

PlatON

Plateforme d'aide à l'apprentissage de l'Orthographe sur le Net

Richard Beaufort, PhD
UCL CENTAL
Groupe Norme



normalink
from correction to learning

Plan

- Pourquoi cette plateforme?
- Démonstration
- Outils et algorithmes
- Evaluation
- Conclusion

Plan

- **Pourquoi cette plateforme?**
- Démonstration
- Outils et algorithmes
- Evaluation
- Conclusion

Un point sur l'ALAO/ELAO

- ALAO/ELAO :
 - Apprentissage/Enseignement
 - des Langues
 - Assisté par **Ordinateur**
 - Cours sur CD-ROM
 - Plateforme en ligne
- Anglais: CALL
 - *Computer Assisted Language Learning*
- Un type de e-learning

Un point sur l'ALAO/ELAO

- Objectif
 - **Améliorer** l'enseignement
 - des langues **secondes**
 - % enseignement **traditionnel**
- Méthode
 - Multimédia (écouter, parler)
 - Liberté horaire (en classe, à distance)
 - Avec ou sans professeur humain
 - Interactivité apprenant-matière
 - Exercices automatisés

Un point sur l'ALAO/ELAO

- Les types exercices
 - **Fermés**
 - Nombre réduit de réponses possibles
 - Ex. : textes à trous, choix multiples
 - Peu de réflexion, connaissances en étude
 - Niveau CEFRL : à pd débutant (A1)

Un point sur l'ALAO/ELAO

- Les types exercices
 - **Ouverts**
 - Réponses libres
 - Ex. : dissertation
 - Beaucoup de réflexion, connaissances acquises
 - Niveau CEFRL : à pd avancé (C1)

Un point sur l'ALAO/ELAO

- Les types exercices
 - **Semi-ouverts**
 - Large éventail de réponses, limitées *d'une certaine manière*
 - Ex. : traduction guidée
 - Une certaine réflexion, connaissances acquises
 - Niveau CEFRL : à pd intermédiaire (B1)

Un point sur l'ALAO/ELAO

- Les types exercices

	Nb réponses	Difficulté	Niveau (àpd)
Fermés	1 à qq unes	Réfl ---, en étude	A1
Semi-ouverts	Plusieurs	Réfl +-, acquis	B1
Ouverts	Illimité	Réfl +++, acquis	C1

Un point sur l'ALAO/ELAO

- Les types exercices

	Nb réponses	Degré prédictibilité réponses
Fermés	1 à qq unes	Maximum
Semi-ouverts	Plusieurs	Important
Ouverts	Illimité	Limité

Desmet & Héroguel (2005)

Un point sur l'ALAO/ELAO

- **Limite** actuelle en ALAO/ELAO

	Nb réponses	Difficulté	Niveau (àpd)
Fermés	1 à qq unes	Réfl ---, en étude	A1
Semi-ouverts	Plusieurs	Réfl +-, acquis	B1
Ouverts	Illimité	Réfl +++, acquis	C1

Un point sur l'ALAO/ELAO

- **Tendance** en ALAO/ELAO

	Nb réponses	Difficulté	Niveau (àpd)
Fermés	1 à qq unes	Réfl ---, en étude	A1
Semi-ouverts	Plusieurs	Réfl +-, acquis	B1
Ouverts	Illimité	Réfl +++, acquis	C1

Un point sur l'ALAO/ELAO

- **Tendance** en ALAO/ELAO

	Nb réponses	Degré prédictibilité réponses
Fermés	1 à qq unes	Maximum
Semi-ouverts	Plusieurs	Important
Ouverts	Illimité	Limité

Desmet & Héroguel (2005)

PlatON

- Plateforme d'aide à l'apprentissage de **l'Orthographe** sur le Net
- Public
 - Langue seconde :
 - C1-C2
 - Langue maternelle :
 - 2^e partie des primaires (\pm 10 ans)

PlatON

- Plateforme « collaborative »
 - Enseignants : mettent le contenu
 - Apprenants : utilisent le contenu

Pourquoi cette plateforme?

PlatON

- A terme
 - Cours/matière
 - Exercices variés
 - De tous types
 - Exercices générés
 - Communication enseignant/apprenant

Pourquoi cette plateforme?

PlatON

- **Objectif actuel**
 - **Dépasser** les exercices fermés
 - Proposer des exercices **semi-ouverts**
 - Hypothèse : **ensemble de réponses gérable**
 - **Dictée**
 - Semi-ouvert grâce à l'original

Plan

- Pourquoi cette plateforme?
- **Démonstration**
- Outils et algorithmes
- Evaluation
- Conclusion

En « ligne »...

- <http://www.normalink.com/platon/>

Plan

- Pourquoi cette plateforme?
- Démonstration
- **Outils et algorithmes**
- Evaluation
- Conclusion

Besoins

- Correction automatique d'une copie
 - Comparer...
 - Des versions d'un texte (original/copie)
 - Savoir où on se trouve...
 - Dans les deux textes
 - Proposer une analyse linguistique
 - Des erreurs

