

Interopérabilité des Données et des Connaissances

Mauro Gaio

Université de Pau et des Pays de l'Adour  
UFR S&T de Pau  
mauro.gaio@univ-pau.fr

Navigation icons: back, forward, search, etc.

## Plan

- Introduction

- Introduction
  - De quoi parle-t-on ?
  - le Web : quelles particularités ?

## 2 Rappels

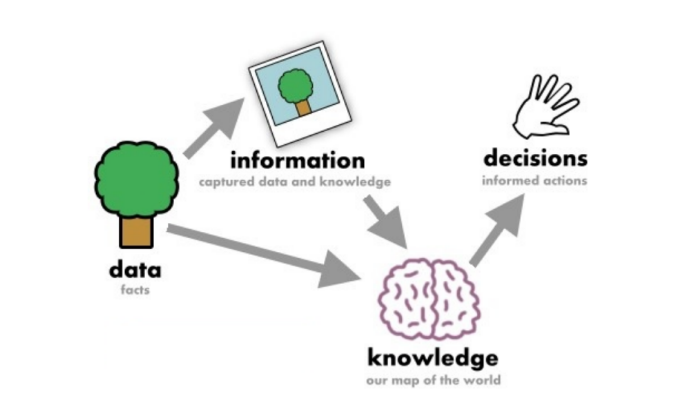
- Les langages formels

Navigation icons: back, forward, search, etc.

## Plan

De quoi parle-t-on ?

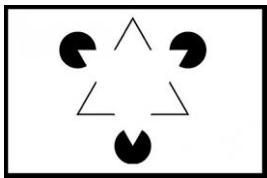
Données vs. Informations vs. Connaissances



Mauro Galo (UPPA) IDC 4/29

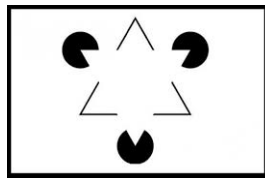
## Données vs. Informations vs. Connaissances

Percevoir la difficulté



## Données vs. Informations vs. Connaissances

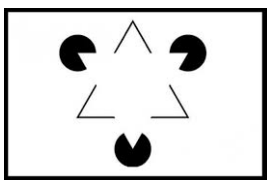
Percevoir la difficulté



- **données** : un ensemble unique de valeurs numériques

## Données vs. Informations vs. Connaissances

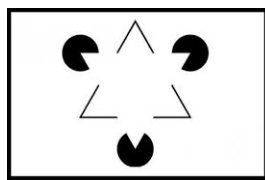
Percevoir la difficulté



- **données** : un ensemble unique de valeurs numériques
- **informations** : des ensembles de valeurs numériques catégorisées

## Données vs. Informations vs. Connaissances

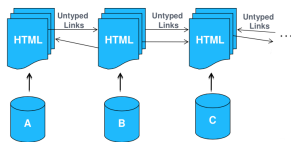
Percevoir la difficulté



- **données** : un ensemble unique de valeurs numériques
- **informations** : des ensembles de valeurs numériques catégorisées
- **connaissances** : un ensembles de mécanismes permettant d'affecter une catégorie et/ou de transformer une catégorie dans une autre

## Le Web : principes fondateurs

Phase 1 : Le Web de l'hypertexte

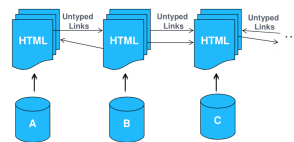


Quels sont les principaux constituants ?

```
<!DOCTYPE html>
<html>
  <head>
    <title>Titre </title>
  </head>
  <body>
    <p> Le contenu textuel, plusieurs lignes de texte </p>
  </body>
</html>
```

