G. & Montemagni, S. (2014).
Defining an annotation scheme with a view to automatic text simplification.
In *CLICIT*, pp. 87–92.


Brunato, D., Dell'Orletta, F., Venturi, G. & Montemagni, S. (2015).
Design and annotation of the first Italian corpus for text simplification.
In *Proc of The 9th Linguistic Annotation Workshop*, pp. 31–41, Denver, Colorado, USA.


Cardon, R. (2021).

Simplification Automatique de Textes Techniques et Spécialisés.
Phd, Université de Lille, Lille, France.

Caseli, H. M., Pereira, T. F., Specia, L., Pardo, T. A. S., Gasperin, C. & Aluisio, S. M. (2009).
Building a Brazilian Portuguese parallel corpus of original and simplified texts.
In *CICLING*, pp. 1–12.

Clercq, O. D., Hoste, V., Desmet, B., van Oosten, P., Cock, M. D. & Macken, L. (2014).
Using the crowd for readability prediction.
Natural Language Engineering, **20**, 293–325.

Collados, J. (2013).
Splitting complex sentences for natural language processing applications: Building a simplified Spanish corpus.
Procedia-Social and Behavioral Sciences, **95**, 464–472.

Coster, W. & Kauchak, D. (2011).
Simple English wikipedia: A new text simplification task.
In *Annual Meeting of the Association for Computational Linguistics*, pp. 665–669.

Dmitrieva, A. & Tiedemann, J. (2018).
Creating an aligned Russian text simplification dataset from language learner data.
In *Proc of the 8th Workshop on Balto-Slavic Natural Language Processing*, pp. 73–79, Kiyv, Ukraine.

Dong, Y., Li, Z., Rezagholizadeh, M. & Cheung, J. (2019).
Editnts: An neural programmer-interpreter model for sentence simplification through explicit editing.
In *Proc of the Eighth Inter Joint Conf on Natural Language Processing*, pp. 3393–3402, Florence, Italy.

Flesch, R. (1948).
A new readability yardstick.

Journ Appl Psychol, **23**, 221–233.

François, T. & Fairon, C. (2013).
Les apports du TAL à la lisibilité du français langue étrangère.
TAL, **54**(1), 171–202.

Gala, N., François, T. & Fairon, C. (2013).
Towards a French lexicon with difficulty measures: NLP helping to bridge the gap between traditional dictionaries and specialized lexicons.
In *eLEX-2013*.

Gala, N., Tack, A., François, L. J.-D. T. & Ziegler, J. (2020).
Alector : A parallel corpus of simplified french texts with alignments of misreadings by poor and dyslexic readers.
In *Language Resources and Evaluation for Language Technologies (LREC)*, Marseille, France.

Gonzalez-Dios, I., Aranzabe, M. J. & Díaz de Ilarraza, A. (2018).
The corpus of basque simplified texts (cbst).
Language Resources and Evaluation, **52**, 217–247.

Goto, I., Tanaka, H. & Kumano, T. (2015).
Japanese news simplification: Task design, data set construction, and analysis of simplified text.
In *Proc of MT Summit XV*, pp. 17–31, Miami, USA.

Grabar, N. & Cardon, R. (2018).
Clear – simple corpus for medical French.
In *Workshop on Automatic Text Adaption (ATA)*, pp. 1–11.

Grabar, N., Farce, E. & Sparrow, L. (2018).
Study of readability of health documents with eye-tracking and machine learning approaches.

In *Int Conf on Healthcare Informatics (ICHI)*, pp. 1–2.
Poster.

Grabar, N. & Hamon, T. (2016).
A large rated lexicon with French medical words.
In *LREC (Language Resources and Evaluation Conference)*, pp. 1–12.

Gunning, R. (1973).
The art of clear writing.
New York, NY: McGraw Hill.

Horn, C., Manduca, C. & Kauchak, D. (2014).
Learning a lexical simplifier using Wikipedia.
In *Annual Meeting of the Association for Computational Linguistics*, pp. 458–463.

Klaper, D., Ebling, S. & Volk, M. (2013).
Building a German/simple German parallel corpus for automatic text simplification.
In *Proc of the 2nd Workshop on Predicting and Improving Text Readability for Target Reader Populations*, pp. 11–19, Sofia, Bulgaria.

Klerke, S. & Søgaard, A. (2012).
DSim, a Danish parallel corpus for text simplification.
In *LREC*, pp. 4015–4018.

Koptient, A., Cardon, R. & Grabar, N. (2019).
Simplification-induced transformations: typology and some characteristics.
In *Proc of the 18th BioNLP Workshop and Shared Task*, pp. 309–318, Florence, Italy.

Levenshtein, V. I. (1966).
Binary codes capable of correcting deletions, insertions and reversals.

Soviet physics. Doklady, **707**(10).

Martin, L., Humean, S., Mazaré, P.-E., Bordes, A., de La Clergerie, E. & Sagot, B. (2018).
Reference-less quality estimation of text simplification systems.
In *ATA workshop*, pp. 1–10.