Besoins

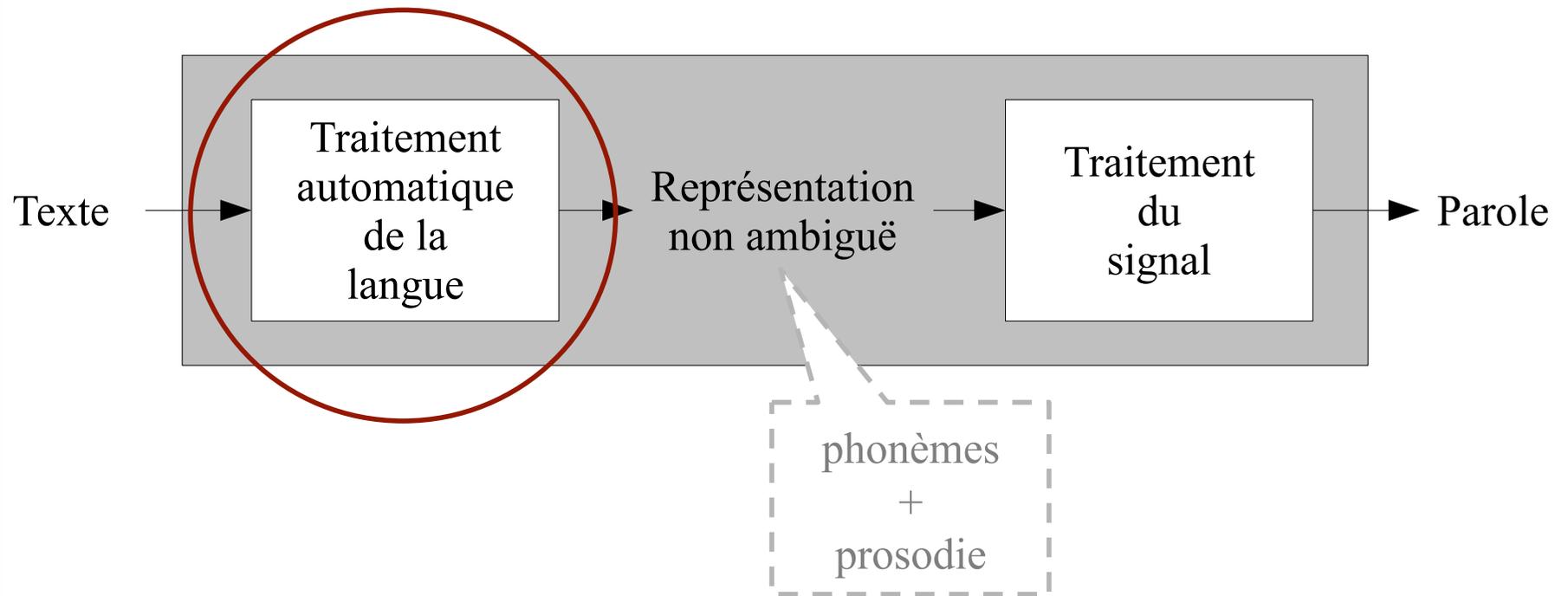
- Réalisation automatique de l'exercice
 - Lire le texte
 - Synthèse
 - Enregistrement
 - Guider l'apprenant...
 - Dans le texte
 - Phrase par phrase

Astuce!

- Préparation **semi-automatique** de la dictée
 - Professeur
 - Choisit le cours et la leçon
 - Introduit l'original
 - Système
 - Analyse l'original
 - Découpe l'original en phrase
 - Synthétise/enregistre la parole

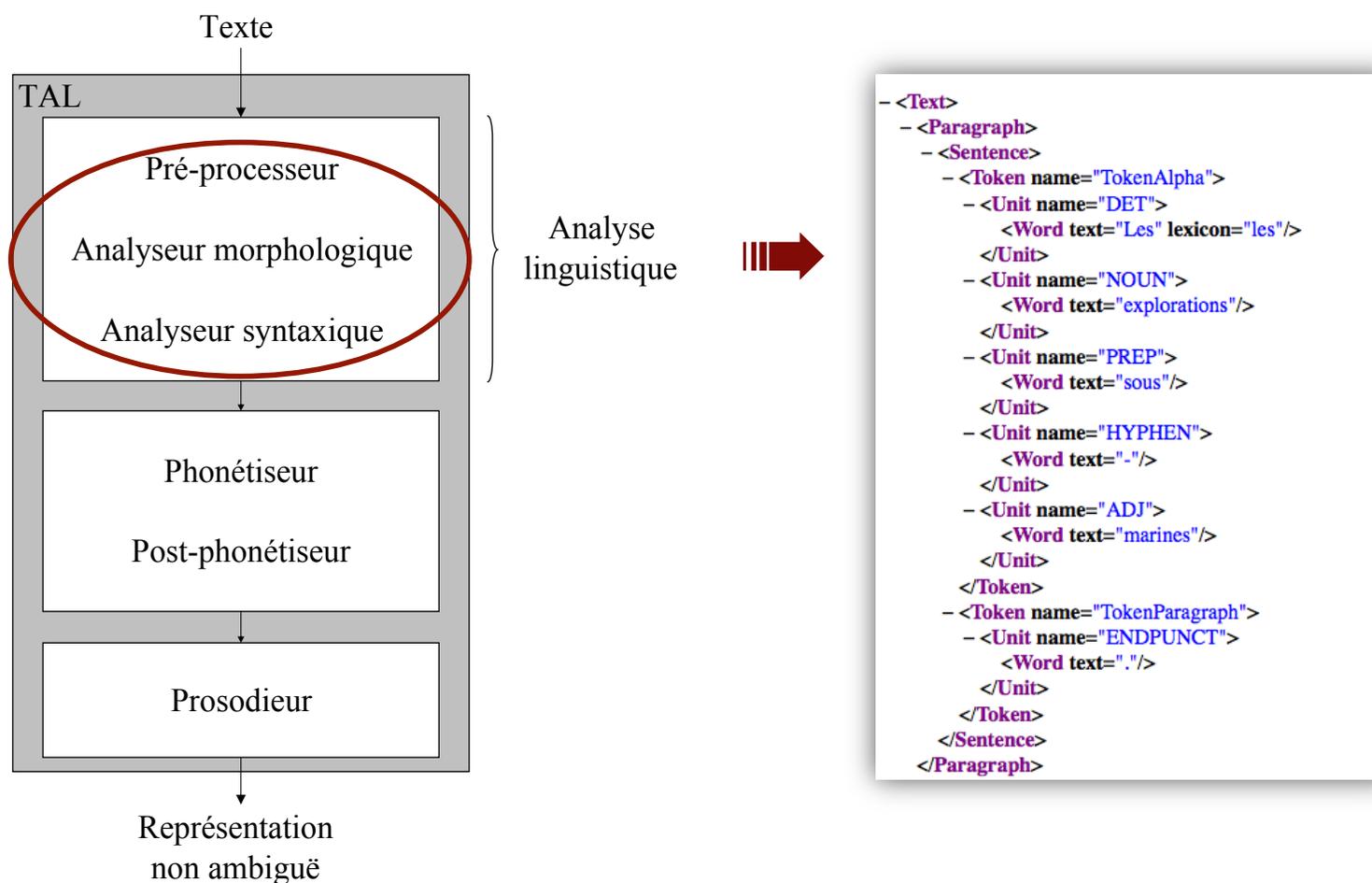
Outils (1)