## Le Web : principes fondateurs

Phase 1 : Le Web de l'hypertexte



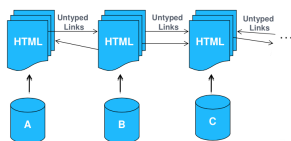
- Le premier constituant : un contenu statique balisé par un langage standardisé (HTML)

```
<!DOCTYPE html>
<html>
  <head>
    <title>Titre </title>
  </head>
  <body>
    <p> Le contenu textuel, plusieurs lignes de texte </p>
  </body>
</html>
```

- Langage de balisage interpréter par la machine rend le contenu très lisible par l'homme (mise en page)

## Le Web : principes fondateurs

Phase 1 : Le Web de l'hypertexte



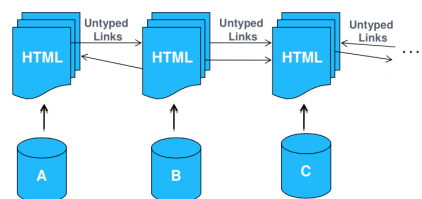
- Le second constituant : une localisation intéropérable (protocole HTTP) et un mécanisme de navigation standard intégré au langage de balisage.

```
<!DOCTYPE html>
<html>
  <head>
    <title>Titre </title>
  </head>
  <body>
    <p> Le <a href="https://fr.wikipedia.org/wiki/Linguistique_textuelle"> contenu textuel </a>,
      peut comporter plusieurs liens non-typés externes</p>
    <p> ... on peut aussi avoir des liens internes, <a href="#section2">Allez à la section
      2</a> et des liens typés
     vers des images (localisées
      ailleurs ou en local)
  </body>
</html>
```

- Mécanismes intégrés dans le langage permettant une navigation hypertexte par l'homme. Navigation allant de l'intra document aux réseaux interconnectés (Web)

## Le Web : principes fondateurs

Phase 1 : Le Web de l'hypertexte

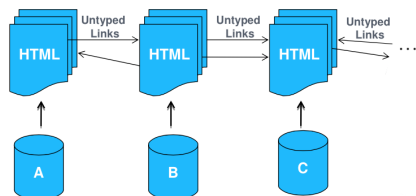


Constats :

- Un langage de balisage exclusivement pour la mise en forme
- Des liens non typés entre documents répartis (ou entre sous-parties d'un même document)

## Le Web : principes fondateurs

Phase 1 : Le Web de l'hypertexte

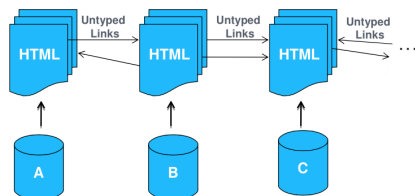


### Constats :

- Un langage de balisage escluisivement pour la mise en forme
- Des liens non typés entre documents répartis (ou entre sous-parties d'un même document)
- Degré de typage des éléments/données contenus dans les doc ?

## Le Web : principes fondateurs

Phase 1 : Le Web de l'hypertexte

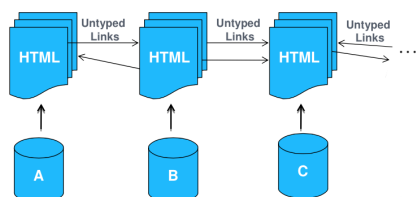


### Constats :

- Un langage de balisage escluisivement pour la mise en forme
- Des liens non typés entre documents répartis (ou entre sous-parties d'un même document)
- Degré de typage des éléments/données contenus dans les doc ? Très faible

## Le Web : principes fondateurs

Phase 1 : Le Web de l'hypertexte

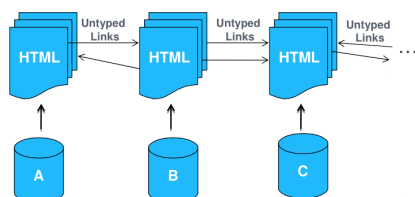


### Constats :

- Un langage de balisage escluisivement pour la mise en forme
- Des liens non typés entre documents répartis (ou entre sous-parties d'un même document)
- Degré de typage des éléments/données contenus dans les doc ? Très faible
- Il y a-t-il une sémantique explicite du contenu ?

