

## Table des matières

<b>Chapitre 1. Questions-réponses interlingues</b> . . . . .	11
Anne-Laure LIGOZAT , Isabelle ROBBA	
1.1. Les systèmes de questions-réponses du LIMSI . . . . .	11
1.2. Architecture du système QALC . . . . .	12
1.2.1. Analyse de la question . . . . .	12
1.2.2. Sélection des documents . . . . .	13
1.2.3. Analyse des documents . . . . .	14
1.2.4. Sélection de la réponse . . . . .	16
1.3. Le système monolingue français FRASQUES . . . . .	18
1.4. Le système interlingue MUSCLEF . . . . .	19
1.5. Performances des systèmes . . . . .	20
1.5.1. Système monolingue anglais QALC . . . . .	22
1.5.2. Système interlingue français vers anglais MUSCLEF . . . . .	23
1.6. Conclusion . . . . .	25
1.7. Bibliographie . . . . .	25



## Chapitre 1

# Questions-réponses interlingues

### 1.1. Les systèmes de questions-réponses du LIMSI

Ce chapitre a pour objet la présentation d'une architecture particulière de réponse aux questions, celle des systèmes du groupe LIR du LIMSI. Le premier système développé dans ce groupe a été QALC, qui travaille sur l'anglais. Sa première version a participé à l'évaluation TREC-8 en 1999 et a constitué la base des systèmes suivants.

Afin d'extraire l'information attendue de la collection de documents, un système de questions-réponses doit analyser les documents et apparier questions et réponses. L'extraction de la réponse peut être réalisée selon différentes techniques, s'appuyant principalement sur l'annotation en entités nommées des documents et leur analyse syntaxique ; l'appariement questions-réponses utilise alors soit des représentations des phrases entières ou même des documents, soit des informations locales sur la réponse. La première solution nécessite de larges bases de connaissances et/ou des analyseurs syntaxiques très fiables. L'approche que nous proposons est plutôt fondée sur des informations syntaxiques locales, utilisant l'analyse morphosyntaxique des documents.

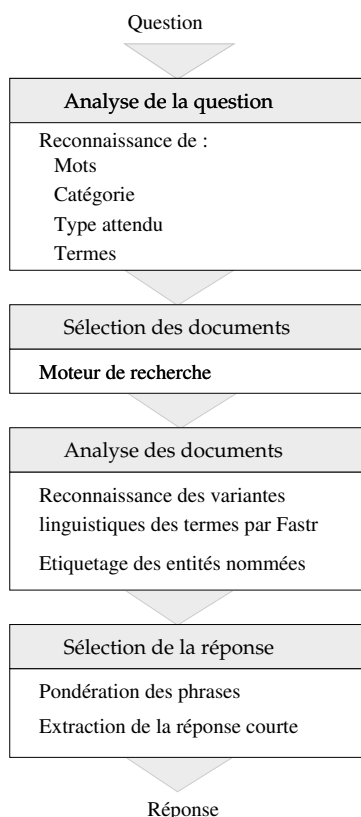
Nous présenterons d'abord ce système monolingue anglais QALC, puis nous donnerons une description des différents systèmes pour le français et des systèmes interlingues, en précisant leurs particularités.

---

Chapitre rédigé par Anne-Laure LIGOZAT et Isabelle ROBBA.

## 1.2. Architecture du système QALC

L'architecture du système QALC est présentée figure 1.1 (pour plus de détails sur le fonctionnement du système, voir par exemple [FER 02]).



**Figure 1.1.** Architecture du système QALC

### 1.2.1. Analyse de la question

Tout d'abord, l'analyse de la question détermine un certain nombre de caractéristiques de la question qui seront utilisées dans la suite de la chaîne, dont en particulier :

– **le type attendu de la réponse**, sous la forme d'un type d'entité nommée ou d'un type général. Ainsi, pour la question «*Name a city in which Giulio Andreotti was put on trial.*», le type attendu est une entité nommée de type *ville*, soit *LOCATION-CITY* pour le système ; pour la question «*Which contract runs from 1995 to 2004 ?*», le type

attendu est «*contract*». L'ensemble des types d'entités nommées sera présenté dans le paragraphe 1.2.2 ;

– **la catégorie de la question.** «*What is the OSCE ?*» par exemple est une question de définition. Ces catégories seront utilisées pour extraire les réponses courtes des phrases comme nous le verrons dans le paragraphe 1.2.4 ;

– **les entités nommées de la question.** La date «*1992*» est ainsi repérée dans la question «*How big was the Japanese share of the American car market in 1992 ?*», et sera considérée comme le contexte temporel de l'information cherchée ;

– **le focus de la question,** c'est-à-dire l'objet sur lequel porte la question, et qui sera probablement présent dans la phrase réponse. Dans la question «*When was Cyprus divided into two parts ?*», le focus sera «*Cyprus*».

Afin de pouvoir déterminer ces caractéristiques, la question est analysée par un étiqueteur morpho-syntaxique, le TreeTagger <sup>1</sup>, puis une segmentation des questions en constituants est assurée par l'analyseur syntaxique Cass <sup>2</sup>.

