

Proposition de recommandation pour l'annotation syntaxique des corpus du CCFM

Nicolas Mazziotta, Université de Liège
`nicolas.mazziotta@ulg.ac.be`

5-6 octobre 2006

Plan

1 Principes

Plan

- 1 Principes
- 2 Proposition de recommandation

Plan

- 1 Principes
- 2 Proposition de recommandation
- 3 Conclusion

- 1 Principes
 - Séparer l'analyse de l'édition
 - Intérêt
 - Mise en œuvre : le pointeur
 - Minimalisme des recommandations
 - Faut-il chercher un jeu d'étiquettes consensuel ?
 - Problèmes d'ergonomie
- 2 Proposition de recommandation
- 3 Conclusion

Séparer l'analyse de l'édition

Stand-off markup

Les créateurs de l'*American National Corpus* recommandent que les données issues de l'analyse soient détachées de celles fournies par les matériaux eux-mêmes.

Ce procédé est connu sous le nom de *stand-off markup*

Séparer l'analyse de l'édition

Intérêt

Relations entre édition et analyse

Séparer l'analyse de l'édition

Intérêt

Relations entre édition et analyse

- 1 Les analyses viennent **après** l'édition des textes ;

Séparer l'analyse de l'édition

Intérêt

Relations entre édition et analyse

- 1 Les analyses viennent **après** l'édition des textes ;
- 2 les analyses ne sont **pas complètement consensuelles** ;

Séparer l'analyse de l'édition

Intérêt

Relations entre édition et analyse

- 1 Les analyses viennent **après** l'édition des textes ;
- 2 les analyses ne sont **pas complètement consensuelles** ;
- 3 il y a une **infinité** d'analyses possibles.

Séparer l'analyse de l'édition

Intérêt

Les analyses viennent après l'édition

Séparer l'analyse de l'édition

Intérêt

Les analyses viennent après l'édition

- Édition et analyse sont des objets **distincts** ;

Séparer l'analyse de l'édition

Intérêt

Les analyses viennent après l'édition

- Édition et analyse sont des objets **distincts** ;
- être intéressé par une **édition** n'est pas être intéressé par l'**analyse** ;

Séparer l'analyse de l'édition

Intérêt

Les analyses viennent après l'édition

- Édition et analyse sont des objets **distincts** ;
- être intéressé par une **édition** n'est pas être intéressé par l'**analyse** ;
- l'**ordre de surface** n'est pas l'ordre de l'analyse.

Séparer l'analyse de l'édition

Intérêt

Les analyses ne sont pas consensuelles

Séparer l'analyse de l'édition

Intérêt

Les analyses ne sont pas consensuelles

- Tout le monde n'est pas d'accord ;

Séparer l'analyse de l'édition

Intérêt

Les analyses ne sont pas consensuelles

- Tout le monde n'est pas d'accord ;
- l'analyse peut être **fausse**.

Séparer l'analyse de l'édition

Intérêt

Il y a une infinité d'analyses possibles

Séparer l'analyse de l'édition

Intérêt

Il y a une infinité d'analyses possibles

- Encoder plusieurs analyses arborescentes sur un même texte le **surcharge** ;

Séparer l'analyse de l'édition

Intérêt

Il y a une infinité d'analyses possibles

- Encoder plusieurs analyses arborescentes sur un même texte le **surcharge** ;
- les **hiérarchies multiples** ne sont pas prévues par XML.

Séparer l'analyse de l'édition

Mise en œuvre : le pointeur

Pointeur

Le pointeur est un élément qui n'a pas d'autre valeur que celle de renvoyer à une autre unité

Séparer l'analyse de l'édition

Mise en œuvre : le pointeur

Pointeur

Le pointeur est un élément qui n'a pas d'autre valeur que celle de renvoyer à une autre unité

XML et les pointeurs

Séparer l'analyse de l'édition

Mise en œuvre : le pointeur

Pointeur

Le pointeur est un élément qui n'a pas d'autre valeur que celle de **renvoyer** à une **autre** unité

XML et les pointeurs

Chaque unité minimale de l'analyse porte un **identifiant** unique (@id), qui les **distingue** des autres unités.

Séparer l'analyse de l'édition

Mise en œuvre : le pointeur

Pointeur

Le pointeur est un élément qui n'a pas d'autre valeur que celle de renvoyer à une autre unité

XML et les pointeurs

Chaque unité minimale de l'analyse porte un identifiant unique (@id), qui les distingue des autres unités.

➔ Tout autre élément peut référer aux unités identifiées

· · **A** tos ceaz ki ces prefenf letres veront z oront [...] (AÉL, Liège, Saint-Martin 1277, n° 98)

.. **A** tos ceaz ki ces prefenf letres veront ꝛ oront [...] (AÉL, Liège, Saint-Martin 1277, n° 98)

Texte avec identifiants

```
<lb id='_1234' />  
<seg type='ponct' id='_381'>...</seg>  
<w id='_1'><c id='_1236'>A</c></w>  
<w id='_2'>tos</w>  
<w id='_3'>ceaz</w>  
<w id='_4'>ki</w>  
<w id='_5'>ces</w>  
<w id='_6'>pre&#x017F ;en&#x017F i</w>  
<w id='_7'><c id='_1243'>l</c>etres</w>  
<w id='_8'>veront</w>  
<w id='_9'><expan>et</expan></w>  
<w id='_10'>oront</w>
```