Narayan, S., Gardent, C., Cohen, S. & Shimorina, A. (2017).
Split and rephrase.
In *Int Conf on Empirical Methods in Natural Language Processing*, pp. 617–627.

Nisioi, S., Stajner, S., Ponzetto, S. P. & Dinu, L. P. (2017).
Exploring neural text simplification models.
In *Ann Meeting of the Assoc for Comp Linguistics*, pp. 85–91.

OCDE (2015).
Guide de style de l'OCDE Troisième édition: Troisième édition.
OECD Publishing.

OECD (2019).
Skills matter.
Technical report, PIAAC, OECD.

Paetzold, G. H. & Specia, L. (2016).
Benchmarking lexical simplification systems.
In *LREC*, pp. 3074–3080.

Papineni, K., Roukos, S., Ward, T., Henderson, J. & Reeder, F. (2002).
BLEU: a method for automatic evaluation of machine translation.
In *Annual Meeting of the Association for Computational Linguistics*, pp. 311–318.

Ruel, J., Kassi, B., Moreau, A. & Mbida-Mballa, S. (2011).

Guide de rédaction pour une information accessible.
Gatineau: Pavillon du Parc.

Saggion, H. (2017).

Automatic Text Simplification, volume 32 of *Synthesis Lectures on Human Language Technologies*.
University of Toronto: Morgan & Claypool.

Siddharthan, A. (2014).

A survey of research on text simplification.
Int J of Applied Linguistics, **165**(2), 259–298.

Snover, M. G., Madnani, N., Dorr, B. & Schwartz, R. (2009).

TER-Plus: paraphrase, semantic, and alignment enhancements to translation edit rate.
Machine Translation, **23**(2-3), 117–127.

Stodden, R. (2021).

When the scale is unclear – analysis of the interpretation of rating scales in human evaluation of text simplification.
In *Proc of the First Workshop on Current Trends in Text Simplification (CTTS 2021)*, pp. 1–12.

Säuberli, A., Ebling, S. & Volk, M. (2020).

Benchmarking data-driven automatic text simplification for German.
In *Proc of the 1st Workshop on Tools and Resources to Empower People with REAding DIfficulties (READI2020)*, pp. 41–48.

Tanprasert, T. & Kauchak, D. (2021).

Flesch-kincaid is not a text simplification evaluation metric.
In A. for Computational Linguistics, Ed., *Proc of the 1st Workshop on Natural Language Generation, Evaluation, and Metrics (GEM 2021)*, pp. 1–14.


Tonelli, S., Palmero Aprosio, A. & Saltori, F. (2016).
SIMPITIKI: a simplification corpus for Italian.
In *Proc of Third Italian Conf on Comp Linguistics (CLiC-it 2016)*, pp. 1–6.


UNAPEI (2019).
L'information pour tous.
UNAPEI.


Vu, T. T., Tran, G. B. & Pham, S. B. (2014).
Learning to simplify children stories with limited data.
In L. . Springer, Ed., *Intelligent Information and Database Systems*, pp. 31–41.


Vásquez-Rodríguez, L., Shardlow, M., Przybyła, P. & Ananiadou, S. (2021).
The role of text simplification operations in evaluation.
In *Proc of the First Workshop on Current Trends in Text Simplification (CTTS 2021)*, pp. 1–13.


Woodsend, K. & Lapata, M. (2011).
Learning to simplify sentences with quasi-synchronous grammar and integer programming.
In *EMNLP*, pp. 409–420.


Wubben, S., van den Bosch, A. & Krahmer, E. (2012).
Sentence simplification by monolingual machine translation.
In *Annual Meeting of the Association for Computational Linguistics*, pp. 1015–1024.


Xu, W., Callison-Burch, C. & Napoles, C. (2015).
Problems in current text simplification research: New data can help.
Transactions of the Association for Computational Linguistics, **3**(1), 283–297.


Xu, W., Napoles, C., Pavlick, E., Chen, Q. & Callison-Burch, C. (2016).
Optimizing statistical machine translation for text simplification.

Transactions of the Association for Computational Linguistics, **4**, 401–415.

Yaneva, V., Temnikova, I. & Mitkov, R. (2015).
Accessible texts for autism: An eye-tracking study.
In ACM, Ed., *Int ACM SIGACCESS Conference on Computers & Accessibility*, pp. 49–57.

Yatskar, M., Pang, B., Danescu-Niculescu-Mizil, C. & Lee, L. (2010).
For the sake of simplicity: Unsupervised extraction of lexical simplifications from Wikipedia.
In *NAACL*, pp. 365–368.

Zhang, X. & Lapata, M. (2017).
Sentence simplification with deep reinforcement learning.
In ACL, Ed., *Proc of the Conf on Empirical Methods in Natural Language Processing*, pp. 584–594, Copenhagen, Denmark.

Zheng, J. & Yu, H. (2017).
Readability formulas and user perceptions of electronic health records difficulty: A corpus study.
Journal of Medical Internet Research, **19**(3), 1–15.

Zhu, Z., Bernhard, D. & Gurevych, I. (2010).
A monolingual tree-based translation model for sentence simplification.
In *COLING 2010*, pp. 1353–1361.