### 1.2.2. Sélection des documents

Le moteur de recherche MG <sup>3</sup> effectue ensuite une sélection de documents, en recherchant les mots de la question, ou des synonymes. La collection de documents est préalablement indexée par MG en utilisant sa fonction de stemming <sup>4</sup>, et est également découpée en paragraphes, afin que les textes en sortie du moteur de recherche soient relativement courts. MG peut travailler avec des requêtes booléennes ou vectorielles, et cette caractéristique a été exploitée : des requêtes booléennes sont d'abord construites, en utilisant les caractéristiques de la question, et leurs synonymes ; si trop peu de documents sont retournés, une requête vectorielle est utilisée. Le nombre de documents sélectionnés est un paramètre de la chaîne. Dans nos évaluations, il est fixé à 100 expérimentalement : ce nombre relativement faible offre un bon compromis entre le temps de recherche des documents et les performances de sélection des documents.

1. Étiqueteur morphosyntaxique probabiliste, <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>

2. Analyseur syntaxique robuste, [ABN 96]

3. Managing Gigabytes, <http://www.cs.mu.oz.au/mg/>

4. Le stemming ou racinisation est un processus qui consiste à retrouver la racine d'un mot par troncature, permettant de relier des mots ayant la même racine comme «connect» et «connection».

### 1.2.3. Analyse des documents

La forme des phrases réponses pouvant être très éloignée de la question, il est nécessaire de gérer la variation linguistique ; c'est pourquoi les documents retournés sont analysés par Fastr [JAC 96]. Fastr est un outil permettant de reconnaître des variantes d'expressions contenant un ou plusieurs mots, à partir de lexiques de variantes morphologiques et sémantiques, et de règles de reformulation syntaxique. Ainsi, l'expression «*Europe's sole currency*» sera reconnue comme une variante de «*single European currency*» car «*Europe*» et «*European*» sont des variantes morphologiques, «*sole*» et «*single*» des variantes sémantiques, et que l'adjectif «*European*» peut se transformer en possessif du nom «*Europe*». La règle «NtoA» ci-dessous montre le passage d'un nom à un adjectif si les deux mots ont la même racine : «*categorial grammar*» sera reconnu comme une variante de «*category grammar*».

EXEMPLE.– Règle de reformulation de Fastr

```
Metarule NtoA( X1 -> N2 N3 ) =
      X1 -> A4 < {A|N|Np|V} 0-3 > N3:
  <N2 root> = <A4 root>
```

La première ligne de cette règle indique qu'il s'agit d'une règle prenant en entrée un groupe formé de deux noms, et la seconde que ce groupe de mots a pour variante possible un adjectif suivi d'un nom, séparés de 0 à 3 mots dont les catégories sont données («A|N|Np|V»). La dernière ligne précise que le premier nom et le nouvel adjectif doivent avoir la même racine. À chaque variante ainsi reconnue, Fastr attribue un poids d'autant plus fort que la variante est proche du terme d'origine : ainsi, une variante morphologique aura un poids plus fort qu'une variante sémantique. Dans le système QALC, Fastr est utilisé pour reconnaître les mots (ou expressions) de la question ou leurs variantes présents dans les documents. Un certain nombre de documents (dans nos évaluations, ce nombre est de 70) sont alors sélectionnés et réordonnés, en fonction des variantes reconnues des mots de la question.

Ces documents sont ensuite découpés en phrases, et seules les phrases contenant au moins un mot de la question ou une variante d'un mot sont conservées. Ils sont également examinés afin d'étiqueter les différentes entités nommées qui s'y trouvent grâce à des modules spécifiques dont la description peut être trouvée dans [ELK 03]. Les types d'entités nommées reconnus sont présentés dans la figure 1.2.

Un premier module en Java est utilisé pour reconnaître les noms propres comme les personnes, organisations, pays ou villes ; un second module en Perl reconnaît les entités numériques comme les années, périodes, heures, dates, mesures (longueur, volume, poids). Le premier module se fonde sur les informations lexicales, les contraintes linguistiques et des informations contextuelles pour détecter les entités qui correspondent