.. **A** tos ceaz ki ces prefenf letres veront z oront [...] (AÉL, Liège, Saint-Martin 1277, n° 98)

Texte avec identifiants

```
<lb id='_1234' />  
<seg type='ponct' id='_381'>...</seg>  
<w id='_1'><c id='_1236'>A</c></w>  
<w id='_2'>tos</w>  
<w id='_3'>ceaz</w>  
<w id='_4'>ki</w>  
<w id='_5'>ces</w>  
<w id='_6'>pre&#x017F ;en&#x017F ;</w>  
<w id='_7'><c id='_1243'>l</c>etres</w>  
<w id='_8'>veront</w>  
<w id='_9'><expan>et</expan></w>  
<w id='_10'>oront</w>
```

Pointeur

```
<strong>  
  <link target='_1' />  
</strong>
```

.. **A** tos ceaz ki ces prefenf letres veront z oront [...]
(AÉL, Liège, Saint-Martin 1277, n° 98)

Texte avec identifiants

```
<lb id='_1234' />  
<seg type='ponct' id='_381'>...</seg>  
<w id='_1'><c id='_1236'>A</c></w>  
<w id='_2'>tos</w>  
<w id='_3'>ceaz</w>  
<w id='_4'>ki</w>  
<w id='_5'>ces</w>  
<w id='_6'>re&#x017F;en&#x017F;i</w>  
<w id='_7'><c id='_1243'>l</c>etres</w>  
<w id='_8'>veront</w>  
<w id='_9'><expan>et</expan></w>  
<w id='_10'>oront</w>
```

Pointeur

```
<strong>  
  <link target='_1' />  
</strong>
```

Minimalisme des recommandations

Faut-il chercher un jeu d'étiquettes consensuel ?

Minimalisme des recommandations

Faut-il chercher un jeu d'étiquettes consensuel ?

Nécessaires ?

Cela permettrait de **partager** des analyses **comparables**.

Minimalisme des recommandations

Faut-il chercher un jeu d'étiquettes consensuel ?

Nécessaires ?

Cela permettrait de **partager** des analyses **comparables**.

Dangereux !

Ce jeu risquerait trois **défauts** inacceptables :

Minimalisme des recommandations

Faut-il chercher un jeu d'étiquettes consensuel ?

Nécessaires ?

Cela permettrait de **partager** des analyses **comparables**.

Dangereux !

Ce jeu risquerait trois **défauts** inacceptables :

- inertie ;

Minimalisme des recommandations

Faut-il chercher un jeu d'étiquettes consensuel ?

Nécessaires ?

Cela permettrait de **partager** des analyses **comparables**.

Dangereux !

Ce jeu risquerait trois **défauts** inacceptables :

- inertie ;
- éclectisme ;

Minimalisme des recommandations

Faut-il chercher un jeu d'étiquettes consensuel ?

Nécessaires ?

Cela permettrait de **partager** des analyses **comparables**.

Dangereux !

Ce jeu risquerait trois **défauts** inacceptables :

- inertie ;
- éclectisme ;
- hétérogénéité.

Minimalisme des recommandations

Faut-il chercher un jeu d'étiquettes consensuel ?

Nécessaires ?

Cela permettrait de **partager** des analyses **comparables**.

Dangereux !

Ce jeu risquerait trois **défauts** inacceptables :

- inertie ;
- éclectisme ;
- hétérogénéité.

En supposant que l'on puisse tomber d'accord . . .

Minimalisme des recommandations

Faut-il chercher un jeu d'étiquettes consensuel ?

Minimalisme des recommandations

Faut-il chercher un jeu d'étiquettes consensuel ?

Inertie

Ces étiquettes seraient exploitées dans des logiciels *ad hoc*

Minimalisme des recommandations

Faut-il chercher un jeu d'étiquettes consensuel ?

Inertie

Ces étiquettes seraient exploitées dans des logiciels *ad hoc*

↳ Faire évoluer le jeu rendrait **obsolètes** les logiciels.

Minimalisme des recommandations

Faut-il chercher un jeu d'étiquettes consensuel ?

Inertie

Ces étiquettes seraient exploitées dans des logiciels *ad hoc*

- ↳ Faire évoluer le jeu rendrait **obsolètes** les logiciels.
- ↳ Cela **imposerait** l'emploi de ces étiquettes.

Minimalisme des recommandations

Faut-il chercher un jeu d'étiquettes consensuel ?

Inertie

Ces étiquettes seraient exploitées dans des logiciels *ad hoc*

↳ Faire évoluer le jeu rendrait **obsolètes** les logiciels.

↳ Cela **imposerait** l'emploi de ces étiquettes.

Éclectisme

Minimalisme des recommandations

Faut-il chercher un jeu d'étiquettes consensuel ?

Inertie

Ces étiquettes seraient exploitées dans des logiciels *ad hoc*

↳ Faire évoluer le jeu rendrait **obsolètes** les logiciels.

↳ Cela **imposerait** l'emploi de ces étiquettes.

Éclectisme

Si le jeu d'étiquettes est choisi en consensus ou extensible. . .

Minimalisme des recommandations

Faut-il chercher un jeu d'étiquettes consensuel ?

Inertie

Ces étiquettes seraient exploitées dans des logiciels *ad hoc*

↳ Faire évoluer le jeu rendrait **obsolètes** les logiciels.

↳ Cela **imposerait** l'emploi de ces étiquettes.

Éclectisme

Si le jeu d'étiquettes est choisi en consensus ou extensible. . .

