

# Outils de transformation de graphes pour l'aide à l'annotation, conversion de formats et vérification de cohérence des annotations

**Bruno Guillaume**

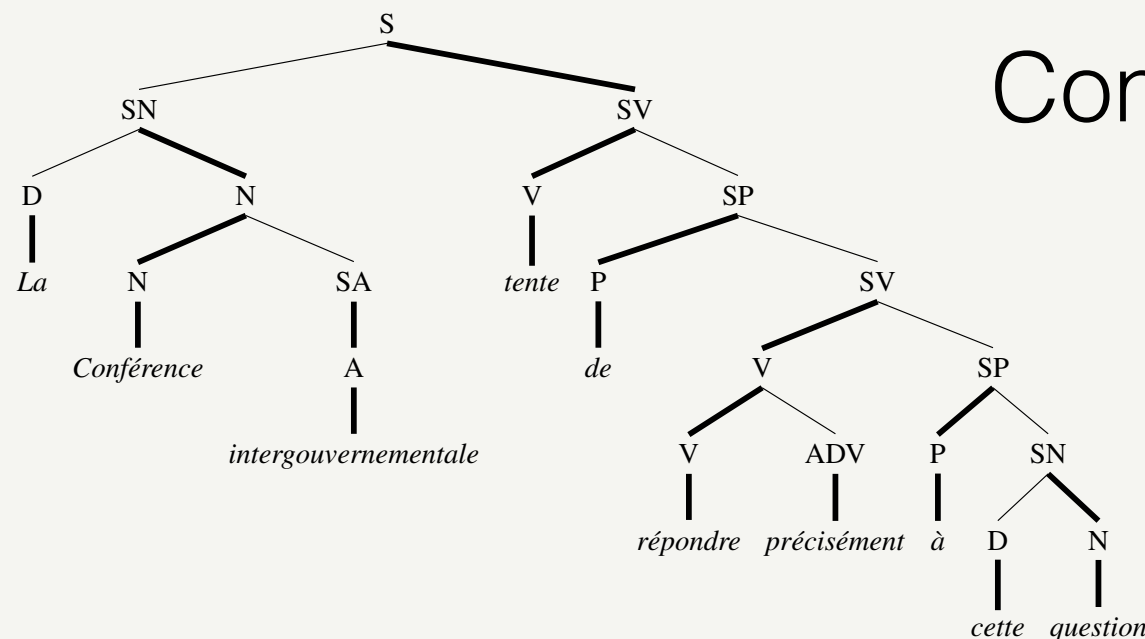
Université de Lorraine, CNRS, Inria, LORIA  
bruno.Guillaume@loria.fr

Consortium PARSEME-FR  
Juin 2019

Collaborations :

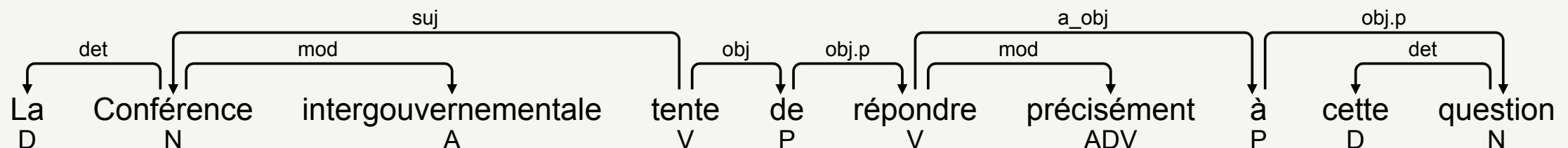
- **Guillaume Bonfante & Guy Perrier** (réécriture de graphes)
- **Marie Candito** (annotations PARSEME)

