



**HAL**  
open science

# Génération de texte sous contraintes pour mesurer des performances de lecture: Une nouvelle approche basée sur les diagrammes de décisions multivalués

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

## ► To cite this version:

Alexandre Bonlarron, Aurelie Calabrese, Pierre Kornprobst, Jean-Charles Régin. Génération de texte sous contraintes pour mesurer des performances de lecture: Une nouvelle approche basée sur les diagrammes de décisions multivalués. MOMI2022 - Le Monde des Mathématiques Industrielles, May 2022, Biot, France. . hal-03673066

**HAL Id: hal-03673066**

**<https://inria.hal.science/hal-03673066v1>**

Submitted on 19 May 2022

**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.



# Génération de texte sous contraintes pour mesurer des performances de lecture

Une nouvelle approche basée sur les diagrammes de décisions multivalués

A. Bonlarron<sup>(1,3)</sup>, A. Calabrèse<sup>(2)</sup>, P. Kornprobst<sup>(1)</sup>, J.-C. Régin<sup>(3)</sup>

Contact : alexandre.bonlarron@inria.fr



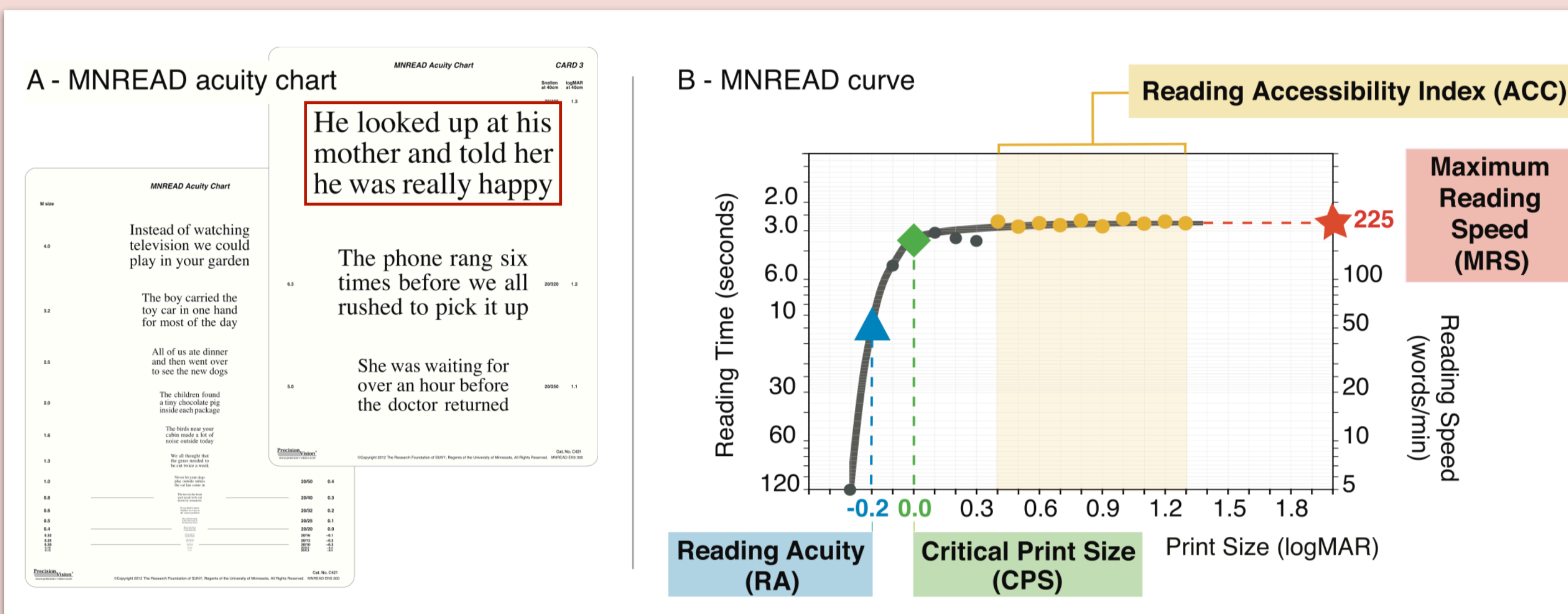
(1) Université Côte d'Azur, Inria, France  
 (2) Aix Marseille Univ, CNRS, LPC, Marseille, France  
 (3) Université Côte d'Azur, I3S, France