↳ il a de fortes chances d'être **éclectique**, donc **asystématique**.

Minimalisme des recommandations

Faut-il chercher un jeu d'étiquettes consensuel ?

Inertie

Ces étiquettes seraient exploitées dans des logiciels *ad hoc*

- ↳ Faire évoluer le jeu rendrait **obsolètes** les logiciels.
- ↳ Cela **imposerait** l'emploi de ces étiquettes.

Éclectisme

Si le jeu d'étiquettes est choisi en consensus ou extensible. . .

- ↳ il a de fortes chances d'être **éclectique**, donc **asystématique**.

Hétérogénéité

Minimalisme des recommandations

Faut-il chercher un jeu d'étiquettes consensuel ?

Inertie

Ces étiquettes seraient exploitées dans des logiciels *ad hoc*

- ↳ Faire évoluer le jeu rendrait **obsolètes** les logiciels.
- ↳ Cela **imposerait** l'emploi de ces étiquettes.

Éclectisme

Si le jeu d'étiquettes est choisi en consensus ou extensible. . .

- ↳ il a de fortes chances d'être **éclectique**, donc **asystématique**.

Hétérogénéité

Il s'en faudrait de beaucoup pour qu'il soit appliqué de la même manière par tous les participants au projet.

Minimalisme des recommandations

Faut-il chercher un jeu d'étiquettes consensuel ?

En conséquence

Minimalisme des recommandations

Faut-il chercher un jeu d'étiquettes consensuel ?

En conséquence

- ① Limitons les recommandations à des considérations d'ordres **méthodologique** et **technique** ;

Minimalisme des recommandations

Faut-il chercher un jeu d'étiquettes consensuel ?

En conséquence

- 1 Limitons les recommandations à des considérations d'ordres **méthodologique** et **technique** ;
- 2 **n'imposons pas** une manière spécifique d'analyser le corpus ;

Minimalisme des recommandations

Faut-il chercher un jeu d'étiquettes consensuel ?

En conséquence

- 1 Limitons les recommandations à des considérations d'ordres **méthodologique** et **technique** ;
- 2 **n'imposons pas** une manière spécifique d'analyser le corpus ;
- 3 il faut des **conventions** pour expliquer les types d'analyse employés.

Problèmes d'ergonomie

Lourdeur des logiciels

Les pointeurs sont difficiles à gérer en XML :

Problèmes d'ergonomie

Lourdeur des logiciels

Les pointeurs sont difficiles à gérer en XML :

- 1 Tout le document doit être chargé en mémoire ;

Problèmes d'ergonomie

Lourdeur des logiciels

Les pointeurs sont difficiles à gérer en XML :

- 1 Tout le document doit être **chargé en mémoire** ;
- 2 les programmes doivent **reconstituer** les données (calculer).

Problèmes d'ergonomie

Lourdeur des logiciels

Les pointeurs sont difficiles à gérer en XML :

- 1 Tout le document doit être **chargé en mémoire** ;
- 2 les programmes doivent **reconstituer** les données (calculer).

➔ Absence de logiciels

Les applications tardent à être disponibles.

Problèmes d'ergonomie

Lourdeur des logiciels

Les pointeurs sont difficiles à gérer en XML :

- 1 Tout le document doit être **chargé en mémoire** ;
- 2 les programmes doivent **reconstituer** les données (calculer).

↳ Absence de logiciels

Les applications tardent à être disponibles.

↳ Nous sommes contraints de **manipuler directement** les données XML.

Problèmes d'ergonomie

Lourdeur des logiciels

Les pointeurs sont difficiles à gérer en XML :

- 1 Tout le document doit être **chargé en mémoire** ;
- 2 les programmes doivent **reconstituer** les données (calculer).

↳ Absence de logiciels

Les applications tardent à être disponibles.

↳ Nous sommes contraints de **manipuler directement** les données XML.

L'abstraction liée aux pointeurs rend l'**encodage impossible** à réaliser :

`<link target='_1' />` est loin d'être
sémantiquement explicite.

- 1 Principes
- 2 Proposition de recommandation
 - Échantillon de travail
 - Format de travail
 - Définition
 - Implications
 - Format d'échange
 - Structure de la « racine » de l'analyse
 - Structure de l'élément anaChunk
 - Structure de l'élément anaNotes
 - Implications.
 - Déclaration du jeu d'étiquettes
- 3 Conclusion

Proposition de recommandation

Deux formats

Nous proposons de définir ici deux formats :

- 1 un pour le travail qui fait la part belle à l'**ergonomie** ;
- 2 et un pour l'**échange**, plus abstrait, générique.

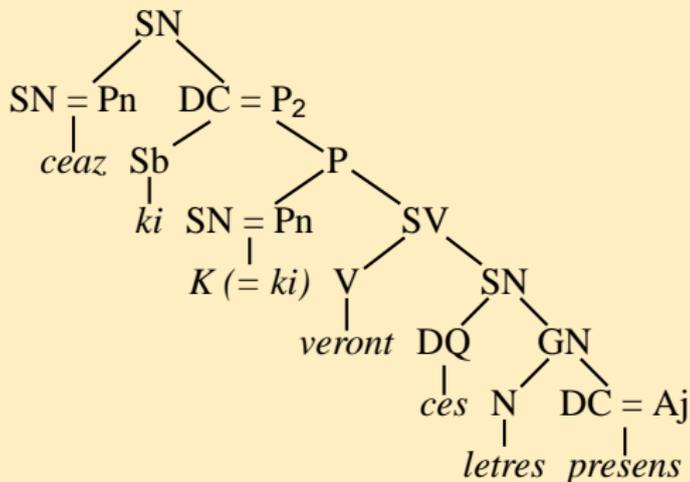
Proposition de recommandation

Échantillon de travail

Échantillon

```
<w id='_3'>ceaz</w>  
<w id='_4'>ki</w>  
<w id='_5'>ces</w>  
<w id='_6'>  
pre&#x17F ;en&#x17F ;  
</w>  
<w id='_7'>  
<c id='_1243'>l</c>etres  
</w>  
<w id='_8'>veront</w>
```