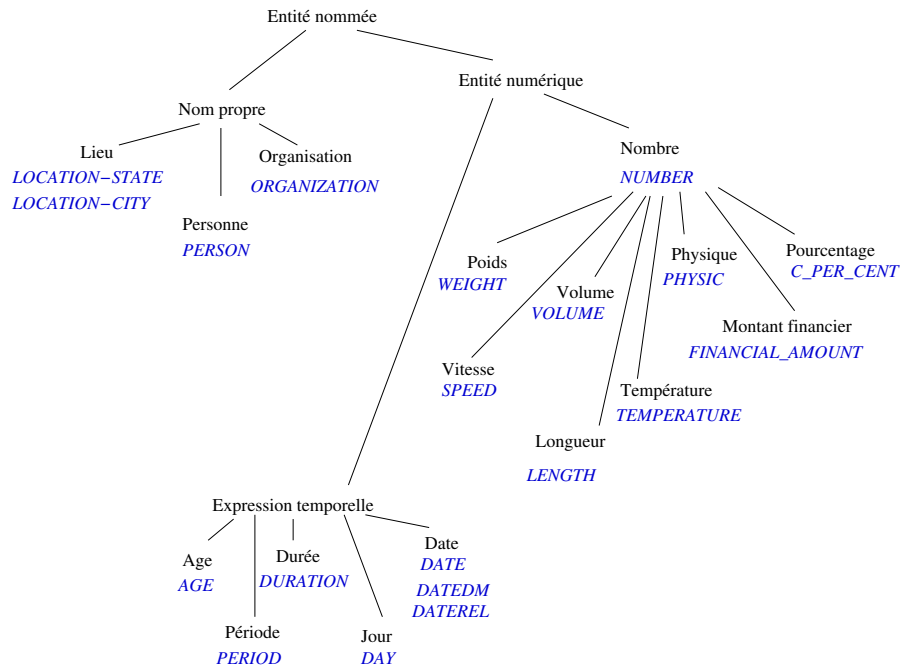


Figure 1.2. Hiérarchie des entités nommées

à des règles particulières. Pour l’anglais, il existe sept règles pour la reconnaissance des personnes, neuf pour les lieux, et sept pour les organisations. Ces règles utilisent principalement des dictionnaires et des listes de déclencheurs : ainsi, un nom propre ou un nom commençant par une majuscule précédé des déclencheurs «*judge*» ou «*ministre*» est considéré comme un nom de personne et étiqueté *PERSON* ; un dictionnaire contenant une liste de noms de villes permet de reconnaître «*Bordeaux*» comme une ville et de l’étiqueter *LOCATION-CITY*.

Le second module contient des règles de réécriture qui identifient les entités numériques à différents niveaux. Les règles sont beaucoup plus nombreuses dans ce module (plusieurs centaines). Un exemple de règle (simplifiée) permettant de reconnaître l’expression «*In (the) 70s*» comme une date est donné ci-dessous : dans un premier temps, des variables sont spécifiées comme les prépositions introduisant des dates, les décennies ou les espaces, puis une expression régulière définit une possibilité d’écriture d’une date.

EXEMPLE.– Règle de reconnaissance d’une date

```
$specifierDate = '(?:in|In|on|On|since|Since|During(...))';
$decad = "(?:10s|20s|30s|40s|50s|60s|70s|80s|90s|00s|(...))";
```

```

$card_num_entier_decad$s)";
$space = '\s';

s/\b(?:$specifierDate ){1,2})(the){0,1} ($decad)($space)/
<b_timex_TYPE="DATE">$1$2 $3<e_timex>$4/gi;

```

Ensuite, les phrases restantes sont pondérées à partir de la présence de mots de la question ou variantes, entités nommées du type attendu, et de la densité des mots. Les réponses courtes possibles sont enfin extraites de ces phrases, en fonction des mots de la question reconnus et de leur distance.

#### 1.2.4. Sélection de la réponse

Pour extraire les réponses courtes, deux stratégies différentes sont appliquées, en fonction du type attendu de la réponse. Si la question attend une entité nommée en réponse, on choisira, dans la phrase candidate, l'entité nommée la plus proche du barycentre des variantes des mots de la question, pondérées par leur poids Fastr. Ainsi pour la question «*Where is the Brandenburg Gate ?*», une phrase réponse est «*This weekend, while we are remembering the end of war in Europe 50 years ago and the demise of Adolf Hitler, let me tell you about a Scottish-based company which, using technology developed in Edinburgh, has been surveying potential secrets still buried in the earth below Berlin's Brandenburg Gate, close to Hitler's wartime bunker.*». Dans cette phrase, «*Edinburgh*» et «*Berlin*» sont étiquetées comme des entités nommées représentant un lieu, et sont donc susceptibles de répondre à la question. L'entité «*Berlin*» sera néanmoins choisie comme réponse car elle est la plus proche des mots de la question «*Brandenburg Gate*».

Dans le cas où la réponse attendue n'est pas une entité nommée, des patrons d'extraction sont appliqués. Une grammaire au format Cass a été écrite, qui permet d'étiqueter des constituants comme réponse s'ils sont entourés de constituants comprenant une caractéristique de la question, et séparés de ceux-ci par des séparateurs prédéfinis. Par exemple, pour la question «*Who is Nick Leeson ?*», la réponse «*Barings' former derivatives trader in Singapore*» est extraite de la phrase réponse «*Mr Norris has come under focus as investigators try to determine how Nick Leeson, Barings' former derivatives trader in Singapore, got hundreds of millions of pounds transferred from London to Singapore to help cover his trading losses.*» : la réponse et le focus «*Nick Leeson*» sont accolés et séparés par une virgule (en réalité, apposés) dans la segmentation en constituants effectuée par la grammaire Cass [LIG 06b].