## Le Web : principes fondateurs

Phase 1 : Le Web de l'hypertexte

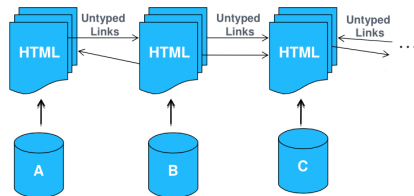


### Constats :

- Un langage de balisage escluisivement pour la mise en forme
- Des liens non typés entre documents répartis (ou entre sous-parties d'un même document)
- Degré de typage des éléments/données contenus dans les doc ? Très faible
- Il y a-t-il une sémantique explicite du contenu ? La sémantique est "implicite"

## Le Web : principes fondateurs

Phase 1 : Le Web de l'hypertexte

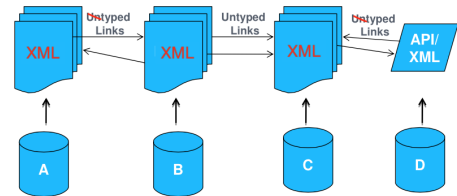


Constats :

- Un langage de balisage exclusivement pour la mise en forme
- Des liens non typés entre documents répartis (ou entre sous-parties d'un même document)
- Degré de typage des éléments/données contenus dans les doc ? Très faible
- Il y a-t-il une sémantique explicite du contenu ? La sémantique est "implicite"
- L'utilisateur principal visé ? Conçu pour être "consommé" principalement par l'humain

## Principes fondateurs

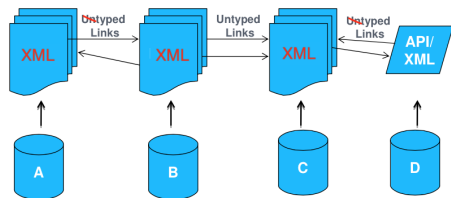
Phase 2 : Le Web dynamique



- Quel est le constituant principal ?

## Principes fondateurs

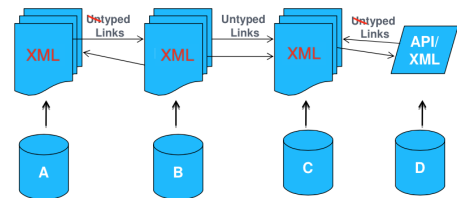
Phase 2 : Le Web dynamique



- Quel est le constituant principal ? des contenus dynamiques et des liens (non)typés entre documents (ou sous-partie)
- Degré de typage des éléments/données contenus ?

## Principes fondateurs

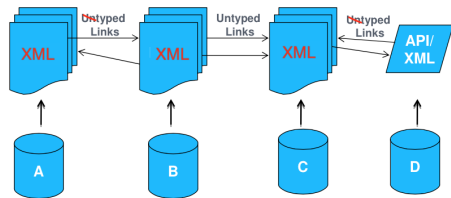
Phase 2 : Le Web dynamique



- Quel est le constituant principal ? des contenus dynamiques et des liens (non)typés entre documents (ou sous-partie)
- Typage des contenus possible mais non obligatoire

## Principes fondateurs

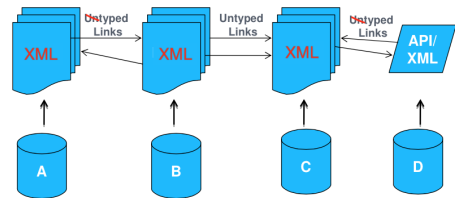
Phase 2 : Le Web dynamique



- **Quel est le constituant principal ?** des contenus dynamiques et des liens (non)typés entre documents (ou sous-partie)
- Typage des contenus possible mais non obligatoire
- **Il y a-t-il une sémantique formelle du contenu ?**