## Contexte

- Évaluation des performances visuelles d'un patient malvoyant en estimant sa vitesse de lecture
- Besoin d'une large base de textes **standardisés** respectant des règles liées à la grammaire, au lexique, à la taille ou à l'affichage
- Limitation : nombre trop restreint de phrases disponibles car les règles sont très contraignantes
- Problématique : **Comment générer automatiquement des textes respectant plusieurs règles ?**

## Cas d'étude : le test MNREAD [1]



Un test psychophysique basé sur des phrases standardisées, disponible en 19 langues (seulement 38 phrases MNREAD en français...)

## Méthodes

### Principe (n-gramme)

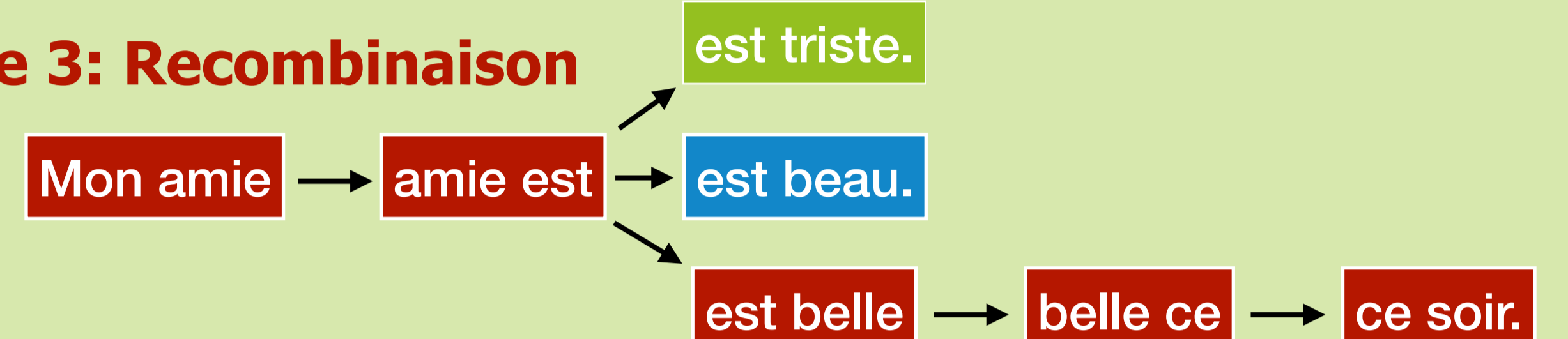
#### Etape 1 : Corpus

Il est beau.  
 Mon amie est triste.  
 Elle est belle ce soir.