# Arbres ou graphes pour les représentations linguistiques ?

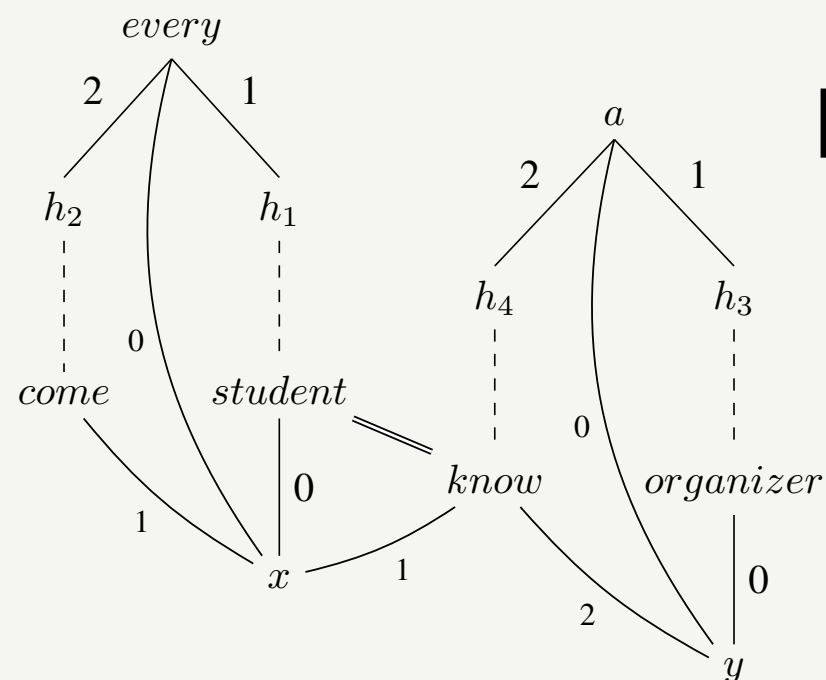


Constituants

Dépendances

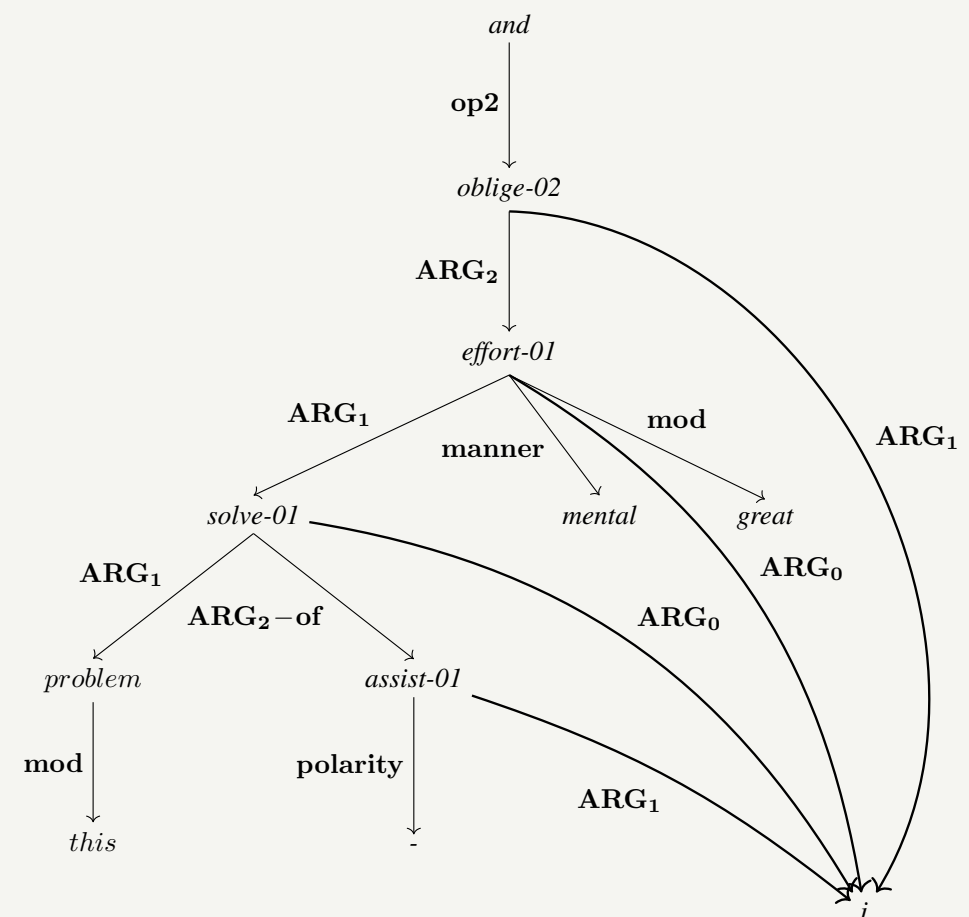