Analyse (ACI)



Format de travail

Définition

Échantillon

```
<div type='analyse syntaxique ACI'>
  <S id='_S1' f='SN'>
    <S id='_S2' f='Pn'><L target='_3' /></S>
    <S id='_S3' f='DC' n='P2'>
      <S id='_S4' f='Sb' n='K'><L target='_4' /></S>
      <S id='_S5' f='P'>
        <S id='_S6' f='SN' n='Pn'><L target='_4' /></S>
        <S id='_S7' f='SV'>
          <S id='_S8' f='V'><L target='_8' /></S>
          <S id='_S9' f='SN'>
            <S id='_S10' f='DQ'><L target='_5' /></S>
            <S id='_S11' f='GN'>
              <S id='_12' f='n'><L target='_7' /></S>
              <S id='_S12' f='DC' n='Aj'><L target='_6' /></S>
            </S></S></S></S></S></S></S>
  </div>
```

Format de travail

Définition

Échantillon

```
<div type='analyse syntaxique ACI'>
  <S id='_S1' f='SN'>
    <S id='_S2' f='Pn'><L target='_3' /></S>
    <S id='_S3' f='DC' n='P2'>
      <S id='_S4' f='Sb' n='K'><L target='_4' /></S>
      <S id='_S5' f='P'>
        <S id='_S6' f='SN' n='Pn'><L target='_4' /></S>
        <S id='_S7' f='SV'>
          <S id='_S8' f='V'><L target='_8' /></S>
          <S id='_S9' f='SN'>
            <S id='_S10' f='DQ'><L target='_5' /></S>
            <S id='_S11' f='GN'>
              <S id='_12' f='n'><L target='_7' /></S>
              <S id='_S12' f='DC' n='Aj'><L target='_6' /></S>
            </S></S></S></S></S></S></S>
  </div>
```

Un élément `div` englobe l'ensemble.

Format de travail

Définition

Échantillon

```
<div type='analyse syntaxique ACI'>
  <S id='_S1' f='SN'>
    <S id='_S2' f='Pn'><L target='_3' /></S>
    <S id='_S3' f='DC' n='P2'>
      <S id='_S4' f='Sb' n='K'><L target='_4' /></S>
      <S id='_S5' f='P'>
        <S id='_S6' f='SN' n='Pn'><L target='_4' /></S>
        <S id='_S7' f='SV'>
          <S id='_S8' f='V'><L target='_8' /></S>
          <S id='_S9' f='SN'>
            <S id='_S10' f='DQ'><L target='_5' /></S>
            <S id='_S11' f='GN'>
              <S id='_12' f='n'><L target='_7' /></S>
              <S id='_S12' f='DC' n='Aj'><L target='_6' /></S>
            </S></S></S></S></S></S></S>
  </div>
```

Cet élément inclut une **hiérarchie** de structures des éléments S.

Format de travail

Définition

Échantillon

```
<div type='analyse syntaxique ACI'>
  <S id='_S1' f='SN'>
    <S id='_S2' f='Pn'><L target='_3' /></S>
    <S id='_S3' f='DC' n='P2'>
      <S id='_S4' f='Sb' n='K'><L target='_4' /></S>
      <S id='_S5' f='P'>
        <S id='_S6' f='SN' n='Pn'><L target='_4' /></S>
        <S id='_S7' f='SV'>
          <S id='_S8' f='V'><L target='_8' /></S>
          <S id='_S9' f='SN'>
            <S id='_S10' f='DQ'><L target='_5' /></S>
            <S id='_S11' f='GN'>
              <S id='_12' f='n'><L target='_7' /></S>
              <S id='_S12' f='DC' n='Aj'><L target='_6' /></S>
            </S></S></S></S></S></S></S>
  </div>
```

Les **nœuds terminaux** de la structure sont des éléments L sont des pointeurs vers les mots du texte.

Format de travail

Définition

Échantillon

```
<div type='analyse syntaxique ACI'>
  <S id='_S1' f='SN'>
    <S id='_S2' f='Pn'><L target='_3' /></S>
    <S id='_S3' f='DC' n='P2'>
      <S id='_S4' f='Sb' n='K'><L target='_4' /></S>
      <S id='_S5' f='P'>
        <S id='_S6' f='SN' n='Pn'><L target='_4' /></S>
        <S id='_S7' f='SV'>
          <S id='_S8' f='V'><L target='_8' /></S>
          <S id='_S9' f='SN'>
            <S id='_S10' f='DQ'><L target='_5' /></S>
            <S id='_S11' f='GN'>
              <S id='_12' f='n'><L target='_7' /></S>
              <S id='_S12' f='DC' n='Aj'><L target='_6' /></S>
            </S></S></S></S></S></S></S>
  </div>
```

Les éléments S reçoivent un **nombre indéterminé** d'attributs, de nom et de valeur laissés à l'appréciation de l'utilisateur.