## Principes fondateurs

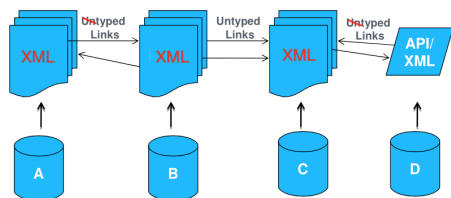
Phase 2 : Le Web dynamique



- **Quel est le constituant principal ?** des contenus dynamiques et des liens (non)typés entre documents (ou sous-partie)
- Typage des contenus possible mais non obligatoire
- Le langage de balisage permet de décrire formellement tout ou partie du contenu

## Principes fondateurs

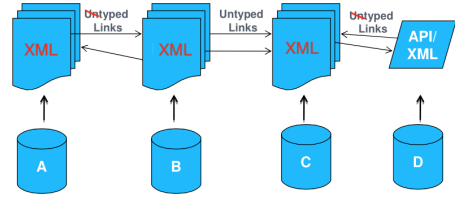
Phase 2 : Le Web dynamique



- **Quel est le constituant principal ?** des contenus dynamiques et des liens (non)typés entre documents (ou sous-partie)
- Typage des contenus possible mais non obligatoire
- Le langage de balisage permet de décrire formellement tout ou partie du contenu
- **Qui est l'utilisateur principal visé ?**

## Principes fondateurs

Phase 2 : Le Web dynamique



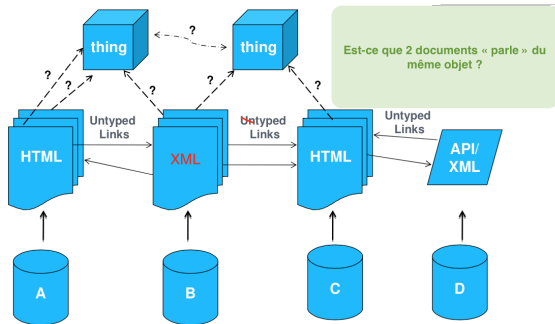
- **Quel est le constituant principal ?** des contenus dynamiques et des liens (non)typés entre documents (ou sous-partie)
- Typage des contenus possible mais non obligatoire
- Le langage de balisage permet de décrire formellement tout ou partie du contenu
- Conçu pour être "consommé" par l'humain et mais "traitable" par la machine

## Principes fondateurs

Phase 3 : Le Web sémantique

Composer des contenus à la volée et selon les trois notions (données, informations et connaissances).

Problème principal :



Mauro Gao (UPPA)

IDC

11 / 29

## Principes fondateurs

Exemple d'interrogation

Si on voulait obtenir des données à partir de cette question

Combien de personnes sont nées depuis 2015 dans les villes de moins de 10.000 habitants ....

?

Mauro Gao (UPPA)

IDC

12 / 29

## Principes fondateurs

Exemple d'interrogation

Si on voulait obtenir des données à partir de cette question

Combien de personnes sont nées depuis 2015 dans les villes de moins de 10.000 habitants ....

Il faudrait :

- Faire la liste de toutes les villes au monde,  
*Résultat* : une ou des listes avec plusieurs centaines de milliers d'éléments
- ne retenir alors que celles de moins de 10.000 habitants  
*Résultat* : une liste avec plusieurs dizaines de milliers d'éléments
- Pour chaque élément de la liste trouver l'info sur les naissances (de préférence par an)  
*Résultat* : un ensemble de données dans différents formats
- Enfin, intégrer toutes ces infos dans un format unique et cohérent...

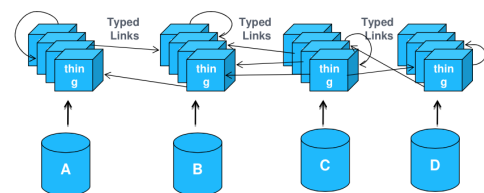
Mauro Gao (UPPA)

IDC

12 / 29

## Principes fondateurs

Phase 4 : le Web des données



- Quel est le constituant principal ?