- Synthétiseur de la parole à partir du texte



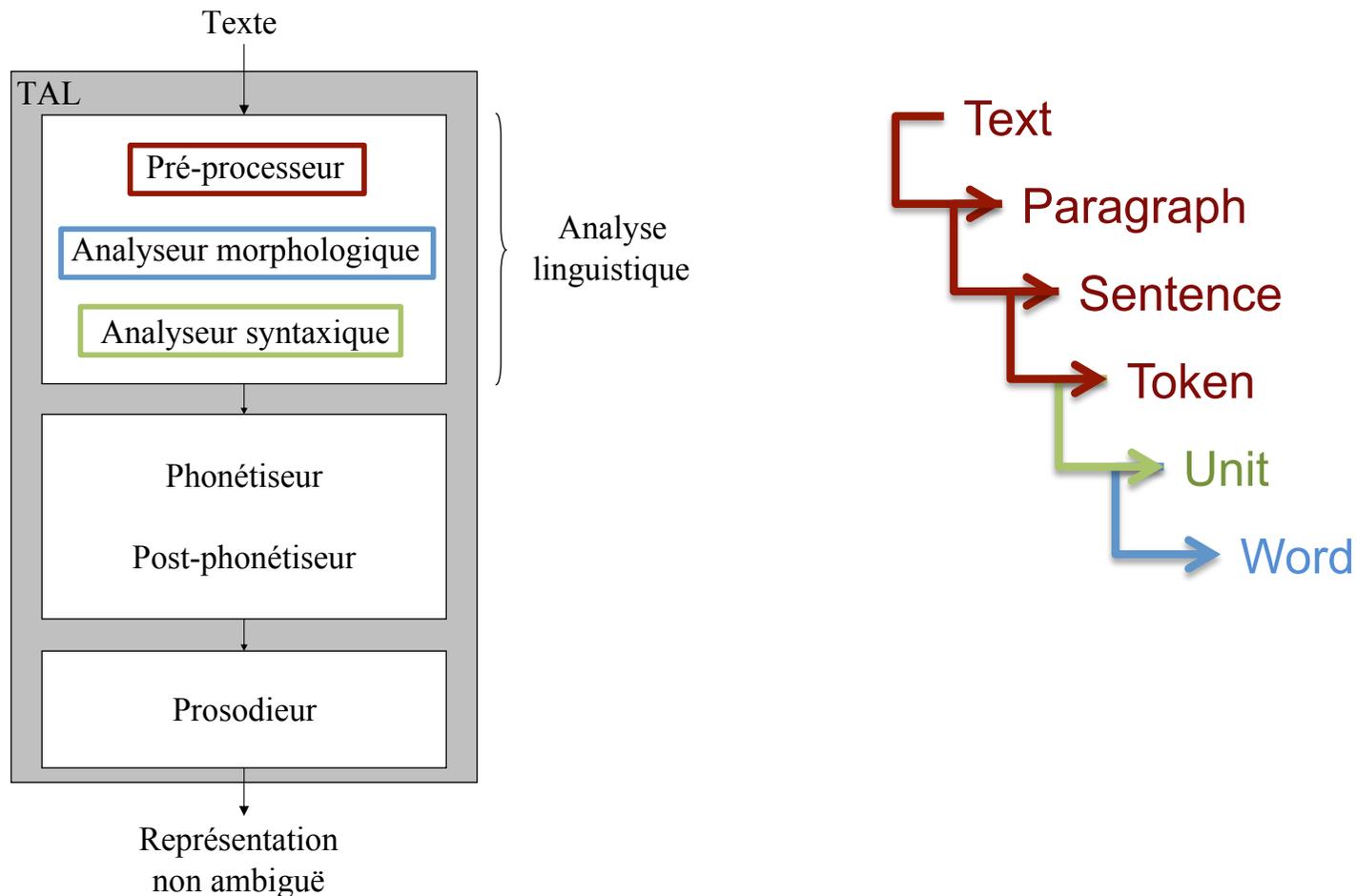
Outils (1)

- Synthétiseur de la parole à partir du texte



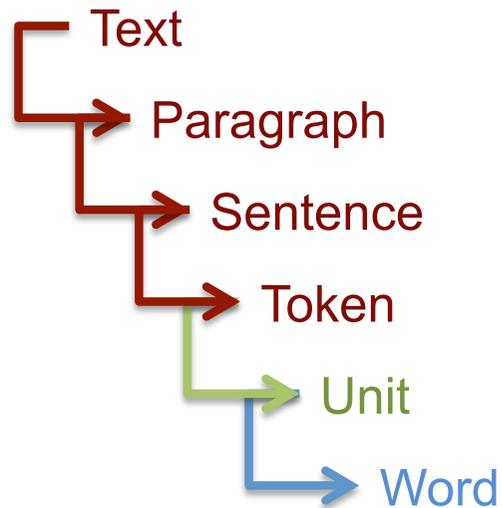
Outils (1)

- Synthétiseur de la parole à partir du texte



Outils (1)

- Synthétiseur de la parole à partir du texte



= URI, unités, monnaies ... num., alpha.

= Mots/verbes composés

= Plus petite partie d'un token

Outils (1)

- Synthétiseur de la parole à partir du texte

eLite-LIONS

- eLite: Beaufort & Ruelle (2006)
- LiONS: Colotte & Beaufort (2005)

Outils (2)

- Machines à états finis (1)
 - Définissent des langages (ensemble de séquences)

$$L = \{"a", "b", "c", "aa", "ab", "abc", "adef"\}$$

- Définissent des relations entre langage

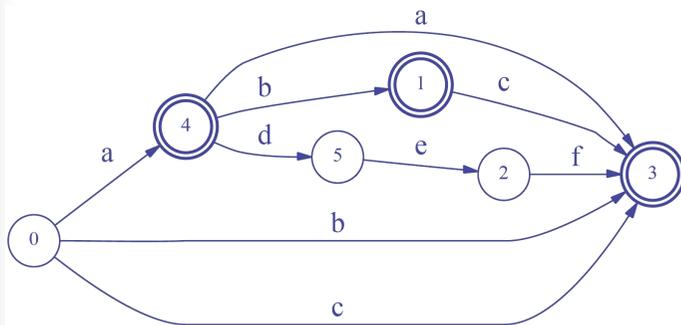
$$L_1 \rightarrow L_2 = \{"a" \rightarrow "e", "b" \rightarrow "f", "c" \rightarrow "g", \\ "aa" \rightarrow "jk", "ab" \rightarrow "lm"\}$$

- Pondèrent des langages et des relations

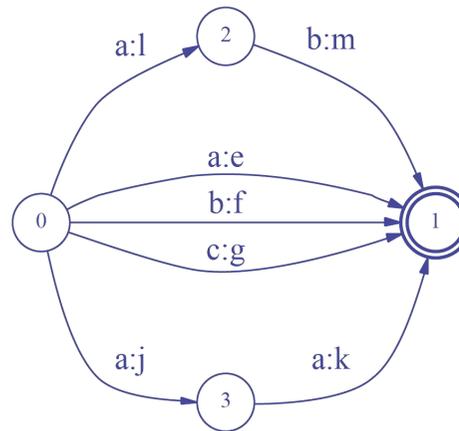
$$L_1 \rightarrow L_2 = \{"aa" \rightarrow "hi" / 1, "aa" \rightarrow "jk" / 2, \\ "ab" \rightarrow "jk" / 1, "ab" \rightarrow "lm" / 2\}$$