Format de travail

Définition

Échantillon

```
<div type='analyse syntaxique ACI'>
  <S id='_S1' f='SN'>
    <S id='_S2' f='Pn'><L target='_3' forme='ceaz' /></S>
    <S id='_S3' f='DC' n='P2'>
      <S id='_S4' f='Sb' n='K'><L target='_4' forme='ki' /></S>
      <S id='_S5' f='P'>
        <S id='_S6' f='SN' n='Pn'><L target='_4' forme='ki' /></S>
        <S id='_S7' f='SV'>
          <S id='_S8' f='V'><L target='_8' forme='veront' /></S>
          <S id='_S9' f='SN'>
            <S id='_S10' f='DQ'><L target='_5' forme='ces' /></S>
            <S id='_S11' f='GN'>
              <S id='_12' f='n'><L target='_7' forme='letres' /></S>
              <S id='_S12' f='DC' n='Aj'><L target='_6' forme='presens' /></S>
            </S></S></S></S></S></S></S>
  </div>
```

Les **formes** peuvent être reportées en attribut (ergonomie).

Format de travail

Implications

Format de travail

Implications

Avantages

Format de travail

Implications

Avantages

- 1 Accès facile aux attributs ;

Format de travail

Implications

Avantages

- 1 Accès facile aux attributs ;
- 2 minimisation de la mémoire nécessaire au chargement ;

Format de travail

Implications

Avantages

- 1 Accès facile aux attributs ;
- 2 minimisation de la mémoire nécessaire au chargement ;
- 3 accès possible pour l'humain.

Format de travail

Implications

Avantages

- 1 Accès facile aux attributs ;
- 2 minimisation de la mémoire nécessaire au chargement ;
- 3 accès possible pour l'humain.

Inconvénients

Format de travail

Implications

Avantages

- 1 Accès facile aux attributs ;
- 2 minimisation de la mémoire nécessaire au chargement ;
- 3 accès possible pour l'humain.

Inconvénients

- 1 Validation difficile ;

Format de travail

Implications

Avantages

- 1 Accès facile aux attributs ;
- 2 minimisation de la mémoire nécessaire au chargement ;
- 3 accès possible pour l'humain.

Inconvénients

- 1 Validation difficile ;
- 2 faible structuration de l'information contenue dans l'attribut (peu important).

Format d'échange

Structure de la « racine » de l'analyse

Format d'échange

Structure de la « racine » de l'analyse

XML

```
<anaGrp type='analyse syntaxique ACI'>  
  <anaChunk>...</anaChunk>  
  <anaNotes>...</anaNotes>  
</anaGrp>
```

Format d'échange

Structure de la « racine » de l'analyse

XML

```
<anaGrp type='analyse syntaxique ACI'>  
  <anaChunk>...</anaChunk>  
  <anaNotes>...</anaNotes>  
</anaGrp>
```

L'élément qui contient l'analyse arborescente est nommé `anaGrp`. Cette « racine » comporte un attribut `@type`, qui précise à quel type d'analyse on a affaire.

Format d'échange

Structure de la « racine » de l'analyse

XML

```
<anaGrp type='analyse syntaxique ACI'>  
  <anaChunk>...</anaChunk>  
  <anaNotes>...</anaNotes>  
</anaGrp>
```

anaChunk contient la **représentation arborescente** des constituants.

Format d'échange

Structure de la « racine » de l'analyse

XML

```
<anaGrp type='analyse syntaxique ACI'>  
  <anaChunk>...</anaChunk>  
  <anaNotes>...</anaNotes>  
</anaGrp>
```

anaNotes contient l'**annotation** des constituants.

Format d'échange

Structure de l'élément anaChunk

Encodage de l'analyse

```
<anaChunk>
  <seg id='_S1'>
    <seg id='_S2'><link target='_3' /></seg>
    <seg id='_S3'>
      <seg id='_S4'><link target='_4' /></seg>
      <seg id='_S5'>
        <seg id='_S6'><link target='_4' /></seg>
        <seg id='_S7'>
          <seg id='_S8'><link target='_8' /></seg>
          <seg id='_S9'>
            <seg id='_S10'><link target='_5' /></seg>
            <seg id='_S11'>
              <seg id='_S12'><link target='_7' /></seg>
              <seg id='_S13'><link target='_6' /></seg>
            </seg></seg></seg></seg></seg></seg>
          </seg></seg></seg></seg></seg></seg>
        </seg></seg></seg></seg></seg></seg>
      </seg></seg></seg></seg></seg></seg>
    </seg></seg></seg></seg></seg></seg>
  </anaChunk>
```

Format d'échange

Structure de l'élément anaChunk

Encodage de l'analyse

```
<anaChunk>
  <seg id='_S1'>
    <seg id='_S2'><link target='_3' /></seg>
    <seg id='_S3'>
      <seg id='_S4'><link target='_4' /></seg>
      <seg id='_S5'>
        <seg id='_S6'><link target='_4' /></seg>
        <seg id='_S7'>
          <seg id='_S8'><link target='_8' /></seg>
          <seg id='_S9'>
            <seg id='_S10'><link target='_5' /></seg>
            <seg id='_S11'>
              <seg id='_S12'><link target='_7' /></seg>
              <seg id='_S13'><link target='_6' /></seg>
            </seg></seg></seg></seg></seg></seg></seg>
          </seg></seg></seg></seg></seg></seg></seg></seg></seg>
        </seg></seg></seg></seg></seg></seg></seg></seg></seg></seg>
      </seg></seg></seg></seg></seg></seg></seg></seg></seg></seg>
    </seg></seg></seg></seg></seg></seg></seg></seg></seg></seg>
  </anaChunk>
```

anaChunk comprend un **élément racine** seg.

