

Exploitation d'un corpus arboré pour non spécialistes

Achille Falaise (LIG-GETALP), Agnès Tutin (LIDILEM), Olivier Kraif (LIDILEM)

Situation

- Des corpus...
 - Annotés : lemmes, parties du discours, flexion
 - Arborés (*treebanks*) : relations syntaxiques entre mots
- Des outils...
 - Peu nombreux (pour les corpus arborés)
 - Complexes pour des non-informaticiens

```
<cat=V,#1> && <lemma=hypothèse,#2> :: (OBJ,#2,#1)
```

Exemple : langage de requête de *ConcQuest*

Scientext

- Un environnement en ligne pour non-informaticiens
- Développé au-dessus du moteur de recherche ConcQuest
- 3 corpus de textes librement consultables
 - Publications scientifiques : français (5M mots) et anglais (14M mots)
 - Mémoires d'apprenants (anglais, 1M mots)

Exploitation d'un corpus arboré pour non spécialistes

Achille Falaise (LIG-GETALP), Agnès Tutin (LIDILEM), Olivier Kraif (LIDILEM)

<http://scientext.msh-alpes.fr>

The screenshot shows the search interface of the scientext.msh-alpes.fr website. On the left is a dark red sidebar with navigation links: Accueil, Choix des textes (1. Types de textes, 2. Liste de textes), Recherche dans les textes (3. Mode de recherche, 4. Recherche), Résultats (5. Concordancier, 6. Statistiques), Mode d'emploi (with a magnifying glass icon), and Compte utilisateur (Connexion). The main content area has a dark red header with the title '4. Recherche' and a 'Recherche' button. Below the header, there is a search form with a 'Mots:' label, a 'Mot 1' dropdown menu, a 'Forme' dropdown menu, and an empty text input field. A '+' button is next to the input field. Below the form, there is another 'Recherche' button and a text label: 'Interrompt arbitrairement la recherche à environ 20 occurrences.' with a dropdown menu set to '20'.

Exploitation d'un corpus arboré pour non spécialistes

Achille Falaise (LIG-GETALP), Agnès Tutin (LIDILEM), Olivier Kraif (LIDILEM)

<http://scientext.msh-alpes.fr>

The screenshot shows the search interface of the scientext.msh-alpes.fr website. The interface is divided into a left sidebar and a main content area. The sidebar contains navigation links: Accueil, Choix des textes (1. Types de textes, 2. Liste de textes), Recherche dans les textes (3. Mode de recherche, 4. Recherche), Résultats (5. Concordancier, 6. Statistiques), and Compte utilisateur (Mode d'emploi, Connexion). The main content area is titled '4. Recherche' and includes a search bar with 'Mot 1' (Catégorie: Verbe (V), Tous) and 'Mot 2' (Lemme: hypothèse). It also shows 'Relations syntaxiques' with 'Relation 1' (Mot 2: objet direct de (OBJ), Mot 1). A warning message states 'Attention, l'ordre des mots n'est plus pris en compte'. The search results show 20 occurrences, with the first four listed in a table.

▼ N°	▼ Contexte gauche: 10 mots	▼ Occurrence:	▼ Contexte droit: 10 mots	▼ Réf. texte
<input checked="" type="checkbox"/> 1	Afin de	valider nos hypothèses	et d' ouvrir de nouvelles perspectives , l' autre objectif	[tal-the-21-body]
<input checked="" type="checkbox"/> 2	voulons en réaliser une maquette afin , premièrement , de	vérifier nos hypothèses	, deuxièmement , de mettre à jour les difficultés inhérentes à chaque	[tal-the-21-body]
<input checked="" type="checkbox"/> 3	Pour choisir entre les différents ensembles , on	utilise les hypothèses	suivantes :	[tal-the-21-body]
<input checked="" type="checkbox"/> 4	de rapports sur les activités économiques d' un pays	confirme les hypothèses	prises pour la conception générale du générateur noyau et permet	[tal-the-21-body]