Les règles de cette grammaire sont organisées par niveaux ; un premier niveau permet de construire les groupes syntaxiques. Dans la figure 1.3, les trois règles du niveau 1 construisent le groupe nominal contenant le focus, celui contenant le nom



**Niveau 1 : Constitution des groupes**  
 GNFOC -> DT ? RB ? (ADJ (CC ADJ) ?) ? FC RB ? ;  
 GN -> DT ? RB\* ADJ\* (NN|NNS)+ RB\* ADJ\* ;  
 GNNP -> DT ? RB\* ADJ\* (NP|NPS)+ ;  
**Niveau 2 : Marquage des réponses**  
 RgFsep -> b= (GN|GNNP) SEP GNFOC ;

**Figure 1.3.** Exemples de règles pour l'analyseur Cass

commun et celui contenant le nom propre. Ensuite la règle de niveau 2 permet de marquer comme réponse finale candidate tout groupe nominal (GN) qui serait à gauche du GN contenant le focus et qui en serait séparé à l'aide d'un séparateur quelconque (la virgule, le tiret... notés «SEP»).

Grâce à ces quatre règles, la réponse correcte «*oncogenes*» à la question «*Which genes cause cancer ?*», est marquée dans la phrase candidate : «*G proteins have assumed critical importance in recent years as researchers have discovered that several of them are produced by oncogenes, genes that cause cancer.*». Le module final d'extraction retourne le groupe marqué de l'étiquette *b=* :

EXEMPLE.– Réponse reconnue par la grammaire Cass

[RgFsep *b*=[GN [NNS *oncogenes*]] [VIRG ,] [GNFOC [FC *genes*]]]

Ainsi au lieu d'être utilisé pour analyser les phrases candidates, Cass est utilisé pour localiser puis étiqueter les réponses courtes à l'intérieur de ces phrases candidates.

Cass prend en entrée 2 éléments : la phrase à traiter et un fichier de termes étiquetés qui proviennent de l'analyse de la question. Ces termes particuliers qui sont utiles à Cass sont le focus (marqué FC dans les règles données en exemple ci-dessus), le verbe principal et le type général de la réponse. Les règles Cass peuvent alors exprimer la proximité de la réponse par rapport à ces termes particuliers.

L'analyse de la question associe également à chaque question une catégorie. Là encore, cette information est utilisée par le module d'extraction de la réponse : à chaque catégorie correspond un ensemble de patrons différents. En effet, selon ces différentes catégories, la réponse pourra être positionnée de façon différente par rapport au focus ou au type général.

Cinq catégories ont été établies par le module d'analyse des questions :

– définition : la réponse doit être une définition de l'objet ou de la personne contenue dans la question («*What is Eurostat ?*», «*Who is Javier Solana ?*»);

- acronyme : la réponse doit définir l’acronyme contenu dans la question («*What is the SLPN ?*»);
- instance : la réponse doit être une instance du focus («*Name a German car producer.*» «*Which space probe set off for the Moon on 25 January 1994 ?*»);
- combien : la réponse est une entité numérique accolée au focus («*How many countries joined the international coalition to restore democratic government in Haiti ?*»);
- autre : quand aucune des catégories précédentes n’a été reconnue

Les résultats de cette étape de sélection de la réponse dépendent donc en partie des résultats de l’étape d’analyse de la question.

Enfin, ces réponses peuvent être comparées à celles obtenues en faisant passer un système identique sur le Web [BER 03]. Cette fusion de réponses repose sur l’hypothèse que la redondance d’une réponse est un indice de sa validité : si une même réponse est trouvée à la fois dans la collection de documents et sur le Web, son score est augmenté.

### 1.3. Le système monolingue français FRASQUES

Le système QALC a été adapté au français à l’occasion de la campagne d’évaluation de systèmes de questions-réponses sur le français EQueR [GRA 06c]. Pour chaque module de la chaîne, la question de l’adaptation au français s’est donc posée. Les outils sur lesquels repose l’analyse des questions (étiqueteur morphosyntaxique et analyseur syntaxique) ont ainsi été modifiés. Les sorties de ces outils ont été projetées sur des formats communs en français et en anglais, afin que le module d’analyse des questions puisse être le même dans les deux langues. Ainsi le cœur de ce module est similaire, mais des fichiers de lexiques et de règles séparés ont été écrits.

Le moteur de recherche MG a laissé la place au moteur de recherche booléen Lucene<sup>5</sup>, qui, contrairement à MG, n’est pas perturbé par la présence d’accents dans les documents. Lucene ne disposant pas de fonctionnalité de stemming, les collections françaises sont étiquetées en catégories morpho-syntaxiques et lemmes par l’analyseur XIP [AIT 02] combiné au TreeTagger français, avant leur indexation. Seules des requêtes booléennes sont construites, ce qui entraîne une baisse des performances de l’étape de recherche documentaire.

Les règles et lexiques de Fastr [JAC 96] et du module de reconnaissance des entités nommées ont été adaptées au français, ainsi que les grammaires Cass comprenant les patrons d’extraction des réponses.

