

## Question Generation for French: Collating Parsers and Paraphrasing Questions

**Delphine Bernhard**

*LiLPa, Université de Strasbourg, France*

DBERNHARD@UNISTRA.FR

**Louis de Viron**

*EarlyTracks SA, Belgium*

LOUIS@EARLYTRACKS.COM

**Véronique Moriceau**

*LIMSI-CNRS, Univ. Paris-Sud, Orsay, France*

VERONIQUE.MORICEAU@LIMSI.FR

**Xavier Tannier**

*LIMSI-CNRS, Univ. Paris-Sud, Orsay, France*

XAVIER.TANNIER@LIMSI.FR

**Editors:** Paul Piwek and Kristy Elizabeth Boyer

### Abstract

This article describes a question generation system for French. The transformation of declarative sentences into questions relies on two different syntactic parsers and named entity recognition tools. This makes it possible to further diversify the questions generated and to possibly alleviate the problems inherent to the analysis tools. The system also generates reformulations for the questions based on variations in the question words, inducing answers with different granularities, and nominalisations of action verbs. We evaluate the questions generated for sentences extracted from two different corpora: a corpus of newspaper articles used for the CLEF Question Answering evaluation campaign and a corpus of simplified online encyclopedia articles. The evaluation shows that the system is able to generate a majority of good and medium quality questions. We also present an original evaluation of the question generation system using the question analysis module of a question answering system.

**Keywords:** Question Generation, Syntactic Analysis, Syntactic Transformation, Paraphrasing, Question Answering

### 1. Introduction

Question Generation (QG) has been addressed recently from different perspectives and for different application domains: dialogue systems, intelligent tutoring, or automatic assessment. Most recent methods perform text to text generation, i.e. they transform declarative sentences into their interrogative counterpart. It thus constitutes the inverse operation to Question Answering (QA) which aims at retrieving answers for a given question based on a collection of text documents. QG is actually a complex task, which relies on a large variety of resources and Natural Language Processing tools: named entity recognition, syntactic analysis, anaphora resolution and text simplification.

However, while several methods have been proposed for the English language, there is no equivalent system for the French language. The system presented in this article aims at closing this gap.

The main contributions of the article are as follows:

- We present a question generation system for French which adapts methods proposed in the context of question generation for the English language
- We evaluate how the use of different syntactic parsers and named entity recognition tools for analysing the source sentence affects the quantity and quality of the generated questions
- We propose two mechanisms for generating different surface forms of the same question: reformulation of the interrogative part and question nominalisation.
- We present a novel evaluation of the question generation system using the question analysis module of a question answering system.

In the next section, we detail related work on the topic of question generation from text. In Section 3, we detail the typology of the questions generated by our system. We then describe the system in Section 4 and explain our method for question paraphrasing in Section 5. We evaluate our question generation system and discuss the results in Section 6. Finally, we conclude and give the perspectives of this work in Section 7.

## 2. State of the Art

### 2.1 Question Generation from Expository Text

Question generation from text consists in automatically transforming a declarative sentence into an interrogative sentence. The question thus generated targets one part of the input sentence, e.g. the subject or an adverbial adjunct. This research domain has been the focus of increasing interest, owing to the recent organisation of an international challenge aimed at comparing question generation systems for English (Rus et al., 2010; Rus et al., this volume).

Automatic question generation from text has two main application domains: (i) dialogue and interactive Question Answering systems and (ii) educational assessment.

In the first application context, question generation has been used for automatically producing dialogues from expository texts (Prendinger et al., 2007; Piwek and Stoyanchev, 2010). The dialogues thus generated may be presented in the form of written text or thanks to virtual agents with speech synthesis. Concrete uses of such dialogues include presenting medical information, such as patient information leaflets and pharmaceutical notices, or educational applications. In a related domain, Question Answering, the quality of the interactions between the system and a user can be improved if the QA system is able to predict some of the questions that the user may wish to ask. Harabagiu et al. (2005) describe a question generation method for interactive QA which first identifies entities relevant for a specific topic and then applies patterns to obtain questions. Automatically generated questions are then presented to the user by selecting those which are most similar to the question asked initially.

The second application context of QG systems, question generation for educational assessment, has been investigated for many years (Wolfe, 1976). Indeed, test writing is a very time-consuming task and the availability of a question generation system thus reduces the workload for instructors. Educational assessment applications rely on question generation methods for producing open or multiple-choice questions for text comprehension (Wolfe, 1976; Gates, 2008) or summative assessment (Mitkov et al., 2006). Depending on the system, the generated questions may correspond to manually-defined templates (Brown et al., 2005; Wang et al., 2008), or be less constrained (Mitkov et al., 2006; Gates, 2008; Heilman and Smith, 2009).

