

cub[AI]<sup>x</sup>

*NeuroSpell*

Le seul correcteur tout-automatique  
pour l'orthographe et la grammaire,  
basé sur une technologie neuronale (Deep Learning)

<http://neurospell.com>



# NeuroSpell

- Conçu pour prendre des décisions de corrections automatiques
  - ➔ Seuil de décision réglable
- Basé sur un modèle statistique neuronal
  - ➔ Entraînable sur des données spécifiques métier
  - ➔ Entraînable pour toute langue
- Les erreurs Speech-To-Text ou OCR peuvent être très différentes des erreurs humaines
  - ➔ Adaptable pour des types d'erreurs spécifiques aux traitements machines
  - ➔ Peut produire des corrections sur des phrases grammaticalement correctes
- Réécrit tous les mots en prenant chaque phrase comme un tout
  - ➔ Peut corriger des erreurs avec des dépendances distantes dans la phrase
  - ➔ Les mots inchangés sont validés

Ce restaurant serre des mais délicieux.  
Le mais pousse au soleil.  
■ Ce restaurant sert des mets délicieux.  
Le maïs pousse au soleil.

# Contrôler n'est pas corriger



Ce restaurant serre des amis délicieux.  
Le mais pousse au soleil.

**! Erreur de grammaire**  
Dans ce contexte, le mot **mais** pourrait être confondu avec son paronyme **amis**.  
**mais -> amis**



Ce restaurant serre **des mais délicieux.**

**Le mais pousse** au soleil.

Normalement, on met une virgule à la gauche de *mais*.



Ce restaurant serre des mais délicieux.

Le mais pousse au soleil.

mais → amis

mais pourrait être confondu avec amis.



mais → **mais**



Ce restaurant serre des mais délicieux.

Le mais pousse au soleil.

mes

La conjonction *mais* est confondue avec le déterminant *mes*.



Ce restaurant serre des mais délicieux.

Le mais pousse au soleil.

**mai**

Dans le groupe nominal « Le mais », le mot « **mais** » ne semble pas accordé avec le reste du groupe : « Le ».

Les systèmes à base de règles sont conçus pour chercher des mots douteux. Ils peuvent produire des suggestions qui nécessitent une décision manuelle.

# Contrôleurs à base de règles

- 1) Analyse grammaticale (verbe, nom, adjectif, etc.)
- 2) Détection d'incohérences lexicales et grammaticales (dictionnaires, règles écrites à la main, modèles statistiques)

- POUR :

- Souvent explicables

- CONTRES :

- Difficiles à utiliser pour de la correction automatique (contrôler n'est pas corriger / doute vs choix)
- Peuvent être perturbés par des erreurs ou des mots inconnus  
L'analyse grammaticale peut être fautive même sur des phrases correctes
- Prennent difficilement en compte les dépendances distantes dans la phrase
- Trouvent difficilement les erreurs produisant une phrase grammaticalement correcte
- Les mots non-signalés ou non-corrigés ne sont pas validés
- Les meilleurs ne sont pas disponibles dans un grand nombre de langues
- Difficiles à adapter sur le vocabulaire et les tournures spécifiques d'un domaine métier
- Non destinés à corriger des erreurs spécifiques aux traitements machines (conçus pour des erreurs humaines)

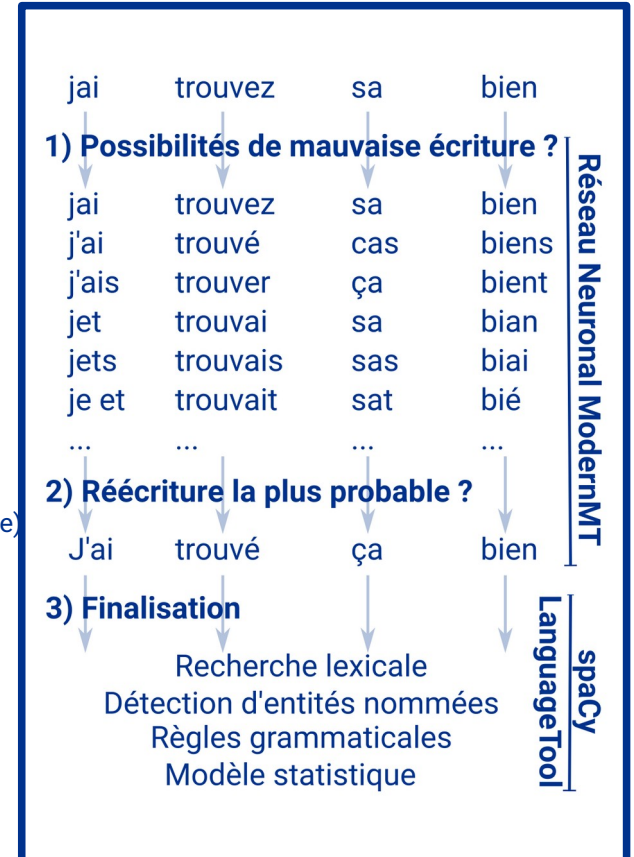
Il prend sa **part**  
et il **part** loin.

Il (PRON)  
prend (VERB)  
sa (DET)  
**part (NOUN)**  
et (CONJ)  
il (PRON)  
**part (VERB)**  
loin (ADV)  
. (PUNCT)



# NeuroSpell

- 1) Évalue toutes les manières de réécrire chaque mot ou groupe de mots d'une phrase (les variations possibles doivent être définies ou collectées)
  - 2) Produit la forme écrite la plus statistiquement probable en prenant en compte la plausibilité statistique de la phrase à corriger (seuil de décision ajustable)
- **POURS :**
    - Conçu pour réellement effectuer des corrections automatiques
    - La qualité produite est (presque) insensible aux erreurs prédéfinies (variations d'écritures), peu importe leur nature ou leur nombre dans la phrase
    - Peut prendre en compte les dépendances distantes dans la phrase (phrase analysée comme un tout)
    - Peut produire des corrections d'erreurs sur des phrases grammaticalement justes (effet statistique)
    - Les mots inchangés sont validés (selon les possibles variantes d'écriture)
    - Peut être créé pour toute nouvelle langue à un coût raisonnable (actuellement 20 langues disponibles)
    - Entraînable sur un vocabulaire et des tournures spécifiques à un domaine métier (grand corpus nécessaire)
    - Peut être entraîné sur des erreurs très spécifiques aux traitements machines
    - Est combiné à un contrôleur à base de règles (système hybride)
  - **CONTRE :**
    - Souvent non-explicable



# cub[AI]<sup>x</sup>

## Etienne Monneret

- 10 ans : industrie documentaire multi-langues (Jouve)
- 14 ans : édition de logiciel pour la traduction (Lingua et Machina)
- 3 ans : expert en IA linguistique / développeur logiciel (fondateur de cubAix SASU)
  - Correcteur automatique d'orthographe et de grammaire basé sur une technologie neuronale unique
  - Extraction terminologique automatique
  - Appariement / indexation sémantique
  - Traduction automatique neuronale (NMT) adaptée à un domaine métier
  - Workflow de traduction / CAT tool / Mémoire de traduction
  - ... comparaison multi-lingue de documents, éditeur vidéo stéréoscopique, etc.

<http://cubaix.com>