Format d'échange

Structure de l'élément anaChunk

Encodage de l'analyse

```
<anaChunk>
  <seg id='_S1'>
    <seg id='_S2'><link target='_3' /></seg>
    <seg id='_S3'>
      <seg id='_S4'><link target='_4' /></seg>
      <seg id='_S5'>
        <seg id='_S6'><link target='_4' /></seg>
        <seg id='_S7'>
          <seg id='_S8'><link target='_8' /></seg>
          <seg id='_S9'>
            <seg id='_S10'><link target='_5' /></seg>
            <seg id='_S11'>
              <seg id='_S12'><link target='_7' /></seg>
              <seg id='_S13'><link target='_6' /></seg>
            </seg></seg></seg></seg></seg></seg>
          </seg></seg></seg></seg></seg></seg>
        </seg></seg></seg></seg></seg></seg>
      </seg></seg></seg></seg></seg></seg>
    </seg></seg></seg></seg></seg></seg>
  </anaChunk>
```

Tous les autres **nœuds** de l'arbre syntaxique sont des seg.

Format d'échange

Structure de l'élément anaChunk

Encodage de l'analyse

```
<anaChunk>
  <seg id='_S1'>
    <seg id='_S2'><link target='_3' /></seg>
    <seg id='_S3'>
      <seg id='_S4'><link target='_4' /></seg>
      <seg id='_S5'>
        <seg id='_S6'><link target='_4' /></seg>
        <seg id='_S7'>
          <seg id='_S8'><link target='_8' /></seg>
          <seg id='_S9'>
            <seg id='_S10'><link target='_5' /></seg>
            <seg id='_S11'>
              <seg id='_S12'><link target='_7' /></seg>
              <seg id='_S13'><link target='_6' /></seg>
            </seg></seg></seg></seg></seg></seg>
          </seg></seg></seg></seg></seg></seg>
        </seg></seg></seg></seg></seg></seg>
      </seg></seg></seg></seg></seg></seg>
    </seg></seg></seg></seg></seg></seg>
  </anaChunk>
```

Les **éléments terminaux** sont rendus par des éléments `link` (pointeurs).

Format d'échange

Structure de l'élément anaChunk

Encodage de l'analyse

```
<anaChunk>
  <seg id='_S1'>
    <seg id='_S2'><link target='_3' /></seg>
    <seg id='_S3'>
      <seg id='_S4'><link target='_4' /></seg>
      <seg id='_S5'>
        <seg id='_S6'><link target='_4' /></seg>
        <seg id='_S7'>
          <seg id='_S8'><link target='_8' /></seg>
          <seg id='_S9'>
            <seg id='_S10'><link target='_5' /></seg>
            <seg id='_S11'>
              <seg id='_S12'><link target='_7' /></seg>
              <seg id='_S13'><link target='_6' /></seg>
            </seg></seg></seg></seg></seg></seg>
          </seg></seg></seg></seg></seg></seg>
        </seg></seg></seg></seg></seg></seg>
      </seg></seg></seg></seg></seg></seg>
    </seg></seg></seg></seg></seg></seg>
  </anaChunk>
```

Le **contenu** de l'analyse est absent de cet arbre.

Format d'échange

Structure de l'élément anaNotes

Définition

Pour comprendre, repartons de la structure de travail :

Format d'échange

Structure de l'élément anaNotes

Définition

Pour comprendre, repartons de la structure de travail :

```
<S id='_S12' f='DC' n='Aj'><L target='_6' /></S>
```

Format d'échange

Structure de l'élément anaNotes

Définition

Pour comprendre, repartons de la structure de travail :

```
<S id='_S12' f='DC' n='Aj'><L target='_6' /></S>
```

Deux analyses sont appliquées au constituant identifié par l'élément S : @f = 'DC' et @n = 'Aj'.

Format d'échange

Structure de l'élément anaNotes

Définition

Pour comprendre, repartons de la structure de travail :

```
<S id='_S12' f='DC' n='Aj'><L target='_6' /></S>
```

Deux analyses sont appliquées au constituant identifié par l'élément S : @f = 'DC' et @n = 'Aj'.

Format d'échange

Structure de l'élément anaNotes

Définition

Pour comprendre, repartons de la structure de travail :

```
<S id='_S12' f='DC' n='Aj'><L target='_6' /></S>
```

Deux analyses sont appliquées au constituant identifié par l'élément S : @f = 'DC' et @n = 'Aj'.

Format d'échange

Structure de l'élément anaNotes

Définition

Pour comprendre, repartons de la structure de travail :

```
<S id='_S12' f='DC' n='Aj'><L target='_6' /></S>
```

Pour optimiser le format d'échange,

Format d'échange

Structure de l'élément anaNotes

Définition

Pour comprendre, repartons de la structure de travail :

```
<S id='_S12' f='DC' n='Aj'><L target='_6' /></S>
```

Pour optimiser le format d'échange,

- 1 détachons la caractérisation des segments de leur hiérarchisation ;

Format d'échange

Structure de l'élément anaNotes

Définition

Pour comprendre, repartons de la structure de travail :

```
<S id='_S12' f='DC' n='Aj'><L target='_6' /></S>
```

Pour optimiser le format d'échange,

- 1 détachons la caractérisation des segments de leur hiérarchisation ;
- 2 classons les analyses par type, puis par valeur.