Outils (2)

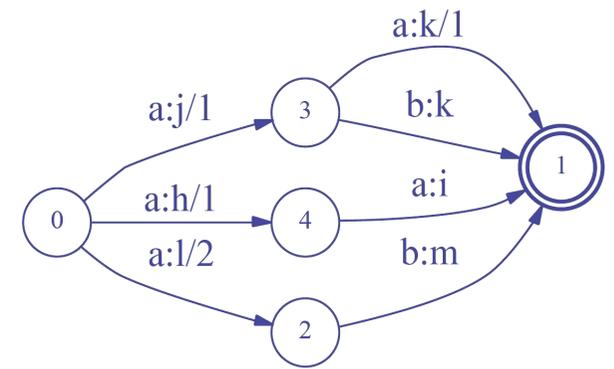
- Machines à états finis (2)
 - Représentables par des graphes orientés étiquetés



Automate
(langage)



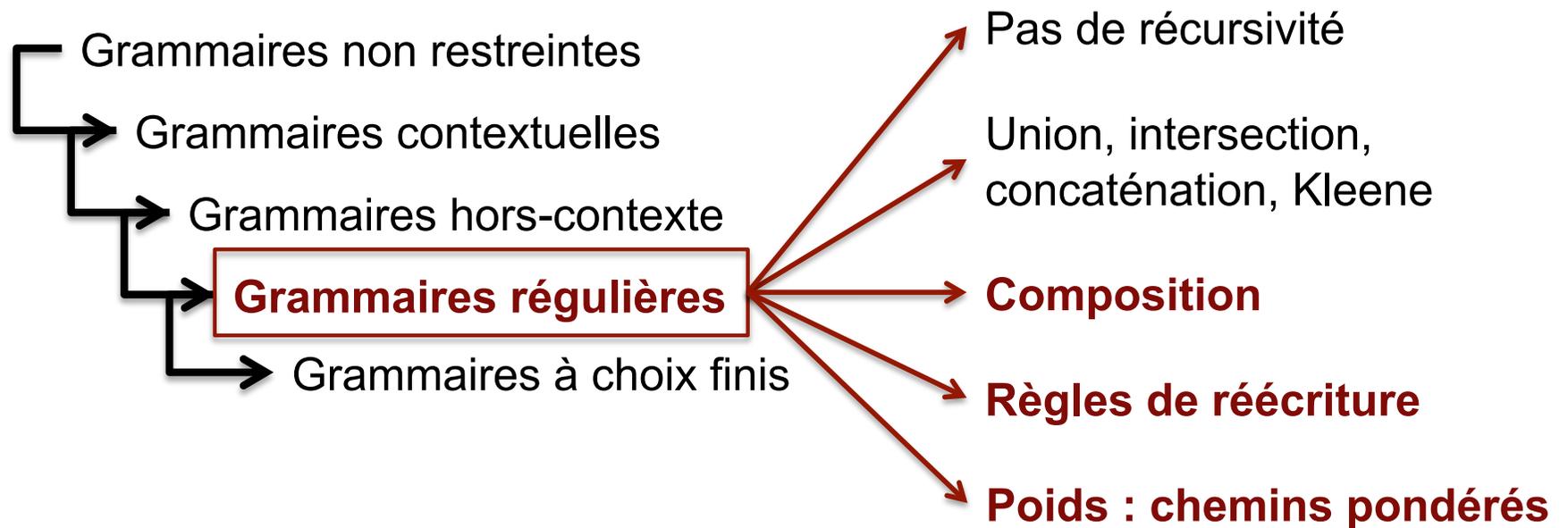
Transducteur
(relation)



Transducteur pondérée
(relation pondérée)

Outils (2)

- Machines à états finis (3)
 - Langages réguliers (Chomsky 1956, Schützenberger 1961)



Outils (2)

- Machines à états finis (3)
 - Langages réguliers (Chomsky 1956, Schützenberger 1961)

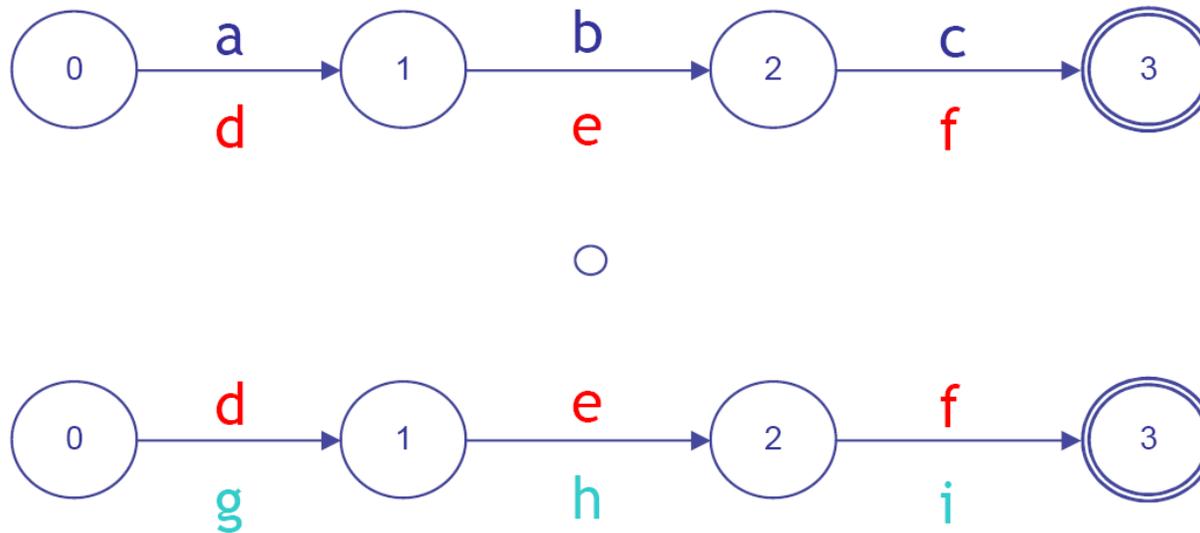
Composition

Création de relations complexes
à partir de
relations simples

Outils (2)

- Machines à états finis (3)
 - Langages réguliers (Chomsky 1956, Schützenberger 1961)