---

5. <http://lucene.apache.org>

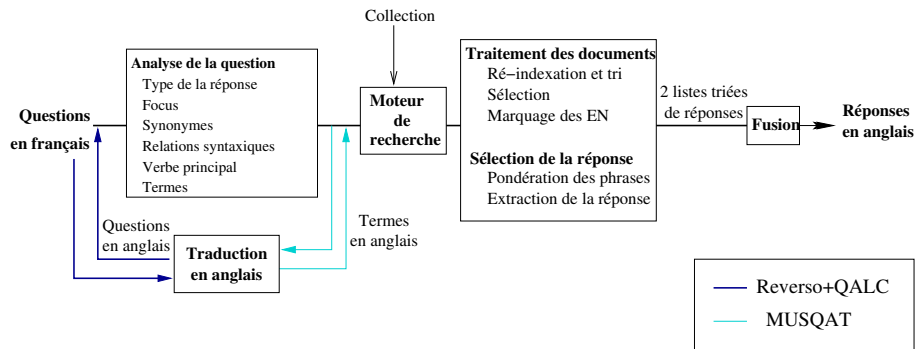


Figure 1.4. Architecture du système MUSCLEF

#### 1.4. Le système interlingue MUSCLEF

Depuis 2005, un système interlingue a également été développé, et a participé aux campagnes d'évaluation QA@CLEF en 2005 et 2006 [LIG 06a, GRA 06a]. Ce système, MUSCLEF, prend en entrée des questions en français, et recherche leurs réponses dans des documents en anglais. Il correspond en réalité à deux sous-systèmes, utilisant deux stratégies parallèles pour passer d'une langue à l'autre <sup>6</sup> :

- la **traduction de la question** par un traducteur automatique, en l'occurrence Reverso <sup>7</sup>, qui est ensuite passée en entrée du système monolingue anglais QALC. Ce système sera désormais appelé Reverso+QALC ;
- et la **traduction des termes de la question**, qui est implémentée dans un système particulier appelé MUSQAT.

Ces deux stratégies sont présentées figure 1.4. Leurs résultats sont ensuite combinés, et les réponses résultant de cette fusion sont celles de MUSCLEF.

La première stratégie, celle de Reverso+QALC, fait appel au traducteur automatique Reverso qui présente l'avantage de fournir une seule traduction, généralement assez fiable, des mots de la question. Par contre, la grammaticalité ou le réalisme de la question ne sont pas assurés, et par conséquent, l'analyse de la question risque d'être perturbée. Cependant, cette stratégie donne les meilleurs résultats.

La seconde stratégie exploitée par le système MUSQAT analyse la question en français, puis traduit chaque terme de la question en recherchant leurs traductions

6. Les systèmes participant à QA@CLEF implémentent généralement l'une ou l'autre des stratégies. On pourra trouver des descriptions des systèmes participants à l'adresse <http://www.clef-campaign.org/>.

7. <http://www.reverso.net>

dans des dictionnaires. Les dictionnaires utilisés sont Magic-Dic <sup>8</sup> et FreeDict <sup>9</sup>, qui sont sous licence GPL. Un des intérêts de Magic-Dic est son évolutivité, puisque ses utilisateurs peuvent soumettre de nouvelles traductions qui sont vérifiées avant intégration éventuelle. L'ajout de FreeDict à Magic-Dic dans le système MUSQAT a permis de combler les lacunes de Magic-Dic : sur le corpus de questions de CLEF 2005, FreeDict a ajouté 424 traductions aux 690 termes déjà traduits par Magic-Dic.

En plus des traductions mot à mot, des expressions formées de plusieurs mots (qui seront appelées *multitermes*) sont également traduites. Dans les questions, de telles expressions sont recherchées en fonction de l'étiquetage des mots : ainsi, dans la question «*Combien d'États sont membres du Conseil de l'Europe ?*», «*Conseil de l'Europe*» est considéré comme une expression multiterme grâce à un patron «GN Prep GN». Ces multitermes sont d'abord traduits terme à terme. Puis, ils sont validés dans la collection : chaque traduction possible du multiterme est cherchée dans la collection, et les traductions trouvées sont ensuite privilégiées. Ainsi, pour le multiterme français «*Conseil de l'Europe*», chaque mot est traduit grâce aux dictionnaires : «*conseil*» <sup>10</sup> a comme traductions les termes anglais «*advice*», «*consel*» et «*council*» ; «*Europe*» a pour traductions «*Europe*» et «*Europa*». Le seul multiterme trouvé dans le corpus combinant ces différentes traductions est «*Council of Europe*», qui sera gardé comme traduction. «*council*» et «*Europe*» seront également conservés comme uniques traductions dans ce contexte de «*conseil*» et «*Europe*».

