



# ANNODIS

## Annotation discursive

*Corpus de référence pour le français  
et*

*Outils d'aide à l'annotation et à l'exploitation*

-

Projet ANR, programme SHS

*Corpus et outils pour la recherche en SHS*

Marie-Paule Péry-Woodley & Lydia-Mai Ho-Dac

CLLE-ERSS, Toulouse; IRIT, Toulouse; GREYC, Caen





# Sommaire

## Présentation

## Approche ascendante

Mise en application des théories formelles du discours

La SDRT mise en oeuvre

Adaptation de la SDRT à la tâche d'annotation

Procédures d'annotation

## Approche macro

Définir l'objet : types de structures à annoter (et annotables)

Un guide d'annotation des structures énumératives

Procédures d'annotation

## Corpus

## Interfaces

## Annotation exploratoire

## Bilan





# ANNODIS : annotations discursives d'un corpus de référence du français écrit

1. Constituer un corpus de référence du français écrit principalement dédié à l'étude de l'organisation du discours (Corpus diversifié)
2. Articuler des approches discursives ascendantes et macros
3. Annoter des phénomènes discursifs variés et relatifs à différents niveaux de grain
4. Recourir à des prétraitements automatiques pour guider les annotations
5. Développer des outils d'aide à l'annotation et à la navigation textuelle
6. Livrer un corpus annoté et une interface de consultation





## Volonté de centraliser l'éparpillement discursif

### Linguistique du discours

- Analyses dispersées, ponctuelles, théoriquement hétérogènes, difficiles à réutiliser, à valider, à généraliser
- Modèles théoriques de l'organisation du discours non évalués à *grande échelle*

### TAL

- Nécessaire prise en compte des structures discursives pour les applications : les textes ne sont pas des paquets de phrases
- Actuellement, approches largement artisanales, dispersées et athéoriques





## Originalité : articuler deux approches

**Une approche ascendante** : caractériser les relations de discours entre les unités du discours

1. Segmenter le texte en unités de discours élémentaires
2. Construire des structures plus complexes via les *relations de discours*

**Une approche macro** : identifier dans le texte des modes de structuration qui organisent *par le haut* de grands empan de textes

1. Détecter des faisceaux d'indices de surface
2. Identifier et délimiter des structures discursives de haut niveau (macro-structures)





# Sommaire

## Présentation

### Approche ascendante

Mise en application des théories formelles du discours

La SDRT mise en oeuvre

Adaptation de la SDRT à la tâche d'annotation

Procédures d'annotation

### Approche macro

Définir l'objet : types de structures à annoter (et annotables)

Un guide d'annotation des structures énumératives

Procédures d'annotation

### Corpus

### Interfaces

### Annotation exploratoire

### Bilan





## La SDRT en question

### Objectif des théories formelles du discours

- décrire et prédire (générer) des discours cohérents
- principe : un discours est cohérent si tout segment qui le constitue entretient une relation de discours avec un autre segment

### Enjeux dans ANNODIS :

- évaluer le modèle sur des données réelles : son applicabilité et ses règles (ex : la règle de la frontière droite)
- décrire en corpus la réalité textuelle des relations de discours : quels indices textuels marquent telle relation ? (voir PDTB)





## La SDRT mise en oeuvre

1. Délimitation des Unités de Discours Élémentaires (UDE)
2. Caractérisation des relations de discours entre UDE
  - distinction entre relations coordonnantes et subordonnantes
  - Composition de constituants simples et complexes
3. puis entre UDE reliées ... jusqu'à la représentation de la *structure* du texte entier





## l'expérience de DISCOR - discourse structure and coreference resolution

**différences avec ANNODIS** : annotateurs experts, langue anglaise, corpus de presse

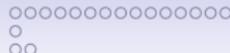
### remarques sur la tâche d'annotation

- fort désaccord
- mise en cause principale : la segmentation en UDE

**conclusions pour ANNODIS** : une segmentation homogène en amont de l'annotation des relations

**une SDRT adaptée** : (1) définir les UDE (2) un guide pour les naïfs





## Une SDRT adaptée - définir les Unités de Discours Elémentaires (UDE)

- Définition opérationnelle d'une UDE (Guide de segmentation)
- Automatisation de la segmentation basée sur une analyse syntaxique automatique (Syntex)
- Débat : peut-on segmenter sans penser relations ?





## Une SDRT adaptée - Faire face à des textes réels et des annotateurs naïfs

**Confrontation aux données** => ajout de relations de discours inexistantes dans le modèle

- relation de flashback (++) , élaboration d'objet
- Question en cours : la relation d'**énumération**

**Confrontation aux naïfs** => ajout de relations pour pallier au caractère non intuitif des définitions des relations dans la théorie

- relation d'**encadrement** (relation d'élaboration intuitivement annotée background)
- relation de **but** (plutôt qu'explication)

**Résultat** : des relations simplifiées dans leur définition et l'abandon des règles de la SDRT (inférences)





## Procédures d'annotation pour l'approche ascendante

1. Segmentation automatique en unités de discours élémentaires
2. L'annotateur valide et corrige si besoin est la segmentation
3. L'annotateur cherche la relation qui lie la première UDE lue aux UDE alentours
4. L'annotateur caractérise cette relation