Format d'échange

Structure de l'élément anaNotes

Hiérarchie types/valeurs

```
<anaNotes>  
  <anaTag type='f'>  
    <anaValue type='DC'>...</anaValue>  
  </anaTag>  
  <anaTag type='n'>  
    <anaValue type='Aj'>...</anaValue>  
    <anaValue type='N'>...</anaValue>  
    <anaValue type='Pn'>...</anaValue>  
  </anaTag>  
</anaNotes>
```

Format d'échange

Structure de l'élément anaNotes

Hiérarchie types/valeurs

```
<anaNotes>
  <anaTag type='f'>
    <anaValue type='DC'>...</anaValue>
  </anaTag>
  <anaTag type='n'>
    <anaValue type='Aj'>...</anaValue>
    <anaValue type='N'>...</anaValue>
    <anaValue type='Pn'>...</anaValue>
  </anaTag>
</anaNotes>
```

Classement pas **type**, puis par valeur.

Format d'échange

Structure de l'élément anaNotes

Hiérarchie types/valeurs

```
<anaNotes>
  <anaTag type='f'>
    <anaValue type='DC'>...</anaValue>
  </anaTag>
  <anaTag type='n'>
    <anaValue type='Aj'>...</anaValue>
    <anaValue type='N'>...</anaValue>
    <anaValue type='Pn'>...</anaValue>
  </anaTag>
</anaNotes>
```

Classement pas type, puis par **valeur**.

Format d'échange

Structure de l'élément anaNotes

Hiérarchie types/valeurs

```
<anaNotes>  
  <anaTag type='f'>  
    <anaValue type='DC'>...</anaValue>  
  </anaTag>  
  <anaTag type='n'>  
    <anaValue type='Aj'>...</anaValue>  
    <anaValue type='N'>...</anaValue>  
    <anaValue type='Pn'>...</anaValue>  
  </anaTag>  
</anaNotes>
```

Ces éléments peuvent contenir des pointeurs.

Format d'échange

Structure de l'élément anaNotes

Encodage hiérarchique

```
<anaChunk>
  <seg id='_S1'>
    <seg id='_S2'><link target='_3' /></seg>
    <seg id='_S3'>
      <seg id='_S4'><link target='_4' /></seg>
      <seg id='_S5'>
        <seg id='_S6'><link target='_4' /></seg>
        <seg id='_S7'>
          <seg id='_S8'><link target='_8' /></seg>
          <seg id='_S9'>
            <seg id='_S10'><link target='_5' /></seg>
            <seg id='_S11'>
              <seg id='_S12'><link target='_7' /></seg>
              <seg id='_S13'><link target='_6' /></seg>
            </seg></seg></seg></seg></seg></seg></seg>
          </seg></seg></seg></seg></seg></seg>
        </seg>
      </seg>
    </seg>
  </anaChunk>
```

Format d'échange

Structure de l'élément anaNotes

Encodage hiérarchique

```
<anaChunk>
  <seg id='_S1'>
    <seg id='_S2'><link target='_3' /></seg>
    <seg id='_S3'>
      <seg id='_S4'><link target='_4' /></seg>
      <seg id='_S5'>
        <seg id='_S6'><link target='_4' /></seg>
        <seg id='_S7'>
          <seg id='_S8'><link target='_8' /></seg>
          <seg id='_S9'>
            <seg id='_S10'><link target='_5' /></seg>
            <seg id='_S11'>
              <seg id='_S12'><link target='_7' /></seg>
              <seg id='_S13'><link target='_6' /></seg>
            </seg></seg></seg></seg></seg></seg>
          </seg></seg></seg></seg></seg></seg>
        </seg>
      </seg>
    </seg>
  </anaChunk>
```

Annotations

```
<anaNotes>
  <anaTag type='f'>
    <anaValue type='DC'>
      <linkGrp>
        <link target='_S3' />
        <link target='_S13' />
      </linkGrp>
    </anaValue>
    <anaValue type='DQ'>
      <linkGrp>
        <link target='_S10' />
      </linkGrp>
    </anaValue>
    ...
  </anaNotes>
```

Format d'échange

Structure de l'élément anaNotes

Encodage hiérarchique

```
<anaChunk>
  <seg id='_S1'>
    <seg id='_S2'><link target='_3' /></seg>
    <seg id='_S3'>
      <seg id='_4'><link target='_4' /></seg>
      <seg id='_5'>
        <seg id='_6'><link target='_4' /></seg>
        <seg id='_7'>
          <seg id='_8'><link target='_8' /></seg>
          <seg id='_S9'>
            <seg id='_S10'><link target='_5' /></seg>
            <seg id='_S11'>
              <seg id='_S12'><link target='_7' /></seg>
              <seg id='_S13'><link target='_6' /></seg>
            </seg></seg></seg></seg></seg></seg></seg>
  </anaChunk>
```

Annotations

```
<anaNotes>
  <anaTag type='f'>
    <anaValue type='DC'>
      <linkGrp>
        <link target='_S3' />
        <link target='_S13' />
      </linkGrp>
    </anaValue>
    <anaValue type='DQ'>
      <linkGrp>
        <link target='_S10' />
      </linkGrp>
    </anaValue>
    ...
  </anaNotes>
```

Format d'échange

Implications.

Format d'échange

Implications.

Inconvénients

Format d'échange

Implications.