L'avantage de cette seconde stratégie est que l'analyse des questions est plus fiable car elle travaille sur des questions non traduites, et par conséquent correctement exprimées. Cependant, plusieurs traductions sont alors générées pour chaque mot de la question, qui ne prennent pas en compte le contexte de la question, et les traductions des mots comprennent beaucoup de bruit. La validation des multitermes dans le corpus a pour but de valider ces traductions monoterme, mais peu de multitermes sont en réalité validés. La raison principale de ce faible taux de validation semble être le manque de traductions monoterme, mais l'analyse des résultats doit encore être approfondie.

### 1.5. Performances des systèmes

Trois systèmes ont ainsi été construits au fil du temps : deux systèmes monolingues, QALC pour l'anglais et FRASQUES pour le français, et un système interlingue français vers anglais, MUSCLEF. Le tableau 1.5 résume le fonctionnement de chaque système.

8. <http://magic-dic.homeunix.net>

9. <http://www.freedict.com>

10. «*Conseil*» est passé en minuscules car il est considéré comme un nom commun par l'étiqueteur.

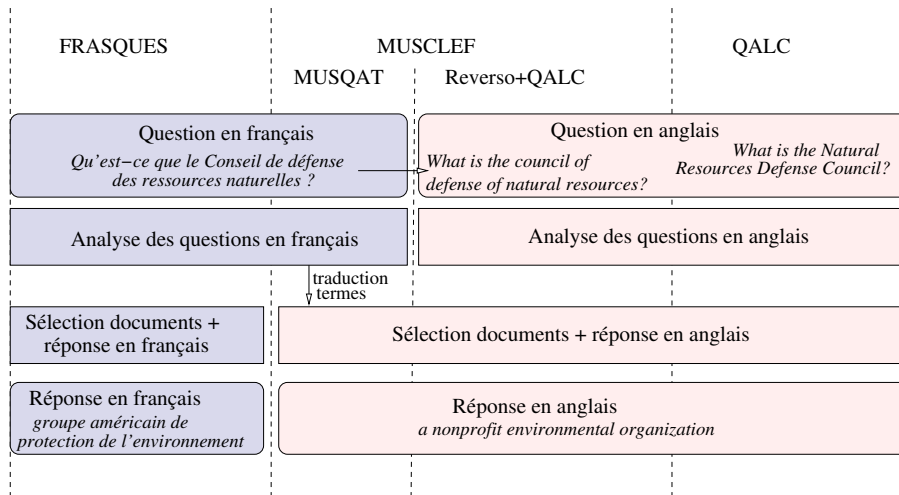


Figure 1.5. Exemples de traitement d'une question par les différents systèmes

Les systèmes présentés sont en constante évolution, et afin de pouvoir évaluer les améliorations qui leur sont apportées, il est nécessaire de disposer de moyens d'évaluation automatique. En effet, d'une part, l'évaluation manuelle des réponses est longue et fastidieuse, et d'autre part, il est important de pouvoir évaluer chaque module des systèmes, afin de cibler les améliorations à leur apporter. Pour pouvoir évaluer automatiquement, il faut avoir un ensemble aussi exhaustif que possible de réponses aux questions.

Notre groupe dispose notamment du corpus de questions et de documents de la tâche français vers anglais de l'évaluation CLEF 2005. Pour cette tâche, un fichier de patrons de réponses (c'est-à-dire un ensemble de réponses possibles pour chaque question) a été constitué, en rassemblant les différentes réponses trouvées par les systèmes ayant participé à cette tâche, ainsi que les nouvelles réponses trouvées par les systèmes QALC, Reverso+QALC et MUSQAT au fil du temps. En plus des questions en français, les équivalents de ces questions en anglais sont également disponibles <sup>11</sup>. Il est donc possible de comparer le système QALC au système MUSCLEF sur ce corpus.

11. Ces questions sont en réalité les questions originales. Car dans les évaluations interlingues des campagnes CLEF, les questions sont tout d'abord écrites en anglais avant d'être traduites dans les diverses langues sources proposées aux participants. Par la suite, les résultats sont comparés entre participants qui répondent aux questions dans la même langue cible.

### 1.5.1. Système monolingue anglais QALC

Les performances de QALC en monolingue anglais sur ce corpus de l'évaluation CLEF 2005 sont présentées dans le tableau 1.1. Cette évaluation a été effectuée par comparaison des documents, phrases ou réponses au fichier de patrons de réponses. Elle est donc légèrement imprécise car certaines phrases peuvent contenir la réponse mais dans un contexte totalement différent de celui de la question. Cependant, les évaluations manuelles montrent que cette imprécision ne dépasse pas 5%.

Étape	Nombre de questions ayant une bonne réponse
Documents	94%
Après sélection (Fastr)	93%
Phrases aux 5 premiers rangs	62,5%
Phrases au premier rang	45%
Réponses aux 5 premiers rangs	51,5%
Réponses au premier rang	39,5%

**Tableau 1.1.** Résultats du système QALC sur les données de CLEF 2005 en monolingue anglais

Ce tableau se lit de la façon suivante : pour 94% des questions, le moteur de recherche retourne au moins un document qui contient une réponse juste ; après la sélection à l'aide de Fastr, 93% etc. On peut donc constater que les performances de la phase de recherche d'information sont tout à fait satisfaisantes, puisque, même après la sélection des documents selon les variantes reconnues par Fastr, 93% des questions ont un document réponse. Les erreurs dans la sélection des documents correspondent soit à des problèmes de sélection des mots-clefs pour la formation de la requête, soit à la non-sélection des documents contenant des variantes non reconnues des mots-clefs.