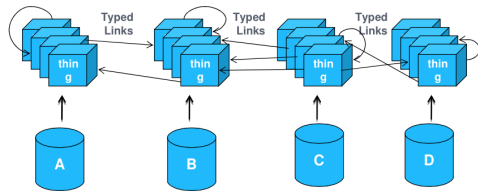
Mauro Gao (UPPA)

IDC

13 / 29

## Principes fondateurs

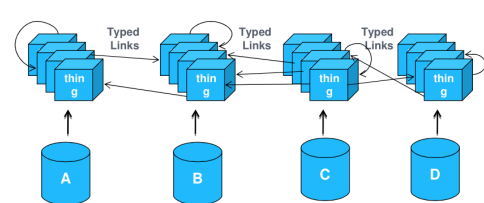
Phase 4 : le Web des données



- Le principal constituant : l'objet (ou plutôt ses descriptions) et des liens vers d'autres objets.  
Exemple : pour le niveau sémantique, un langage supplémentaire (RDF) permet de décrire des relations sémantiques telle que :  
`<personnage05> <membre> <empereur>`

## Principes fondateurs

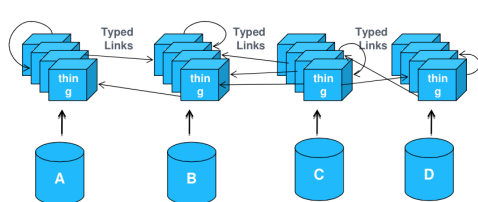
Phase 4 : le Web des données



- `<personnage05> <membre> <empereur>`
- Formalisation de la sémantique** : basée sur un modèle de donnée unique le triplet RDF:  
Sujet -- Prédicat -- Objet

## Principes fondateurs

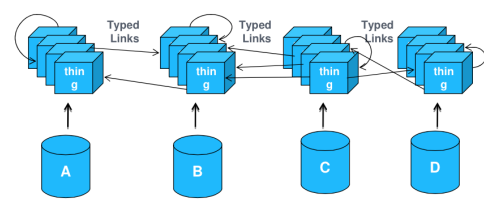
Phase 4 : le Web des données



- `<personnage05> <membre> <empereur>`
- basée sur un modèle de donnée unique le triplet RDF:  
Sujet -- Prédicat -- Objet
- Qui est l'utilisateur principal visé?

## Principes fondateurs

Phase 4 : le Web des données



- `<personnage05> <membre> <empereur>`
- basée sur un modèle de donnée unique le triplet RDF:  
Sujet -- Prédicat -- Objet
- Qui est l'utilisateur principal visé?  
Conçu pour la machine, mais reste lisible par l'humain



## Principes fondateurs

### Les contenus du Web :

- produits par le biais d'une application doivent pouvoir être traités par d'autres applications ;
- mais doivent rester compréhensibles par l'homme.

Cela implique ?

## Conséquences

Pour pouvoir être "traitée" indépendamment de l'application source, il est préférable de disposer :

- d'une représentation indépendante ;
- d'outils de validation génériques ;
- d'une version sérialisée des structures de données internes aux applications.

## Conséquences

Pour pouvoir être "traitée" indépendamment de l'application source, il est préférable de disposer :

- d'une représentation indépendante ;
- d'outils de validation génériques ;
- d'une version sérialisée des structures de données internes aux applications.

Pour faciliter la "compréhension", il est préférable de disposer d'une description explicite :

- "de la forme" des éléments constitutifs du contenu ;
- "de la manière" dont ces éléments s'organisent entre eux ;
- "des propriétés et des contraintes" de chaque élément ou de chaque combinaison d'éléments.

## Conséquences

En gros un langage où l'on souhaite :

- éviter la polysémie,
- c.-à-d. réduire les ambiguïtés ... voire les supprimer

## Conséquences

En gros un langage où l'on souhaite :

- éviter la polysémie,
- c.-à-d. réduire les ambiguïtés ... [voire les supprimer](#)

Autrement dit un *langage formel*

## Conséquences

En gros un langage où l'on souhaite :

- éviter la polysémie,
- c.-à-d. réduire les ambiguïtés ... [voire les supprimer](#)

Autrement dit un *langage formel*

Quelle définition pouvez vous donner à : *langage formel* ?