Automatic question generation systems for the English language usually proceed according to the following steps: (i) perform a morphosyntactic, syntactic, semantic and/or discourse analysis of the source sentence, (ii) identify the target phrase for the question in the source sentence, (iii) replace the target phrase with an adequate question word, (iv) make the subject agree with the verb and invert their positions, (v) post-process the question to generate a grammatical and well-formed question (Wolfe, 1976; Gates, 2008; Heilman and Smith, 2009; Kalady et al., 2010). Given this procedure, question generation requires that the input sentence be at least morpho-syntactically analysed. It is often also useful to have additional information about semantics, such as named entities (person, organisation, country, town) or the distinction between animates and inanimates.

This approach has been shown to yield good results for the English language. We therefore apply the same method to the French language. Prior to question generation, we use two different syntactic analysers and two different named entity recognition methods to analyse the input text. Our goal in doing so is to evaluate the impact of the prior analysis tools on the quality and quantity of generated questions.

## 2.2 Question Paraphrasing

As a related task to question generation from text we consider question paraphrasing. Indeed, the same question may be asked with different surface forms. For instance, the following questions can be considered as paraphrases, since they have the same meaning and expect the same answer, while presenting alternate wordings: *How many ounces are there in a pound?*, *What's the number of ounces per pound?*, *How many oz. in a lb.?*

The ability to identify question paraphrases is useful for question answering on databases of question-answer pairs, e.g. FAQs or social Q&A sites (Tomuro and Lytinen, 2004; Zhao et al., 2007; Bernhard and Gurevych, 2008). In this case, the task of question answering is boiled down to the problem of finding question paraphrases in a database of answered questions.

In the context of question generation, the potential uses of automatic question paraphrasing are manifold. We will detail three of these applications: (i) Vary the form of the question for educational assessment, (ii) Induce different phrasings of answers and (iii) Improve recall in automatic QA.

First, one drawback of automatic question generation from text is that generated questions may be too close to the original sentence, as generation usually relies on transforming the syntactic structure from declarative to interrogative, without changing the words used (apart maybe from modifications in the main verb's inflection). However, for the automatic generation of educational assessment questions from study material, it is necessary that the generated questions be different from the surface form of the input clause, so as not to give extraneous cues to students and make questions too easy to answer (Karamanis et al., 2006).

Second, different formulations of the same question induce different phrasings of answers, varying in their level of precision (granularity) or elaborateness. This is due to the phenomenon of *lexical entrainment* which has been observed in human-human dialogues (Garrod and Anderson, 1987) and human-machine dialogues (Gustafson et al., 1997; Stoyanchev and Stent, 2009; Parent and Eskenazi, 2010): people tend to adapt to the vocabulary used by their interlocutor, be it a human or a machine. What is more, interrogative words determine the semantic type of the expected answer. They may be rather vague (e.g., “where”) and thus determine a large spectrum of potential answers subsumed by the broad semantic type related to the question word (e.g., “location”), or more precise, in which case they determine a restricted type of answer (e.g., “town”, “road”, “region”, etc.)

(Garcia-Fernandez, 2010; Garcia-Fernandez et al., 2010). As an example, the two questions below have different interrogative words and request answers with differing degrees of precision:

**Adverbial question:** Où est-ce que se trouve la Joconde ? (*Where is the Mona Lisa?*)

- **Vague answer:** en France
- **Mid-range answer:** à Paris
- **Precise answer:** au Musée du Louvre

**Determinative question:** Dans quelle ville se trouve la Joconde ? (*In what city is the Mona Lisa located?*)

- **Answer:** à Paris, en France

A basic question can thus be reformulated to induce different answers. We make use of this observation in our question generation system by varying the interrogative part of the question (see Section 5.1).

Finally, automatic question paraphrasing can be used to reformulate questions given as an input to automatic Question Answering systems, in order to facilitate answer retrieval and improve recall. Tomuro (2003) present paraphrasing patterns for questions, aimed at identifying question reformulations for answer retrieval from FAQs. Rinaldi et al. (2003) maps user questions to a logical form, which makes it possible to resolve syntactic variations between question and answer. Term variations involving synonyms and morpho-syntactic variants are associated with a unique concept identifier. Duboue and Chu-Carroll (2006) describe a more shallow technique based on machine translation to automatically paraphrase questions, which were then fed to their QA system.

In our QG system, we perform automatic question paraphrasing by transforming verbal constructions involving action verbs into the corresponding nominal construction (see Section 5.2).