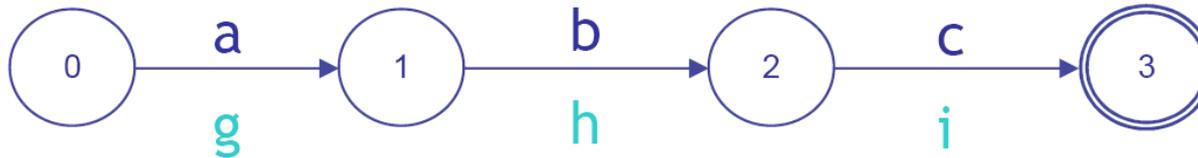
Composition



Outils (2)

- Machines à états finis (3)
 - Langages réguliers (Chomsky 1956, Schützenberger 1961)

Composition



Outils (2)

- Machines à états finis (3)
 - Langages réguliers (Chomsky 1956, Schützenberger 1961)

Règles de réécriture

$$\phi \rightarrow \psi :: \lambda _ \rho$$

$$\phi \rightarrow \psi :: \lambda _ \rho / \omega$$

$$\phi \xrightarrow{?} \psi :: \lambda _ \rho$$

$$\phi \xrightarrow{?} \psi :: \lambda _ \rho / \omega$$

Outils (2)

- Machines à états finis (3)
 - Langages réguliers (Chomsky 1956, Schützenberger 1961)

Poids : chemins pondérés

Choix possible pour problèmes du type « n-best »

Outils (2)

- Machines à états finis

Bibliothèque FSM & compilateur Ovide

- Beaufort (2008)

Algorithmes

- Ce que nous avons
 - La copie de l'étudiant et l'original de la dictée
 - L'analyse linguistique de l'original
- Algorithme en 3 temps
 1. Détection des erreurs
 2. Parsing parallèle erreurs/analyse
 3. Génération de diagnostics

Algorithmes (1)

- Détection des erreurs
 - Comparaison : alignement

pourraient demandé
pourrait demander

Algorithmes (1)

- Détection des erreurs
 - Comparaison : alignement

```
pourraient demandé_  
pourrai__t demander
```

Algorithmes (1)

- Détection des erreurs
 - Comparaison : alignement

pourraient demandé_
pourrait__ demander

Algorithmes (1)

- Détection des erreurs
 - Comparaison : alignement

... c e l u i d e s a b i s s e s ...

... c e l u i d e s a b y s s e s ...

Algorithmes (1)

- Détection des erreurs
 - Comparaison : alignement

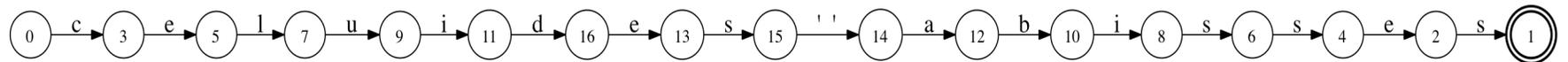
... c e l u i d e s a b i s s e s ...

idem: 0
substitution: 1
insertion: 1
deletion: 1

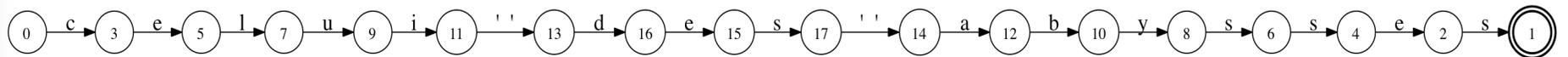
... c e l u i d e s a b y s s e s ...

Algorithmes (1)

- Détection des erreurs
 - Comparaison : alignement

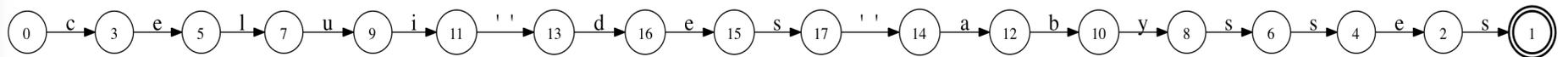
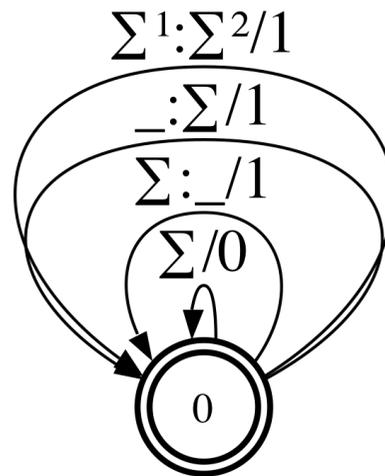
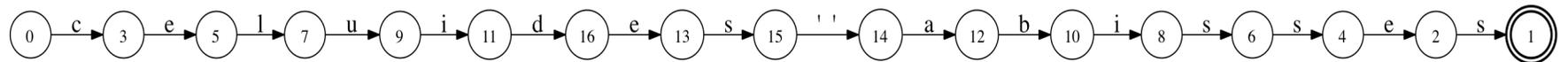


idem: 0
substitution: 1
insertion: 1
deletion: 1



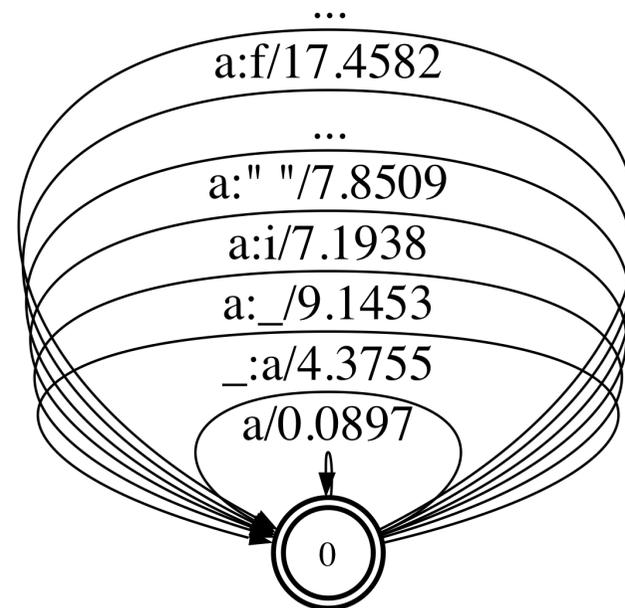
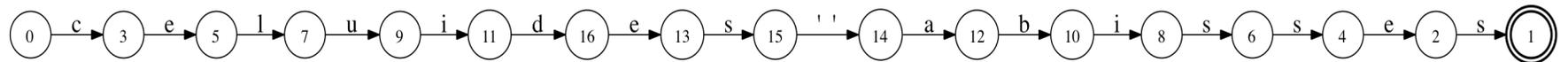
Algorithmes (1)