La sélection des phrases et l'extraction de la réponse sont sources de plus importantes pertes de réponses. Plusieurs facteurs expliquent les pertes à ces étapes. Pour la sélection des phrases, les principales causes d'erreurs sont la présence de variantes de mots de la question non reconnues («*to dress*» n'est pas reconnu comme une variante de «*to wear*») et l'absence d'une information importante dans la phrase réponse (anaphore, date donnée quelques phrases avant...).

Enfin, pour l'extraction de la réponse, les erreurs ont des causes diverses. Ainsi pour les questions attendant une entité nommée deux cas d'erreurs sont possibles. En amont, lors du marquage des entités nommées, la réponse peut ne pas avoir été étiquetée avec le bon type d'entité nommée, ou ne pas avoir été étiquetée du tout. Au moment de l'extraction de la réponse, une mauvaise entité nommée peut être choisie lorsque plusieurs entités nommées du bon type sont présentes dans la phrase réponse. Pour

les questions traitées à l'aide des patrons, les erreurs peuvent provenir de l'analyse des questions : le focus ou le verbe principal peuvent être mal détectés. Elles peuvent être dues à des variantes linguistiques des termes de la question qui ne sont pas reconnues dans les phrases candidates. Ainsi dans la question «*Which computer virus was confirmed as a hoax by the US National Computer Security Association ?*», l'analyse des questions marque à raison «*virus*» comme type général, mais dans les phrases candidates seul le terme «*hoax*» apparaît. Enfin les erreurs peuvent être dues aux patrons eux-mêmes, il arrive par exemple qu'aucun patron ne parvienne à s'appliquer, ou encore que plusieurs réponses soit marquées mais que celle retenue finalement ne soit pas la bonne.

Néanmoins, comme le montre la table 1.1, l'extraction de la réponse n'est pas l'étape qui obtient les plus faibles résultats. En particulier, nous avons observé que sur certaines catégories de question, comme la catégorie *définition*, les patrons obtiennent de bonnes performances : 84 % de bonnes réponses dans les 5 premiers rangs et 62 % de bonnes réponses au rang 1.

Pour améliorer encore les résultats de cette étape, il faudrait mettre place un mécanisme de validation de la réponse. Celui-ci contrôlerait que la réponse finalement retournée est bien une réponse possible à la question en utilisant des techniques simples de vérification. Par exemple pour des questions telles que «*Qui est l'épouse de Bill Clinton ?*», ou «*Donner le nom d'un producteur de voitures allemand.*», il est envisageable de consulter une base de connaissances, pour vérifier la véracité de la réponse donnée.

Pour plus de précisions sur l'évaluation de ce système ou de ses équivalents, voir [GRA 06c, LIG 06b].

### 1.5.2. Système interlingue français vers anglais MUSCLEF

Pour ce même corpus, le système interlingue a des résultats inférieurs, présentés dans le tableau 1.2.

Étape	Nombre de questions ayant une bonne réponse	
	Reverso + QALC	MUSQAT
Phrases aux 5 premiers rangs	55,5%	48,5%
Phrases au premier rang	34,5%	32%
Réponses aux 5 premiers rangs	43,5%	40,5%
Réponses au premier rang	28,5%	25,5%

**Tableau 1.2.** Résultats du système MUSCLEF sur les données de CLEF 2005 pour la tâche français vers anglais

Les résultats présentés ici sont ceux de la version actuelle du système, et diffèrent par conséquent des résultats obtenus à l'évaluation CLEF 2005. À CLEF 2005, le meilleur système sur la tâche ayant pour cible l'anglais avait obtenu un score de 25,5% de bonnes réponses au premier rang, ce qui situe MUSCLEF au niveau de l'état de l'art en questions-réponses interlingues. Néanmoins, on peut constater que lors du passage du monolingue à l'interlingue, les performances du système chutent d'environ 10% à chaque étape. Cette différence est bien entendu due à des difficultés de traduction.

Pour le système Reverso+QALC, certains noms propres ou acronymes de la question ne sont pas ou mal traduits, et les traductions posent parfois problème :

– soit pour leur forme syntaxique :

- «*Combien y a-t-il d'habitants en France ?*»  
«*How much is there of inhabitants in France ?*»
- «*Quand est mort Gengis Khan ?*»  
«*When died Gengis Khan ?*»
- «*Qu'est-ce que le GATT ?*»  
«*What the GATT*»
- «*Depuis quand Israël et la Jordanie sont-ils en guerre ?*»  
«*Since when Israel and Jordan they are at war ?*»

– soit pour les choix lexicaux de traduction :

- «*Quels gènes causent le cancer ?*»  
«*What hamper cause the cancer ?*»
- «*Dans combien de scandales fut impliqué Tapie, lorsqu'il était patron de Marseille ?*»  
«*In how many scandals was involved Hidden, when he was a boss of Marseille ?*»
- «*Quel constructeur a lancé la Taurus ?*»  
«*What builder threw Taurus ?*»
- «*Quel montant Selten, Nash et Harsanyi ont-ils reçu pour le prix Nobel d'économie ?*»  
«*What going up Selten, Nash and Harsanyi did they receive for the Nobel prize of economy ?*».