### 3. Typology of Generated Questions

Numerous studies have been devoted to interrogative sentences, especially in linguistics, cognitive psychology and information science (Pomerantz, 2005). As a consequence, several taxonomies of question types have been proposed, depending on the target application or the type of discourse under study.

In the field of NLP, research on this topic has been fostered by the development of Question Answering systems. Question types defined within large-scale evaluation campaigns for Question Answering, namely TREC<sup>1</sup> or CLEF<sup>2</sup>, focus on factoid questions (person, time, location, etc.), definition questions and list questions (Dang et al., 2006; Giampiccolo et al., 2007). More complex questions types such as *why* or *how* have been more seldomly dealt with, even though they have been studied in recent research (Diekema et al., 2003; Moriceau et al., 2010).

In our system, we focus on *factoid questions* (*what, when, where, who*) targeted at concept completion and knowledge elicitation, *closed questions*, which require a yes/no answer, and some *definition questions* (Lehnert, 1978; Graesser and Person, 1994). The expected response for our questions is either a boolean (yes/no) or a short phrase from the input sentence. For the time being,

---

1. <http://trec.nist.gov/>

2. <http://www.clef-campaign.org/>

we do not deal with quantity (*How much? How many?*) or measurement (*What size? What length?*) questions.

Questions are generated based on syntactic transformations applied to a source sentence. We have therefore identified which syntactic constituents are best suited as targets for automatically generated questions, based both on our own intuition and linguistic studies (Langacker, 1965; Grevisse, 1975). These syntactic constituents are the expected answers for the generated questions. Next, we have classified these constituents based on their grammatical function. Table 1 details the results of this preliminary study. In addition to the question categories described in this typology, the QG system also generates boolean questions which do not target a specific constituent in the source sentence.

Table 1: Typology of question types and related grammatical functions.

Function	Example	Constituent
Subject	S: Terminer ses études est l'objectif de John. <i>To complete his studies is John's goal.</i> Q: Quel est l'objectif de John ? <i>What is John's goal?</i>	Noun Phrase Pronoun Infinitive clause That/What clause
Direct object	S: J'ai vu le père de Marie. <i>I saw Mary's father.</i> Q: Qui as-tu vu ? <i>Whom did you see?</i>	Noun phrase Pronoun Subordinate clause
Indirect object	S: J'ai offert ce cadeau à mon frère. <i>I offered this present to my brother.</i> Q: À qui as-tu offert ce cadeau ? <i>To whom did you offer this present?</i>	Noun phrase Pronoun
Subject complement	S: La peinture est très belle. <i>The painting is beautiful.</i> Q: Comment est la peinture ? <i>What is the painting?</i>	Noun phrase Adjective
Locative adverbial adjunct	Q: Je viens du cinéma. <i>I come from the movie theater.</i> Q: D'où viens-tu ? <i>Where you do come from?</i>	Noun phrase Pronoun
Temporal adverbial adjunct	S: Mon avion décolle à 13h. <i>My plane takes off at 1 pm.</i> Q: À quelle heure décolle ton avion ? <i>When does your plane take off?</i>	Noun phrase Pronoun
Appositive	S: Le président français, Nicolas Sarkozy, ... <i>The French president, Nicolas Sarkozy, ...</i> Q: Qui est Nicolas Sarkozy ? <i>Who is Nicolas Sarkozy?</i>	Noun phrase
Parenthesised acronym	S: L'Organisation des Nations unies (ONU) ... <i>The United Nations (UN) ...</i> Q: Que signifie ONU ? <i>What does UN mean?</i>	Noun phrase