File Edit View Insert Format Tools Slide Show Window Help

Normal Outline Notes Handout Slide Sorter

annodis:annotation\_des\_pat... annodis:les\_pates\_de\_verre

[La restitution d'oeuvres d'art volées en France.]\_1 [Le juge Thiel dit non aux Japonais.]\_2

[Les pâtes de verre historiques du Musée de l'Ecole de Nancy, {volées en 1985}\_3 {et récupérées en juin 1990 au Japon par la police française}\_4, vont-elles repartir dans ce pays ?}\_5 [Les autorités japonaises réclament la restitution1 des trois vases et de la coupe]\_6 [qu'elles avaient " prêtées " ]\_7 [afin de servir de pièces à conviction dans l'enquête en cours en France.]\_8

[Le juge Gilbert Thiel, {qui sera installé le 7 janvier substitut général à la cour d'appel de Metz}\_9 {après treize années passées à l'instruction à Nancy,}\_10 a refusé d'accressante des Japonais.]\_11 [Pour son dernier jour de magistrat instruc décembre,]\_13 [M. Thiel { qui fut chargé du dossier de Mme Simone W factures de Nancy \_ }\_14 a rendu une ordonnance de refus de restitution trois Gallé et un Daum.]\_16 [Le procureur de la République de Nancy av allant dans le même sens.]\_17

[Cette fermeté des magistrats nancéiens embarrasse fortement le minist étrangères.]\_18 [Ce n'est en effet qu'au prix d'engagements pris par cerl Japon { et sur lesquels une grande discrétion a été de mise \_ }\_19 que Main aux algues d'Emile Gallé, le vase parlant le Souci de plaire du mër Daum orné de prunelles, avaient pu être récupérés en juin par les polici

Elaboration (1-2,5)

Narration (3,4)

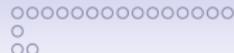
Background (5,3-4)

Explication\* (5,6)

Background (6,7)

But (7,8)





## Questionnements

- Prise en compte des *discourse pops* : relations macro (titres) ?
- Incertitude des définitions des relations
- Temps moyen (nombre de relations annotées en X temps)





# Sommaire

## Présentation

### Approche ascendante

Mise en application des théories formelles du discours

La SDRT mise en oeuvre

Adaptation de la SDRT à la tâche d'annotation

Procédures d'annotation

### Approche macro

Définir l'objet : types de structures à annoter (et annotables)

Un guide d'annotation des structures énumératives

Procédures d'annotation

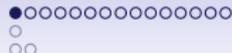
### Corpus

### Interfaces

### Annotation exploratoire

### Bilan



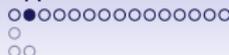


## Des structures macro annotables

les structures discursives macro se perçoivent par un regard global sur l'objet texte ou des portions de l'objet texte et non par l'analyse sémantique de l'enchaînement effectif des énoncés.

- Décrire les phénomènes de mise en texte du discours :  
linéarisation et séquentialité
- Délimiter les segments textuels définis en termes de continuité et de discontinuité (séquentialité complexe)
- Découvrir quels indices (signaux, traces, marques) permettent d'identifier et de caractériser (d'interpréter ?) les différents segments



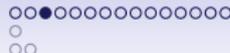


## Des structures macro annotables

### Différences avec l'approche ascendante

- Absence de modèle prédéfini (même si cadre fonctionnaliste de Halliday)
  - pas d'assises théoriques
  - pas de modèle à évaluer
- Une approche guidée par les données (indices lexico-syntaxiques **et** typo-dispositionnels)





File Edit View Insert Format Tools Slide Show Window Help

Normal Outline Notes Handout Slide Sorter

## Principes de la sélection naturelle (suite)

La théorie de la sélection naturelle telle qu'elle a été initialement décrite par [Charles Darwin](#), repose sur trois principes<sup>[1]</sup>:

1. le principe de variation
2. le principe d'adaptation
3. le principe d'hérédité

### Principe 1 : Les individus diffèrent les uns des autres (suite)



En général, dans une [population](#) d'individus d'une même [espèce](#), il existe des différences plus ou moins importantes entre ces individus. En biologie, on appelle [caractère](#), tout ce qui est visible et peut varier d'un individu à l'autre. On dit qu'il existe plusieurs [traits](#) pour un même caractère. Par exemple, chez l'[être humain](#), la [couleur de la peau](#), la [couleur des yeux](#) sont des caractères pour lesquels il existe de multiples variations ou traits. La variation d'un caractère chez un individu donné constitue son [phénotype](#). C'est là, la première condition pour qu'il y ait sélection naturelle : au sein d'une population, certains caractères doivent présenter des variations, c'est le *principe de variation*.

### Principe 2 : Les individus les plus adaptés au milieu survivent et se reproduisent davantage (suite)



Certains individus portent des variations qui leur permettent de se reproduire davantage que les autres, dans un [environnement](#) précis. On dit qu'ils disposent d'un [avantage sélectif](#) sur leurs congénères :

- La première possibilité est, par exemple, qu'en échappant mieux aux [prédateurs](#), en étant moins malades, en accédant plus facilement à la nourriture, ces individus atteignent plus facilement l'âge adulte, pour être aptes à la [reproduction](#). Ceux qui ont une meilleure capacité de survie pourront donc se reproduire davantage.
- Dans le cas particulier de la reproduction sexuée, les individus ayant survécu peuvent être porteurs d'un caractère particulièrement attirant pour les partenaires de [sexes](#) opposés. Ceux-ci seront capables d'engendrer une plus grande descendance en [copulant](#) davantage.

