

## Manipulations

1. Sur le site des ressources RTE, télécharger un jeu de données : on choisira le *development set* de la campagne RTE3 (pour les campagnes plus récentes, les données sont accessibles, mais il faut demander un accord et signer un engagement).  
[http://aclweb.org/aclwiki/index.php?title=Textual\\_Entailment\\_Resource\\_Pool](http://aclweb.org/aclwiki/index.php?title=Textual_Entailment_Resource_Pool)
2. Ecrire un premier script pour réaliser une baseline : on va mesurer pour chaque paire le recouvrement lexical en se basant uniquement sur les formes présentes (pas de lemmatisation). Sur le jeu de développement, calculer le seuil  $\alpha$  tel que en répondant YES chaque fois que le recouvrement est  $> \alpha$ , et NO sinon, on maximise le score. Faire ensuite tourner cette baseline sur le *test set*, et noter la performance.
3. Version 2 de la baseline : réaliser la même mesure après une lemmatisation des paires.
4. Version 3 de la baseline : remplacer le recouvrement lexical strict par une mesure de similarité lexicale (par exemple similarité par chemin dans WordNet, ou celle de Resnik).
5. Faire une analyse d'erreur : proposer, sur la base d'une étude manuelle d'un nombre significatif d'erreurs dans le *test set* ( $\approx 40$ ), des heuristiques qui, sans passer par une analyse syntaxique et une représentation sémantique, devraient permettre d'améliorer sensiblement la baseline.

On demande le code des trois baselines, et un fichier texte (ou pdf) pour la discussion et l'analyse d'erreurs.

Exceptionnellement, ce TP peut être réalisé par groupes de 2 personnes au plus.