## Conséquences

En gros un langage où l'on souhaite :

- éviter la polysémie,
- c.-à-d. réduire les ambiguïtés ... [voire les supprimer](#)

Autrement dit un *langage formel*

Un *langage formel* est un ensemble de **mots** construits à partir d'un **alphabet**.

L'**alphabet** d'un langage formel est l'ensemble des **symboles** (lettres ou lexèmes) qui servent à construire les mots du langage.

Les mots sont donc des suites d'éléments de cet alphabet ;

les mots qui appartiennent à un langage formel particuliers sont appelés **mots bien formés**.

## Conséquences

En gros un langage où l'on souhaite :

- éviter la polysémie,
- c.-à-d. réduire les ambiguïtés ... [voire les supprimer](#)

Autrement dit un *langage formel*

Un *langage formel* est un ensemble de **mots** construits à partir d'un **alphabet**.

L'**alphabet** d'un langage formel est l'ensemble des **symboles** (lettres ou lexèmes) qui servent à construire les mots du langage.

Les mots sont donc des suites d'éléments de cet alphabet ;

les mots qui appartiennent à un langage formel particuliers sont appelés **mots bien formés**.

avec l'alphabet suivant a,b,c :

- 1 peut-on construire les mots suivants : **ab, abc**

## Conséquences

En gros un langage où l'on souhaite :

- éviter la polysémie,
- c.-à-d. réduire les ambiguïtés ... voire les supprimer

Autrement dit un *langage formel*

Un *langage formel* est un ensemble de **mots** construits à partir d'un **alphabet**.

L'**alphabet** d'un langage formel est l'ensemble des **symboles** (lettres ou lexèmes) qui servent à construire les mots du langage.

Les mots sont donc des suites d'éléments de cet alphabet ;

les mots qui appartiennent à un langage formel particuliers sont appelés **mots bien formés**.

avec l'alphabet suivant a,b,c :

- 1 peut-on construire les mots suivants : **ab, abc**
- 2 uniquement ces mots ?

## Conséquences

En gros un langage où l'on souhaite :

- éviter la polysémie,
- c.-à-d. réduire les ambiguïtés ... voire les supprimer

Autrement dit un *langage formel*

Un *langage formel* est un ensemble de **mots** construits à partir d'un **alphabet**.

L'**alphabet** d'un langage formel est l'ensemble des **symboles** (lettres ou lexèmes) qui servent à construire les mots du langage.

Les mots sont donc des suites d'éléments de cet alphabet ;

les mots qui appartiennent à un langage formel particuliers sont appelés **mots bien formés**.

avec l'alphabet suivant a,b,c :

- 1 peut-on construire les mots suivants : **ab, abc**
- 2 uniquement ces mots ?
- 3 pour ces mêmes mots, peut-on imaginer d'autres alphabets ?

## Conséquences

En gros un langage où l'on souhaite :

- éviter la polysémie,
- c.-à-d. réduire les ambiguïtés ... voire les supprimer

Autrement dit un *langage formel*

Un *langage formel* est un ensemble de **mots** construits à partir d'un **alphabet**.

L'**alphabet** d'un langage formel est l'ensemble des **symboles** (lettres ou lexèmes) qui servent à construire les mots du langage.

Les mots sont donc des suites d'éléments de cet alphabet ;

les mots qui appartiennent à un langage formel particuliers sont appelés **mots bien formés**.

avec l'alphabet suivant a,b,c :

- 1 peut-on construire les mots suivants : **ab, abc**
- 2 uniquement ces mots ?
- 3 pour ces mêmes mots, peut-on imaginer d'autres alphabets ?
- 4 que faut-il d'autre pour que à partir d'un alphabet on puisse produire que les mots souhaités ?

