



HAL
open science

A new method to generate automatically highly standardized reading chart text in multiple languages: the case of MNREAD

Aurelie Calabrese, Alexandre Bonlarron, Jean-Charles Régin, Pierre Kornprobst

► To cite this version:

Aurelie Calabrese, Alexandre Bonlarron, Jean-Charles Régin, Pierre Kornprobst. A new method to generate automatically highly standardized reading chart text in multiple languages: the case of MNREAD. ARVO 2024 - Annual Meeting of the Association for Research in Vision and Ophthalmology, May 2024, Seattle (USA), Washington, United States. hal-04576614v2

HAL Id: hal-04576614

<https://hal.science/hal-04576614v2>

Submitted on 16 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



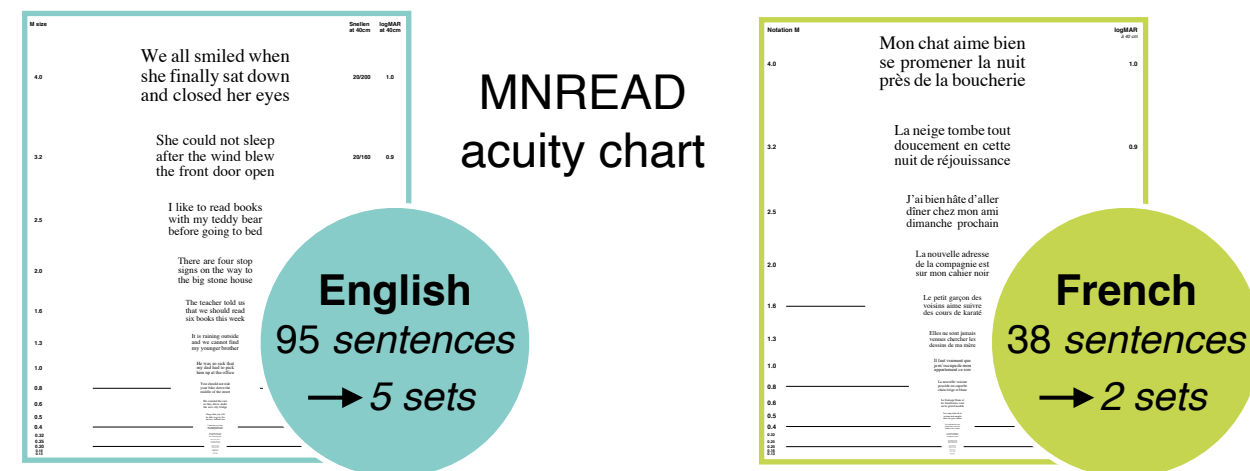
Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

A new method to generate automatically highly standardized reading chart text in multiple languages: the case of MNREAD

Calabrèse Aurélie¹, Bonlarron Alexandre^{2,3}, Régim Jean-Charles³, Kornprobst Pierre²

Background

- Measuring reading performance using reading charts requires highly standardized reading material.
- Such material is hard to produce and most charts are limited by their relatively small number of texts or sentences.



How to generate a large number of standardized sentences?

1 - Naïve method: search for MNREAD-like sentences in a corpus

2,300 → 10,000,000 sentences → ❌

2 - Ad-hoc method, based on hand-defined templates (Mansfield et al., 2019)