# Arbres ou graphes pour les représentations linguistiques ?

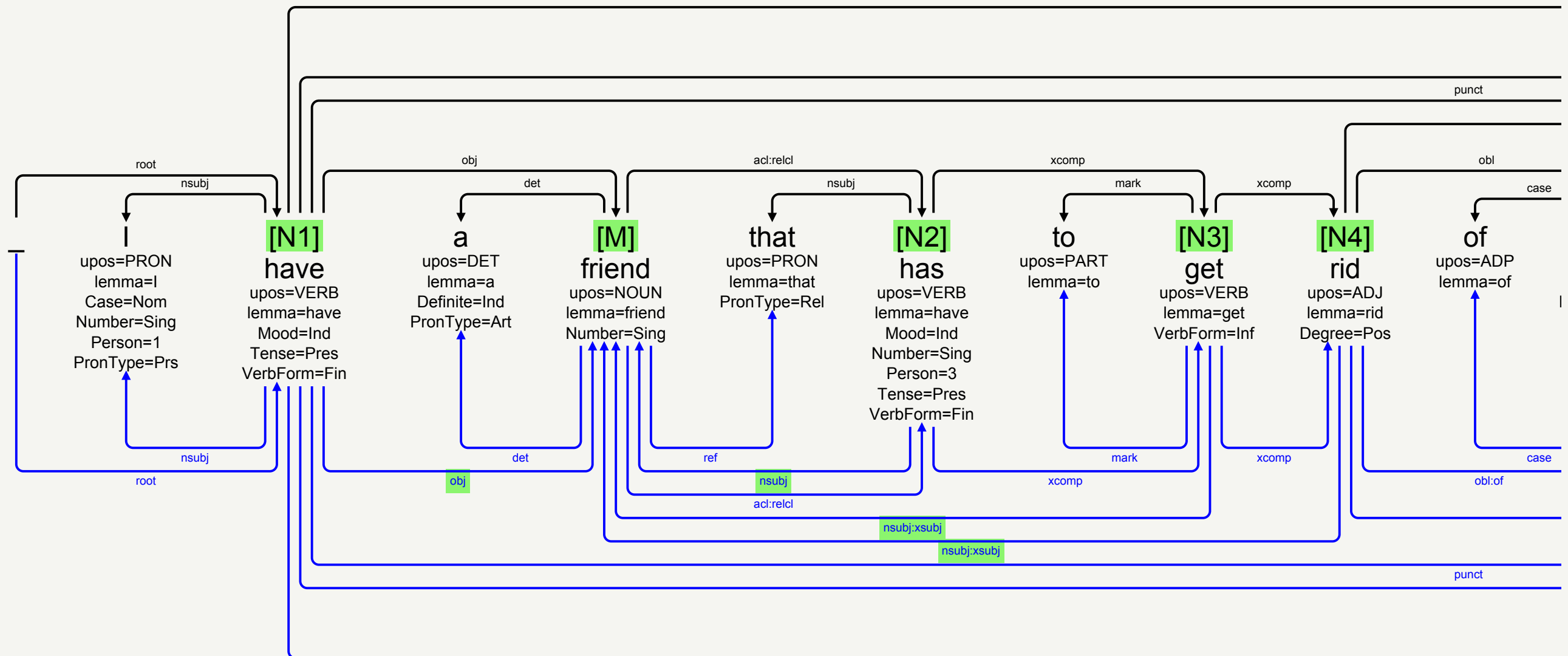


DMRS

AMR



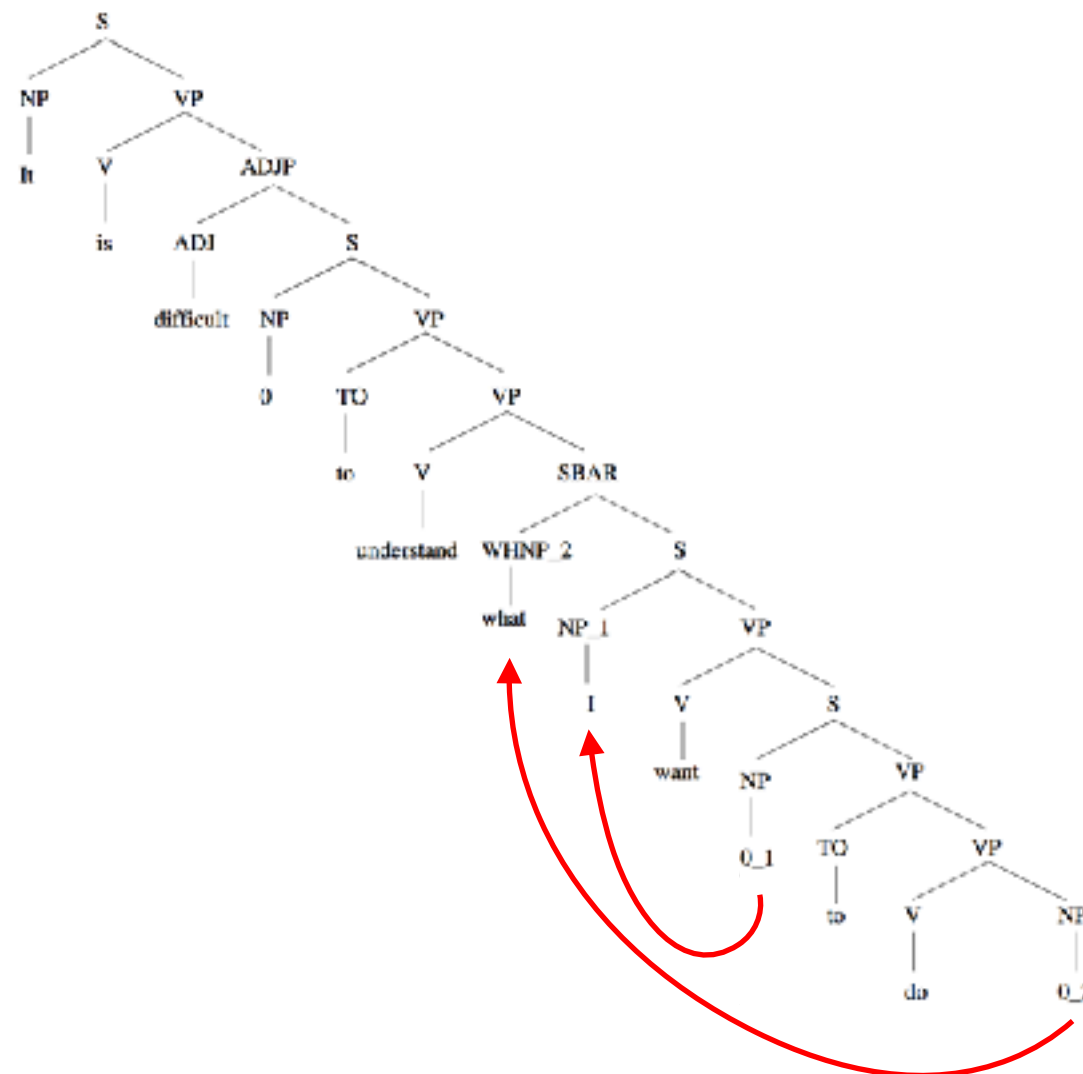
# Arbres ou graphes pour les représentations linguistiques ?