## Conséquences

En gros un langage où l'on souhaite :

- éviter la polysémie,
- c.-à-d. réduire les ambiguïtés ... voire les supprimer

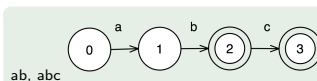
Autrement dit un *langage formel*

Un *langage formel* est un ensemble de **mots** construits à partir d'un **alphabet**.

L'**alphabet** d'un langage formel est l'ensemble des **symboles** (lettres ou lexèmes) qui servent à construire les mots du langage.

Les mots sont donc des suites d'éléments de cet alphabet ;

les mots qui appartiennent à un langage formel particuliers sont appelés **mots bien formés**.



Rappels ou pas...

D'un extrême à l'autre : du "formel" au "naturel"

#### Un langage informatique :

c'est un **langage formel** utilisé lors de la conception, la mise en œuvre, ou l'exploitation d'un *système de traitement automatique de l'information*.

Rappels ou pas...

D'un extrême à l'autre : du "formel" au "naturel"

#### Un langage informatique :

c'est un **langage formel** utilisé lors de la conception, la mise en œuvre, ou l'exploitation d'un *système de traitement automatique de l'information*.

Surtout ne pas le réduire à son sens le plus restrictif :  
**de langage de programmation**

Rappels ou pas...

#### Un langage de programmation

- le modèle de calcul : collection de valeurs et d'opérations (exécutions)

Rappels ou pas...

#### Un langage de programmation

- le modèle de calcul : collection de valeurs et d'opérations (exécutions)

Le calcul est l'application à une valeur – d'une séquence d'opérations – pour la rapporter à une autre valeur.

##### opérations :

- En général arithmétiques pour les valeurs primitives (caractères, nombre, booléens)
- Plus diversifiées pour les valeurs composées (type : tableaux, piles, arbres,...) :  
par exemple, opérations de construction ou d'accès aux valeurs composées de même type

Rappels ou pas...

#### Un langage de programmation

- le modèle de calcul : collection de valeurs et d'opérations (exécutions)
- la syntaxe et la sémantique

Rappels ou pas...

#### Un langage de programmation

- le modèle de calcul : collection de valeurs et d'opérations (exécutions)
- la syntaxe et la sémantique

- La **syntaxe** décrit la structure ou la forme des programmes.
- La **sémantique** décrit la relation entre un programme et le modèle de calcul.

Rappels ou pas...

#### Un langage de programmation

- le modèle de calcul : collection de valeurs et d'opérations (exécutions)
- la syntaxe et la sémantique
- les considérations pragmatiques qui forment le langage

Rappels ou pas...

#### Un langage de programmation

- le modèle de calcul : collection de valeurs et d'opérations (exécutions)
- la syntaxe et la sémantique
- les considérations pragmatiques qui forment le langage

La **pragmatique** décrit le degré de succès avec lequel le langage est conforme à ses buts.

Rappels ou pas...

#### Un langage de programmation

- le modèle de calcul : collection de valeurs et d'opérations (exécutions)
- la syntaxe et la sémantique
- les considérations pragmatiques qui forment le langage

Un programme ?

Rappels ou pas...

#### Un langage de programmation

- le modèle de calcul : collection de valeurs et d'opérations (exécutions)
- la syntaxe et la sémantique
- les considérations pragmatiques qui forment le langage

Un programme peut être considéré comme une fonction : les valeurs des données de sortie sont une fonction des valeurs des données d'entrée.

Rappels ou pas...

#### La langue naturelle (LN) :

Principal mode d'expression utilisé par les humains.  
Il y en a plusieurs *implémentations* : elles contiennent toutes de la *polysémie* et donc elles peuvent contenir des ambiguïtés ?

Rappels ou pas...

#### La langue naturelle (LN) :

Principal mode d'expression utilisé par les humains.  
Il y en a plusieurs *implémentations* : elles contiennent toutes de la *polysémie* et donc elles peuvent contenir des ambiguïtés ?