Pour le système MUSQAT, les difficultés tiennent au bruit contenu dans les traductions des termes, ou au contraire à l'absence de traduction pour certains mots. Il conviendrait ainsi de compléter les dictionnaires, et d'élargir la validation des traductions, afin de supprimer du bruit au moment de la traduction des termes. Au cours de l'évaluation CLEF 2006 [GRA 06b], l'introduction de traductions issues du Web a été expérimentée, mais cette piste reste à développer.



## 1.6. Conclusion

Ce chapitre a présenté plusieurs systèmes de questions-réponses au niveau de l'état de l'art : deux systèmes monolingues, QALC pour l'anglais, et FRASQUES pour le français, et deux systèmes interlingues, Reverso+QALC et MUSQAT. Le développement parallèle de modules dans deux langues est facilité par la mise en commun de nombreux modules, des fichiers de données séparés assurant le passage d'une langue à l'autre. Certains points restent à développer pour renforcer les stratégies interlingues. Dans le cas de la traduction des questions, c'est-à-dire pour le système Reverso+QALC, des règles de réécriture pourraient être écrites avant l'envoi des questions au traducteur automatique Reverso afin de contourner certaines erreurs systématiques. Dans le cas de la traduction des termes du système MUSQAT, plusieurs pistes d'amélioration sont envisageables : la complétion des ressources de traduction par l'ajout de dictionnaires ou la recherche de traduction sur le Web ; la stratégie de validation des multitermes pourrait être étendue, par exemple en essayant de valider ces termes sur un corpus externe.

## 1.7. Bibliographie

- [ABN 96] ABNEY S., « Partial Parsing via Finite-State Cascades », *ESSLLI '96 Robust Parsing Workshop*, 1996.
- [AIT 02] AIT-MOKHTAR S., CHANOD J.-P., ROUX C., « Robustness beyond shallowness : incremental deep parsing », *Journal of Natural Language Engineering*, vol. 8, n°3-2, 2002.
- [BER 03] BERTHELIN J.-B., DE CHALENDAR G., ELKATEB-GARA F., FERRET O., GRAU B., HURAUPT-PLANTET M., ILLOUZ G., MONCEAUX L., ET ANNE VILNAT I. R., « Getting reliable answers by exploiting results from several sources of information », *CoLogNET-ElsNET Symposium (Question and Answers : Theoretical and Applied Perspectives)*, Amsterdam, 2003.
- [ELK 03] ELKATEB F., « Extraction d'entités nommées pour la recherche d'informations précises », *4ème congrès ISKO-France, L'organisation des connaissances*, Grenoble, France, 2003.
- [FER 02] FERRET O., GRAU B., HURAUPT-PLANTET M., ILLOUZ G., JACQUEMIN C., MONCEAUX L., ROBBA I., VILNAT A., « How NLP Can Improve Question Answering », *Knowledge Organization*, vol. 29, n°3-4, p. 135-155, 2002.
- [GRA 06a] GRAU B., LIGOZAT A.-L., ROBBA I., VILNAT A., « Term Translation Validation by Retrieving Bi-terms », *LNCS*, 2006.
- [GRA 06b] GRAU B., LIGOZAT A.-L., ROBBA I., VILNAT A., BAGUR M., SÉJOURNÉ K., « The bilingual system MUSCLEF at QA@CLEF 2006 », *Working Notes, CLEF Cross-Language Evaluation Forum*, Alicante, Espagne, 2006.
- [GRA 06c] GRAU B., LIGOZAT A.-L., ROBBA I., VILNAT A., MONCEAUX L., « FRASQUES : A Question-Answering System in the EQueR Evaluation Campaign », *LREC 2006*, 2006.

- [JAC 96] JACQUEMIN C., « *A symbolic and surgical acquisition of terms through variation* », p. 425–438, Springer, Heidelberg, 1996.
- [LIG 06a] LIGOZAT A.-L., GRAU B., ROBBA I., VILNAT A., « Evaluation and Improvement of Cross-Lingual Question Answering Strategies », *Workshop MLQA, EACL*, Trente, Italie, 2006.
- [LIG 06b] LIGOZAT A.-L., GRAU B., ROBBA I., VILNAT A., « L'extraction des réponses dans un système de question-réponse », *Actes de la 13<sup>e</sup> Conférence sur le Traitement Automatique des Langues Naturelles (TALN 2006)*, Louvain, Belgique, 2006.