## Enhanced Universal Dependencies

# Arbres ou graphes

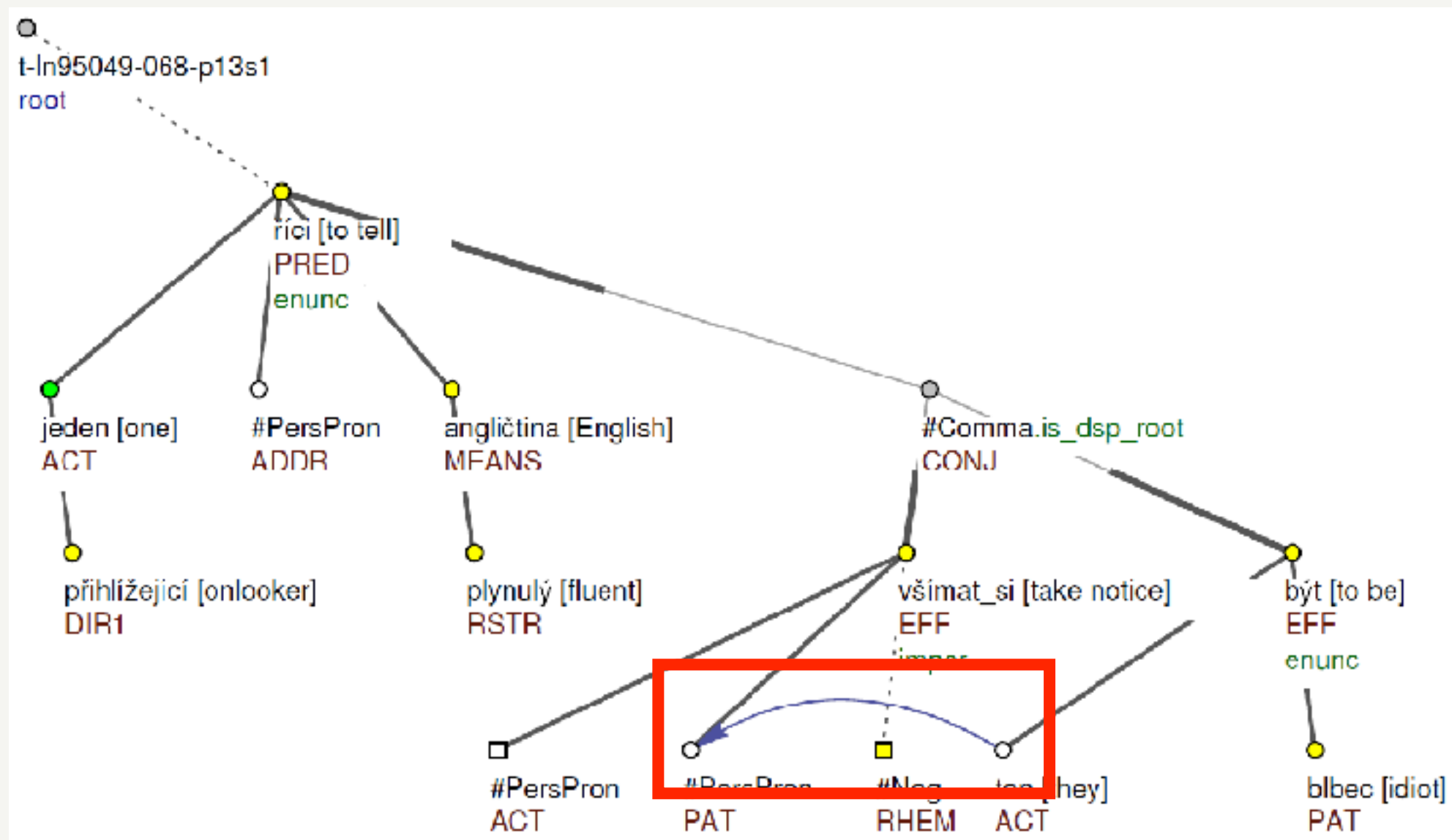
## *An Example Penn Treebank Tree*



Stephen Clark ([http://www.cs.ox.ac.uk/files/552/stat\\_parse2.pdf](http://www.cs.ox.ac.uk/files/552/stat_parse2.pdf))

# Arbres ou graphes

La première figure de la documentation du Prague Dependency Treebank 3.0



# Arbres ou graphes pour les représentations linguistiques ?

- Un **arbre** est un **graphe** !
- On considère toutes les structures comme des **graphes**

# Représentation des MWE & NE

1	Rémy	_	N	_	4	subj	_	2:NPP NE-PERS.final _
2	Halbwax	_	N	_	_	1	mod	_ 2
3	est	_	V	_	9	aux.pass	_	_ *
4	mis	_	V	_	0	root	_	1:_ MWE-VID _
5	en	_	P	_	4	p_obj.o	_	_ 1
6	examen	_	N	_	5	obj.p	_	_ 1
7	.	_	PONCT	_	4	ponct	_	_ *

Rémy Halbwax est mis en examen.

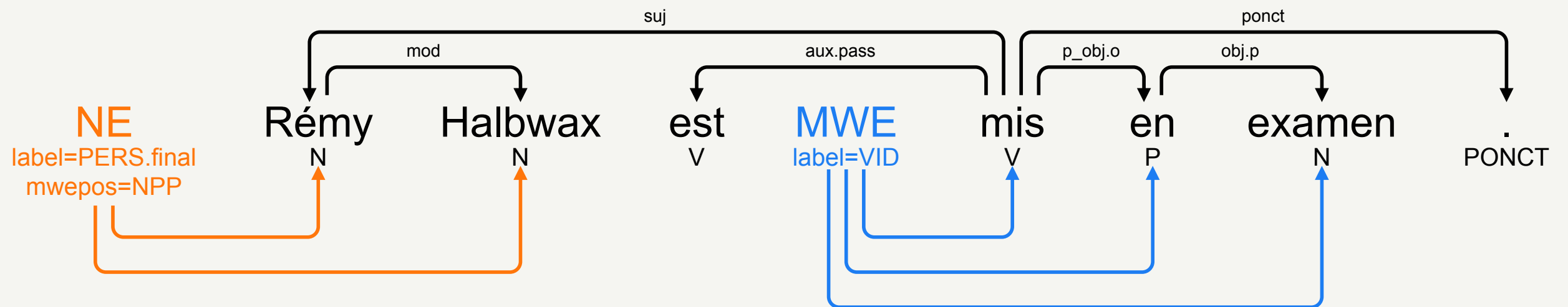
NE MWE



# Représentation des MWE & NE

Rémy Halbwax est mis en examen.