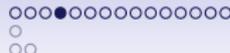
Dans les deux cas, l'augmentation de la capacité à survivre et à se reproduire se traduit par une augmentation du [taux de reproduction](#) et donc par une descendance plus nombreuse, pour les individus porteurs de ces caractéristiques. On dit alors que ce trait de caractère donné offre un [avantage sélectif](#) par rapport à d'autres. C'est dans ce *principe d'adaptation* uniquement, qu'intervient le [milieu de vie](#).

### Principe 3 : Les caractéristiques avantageuses doivent être héréditaires (suite)



La troisième condition pour qu'il y ait sélection naturelle est que les caractéristiques des individus doivent être [héréditaires](#), c'est-à-dire qu'elles puissent être transmises à leur descendance. En effet certains caractères, comme le bronzage ou la culture, ne dépendent pas du [génotype](#), c'est-à-dire l'ensemble des [gènes](#) de l'individu. Lors de la [reproduction](#), ce sont donc les gènes qui, transmis aux descendants, entraîneront le passage de certains





File Edit View Insert Format Tools Slide Show Window Help

Normal Outline Notes Handout Slide Sorter

## Principes de la sélection naturelle (Lecture)

La théorie de la sélection naturelle telle qu'elle a été initialement décrite par [Charles Darwin](#), repose sur trois principes :

1. le principe de variation
2. le principe d'adaptation
3. le principe d'hérédité

### Principe 1 : Les individus diffèrent les uns des autres (Lecture)



En général, dans une [population](#) d'individus d'une même [espèce](#), il existe des différences plus ou moins importantes entre ces individus. En biologie, on appelle [caractère](#), tout ce qui est visible et peut varier d'un individu à l'autre. On dit qu'il existe plusieurs [traits](#) pour un même caractère. Par exemple, chez l'[être humain](#), la [couleur de la peau](#), la [couleur des yeux](#) sont des caractères pour lesquels il existe de multiples variations ou traits. La variation d'un caractère chez un individu donné constitue son [phénotype](#). C'est là, la première condition pour qu'il y ait sélection naturelle : au sein d'une population, certains caractères doivent présenter des variations, c'est le [principe de variation](#).

### Principe 2 : Les individus les plus adaptés au milieu survivent et se reproduisent davantage (Lecture)



Certains individus portent des variations qui leur permettent de se reproduire davantage que les autres, dans un [environnement](#) précis. On dit qu'ils disposent d'un [avantage sélectif](#) sur leurs congénères :

- La première possibilité est, par exemple, qu'en échappant mieux aux [prédateurs](#), en étant moins malades, en accédant plus facilement à la nourriture, ces individus atteignent plus facilement l'âge adulte, pour être aptes à la [reproduction](#). Ceux qui ont une meilleure capacité de survie pourront donc se reproduire davantage.
- Dans le cas particulier de la reproduction sexuée, les individus ayant survécu peuvent être porteurs d'un caractère particulièrement attirant pour les partenaires de [sexes](#) opposés. Ceux-ci seront capables d'engendrer une plus grande descendance en [copulant](#) davantage.

Dans les deux cas, l'augmentation de la capacité à survivre et à se reproduire se traduit par une augmentation du [taux de reproduction](#) et donc par une descendance plus nombreuse, pour les individus porteurs de ces caractéristiques. On dit alors que ce trait de caractère donné offre un [avantage sélectif](#) par rapport à d'autres. C'est dans ce [principe d'adaptation](#) uniquement, qu'intervient le [milieu de vie](#).

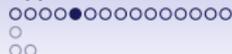
### Principe 3 : Les caractéristiques avantageuses doivent être héréditaires (Lecture)



La troisième condition pour qu'il y ait sélection naturelle est que les caractéristiques des individus doivent être [héréditaires](#), c'est-à-dire qu'elles puissent être transmises à leur descendance. En effet certains caractères, comme le bronzage ou la culture, ne dépendent pas du [phénotype](#), c'est-à-dire l'ensemble des [osés](#) de l'individu. Lors de la [reproduction](#), ce sont donc les gènes qui, transmis aux descendants, entraîneront le passage de certains

## Organisation en sections et récurrence lexicale





File Edit View Insert Format Tools Slide Show Window Help

Normal Outline Notes Handout Slide Sorter

## Principes de la sélection naturelle (auteur)

La théorie de la sélection naturelle telle qu'elle a été initialement décrite par [Charles Darwin](#), repose sur trois principes :

1. le principe de variation
2. le principe d'adaptation
3. le principe d'hérédité

### Principe 1 : Les individus diffèrent les uns des autres (auteur)



En général, dans une [population](#) d'individus d'une même [espèce](#), il existe des différences plus ou moins importantes entre ces individus. En biologie, on appelle [caractère](#), tout ce qui est visible et peut varier d'un individu à l'autre. On dit qu'il existe plusieurs [traits](#) pour un même caractère. Par exemple, chez l'[être humain](#), la [couleur de la peau](#), la [couleur des yeux](#) sont des caractères pour lesquels il existe de multiples variations ou traits. La variation d'un caractère chez un individu donné constitue son [phénotype](#). C'est là, la première condition pour qu'il y ait sélection naturelle : au sein d'une population, certains caractères doivent présenter des variations, c'est le [principe de variation](#).