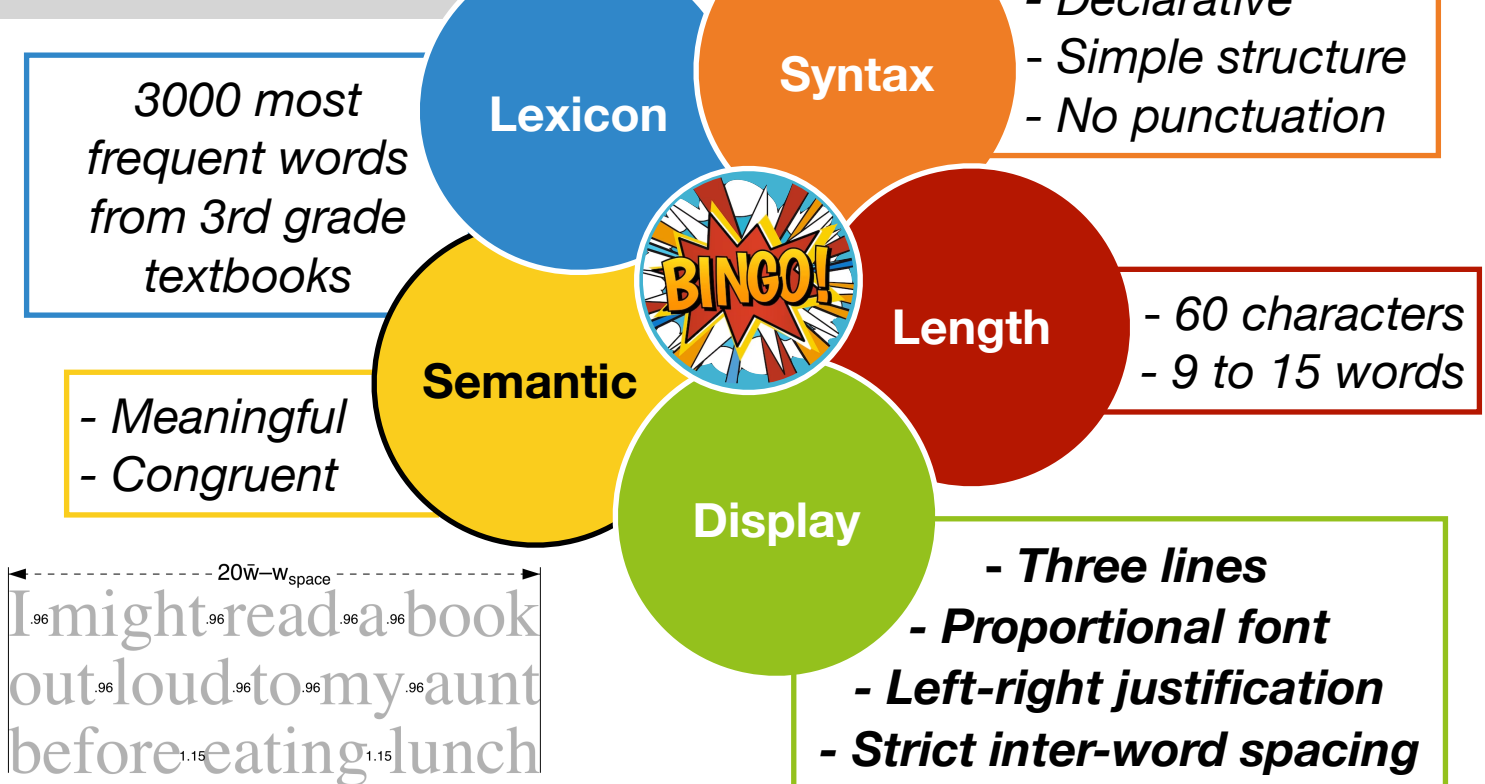
CONS

- Semi-automatic method
- Must generate 8,000 sentences to get 1 match
- Non applicable to Latin languages (e.g., French) ❌

