Desambiguïsation morpho-syntaxique à partir de transducteurs

1 Objectif

L'objectif de ce projet est d'effectuer de la désambiguïsation morpho-syntaxique à partir d'automates.

L'idée est de créer des automates qui représentent des suites interdites de catégories (et / ou) de mots. Ces automates seront combinés avec un automate A représentant toutes les analyses morpho-syntaxiques d'une phrase. A l'issue de la combinaison, les chemins correspondant aux combinaisons interdites auront été éliminés de A. Des détails sur la méthode sont donnés dans la section 2.

Pour élaborer vos règles de désambiguïsation, vous utiliserez un corpus dans lequel chaque mot potentiellement ambigu est utilisé avec une catégorie morpho-syntaxique particulière. Vous vous servirez de ce corpus pour "découvrir" des règles de désambiguïsation.

2 Désambiguïsation par automates

Le lexique Le lexique peut être représenté par une série de transducteurs qui associent à une forme fléchie cette même forme fléchie suivie de toutes ses catégories possibles :

```
macro(lexique, {(la x [la, @, {pro,n,det}, @]), (belle x [belle,
@, {adj,n}, @]), (ferme x [ferme, @, {n,v,adj}, @]), (porte x
[porte, @, {n,v}, @])}* ).
```

L'analyse morphologique de la phrase L'analyse morpho-syntaxique de la phrase consiste à appliquer le lexique décrit ci-dessus à la suite de formes fléchies composant la phrase.

```
macro(ana(Phrase), [range(Phrase o lexique)]).}
```

A l'issue de cette opération, chaque forme est suivie par ses catégories possibles :

```
[la, @, {pro,n,det}, @, belle, @, {adj,n}, @, ferme, @, {n,v,adj},
@,la, @, {pro,n,det}, @, porte, @, {n,v}, @]
```

Les contraintes négatives Chaque règles de désambiguïsation peut être représentée par un automate décrivant la séquence interdite. Cette dernière peut être composée uniquement de catégories :

```
macro(c1, [(? - @)*, @, pro , @, (? - @)*, @, n]).
```

ou d'un mélange de catégories et de formes :

```
macro(c2, [la, @, n , @, (? - @)*, @, n]).
```

Les contraintes sont ensuite combinées au sein d'un automate qui reconnaît toute suite de formes fléchies et de catégories contenant une occurrence d'une configuration interdite :

```
macro(contraintes, [[(? - @)*, @, (? - @)*, @]*, {c1,c2}, [@,
(? - @)*, @, (? - @)*]*]).
```

Désambiguïsation La désambiguïsation s'effectue en faisant la différence de deux automates : l'automate correspondant à l'analyse de la phrase et l'automate des contraintes :

```
macro(desamb(Phrase), [ana(Phrase),@] - contraintes).
```