### Principe 2 : Les individus les plus adaptés au milieu survivent et se reproduisent davantage (auteur)



Certains individus portent des variations qui leur permettent de se reproduire davantage que les autres, dans un [environnement](#) précis. On dit qu'ils disposent d'un [avantage sélectif](#) sur leurs congénères :

- La première possibilité est, par exemple, qu'en échappant mieux aux [prédateurs](#), en étant moins malades, en accédant plus facilement à la nourriture, ces individus atteignent plus facilement l'âge adulte, pour être aptes à la [reproduction](#). Ceux qui ont une meilleure capacité de survie pourront donc se reproduire davantage.
- Dans le cas particulier de la reproduction sexuée, les individus ayant survécu peuvent être porteurs d'un caractère particulièrement attrayant pour les partenaires de [sexes](#) opposés. Ceux-ci seront capables d'engendrer une plus grande descendance en [copulant](#) davantage.

Dans les deux cas, l'augmentation de la capacité à survivre et à se reproduire se traduit par une augmentation du [taux de reproduction](#) et donc par une descendance plus nombreuse, pour les individus porteurs de ces caractéristiques. On dit alors que ce trait de caractère donne à l'individu un [avantage sélectif](#) par rapport à d'autres. C'est dans ce [principe d'adaptation](#) uniquement, qu'intervient le [milieu de vie](#).

### Principe 3 : Les caractéristiques avantageuses doivent être héréditaires (auteur)



La troisième condition pour qu'il y ait sélection naturelle est que les caractéristiques des individus doivent être [héréditaires](#), c'est-à-dire qu'elles puissent être transmises à leur descendance. En effet certains caractères, comme le bronzage ou la culture, ne dépendent pas du [génotype](#), c'est-à-dire l'ensemble des [gènes](#) de l'individu. Lors de la [reproduction](#), ce sont donc les gènes qui, transmis aux descendants, entraîneront le passage de certains

Organisation en sections et récurrence lexicale

Titres parallèles de forme :  
Principe # : proposition





File Edit View Insert Format Tools Slide Show Window Help

Normal Outline Notes Handout Slide Sorter

## Principes de la sélection naturelle (auteur)

La théorie de la sélection naturelle telle qu'elle a été initialement décrite par [Charles Darwin](#), repose sur trois principes :

1. le principe de variation
2. le principe d'adaptation
3. le principe d'hérédité

### Principe 1 : Les individus diffèrent les uns des autres (auteur)



En général, dans une [population](#) d'individus d'une même [espèce](#), il existe des différences plus ou moins importantes entre ces individus. En biologie, on appelle [caractère](#), tout ce qui est visible et peut varier d'un individu à l'autre. On dit qu'il existe plusieurs [traits](#) pour un même caractère. Par exemple, chez l'[être humain](#), la [couleur de la peau](#), la [couleur des yeux](#) sont des caractères pour lesquels il existe de multiples variations ou traits. La variation d'un caractère chez un individu donné constitue son [phénotype](#). C'est là, la première condition pour qu'il y ait sélection naturelle : au sein d'une population, certains caractères doivent présenter des variations, c'est le [principe de variation](#).

### Principe 2 : Les individus les plus adaptés au milieu survivent et se reproduisent davantage (auteur)



Certains individus portent des variations qui leur permettent de se reproduire davantage que les autres, dans un [environnement](#) précis. On dit qu'ils disposent d'un [avantage sélectif](#) sur leurs congénères :

- La première possibilité est, par exemple, qu'en échappant mieux aux [prédateurs](#), en étant moins malades, en accédant plus facilement à la nourriture, ces individus atteignent plus facilement l'âge adulte, pour être aptes à la [reproduction](#). Ceux qui ont une meilleure capacité de survie pourront donc se reproduire davantage.
- Dans le cas particulier de la reproduction sexuée, les individus ayant survécu peuvent être porteurs d'un caractère particulièrement attrayant pour les partenaires de [sexes](#) opposés. Ceux-ci seront capables d'engendrer une plus grande descendance en [copulant](#) davantage.

Dans les deux cas, l'augmentation de la capacité à survivre et à se reproduire se traduit par une augmentation du [taux de reproduction](#) et donc par une descendance plus nombreuse, pour les individus porteurs de ces caractéristiques. On dit alors que ce trait de caractère donne à l'individu un [avantage sélectif](#) par rapport à d'autres. C'est dans ce [principe d'adaptation](#) uniquement, qu'intervient le [milieu de vie](#).