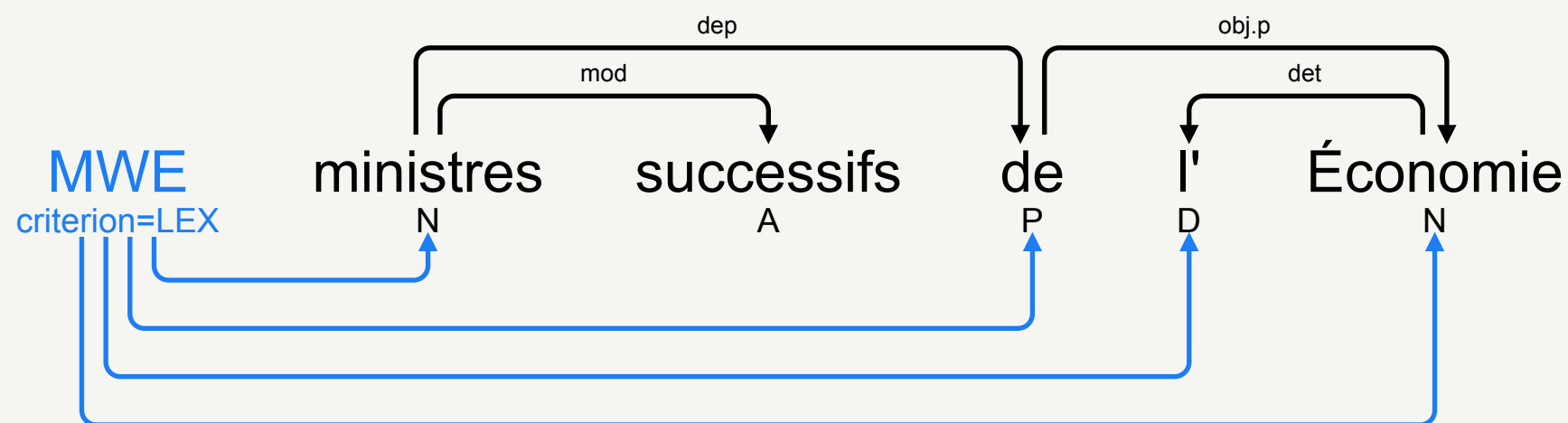
NE MWE



# Représentation des MWE & NE

ministres successifs de l'Économie.

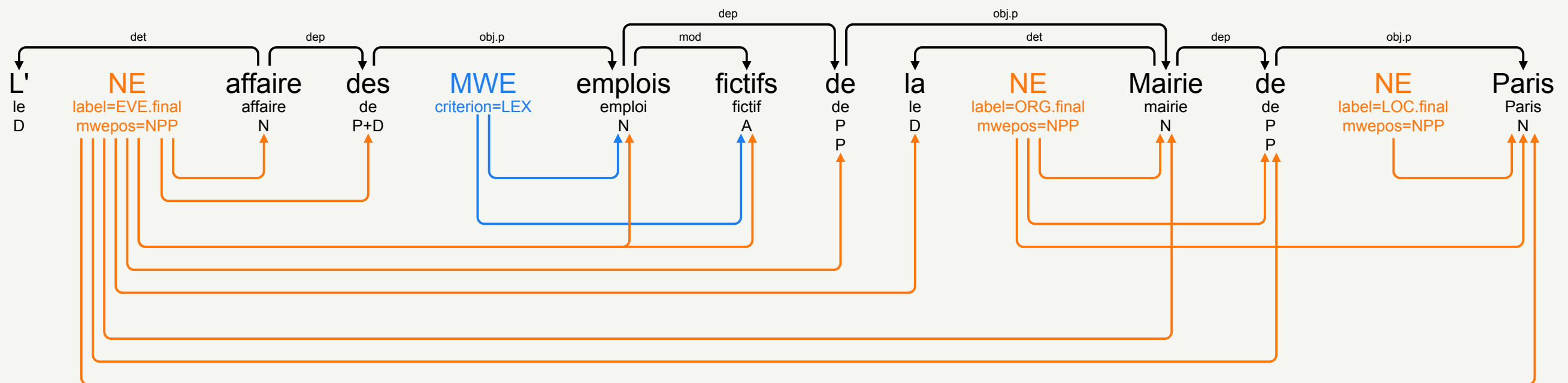
MWE



# Représentation des MWE & NE

L'affaire des emplois fictifs de la Mairie de Paris

MWE NE NE



# Calculer avec des graphes

On utilise le formalisme de la **réécriture de graphes** pour manipuler ces structures :

- Toute transformation peut-être décrite
- Une transformation globale est divisée en une suite d'étapes élémentaires
- On peut décrire des transformations non-déterministes

# La réécriture de graphes

- une transformation élémentaire est décrite par une **règle** :
  - un **motif**
  - des **conditions négatives**
  - une liste de **commandes**
- une transformation globale est décrite par une **stratégie** d'application des règles

On propose une implémentation de la réécriture adaptée au **TAL** :

- **structures de traits**
- règles paramétrées par un **lexique**

<http://grew.fr>

# La recherche de motifs

Un produit dérivé de la réécriture de graphes

La partie **recherche de motifs** peut être utilisée seule :

- exploration de corpus
- fouille d'erreur
- études linguistiques

Disponible en ligne : <http://match.grew.fr>

# Exemples d'exploration de corpus

Recherche d'une arête "MWE" :

```
pattern { MWE -[MWE]-> N }
```



Recherche d'un nœud N qui est dans 2 MWE :

```
pattern { MWE1 -[MWE]-> N; MWE2 -[MWE]-> N; MWE1 << MWE2 }
```



Recherche d'un nœud N qui est dans 3 MWE :

```
pattern {MWE1 -[MWE]-> N; MWE2 -[MWE]-> N; MWE3 -[MWE]-> N; MWE1 << MWE2; MWE2 << MWE3}
```



Recherche d'une MWE discontinue :

```
pattern {  
  MWE -[MWE]-> N1; MWE -[MWE]-> N2;  
  I []; N1 << I; I << N2;  
}  
without { MWE -[MWE]-> I }
```



# Exemples de fouille d'erreurs

Incohérences entre les **dep\_cpd** Sequoia et les **MWE** PARSEME

```
pattern { M -[dep_cpd]-> N }  
without { * -[MWE]-> M }
```



```
pattern {  
    MWE [criterion = IRREG];           % IRREG criterion  
    MWE -[MWE]-> N1; MWE -[MWE]-> N2; % At least 2 items  
    N1 < N2;                           % Fix the order (avoid duplicates)  
}  
without { MWE -[MWE]-> N0; N0 << N1 } % Ensure we have the first 2 items (avoid duplicates)  
without { N1 -[dep_cpd]-> * }         % The MWE is not annotated with dep_cpd
```