Inconvénients

- ❶ Très lourd à implémenter ;

Format d'échange

Implications.

Inconvénients

- 1 Très lourd à implémenter ;
- 2 très lourd à manipuler.

Format d'échange

Implications.

Inconvénients

- 1 Très lourd à implémenter ;
- 2 très lourd à manipuler.

Avantages

Format d'échange

Implications.

Inconvénients

- 1 Très lourd à implémenter ;
- 2 très lourd à manipuler.

Avantages

- 1 **Validation** possible ;

Format d'échange

Implications.

Inconvénients

- 1 Très lourd à implémenter ;
- 2 très lourd à manipuler.

Avantages

- 1 **Validation** possible ;
- 2 **séparation** de la segmentation et de la caractérisation ;

Format d'échange

Implications.

Inconvénients

- 1 Très lourd à implémenter ;
- 2 très lourd à manipuler.

Avantages

- 1 **Validation** possible ;
- 2 **séparation** de la segmentation et de la caractérisation ;
- 3 **archivage** simple (fichiers séparés, etc.).

Déclaration du jeu d'étiquettes

Définition

Chaque annotation peut être déclarée (principe Tiger-XML) :

Déclaration du jeu d'étiquettes

Définition

Chaque annotation peut être déclarée (principe Tiger-XML) :

```
<anaDecl type='analyse syntaxique ACI'>  
  <anaTagDecl type='f'>  
    <anaValueDecl type='SN'>syntagme nominal</anaValueDecl>  
    <anaValueDecl type='SV'>syntagme verbal</anaValueDecl>  
    ...  
  </anaTagDecl>  
  <anaTagDecl type='n'>  
    <anaValueDecl type='Aj'>adjectif</anaValueDecl>  
    ...  
  </anaTagDecl>  
</anaDecl>
```

Déclaration du jeu d'étiquettes

Définition

Chaque annotation peut être déclarée (principe Tiger-XML) :

```
<anaDecl type='analyse syntaxique ACI'>
  <anaTagDecl type='f'>
    <anaValueDecl type='SN'>syntagme nominal</anaValueDecl>
    <anaValueDecl type='SV'>syntagme verbal</anaValueDecl>
    ...
  </anaTagDecl>
  <anaTagDecl type='n'>
    <anaValueDecl type='Aj'>adjectif</anaValueDecl>
    ...
  </anaTagDecl>
</anaDecl>
```

L'élément `anaDecl` englobe l'ensemble de la déclaration.

Déclaration du jeu d'étiquettes

Définition

Chaque annotation peut être déclarée (principe Tiger-XML) :

```
<anaDecl type='analyse syntaxique ACI'>
  <anaTagDecl type='f'>
    <anaValueDecl type='SN'>syntagme nominal</anaValueDecl>
    <anaValueDecl type='SV'>syntagme verbal</anaValueDecl>
    ...
  </anaTagDecl>
  <anaTagDecl type='n'>
    <anaValueDecl type='Aj'>adjectif</anaValueDecl>
    ...
  </anaTagDecl>
</anaDecl>
```

`anaDecl` comprend une liste d'éléments `anaTagDecl` qualifiés d'un attribut `@type`.

Correspond aux types d'`anaTag` dans l'élément `anaNotes`.

Déclaration du jeu d'étiquettes

Définition

Chaque annotation peut être déclarée (principe Tiger-XML) :

```
<anaDecl type='analyse syntaxique ACI'>
  <anaTagDecl type='f'>
    <anaValueDecl type='SN'>syntagme nominal</anaValueDecl>
    <anaValueDecl type='SV'>syntagme verbal</anaValueDecl>
    ...
  </anaTagDecl>
  <anaTagDecl type='n'>
    <anaValueDecl type='Aj'>adjectif</anaValueDecl>
    ...
  </anaTagDecl>
</anaDecl>
```

`anaValueDecl` décrivent par leur `@type` les valeurs possibles pour les `anaValue` regroupés sous les `anaNotes/anaTag`.

Déclaration du jeu d'étiquettes

Définition

Chaque annotation peut être déclarée (principe Tiger-XML) :

```
<anaDecl type='analyse syntaxique ACI'>  
  <anaTagDecl type='f'>  
    <anaValueDecl type='SN'>syntagme nominal</anaValueDecl>  
    <anaValueDecl type='SV'>syntagme verbal</anaValueDecl>  
    ...  
  </anaTagDecl>  
  <anaTagDecl type='n'>  
    <anaValueDecl type='Aj'>adjectif</anaValueDecl>  
    ...  
  </anaTagDecl>  
</anaDecl>
```

Le texte des `anaValueDecl` **explicit**e le sens des abréviations employées comme `@type`.

- 1 Principes
- 2 Proposition de recommandation
- 3 Conclusion

Conclusion

Synthétisons

Conclusion

Synthétisons

- 1 Employer des pointeurs ;

Conclusion

Synthétisons

- 1 Employer des pointeurs ;
- 2 laisser les scientifiques libres de leurs choix ;

Conclusion

Synthétisons

- 1 Employer des pointeurs ;
- 2 laisser les scientifiques libres de leurs choix ;
- 3 simplifier le travail en employant plusieurs formats ;

Conclusion

Synthétisons

- 1 Employer des pointeurs ;
- 2 laisser les scientifiques libres de leurs choix ;
- 3 simplifier le travail en employant plusieurs formats ;
- 4 partager tout de même en explicitant ses choix.

MERCI