- Détection des erreurs
 - Comparaison : alignement

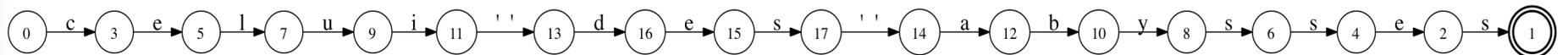


Algorithmes (1)

- Détection des erreurs
 - Comparaison : alignement

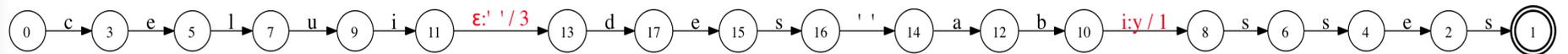


Beaufort (2010)



Algorithmes (1)

- Détection des erreurs
 - Comparaison : alignement



Algorithmes (1)

- Détection des erreurs
 - Comparaison : alignement

```
... c e l u i _ d e s a b i s s e s ...  
... c e l u i   d e s a b y s s e s ...
```

Algorithmes (2)

- Parsing parallèle erreurs/analyse

... c e l u i _ d e s a b i s s e s ...
... c e l u i d e s a b y s s e s ...

```
- <Unit name="PRONDEM">  
  <Word text="celui"/>  
</Unit>  
- <Unit name="DETPREP">  
  <Word text="des"/>  
</Unit>  
- <Unit name="NOUN">  
  <Word text="abysses"/>  
</Unit>
```

Algorithmes (2)

- Parsing parallèle erreurs/analyse

... c e l u i _ d e s a b i s s e s ...
... c e l u i d e s a b y s s e s ...

```
- <Unit name="PRONDEM">  
  <Word text="celui"/>  
</Unit>  
- <Unit name="DETPREP">  
  <Word text="des"/>  
</Unit>  
- <Unit name="NOUN">  
  <Word text="abysses"/>  
</Unit>
```

Algorithmes (2)

- Parsing parallèle erreurs/analyse

... c e l u i _ d e s a b i s s e s ...
... c e l u i d e s a b y s s e s ...

```
- <Unit name="PRONDEM">  
  <Word text="celui"/>  
</Unit>  
- <Unit name="DETPREP">  
  <Word text="des"/>  
</Unit>  
- <Unit name="NOUN">  
  <Word text="abysses"/>  
</Unit>
```

Algorithmes (2)

- Parsing parallèle erreurs/analyse

... c e l u i _ d e s a b i s s e s ...
... c e l u i d e s a b y s s e s ...

```
- <Unit name="PRONDEM">  
  <Word text="celui"/>  
</Unit>  
- <Unit name="DETPREP">  
  <Word text="des"/>  
</Unit>  
- <Unit name="NOUN">  
  <Word text="abysses"/>  
</Unit>
```

Algorithmes (2)

- Parsing parallèle erreurs/analyse

... c e l u i _ d e s a b i s s e s ...
... c e l u i d e s a b y s s e s ...

```
- <Unit name="PRONDEM">  
  <Word text="celui"/>  
</Unit>  
- <Unit name="DETPREP">  
  <Word text="des"/>  
</Unit>  
- <Unit name="NOUN">  
  <Word text="abysses"/>  
</Unit>
```

Algorithmes (2)

- Parsing parallèle erreurs/analyse

... c e l u i _ d e s a b i s s e s ...
... c e l u i d e s a b y s s e s ...

```
- <Unit name="PRONDEM">  
  <Word text="celui"/>  
</Unit>  
- <Unit name="DETPREP">  
  <Word text="des"/>  
</Unit>  
- <Unit name="NOUN">  
  <Word text="abysses"/>  
</Unit>
```

Algorithmes (2)

- Parsing parallèle erreurs/analyse

... c e l u i _ d e s a b i s s e s ...
... c e l u i d e s a b y s s e s ...

```
- <Unit name="PRONDEM">  
  <Word text="celui"/>  
</Unit>  
- <Unit name="DETPREP">  
  <Word text="des"/>  
</Unit>  
- <Unit name="NOUN">  
  <Word text="abysses"/>  
</Unit>
```

Algorithmes (2)

- Parsing parallèle erreurs/analyse

... c e l u i _ d e s a b i s s e s ...
... c e l u i d e s a b y s s e s ...

```
- <Unit name="PRONDEM">  
  <Word text="celui"/>  
</Unit>  
- <Unit name="DETPREP">  
  <Word text="des"/>  
</Unit>  
- <Unit name="NOUN">  
  <Word text="abysses"/>  
</Unit>
```

Algorithmes (2)

- Parsing parallèle erreurs/analyse

```
- <Unit name="PRONDEM">  
  <Word text="celui"/>  
</Unit>  
- <Unit name="DETPREP">  
  <Word text="des"/>  
</Unit>  
- <Unit name="NOUN">  
  <Word text="abysses"/>  
</Unit>
```

Algorithmes (2)

- Parsing parallèle erreurs/analyse

```
- <Unit name="PRONDEM">  
  <Word text="celui"/>  
</Unit>  
- <Unit name="DETPREP">  
  <Word text=" " cat="SPACE" errPr="[_]" corrPr="{ }" errKind="17"/>  
  <Word text="des"/>  
</Unit>  
- <Unit name="NOUN">  
  <Word text="abysses" err="abisses" errPr="ab[[i]]sses" corrPr="ab{{y}}sses" errKind="4"/>  
</Unit>
```

Algorithmes (2)

- Parsing parallèle erreurs/analyse

errKind = étiquettes associées aux erreurs :

- Begin-, End-, InWord
- WordMiss, WordExtra
- BetWord
- OnlySep

Algorithmes (3)

- Génération de diagnostics
 - Etiquettes WordMiss et WordExtra
 - ⇒ « *Mot manquant* » ou « *superflu* »
 - ⇒ « *Ponctuation manquante* » ou « *superflue* »
 - ⇒ « *Espace manquante* » ou « *superflue* »

Algorithmes (3)

- Génération de diagnostics
 - Etiquette BetWord + WordMiss/WordExtra
 - Fusion (quel que → quelque, d'avantage → davantage)
 - Division (quelque → quel que, davantage → d'avantage)
 - Test morphologique réalisé
 - Autres diagnostics seulement si test échoue