#### Etape 2: Ensemble de 2-grammes

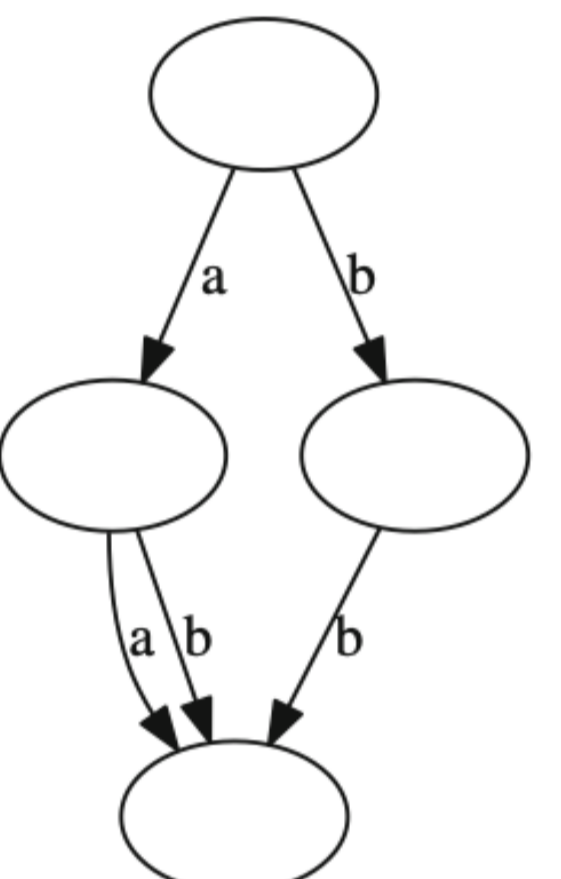
est belle amie est est beau. ce soir.  
 Mon amie belle ce Elle est est triste. Il est

#### Etape 3: Recombinaison



### Paradigme (MDD)

- Structures de données permettant de calculer et stocker les solutions d'un problème de façon compressée à l'aide d'un graphe orienté acyclique
- Avantage : possibilité de combiner les contraintes à l'aide d'intersections entre différents MDD [2]



Le paradigme MDD va permettre de décomposer le problème (engendrer du texte standardisé) en définissant autant de MDD que de règles que le texte doit respecter

## Résultats

### Des règles MNREAD aux MDD (illustration sur un sous-ensemble de règles)

Des phrases 9 à 15 mots

MDD Universel

Des phrases construites uniquement à partir des 3000 lemmes (forme canonique) des mots les plus fréquents dans les manuel de niveau CE2

MDD Lexique

Des phrases de 59 caractères avec les espaces et sans compter le point final

MDD Taille

Des phrases s'inspirant de suite de mots du corpus

MDD Corpus

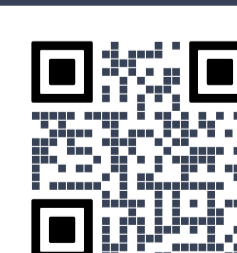
➔ L'intersection des MDD est égale à : **Le sac noir**

### Résultats de génération

Observation générale	Analyse de performance					Exemples de phrases engendrées en 5g		
n des n-grammes ↗						L'homme tourna la tête en direction de la voiture de police	La pluie formait un ruisseau de plus en plus insupportables	Vous lui attrapez la main et vous aide à monter sur son dos
Qualité des phrases ↗	3g	63k	36Go	6min	9899	Qualité : Excellente	Qualité : Moyenne	Qualité : Mauvaise
Nombre de phrases ↘	5g	6,3M	64Go	11h	4335	NB : Validation du sens des phrases a posteriori		

## Conclusion

Pour en savoir plus :



- Méthode prometteuse : plus adaptée que des méthodes génériques pour **gérer les contraintes** (e.g., GPT, Bert) et plus **flexible** que la méthode ad-hoc de Mansfield et al [3].
- Avantages : **modularité** (facile d'ajouter et/ou de retirer des règles), **prise en compte des contraintes dès la génération**, potentiellement applicable à d'autres langues
- Perspectives: contrôler la syntaxe et le sens

## Références

- [1] Mansfield, Ahn, et al. A new reading-acuity chart for normal and low vision. Ophthalmic and Visual Optics/Noninvasive Assessment of the Visual System Technical Digest, 3 :232-235, 1993.
- [2] Perez and Régin. Efficient Operations On MDDs for Building Constraint Programming Models, IJCAI, 2015.
- [3] Mansfield, Atilgan, et al. Extending the MNREAD sentence corpus: Computer-generated sentences for measuring visual performance in reading, 2019