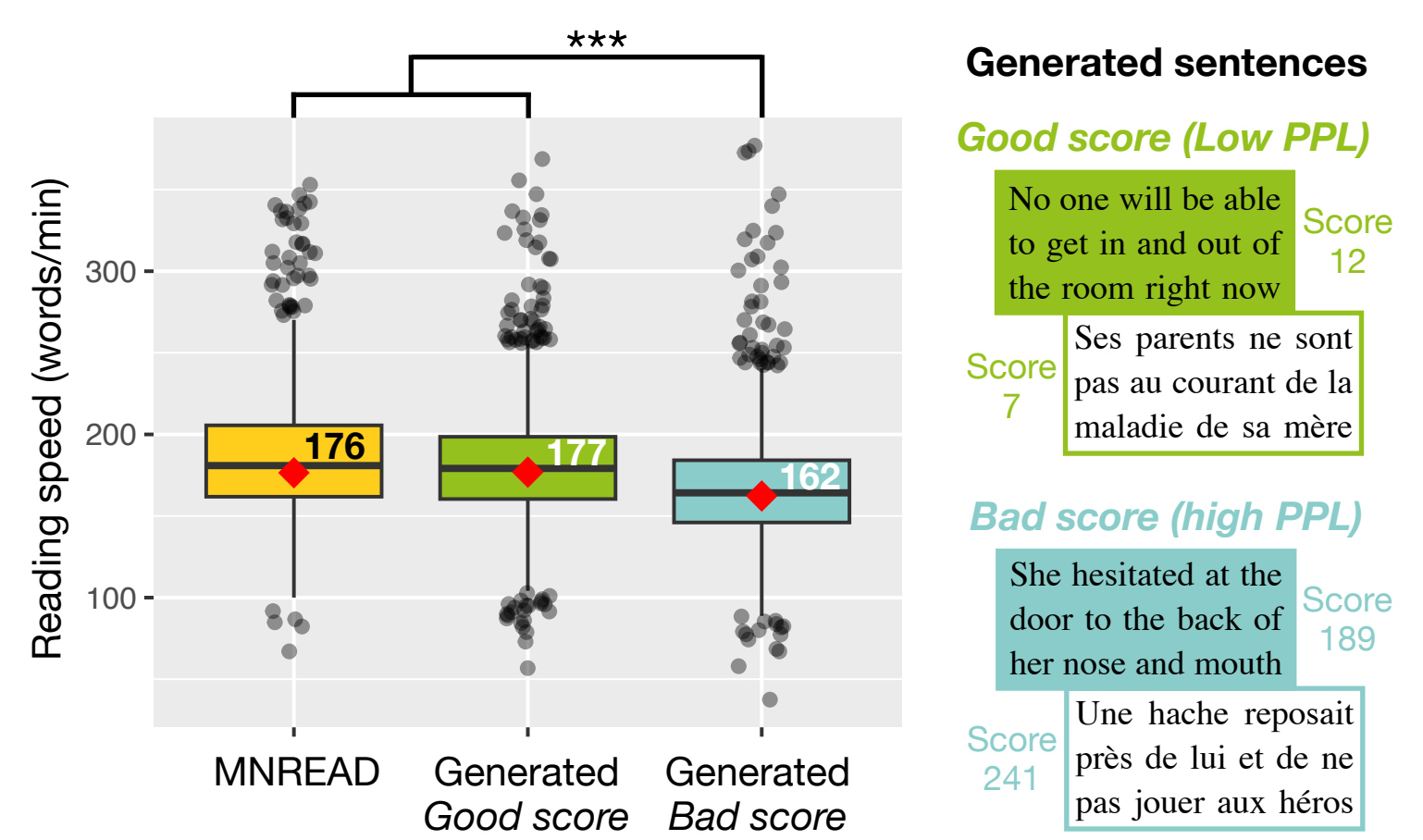
3 - Trending method... ask chatGPT

Mansfield JS, Atilgan N, Lewis AM, Legge GE. Extending the MNREAD sentence corpus: Computer-generated sentences for measuring visual performance in reading. *Vision Res.* 2019 May;158:11-18.

MNREAD rules



Results



Conclusions

- Our method provides valid standardized sentences that follow the MNREAD rules and yield similar performance, at least in French.
- This method is flexible and allows to customise the rules easily.
- This method is applicable to other languages and remains to be validated in English, Spanish, Italian, etc.

Objective

Using the MNREAD as a proof-of-concept, we propose a new automated method to generate highly constrained short sentences, following the strict rules of MNREAD and applicable to different languages.

Methods

1- Creating the sentences

Step 1: Corpus

Step 2: Set of n-grams

Step 3: Recombination

Multi-valued Decision Diagrams (MDD)

- Each MNREAD rule is modeled using a Multi-valued Decision Diagram (MDD).
- The intersection of all diagrams results in MNREAD-like sentences → 7,000 in 72 sec.

2- Scoring the sentences

- GPT-2 was used to assign a perplexity score (PPL) to each sentence.
- The lower the score, the most "natural" the sentence.

3- Validating the sentences

→ Normally sighted subject read French sentences at fixed print size (1.2 logMAR at 40 cm)

Experiment 1	Experiment 2	Experiment 3
N = 14	N = 24	N = 12
Mean age = 33±14	Mean age = 40±14	Mean age = 24±6
VR headset	Computer screen	Computer screen + eye tracker
38 + 30 + 30	38 + 30 + 30	38 + 100 + 100