Algorithmes (3)

- Génération de diagnostics
 - Erreur sur séparateur (ponctuation, espace)
 - Sans WordMiss, WordExtra, BetWord

⇒ « *Erreur de séparateur* »

Algorithmes (3)

- Génération de diagnostics
 - Erreur sur mot
 - Sans WordMiss, WordExtra, BetWord
 - Test morphologique
 - ⇒ « *Erreur lexicale* » (*hors-vocabulaire/erreur de lemme*)
 - ⇒ « *Erreur grammaticale* » (*erreur trait morphologique*)

Algorithmes (3)

- Test morphologique
 - Si BetWord
 - WordMiss: (espace manquante)
 - « *quel que* » → « *quelque* »
 - Erreur: mots autour du séparateur manquant collés sans séparateur (*quelque*)
 - WordExtra: (espace superflue)
 - « *quelque* » → « *quel que* »
 - Erreur: parties du mot testées séparément (*quel* et *que*)

Algorithmes (3)

- Test morphologique
 - Recherche du mot erroné dans le dictionnaire
 - Inconnu → erreur lexicale, mot hors-vocabulaire
 - Connu → **comparaison morphologique**
 - Comparaison morphologique
 - Plus grande intersection morphologique entre erroné/correct
 - Résultat:
 - Lemme différent : erreur lexicale, erreur de lemme
 - Autre erreur : erreur grammaticale (genre, nombre, pers., etc.)

Algorithmes (3)

- Comparaison morphologique

Intersection morphologique

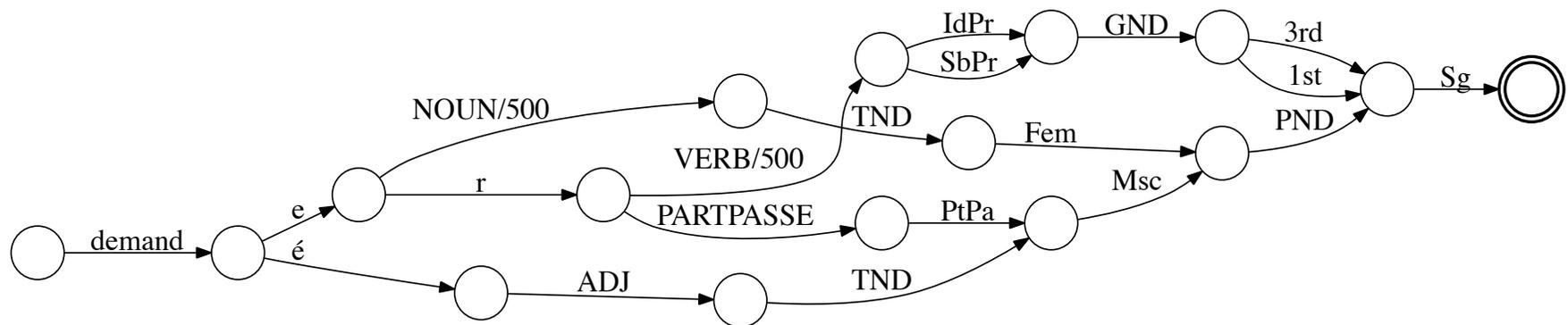
Par

Composition de transducteurs

Algorithmes (3)

- Comparaison morphologique

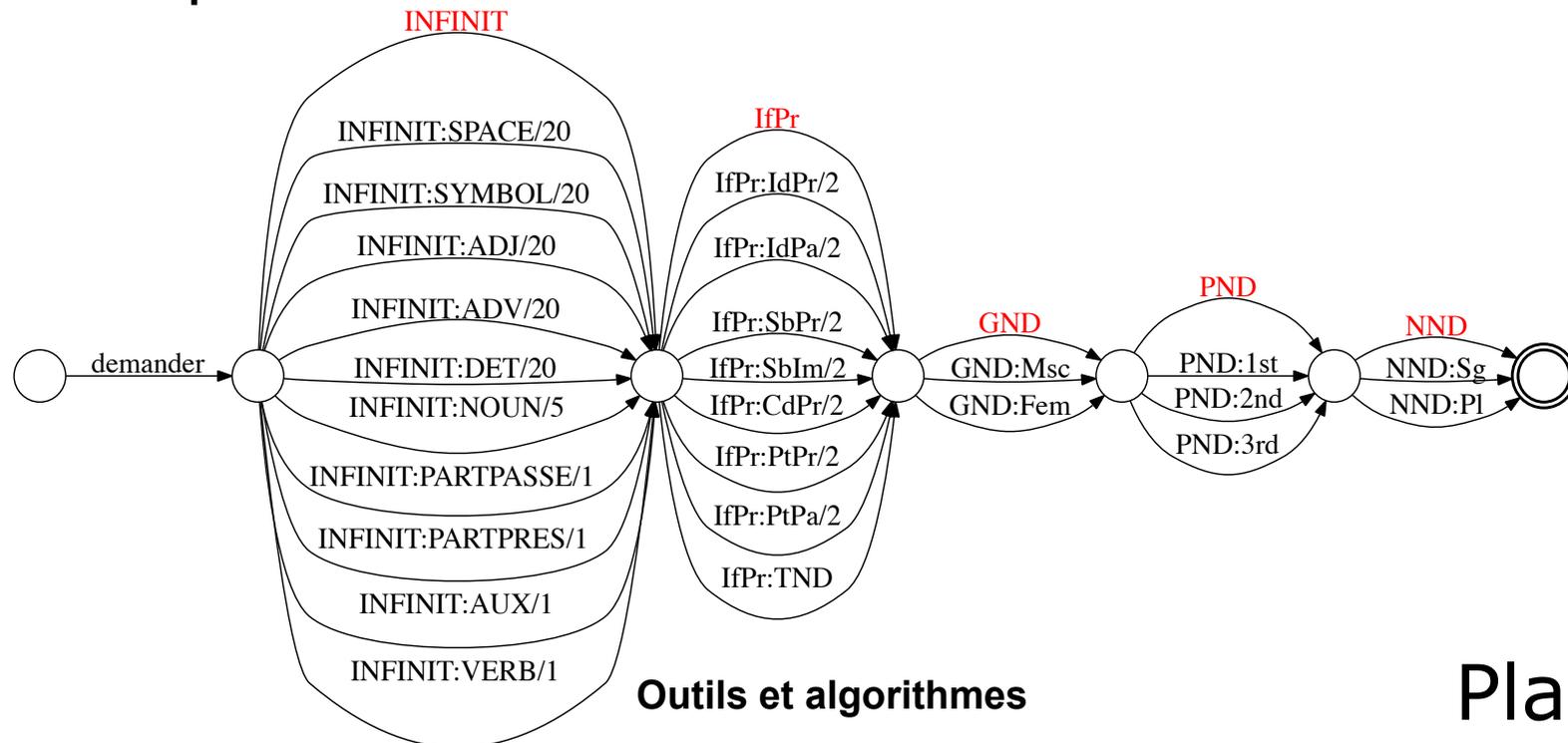
1) Recherche de la forme fautive dans le lexique
« pourrait **demandé** »