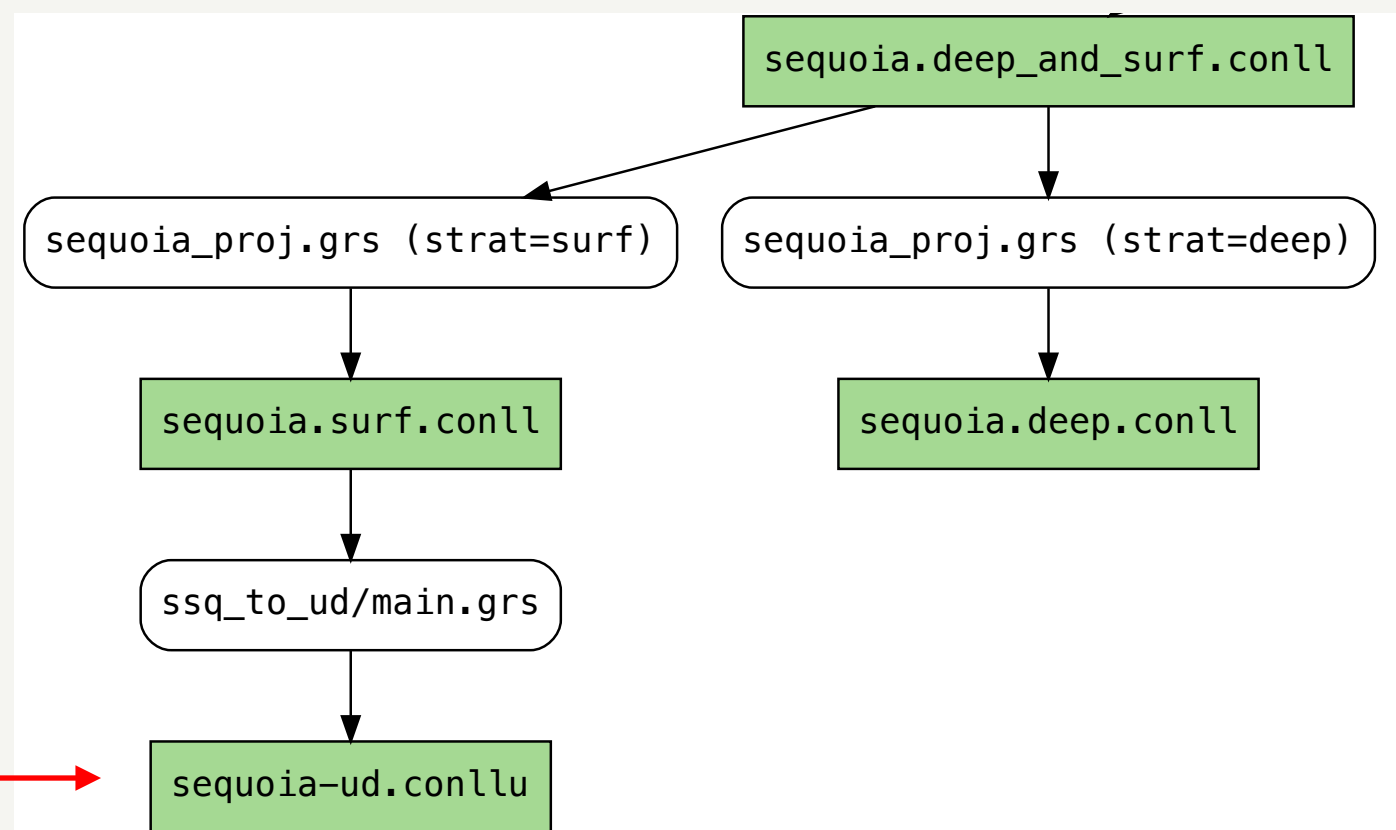




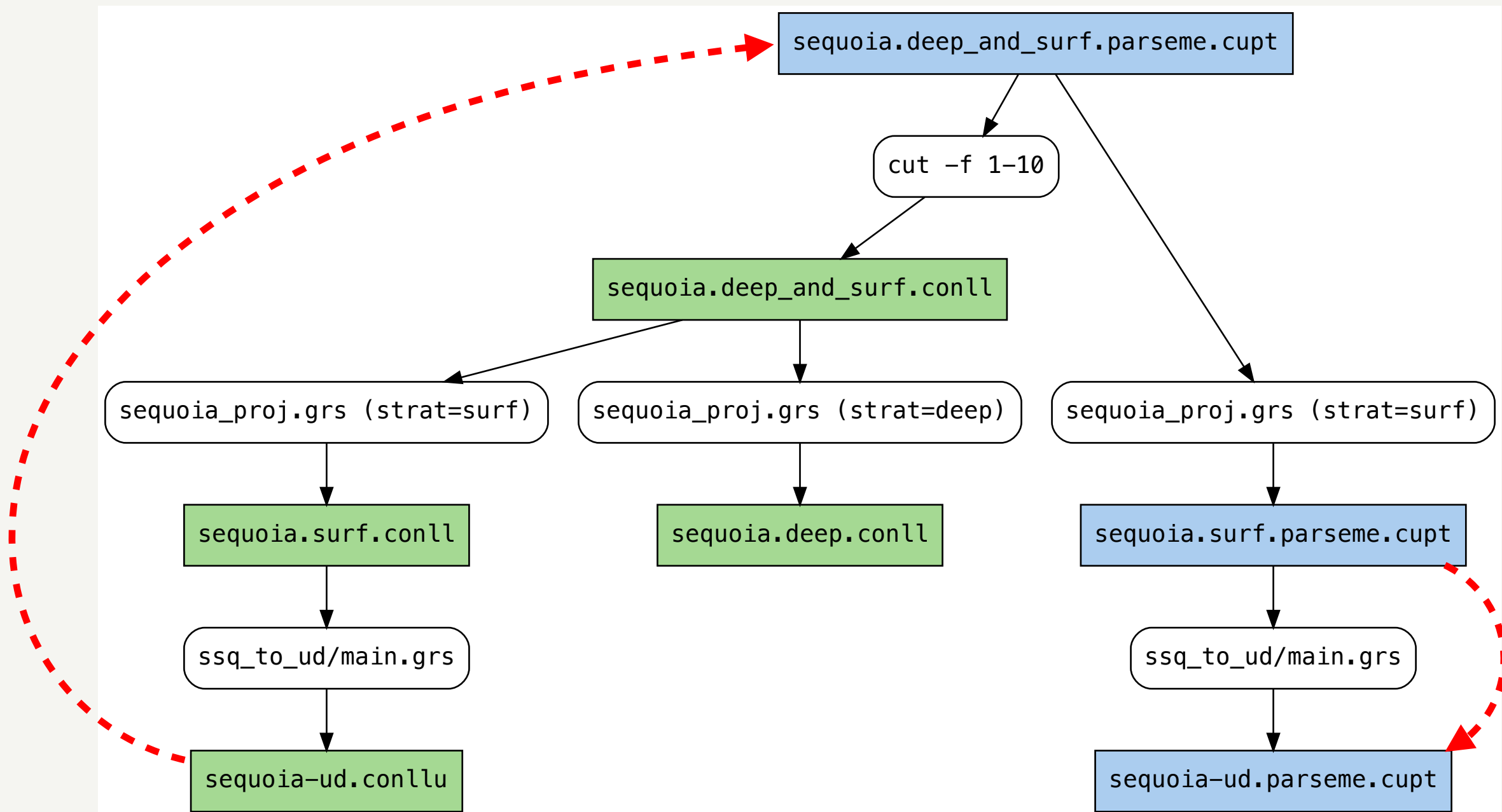
# Annotation PARSEME dans Sequoia

## Le corpus **Sequoia**

Annotations PARSEME

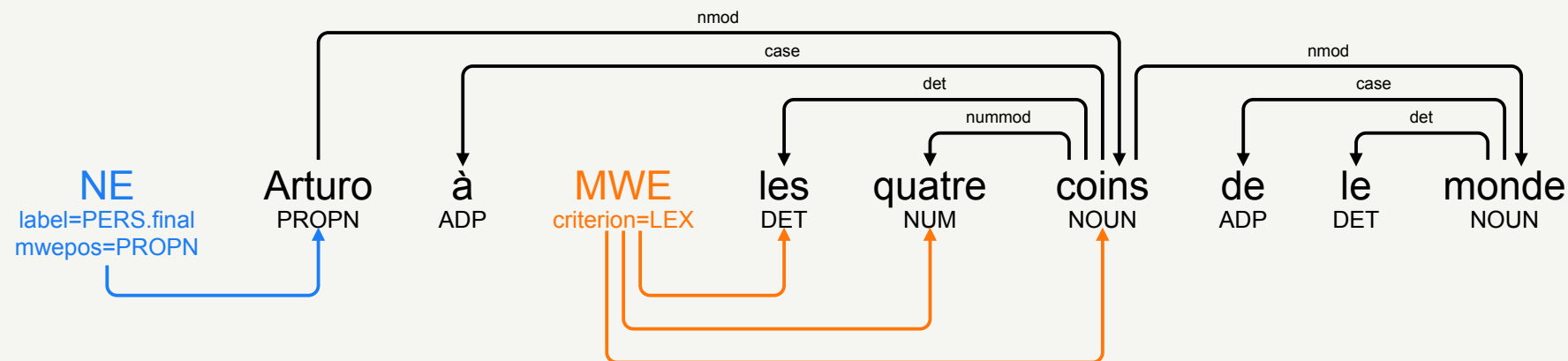


# Procédure de mise à jour de Sequoia

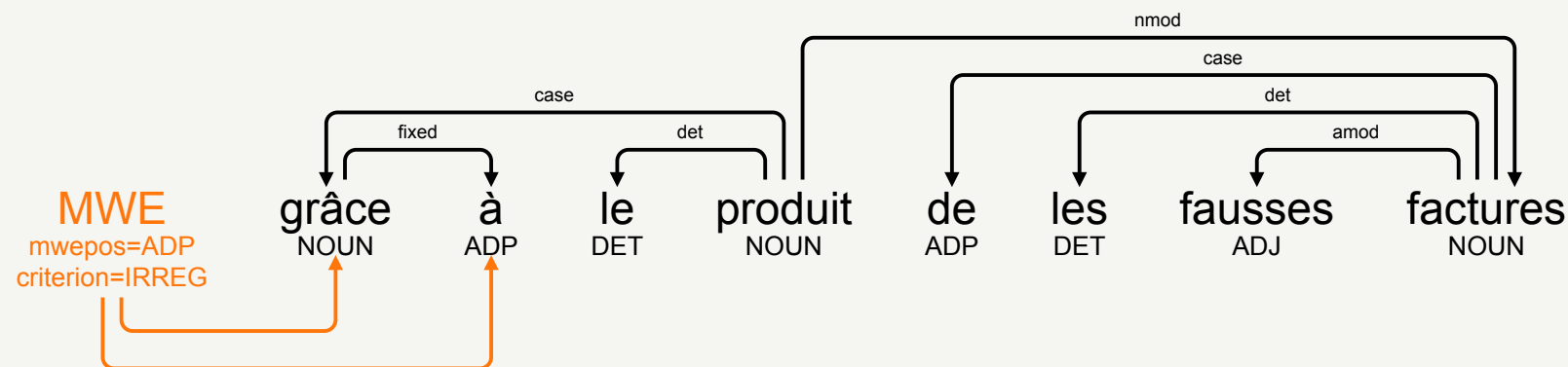


# Procédure de mise à jour de Sequoia

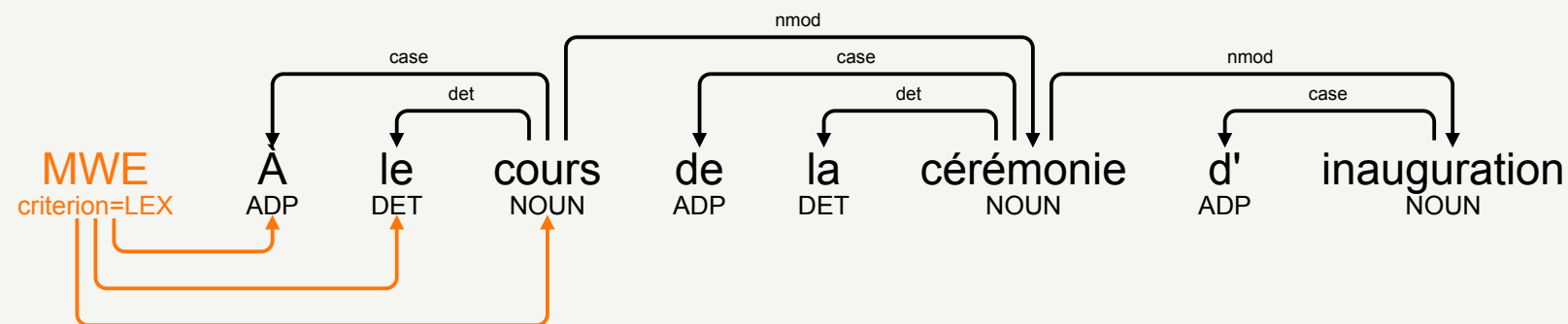