### Principe 3 : Les caractéristiques avantageuses doivent être héréditaires (auteur)



La troisième condition pour qu'il y ait sélection naturelle est que les caractéristiques des individus doivent être [héréditaires](#), c'est-à-dire qu'elles puissent être transmises à leur descendance. En effet certains caractères, comme le bronzage ou la culture, ne dépendent pas du [phénotype](#), c'est-à-dire l'ensemble des [osés](#) de l'individu. Lors de la [reproduction](#), ce sont donc les gènes qui, transmis aux descendants, entraîneront le passage de certains

Organisation en sections et récurrence lexicale

Titres parallèles de forme :  
Principe # : proposition





File Edit View Insert Format Tools Slide Show Window Help

Normal Outline Notes Handout Slide Sorter

## Principes de la sélection naturelle (auto2)

La théorie de la sélection naturelle telle qu'elle a été initialement décrite par [Charles Darwin](#), repose sur trois principes :

1. le principe de variation
2. le principe d'adaptation
3. le principe d'hérédité

### Principe 1 : Les individus diffèrent les uns des autres (auto2)



En général, dans une [population](#) d'individus d'une même [espèce](#), il existe des différences plus ou moins importantes entre ces individus. En biologie, on appelle [caractère](#), tout ce qui est visible et peut varier d'un individu à l'autre. On dit qu'il existe plusieurs [traits](#) pour un même caractère. Par exemple, chez l'[être humain](#), la [couleur de la peau](#), la [couleur des yeux](#) sont des caractères pour lesquels il existe de multiples variations ou traits. La variation d'un caractère chez un individu donné constitue son [phénotype](#). C'est là, la première condition pour qu'il y ait sélection naturelle : au sein d'une population, certains caractères doivent présenter des variations, c'est le [principe de variation](#).

### Principe 2 : Les individus les plus adaptés au milieu survivent et se reproduisent davantage (auto2)



Certains individus portent des variations qui leur permettent de se reproduire davantage que les autres, dans un [environnement](#) précis. On dit qu'ils disposent d'un [avantage sélectif](#) sur leurs congénères :

- La première possibilité est, par exemple, qu'en échappant mieux aux [prédateurs](#), en étant moins malades, en accédant plus facilement à la nourriture, ces individus atteignent plus facilement l'âge adulte, pour être aptes à la [reproduction](#). Ceux qui ont une meilleure capacité de survie pourront donc se reproduire davantage.
- Dans le cas particulier de la reproduction sexuée, les individus ayant survécu peuvent être porteurs d'un caractère particulièrement attrayant pour les partenaires de [sexes](#) opposés. Ceux-ci seront capables d'engendrer une plus grande descendance en [copulant](#) davantage.

Dans les deux cas, l'augmentation de la capacité à survivre et à se reproduire se traduit par une augmentation du [taux de reproduction](#) et donc par une descendance plus nombreuse, pour les individus porteurs de ces caractéristiques. On dit alors que ce trait de caractère donné offre un [avantage sélectif](#) par rapport à d'autres. C'est dans ce [principe d'adaptation](#) uniquement, qu'intervient le [milieu de vie](#).

### Principe 3 : Les caractéristiques avantageuses doivent être héréditaires (auto2)



La troisième condition pour qu'il y ait sélection naturelle est que les caractéristiques des individus doivent être [héréditaires](#), c'est-à-dire qu'elles puissent être transmises à leur descendance. En effet certains caractères, comme le bronzage ou la culture, ne dépendent pas du [phénotype](#), c'est-à-dire l'ensemble des [os](#) de l'individu. Lors de la [reproduction](#), ce sont donc les gènes qui, transmis aux descendants, entraîneront le passage de certains

Organisation en sections et récurrence lexicale

Annnonce d'une énumération en 3 items

Titres parallèles de forme :  
Principe # : proposition





File Edit View Insert Format Tools Slide Show Window Help

Normal Outline Notes Handout Slide Sorter

## Principes de la sélection naturelle 1/24/04

La théorie de la sélection naturelle telle qu'elle a été initialement décrite par [Charles Darwin](#), repose sur trois principes :

1. le principe de variation
2. le principe d'adaptation
3. le principe d'hérédité

### Principe 1 : Les individus diffèrent les uns des autres 1/24/04



En général, dans une **population** d'individus d'une même **espèce**, il existe des différences plus ou moins appelées **caractères** : tout ce qui est visible et peut varier d'un individu à l'autre. On dit qu'il existe chez l'**être humain**, la **couleur de la peau**, la **couleur des yeux** sont des caractères pour lesquels d'un caractère chez un individu donné constitue son **phénotype**. **C'est là la première condition** de la sélection naturelle : certains caractères doivent présenter des variations. **C'est le principe de variation**.

### Principe 2 : Les individus les plus adaptés au milieu survivent et se reproduisent davantage 1/24/04



Certains individus portent des variations qui leur permettent de se reproduire davantage que les autres, dans un **environnement** précis. On dit qu'ils disposent d'un **avantage sélectif** sur leurs congénères :

- La première possibilité est, par exemple, qu'en échappant mieux aux **prédateurs**, en étant moins malades, en accédant plus facilement à la nourriture, ces individus atteignent plus facilement l'âge adulte, pour être aptes à la **reproduction**. Ceux qui ont une meilleure capacité de survie pourront donc se reproduire davantage.
- Dans le cas particulier de la reproduction sexuée, les individus ayant survécu peuvent être porteurs d'un caractère particulièrement attirant pour les partenaires de **sexes** opposés. Ceux-ci seront capables d'engendrer une plus grande descendance en **copulant** davantage.

Dans les deux cas, l'augmentation de la capacité à survivre et à se reproduire se traduit par une augmentation du **taux de reproduction** et donc par une descendance plus nombreuse. **Pour les individus porteurs de ces caractéristiques**. On dit alors que ce trait de caractère donné offre un **avantage sélectif** par rapport à d'autres. **C'est dans ce principe d'adaptation à l'environnement**, qui intervient la **sélection de vie**.