Algorithmes (3)

- Comparaison morphologique

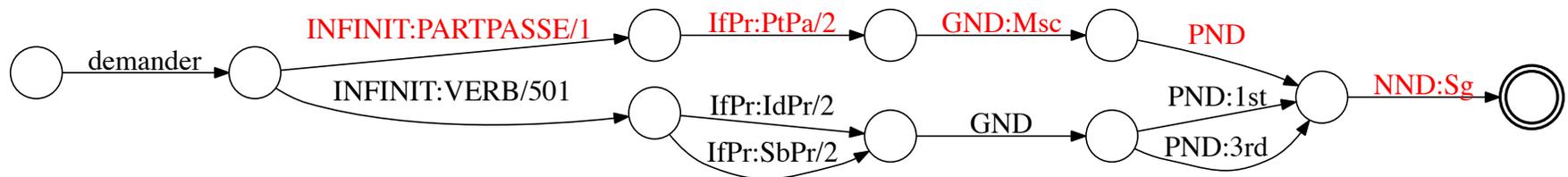
2) Extension de l'analyse de la forme correcte
« pourrait demander »



Algorithmes (3)

- Comparaison morphologique

3) Composition des deux transducteurs



Plan

- Pourquoi cette plateforme?
- Démonstration
- Outils et algorithmes
- **Evaluation**
- Conclusion

Méthodologie

- Corpus utilisé
 - Michèle Lenoble-Pinson (FUSL, Bruxelles)
 - 40 années de dictées (1969–2008)
 - 1^{re} et 2^e bacs
 - 40 dictées
 - 1.300 copies
 - 13.255 erreurs (± 10 erreurs/copie)
 - Numérisé par Fairon & Simon (2009)

Méthodologie

- Corpus de test :
 - 5 dictées, 1 année tous les 5 ans à pd 1969
 - 115 copies
 - 1.532 erreurs (13,3 erreurs/copie)
- Evaluation:
 - Performances
 - Qualité de la détection (alignement)
 - Qualité du diagnostic

Performance

- Environnement
 - Intel 2.66 GHz, 64bits, 4GB RAM
 - Ubuntu 10.04.2
- Résultats
 - 0,17ms/caractère, 1,05ms/mot
 - Ex.: 590ms pour 56 phrases de 566 mots, 3.559 caractères

Détection (alignement)

- Alignement correct à 99% (1521 sur 1532)
 - 11 phrases mal alignées
- Toujours même erreur
- Sur passage manquant ou superflu

... français_____e_____.

... française, quels qu'en puissent être la gravité et le nombre.

Diagnostic

- Sur les 1521 phrases bien alignées

Type of error	Superficial feedback		In-depth feedback		Morpho-syntactic analysis	
<i>Usage</i>	436/441	98,9%	434/441	98,4%	244/278	87,8%
<i>Grammar</i>	428/453	94,5%	411/453	90,7%	401/453	87,7%
<i>Usage and grammar</i>	62/69	89,5%	62/69	89,5%	51/63	80,9%
<i>Punctuation</i>	457/457	100%	457/457	100%	–	
<i>Transcription</i>	96/101	95%	96/101	95%	–	
<i>Total</i>	1479/1521	97,2%	1460/1521	96%	696/794	87,7%

- formes absentes du lexique (*bathysphère*)
- erreurs d'analyse par manque de contexte:
 - ind >< subj, 1^{re} pers >< 3^e pers

Plan

- Pourquoi cette plateforme?
- Démonstration
- Outils et algorithmes
- Evaluation
- **Conclusion**

Objectifs

- Plateforme collaborative
- Centrée actuellement sur dictée
 - Exercice semi-ouvert
- Principe
 - Automatiser
 - Réalisation
 - Correction
 - En semi-automatisant la préparation
- Proposer un diagnostic linguistique des erreurs

Limites & solutions envisagées

- Erreurs d'alignement
 - Calcul de similarité
 - Filiation
 - Bio-informatique
- Erreurs de diagnostic
 - Stratégie moins-disante
 - Appliquer grammaires de contraintes

Travaux futurs

- Exploitation des statistiques
 - Courbes d'évolution
 - Génération d'exercices dédiées

Références (1)

- Beaufort R. (2008), *Application des machines à états finis en synthèse de la parole. Sélection d'unités non uniformes et correction orthographique*, PhD thesis, Facultés Universitaires Notre-Dame de la Paix (FUNDP), Namur, Belgique. 605 pages.
- Beaufort R. (2010), Composition filtrée et marqueurs de règles de réécriture pour une distance d'édition flexible. Application à la correction des mots hors-vocabulaire, *Traitement Automatique des Langues*, **51** (1), pp. 11–40.
- Beaufort R. & Ruelle A (2006), eLite : système de synthèse de la parole à orientation linguistique, *Proceedings of JEP 2006*, pp. 509-512.
- Chomsky N. (1956), Three models for the description of language, *I.R.E. Transactions on Information Theory*, **IT-2** (3), 113–124.
- Colotte V. & Beaufort R (2005), Linguistic Features Weighting for a Text-to-Speech System Without Prosody Model, *Proceedings of Interspeech 2005*, pp. 2549-2552.
- Desmet P. & Héroguel A. (2005), Les enjeux de la création d'un environnement d'apprentissage électronique axé sur la compréhension orale à l'aide du système auteur IDIOMATIC, *Alsic. Apprentissage des Langues et Systèmes d'Information et de Communication*, **8** (1), pp. 281–303.

Références (2)

- Fairon C. & Simon A.C. (2009), Informatisation d'un corpus de dictées: 40 années de pratique orthographique (1967-2008), *Pour l'amour des mots. Glanures lexicales, dictionnairiques, grammaticales et syntaxiques. Hommage à Michèle Lenoble-Pinson*, pp. 131–154.
- Schützenberger M. (1961), On the definition of a family of automata, *Information and Control*, **4**, pp. 245–270.

Merci!

Des questions?