Problèmes pour la conversion Sequoia vers UD :  $au(x) \rightarrow \text{à } le(s)$



aux	2/2: _   MWE   LEX
quatre	2
coins	2

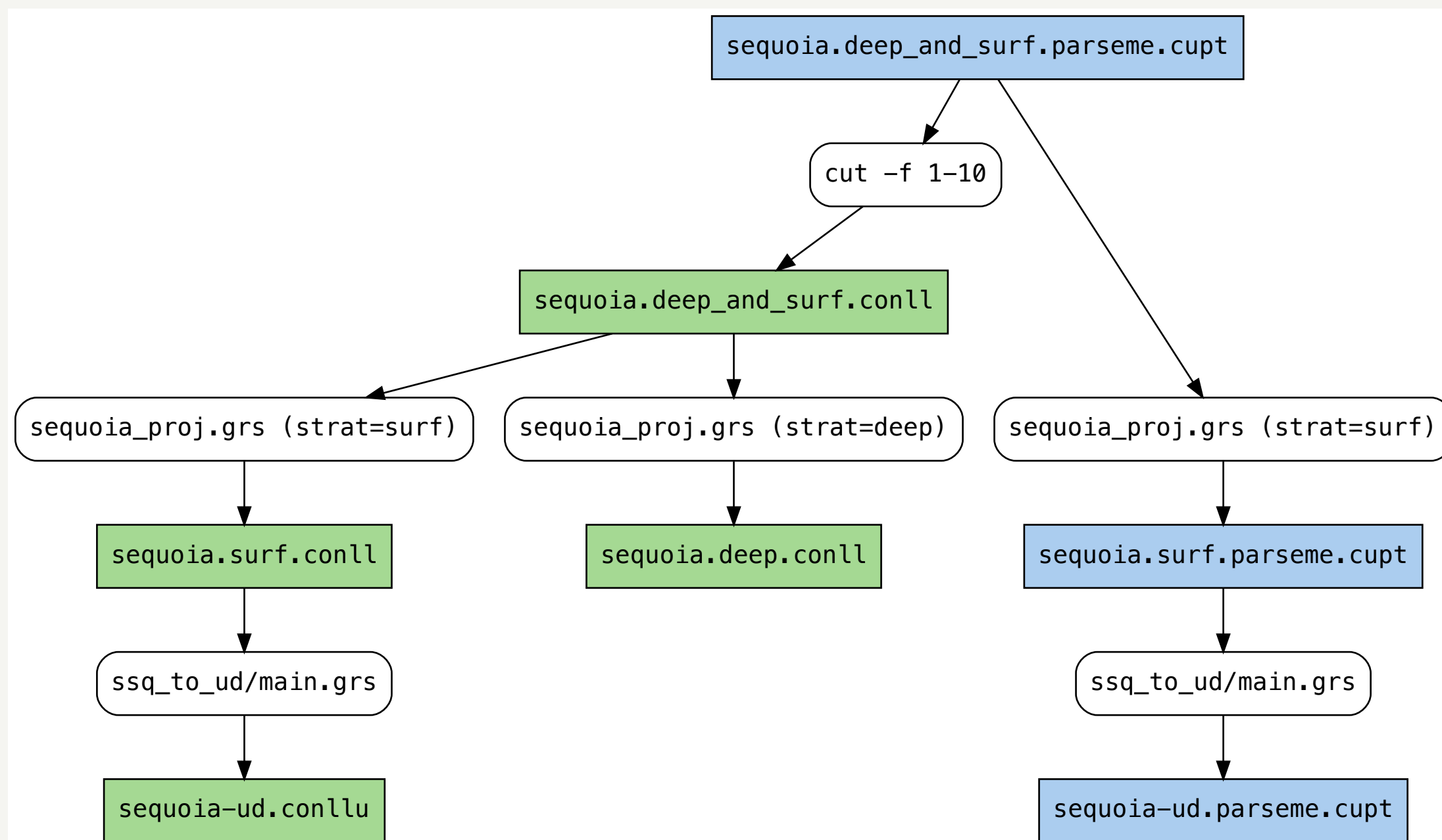


grâce	5: P   MWE   IRREG
au	5/1



Au	1: _   MWE   LEX
cours	1

# Sequoia aujourd'hui





# Conclusion

## Réécriture de graphes :

- Outil générique pour décrire des transformations de graphes
- Utilisé surtout sur la syntaxe en dépendance
- Applications à tous les corpus UD (140 corpus, 80 langues)
- Une bibliothèque Python disponible

# La suite

- Utilisation de la réécriture pour d'autres représentations linguistiques (sémantiques)
- Développer **Grew** :
  - Étendre la bibliothèque Python
  - Arborator + Grew
- Mise en cohérence avec les annotations de UD  

Plus d'infos sur <http://grew.fr>