### Principe 3 : Les caractéristiques avantageuses doivent être héréditaires 1/24/04



La troisième condition pour qu'il y ait sélection naturelle est que les caractéristiques des individus doivent être **héréditaires**, c'est-à-dire qu'elles puissent être transmises à leur descendance. En effet certains caractères, comme le bronzage ou la culture, ne dépendent pas du **génotype**, c'est-à-dire l'ensemble des **gènes** de l'individu. Lors de la **reproduction**, ce sont donc les gènes qui, transmis aux descendants, entraîneront le passage de certains

Organisation en sections et récurrence lexicale

Annnonce d'une énumération en 3 items

Indices plus locaux qui confortent la perception globale première :

- indices d'énumération :
- amorce
- marque d'item
- parallélismes en même position textuelle de la typo. + construction syntaxique

Titres parallèles de forme :

Principe # : proposition





File Edit View Insert Format Tools Slide Show Window Help

Normal Outline Notes Handout Slide Sorter

## Principes de la sélection naturelle 1/20/2010

La théorie de la sélection naturelle telle qu'elle a été initialement décrite par [Charles Darwin](#), repose sur trois principes :

1. le principe de variation
2. le principe d'adaptation
3. le principe d'hérédité

### Principe 1 : Les individus diffèrent les uns des autres 1/20/2010



En général, dans une **population** d'individus d'une même **espèce**, il existe des différences plus ou moins appelées **caractères** : tout ce qui est visible et peut varier d'un individu à l'autre. On dit qu'il existe chez l'**être humain**, la **couleur de la peau**, la **couleur des yeux** sont des caractères pour lesquels d'un caractère chez un individu donné constitue son **phénotype**. C'est là la **première condition** : population, certains caractères doivent présenter des variations. C'est le **principe de variation**.

### Principe 2 : Les individus les plus adaptés au milieu survivent et se reproduisent davantage 1/20/2010



Certains individus portent des variations qui leur permettent de se reproduire davantage que les autres, dans un **environnement** précis. On dit qu'ils disposent d'un **avantage sélectif** sur leurs congénères.

- La première possibilité est, par exemple, qu'en échappant mieux aux **prédateurs**, en étant moins malades, en accédant plus facilement à la nourriture, ces individus atteignent plus facilement l'âge adulte, pour être aptes à la **procréation**. Ceux qui ont une meilleure capacité de survie pourront donc se reproduire davantage.
- Dans le cas particulier de la reproduction sexuée, les individus ayant survécu peuvent être porteurs d'un caractère particulièrement attirant pour les partenaires de **sexes** opposés. Ceux-ci seront capables d'engendrer une plus grande descendance en **copulant** davantage.

Dans les deux cas, l'augmentation de la capacité à survivre et à se reproduire se traduit par une augmentation du **taux de reproduction** et donc par une **descendance plus nombreuse**, pour les individus porteurs de ces caractéristiques. On dit alors que ce trait de caractère donné offre un **avantage sélectif** par rapport à d'autres. C'est dans ce **principe d'adaptation** qu'intervient la **lutte de vie**.

### Principe 3 : Les caractéristiques avantageuses doivent être héréditaires 1/20/2010



La troisième condition pour qu'il y ait sélection naturelle est que les caractéristiques des individus doivent être **héréditaires**, c'est-à-dire qu'elles puissent être transmises à leur descendance. En effet certains caractères, comme le bronzage ou la culture, ne dépendent pas du **génotype**, c'est-à-dire l'ensemble des **gènes** de l'individu. Lors de la **reproduction**, ce sont donc les gènes qui, transmis aux descendants, entraîneront le passage de certains

Organisation en sections et récurrence lexicale

Annnonce d'une énumération en 3 items

Indices plus locaux qui confortent la perception globale première :

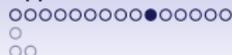
- indices d'énumération :
- amorce
- marque d'item

parallélismes en même position textuelle de la typo. + construction syntaxique

Titres parallèles de forme : Principe # : proposition

Récurtivité du principe d'énumération

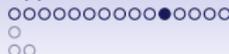




## La structure énumérative - SE : une structure méta-macro ?

- Les structures macro ne se construisent pas dans l'enchaînement des énoncés mais dans l'interaction d'indices de niveau global et local
- La structure énumérative (SE) correspond à une transcription des processus d'agencement des propos selon le format linéaire du texte





File Edit View Insert Format Tools Slide Show Window Help

Normal Outline Notes Handout Slide Sorter

## Principes de la sélection naturelle (Lecture)

La théorie de la sélection naturelle telle qu'elle a été initialement décrite par [Charles Darwin](#), repose sur trois principes (Lecture):

1. le principe de variation
2. le principe d'adaptation
3. le principe d'hérédité

### Principe 1 : Les individus diffèrent les uns des autres (Lecture)



En général, dans une [population](#) d'individus d'une même [espèce](#), il existe des différences plus ou moins importantes entre ces individus. En biologie, on appelle [caractère](#), tout ce qui est visible et peut varier d'un individu à l'autre. On dit qu'il existe plusieurs [traits](#) pour un même caractère. Par exemple, chez l'[être humain](#), la [couleur de la peau](#), la [couleur des yeux](#) sont des caractères pour lesquels il existe de multiples variations ou traits. La variation d'un caractère chez un individu donné constitue son [phénotype](#). C'est là, la première condition pour qu'il y ait sélection naturelle : au sein d'une population, certains caractères doivent présenter des variations, c'est le [principe de variation](#).