Donnez moi des exemples de mots ou compositions de mots polysémiques/ambigus en LN ?

Rappels ou pas...

#### La langue naturelle (LN) :

Principal mode d'expression utilisé par les humains.  
Il y en a plusieurs *implémentations* : elles contiennent toutes de la *polysémie* et donc elles peuvent contenir des ambiguïtés ?

Donnez moi des exemples de mots ou compositions de mots polysémiques/ambigus en LN ?  
avocat, grue, proche, est, grève, courses ...

Rappels ou pas...

#### La langue naturelle (LN) :

Principal mode d'expression utilisé par les humains.  
Il y en a plusieurs *implémentations* : elles contiennent toutes de la *polysémie* et donc elles peuvent contenir des ambiguïtés ?

Donnez moi des exemples de mots ou compositions de mots polysémiques/ambigus en LN ?  
avocat, grue, proche, est, grève, courses ...

Comment ces mots peuvent-ils être désambiguïsés ?

Rappels ou pas...

#### La langue naturelle (LN) :

Principal mode d'expression utilisé par les humains.  
Il y en a plusieurs *implémentations* : elles contiennent toutes de la *polysémie* et donc elles peuvent contenir des ambiguïtés ?

Donnez moi des exemples de mots ou compositions de mots polysémiques/ambigus en LN ?  
avocat, grue, proche, est, grève, courses ...

Comment ces mots peuvent-ils être désambiguïsés ?

*L'avocat est un fruit ou un légume ?  
L'installation de la grue sur le chantier est enfin terminée.  
J'allais volontiers m'asseoir au bord du lac, sur la grève.  
Je ramène les courses à la maison ce soir.*

Rappels ou pas...

#### La langue naturelle (LN) :

Principal mode d'expression utilisé par les humains.  
Il y en a plusieurs *implémentations* : elles contiennent toutes de la *polysémie* et donc elles peuvent contenir des ambiguïtés ?

Donnez moi des exemples de mots ou compositions de mots polysémiques/ambigus en LN ?

*L'avocat est un fruit ou un légume ?  
L'installation de la grue sur le chantier est enfin terminée.  
J'allais volontiers m'asseoir au bord du lac, sur la grève.  
Je ramène les courses à la maison ce soir.*

#### Le contexte d'évocation

*Oui mais... suffit-il, que fait-on avec ?  
... Tu as mangé du lion ce matin. Hier, je lui ai fait graver des sommets. Ne brûler pas les étapes.  
Dès le début, ça m'a crevé les yeux. ....*

Rappels ou pas...

#### La langue naturelle (LN) :

Principal mode d'expression utilisé par les humains.

Il y en a plusieurs *implémentations* : elles contiennent toutes de la *polysémie* et donc elles *peuvent contenir des ambiguïtés* ?

Donnez moi des exemples de mots ou compositions de mots polysémiques/ambigus en LN ?

*L'avocat est un fruit ou un légume ?*

*L'installation de la grue sur le chantier est enfin terminée.*

*J'allais volontiers m'asseoir au bord du lac, sur la grève.*

*Je ramène les courses à la maison ce soir.*

*Le contexte d'évocation*

*Oui mais... suffit-il, que fait-on avec ?*

*... Tu as mangé du lion ce matin. Hier, je lui ai fait gravir des sommets. Ne brûler pas les étapes.*

*Dès le début, ça m'a crevé les yeux ....*

*Quel serait le contexte à considérer pour désambigüiser ?*

Rappels ou pas...

#### La langue naturelle (LN) :

Principal mode d'expression utilisé par les humains.

Il y en a plusieurs *implémentations* : elles contiennent toutes de la *polysémie* et donc elles *peuvent contenir des ambiguïtés* ?

- Peut-on faire correspondre aux expressions en LN un (ou des) langage(s) formel(s) afin de les rendre non-ambiguës ?