### Principe 2 : Les individus les plus adaptés au milieu survivent et se reproduisent davantage (Lecture)



Certains individus portent des variations qui leur permettent de se reproduire davantage que les autres, dans un [environnement](#) donné. Ils disposent d'un [avantage sélectif](#) sur leurs congénères :

- La première possibilité est, par exemple, qu'en échappant mieux aux [prédateurs](#), en étant moins malades, en accédant plus facilement à la nourriture, ces individus atteignent plus facilement l'âge adulte, pour être aptes à la [reproduction](#). Ceux qui ont une meilleure survie pourront donc se reproduire davantage.
- Dans le cas particulier de la reproduction sexuée, les individus ayant survécu peuvent être porteurs d'un caractère particulier qui leur permet de trouver plus facilement des partenaires de [sexes](#) opposés. Ceux-ci seront capables d'engendrer une plus grande descendance en [copulant](#) davantage.

Dans les deux cas, l'augmentation de la capacité à survivre et à se reproduire se traduit par une augmentation du [taux de reproduction](#) et donc par une descendance plus nombreuse, pour les individus porteurs de ces caractéristiques. On dit alors que ce trait de caractère donne un [avantage sélectif](#) par rapport à d'autres. C'est dans ce [principe d'adaptation](#) uniquement, qu'intervient le [milieu de vie](#).

### Principe 3 : Les caractéristiques avantageuses doivent être héréditaires (Lecture)

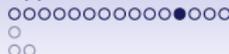


La troisième condition pour qu'il y ait sélection naturelle est que les caractéristiques des individus doivent être [héréditaires](#), c'est-à-dire qu'elles puissent être transmises à leur descendance. En effet certains caractères, comme le bronzage ou la culture, ne dépendent pas du [génotype](#), c'est-à-dire l'ensemble des [gènes](#) de l'individu. Lors de la [reproduction](#), ce sont donc les gènes qui, transmis aux descendants, entraîneront le passage de certains

Diagramme de la droite :

- amorce
- Item
- Item
- Item
- amorce
- Item
- Item





File Edit View Insert Format Tools Slide Show Window Help

Normal Outline Notes Handout Slide Sorter

En général, dans une [population](#) d'individus d'une même [espèce](#), il existe des différences plus ou moins importantes entre ces individus. En biologie, on appelle [caractère](#), tout ce qui est visible et peut varier d'un individu à l'autre. On dit qu'il existe plusieurs [traits](#) pour un même caractère. Par exemple, chez l'[être humain](#), la [couleur de la peau](#), la [couleur des yeux](#) sont des caractères pour lesquels il existe de multiples variations ou traits. La variation d'un caractère chez un individu donné constitue son [phénotype](#). C'est là, la première condition pour qu'il y ait sélection naturelle : au sein d'une population, certains caractères doivent présenter des variations, c'est le [principe de variation](#).

**Principe 2 : Les individus les plus adaptés au [milieu](#) survivent et se reproduisent davantage [\(modifier\)](#)**



Certains individus portent des variations qui leur permettent de se reproduire davantage que les autres, dans un [environnement](#) précis. On dit qu'ils disposent d'un [avantage sélectif](#) sur leurs congénères:

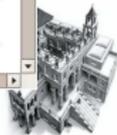
- La première possibilité est, par exemple, qu'en échappant mieux aux [prédateurs](#), en étant moins malades, en accédant plus facilement à la nourriture, ces individus atteignent plus facilement l'âge adulte, pour être apte à la [reproduction](#). Ceux qui ont une meilleure capacité de survie pourront donc se reproduire davantage.
- Dans le cas particulier de la reproduction sexuée, les individus ayant survécu peuvent être porteurs d'un caractère particulièrement attirant pour les partenaires de [sexe](#) opposé. Ceux-là seront capables d'engendrer une plus grande descendance en [copulant](#) davantage.

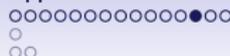
Dans les deux cas, l'augmentation de la capacité à survivre et à se reproduire se traduit par une augmentation du [taux de reproduction](#) et donc par une descendance plus nombreuse, pour les individus porteurs de ces caractéristiques. On dit alors que ce trait de caractère donné offre un [avantage sélectif](#), par rapport à d'autres. C'est dans ce [principe d'adaptation](#) uniquement, qu'intervient le [milieu de vie](#).

**Principe 3 : Les caractéristiques avantageuses doivent être [héréditaires](#) [\(modifier\)](#)**



amorce  
Item  
Item  
Clôture

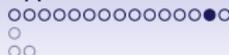




## La structure énumérative - SE : une structure méta-macro ?

- La structure énumérative (SE) n'existe pas que par la mise en forme matérielle mais également par des effets de parallélismes sémantico-syntaxiques et structurels (série de cadres)





File Edit View Insert Format Tools Slide Show Window Help

Normal Outline Notes Handout Slide Sorter

partenaires arabes, est une magnifique vitrine de cette politique avec tout ce qu'elle comporte comme atouts en termes de réalisation mais aussi de difficultés en terme de gestion.

*Les relations nouées depuis des siècles dans la région nous valent assurément estime et considération. Elles suscitent aussi des attentes et des déceptions.* } amorce

**Au Maghreb,** les gouvernements attendent de nous concours et, pour chacun d'entre eux, soutien exclusif. Les populations sont plus attentives à la coopération, à la liberté de circulation et à la situation des immigrés chez nous. } Item

**Au Proche-Orient,** nos prises de parole sont scrutées et analysées dans le détail. Nous y sommes attendus, sollicités et espérés tant l'image d'une France compagne de route des grandes causes arabes demeure encore enracinée. } Item

L'approche est différente dans le Golfe où nous sommes vus comme un partenaire privilégié pour se soustraire à un tête-à-tête trop exclusif avec les Etats-Unis. } Item

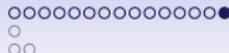
Les perspectives pour la France dans tous les domaines y sont remarquables. En témoignent tout récemment les opérations du Louvre et de la Sorbonne à Abou Dhabi.

**2.2. Un climat désenc** Titre de section  
+ découpage en section  
+ initiales de paragraphes similaires (//sme)

En dépit des relations ... et politique consenti pendant les quinze dernières années, la relation semble désenchantée et incertaine.

Sans doute payons-nous le prix d'une personnalisation excessive et généralisée de la relation politique. Nous avons fait un usage extensif des termes de coopération privilégiée, d'alliances





## La structure énumérative - SE

### Intérêts

- procédé basique d'organisation du discours
- grande variabilité de grain
- un cadre pour représenter (et annoter) l'organisation du discours et les zones de continuité/discontinuité

### Incertitudes

- Quelle couverture (alternatives : les unités référentielles)
- Quels recouvrements avec les relations du discours (relation d'énumération... ouf!)
- Des annotations pour qui et pour quoi faire ?





## Guide d'annotation de SE

### Problèmes et questionnements

- des objets très intuitifs (quels accords ?) à la définition fonctionnelle vague et nécessairement interdépendante
- absence de relations de discours *nobles* et d'ancrage théorique (quelle terminologie)
- comment pousser l'annotateur à voir global (ex : à prendre en compte les titres) ?

### Atouts

- une structure connue (intuitivement) des naïfs (mais y a-t-il besoin de naïfs ?)
- un modèle simple
- un guidage par les indices de surfaces : les structures ne peuvent être définies en dehors de leurs indices



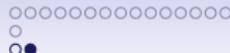


## Procédures d'annotation pour l'approche macro

1. Pré-traitement automatique : marquage des indices exploitant des travaux antérieurs
2. L'annotateur navigue dans le texte, guidé par une visualisation globale colorée des indices marqués.
3. L'annotateur délimite les SE et ses objets (amorce, co-items, clôture)

Exemple





## Questionnements

- Quel accès donner (demander) à l'annotateur : quelle perception globale d'un document sur un écran ?
- Quels guides constituent les indices :
  - Donner des poids différents aux indices (et donc complexifier le modèle)
  - Filtrer (par contraintes) les indices à colorer





## Un corpus multi-genre, des **documents** libres de droits et pertinents

**multi-genre** : tester l'hypothèse de la variation de l'organisation discursive en fonction du type de texte et couvrir différents phénomènes discursifs

**pertinents** : constituer un bon terrain d'investigation pour les deux approches

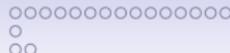
- toutes les relations de discours de l'écrit => profilage
- structures discursives complexes (question de longueur et de titraille, et de genre)

**des documents** : prise en compte des indices de mise en forme matérielle (le texte est avant tout un document)

**libres de droits et éditables** (Geopo, Wikipedia, CMLF, Est républicain)

**Volume visé** : 5000 segments annotés ?





## Besoin d'outils d'aide à l'annotation (GREYC)

### Interface de visualisation des documents

- MFM
- annotations existantes : indices macro et segmentation en UDE

**Délimitation** des zones textuelles de taille quelconque, en gérant discontinuités, chevauchements avec zones déjà délimitées

**Formats** ouverts et interopérables (famille XML) pour réutilisabilité







## Campagne d'annotation exploratoire

**Protocole général** : annotation par les différents membres du projet avec échange d'approche

- des textes courts (ascendants) et longs (macro) déjà connus pour mise au point de l'annotation
- ascendants : une première expérience avec des naïfs (en cours)
- macro : laisser une certaine liberté aux annotateurs et affiner ensuite selon les retours

### Objectifs

- stabiliser les guides et tester l'interface
- Avoir une idée du temps que la tâche d'annotation représente





## Bilan d'une première année

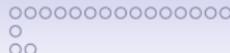
### Fait

- prétraitements en partie achevés : segmentation, MFM et marquage
- besoins en terme d'interface fixés (sauf peut-être pour la visualisation du document)
- une première version des guides ascendants finalisée
- un guide macro en cours de construction (extrait)

### à faire (à court terme)

- homogénéiser les prétraitements
- rencontrer l'interface
- récolter et utiliser les retours de l'annotation exploratoire pour les guides et l'interface
- sélectionner le corpus (pour l'annotation finale)





## Ambitions en TAL

Mise en place de **techniques d'apprentissage automatique**

- trouver corrélations entre configurations de marques calculables automatiquement et les structures discursives (SE, relations de discours) posées par les annotateurs humains

Techniques de **profilage de textes** pour structurer le corpus de référence