## 4. Description of the Question Generation System

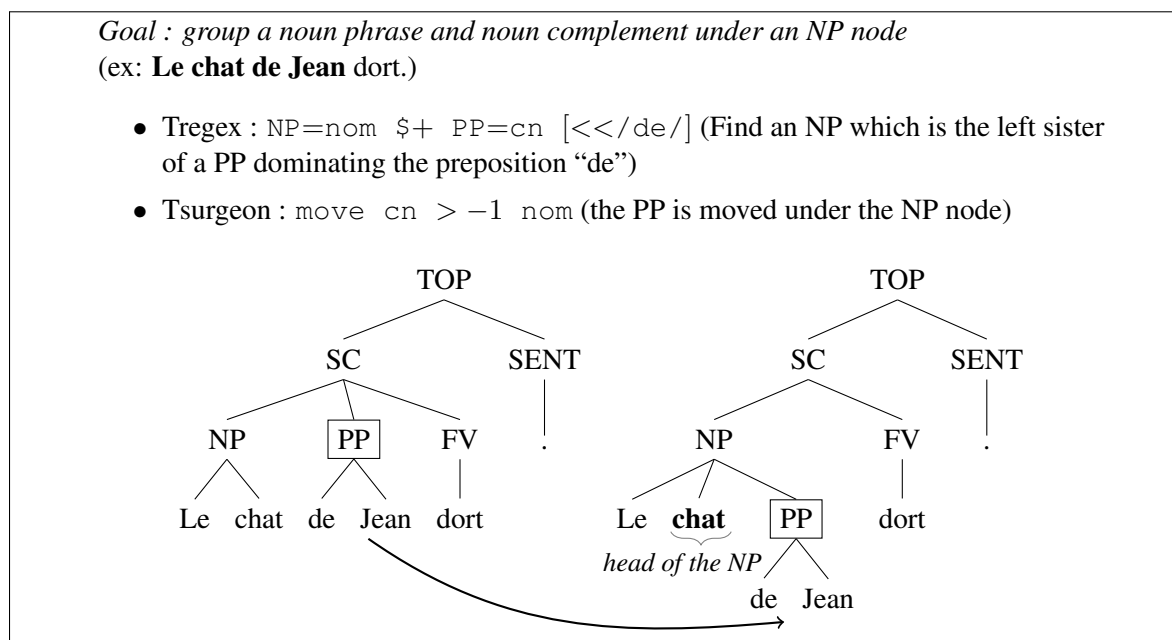
In this section, we detail the question generation system. The input is a French sentence. The output consists of all the questions which were generated by the system for this input sentence.

### 4.1 Syntactic Analysis and Named Entity Recognition

The question generation system proceeds by transforming declarative sentences into questions. The transformations rely on a syntactic analysis of the input sentences. In order to alleviate possible errors stemming from erroneous parses, we use the analyses of two different syntactic parsers: XIP (Aït-Mokhtar et al., 2002) and Bonsai (Candito et al., 2010). XIP (Xerox Incremental Parsing) is a robust parser which also performs named entity recognition and dependency parsing. Bonsai relies on the Berkeley Parser adapted for French and trained on the French TreeBank.

XIP only produces a shallow constituent tree, as the other type of information is available in the form of dependency relationships.<sup>3</sup> We therefore pre-process the tree in order to incrementally group phrases which belong together and thus obtain complete syntactic groups for question generation. In practice, we grouped elements which, together, realize the same grammatical function in the sentence. Figure 1 illustrates the process of grouping constituents in the parse tree produced by XIP.

Figure 1: Example rule for grouping constituents.



While XIP combines syntactic analysis with named entity tagging, Bonsai does not perform any named entity recognition. We therefore combined Bonsai’s output with the following information:

3. The current version of the system does not make use of information about dependency relations. As suggested by one reviewer, this type of information would however be helpful for QG.

- Identification of numerical expressions: day, month, year, date, period, duration, percentage, length, temperature, financial amount, weight, volume, physical measure. Numerical expressions are identified with a tool has been originally developed for a Question Answering system (Ferret et al., 2002) and which relies on regular expressions.
- Identification of named entities: organisations (association, firm, institution), locations (astronym, building, city, country, geonym, hydronym, region, supranational location, way), people (celebrity, dynasty, ensemble, ethnonym, firstname, pseudoanthroponym, job, family), events (feast, history, manifestation). These named entities are identified using a simple list-based approach. We used two different named entity databases: Prolexbase (Tran and Maurel, 2006; Bouchou and Maurel, 2008) which is a lexical resource for proper names, and some of the lists constituted for the system described in Elkateb-Gara (2005). We also used the French Wiktionary<sup>4</sup> to retrieve lists of terms related to family (*mother, father*, etc.) as well as lists of jobs. A further list of jobs was extracted from the ONISEP website.<sup>5</sup> The size of each list is detailed in Appendix C.

We also used the MELt POS tagger combined with the *Lefff* lexicon to obtain morpho-syntactic tags (Denis and Sagot, 2009).

The syntactic trees produced by XIP and Bonsai are processed with Tregex and Tsurgeon (Levy and Andrew, 2006).<sup>6</sup> Tregex makes it possible to explore syntactic trees with a specific regular expression language, relying on node relations in the tree. Tsurgeon performs transformations on the syntactic tree, based on specific nodes identified with Tregex patterns. These tools have been used previously for QG, in other systems (Gates, 2008; Heilman and Smith, 2009; Kalady et al., 2010).

## 4.2 Sentence Simplification

Prior to question generation, the input sentences are simplified. Sentence simplification has been shown to be a key asset for several Natural Language Processing applications, including question generation. Sentence simplification can occur before question generation, in order to ease the generation process. For instance, Heilman and Smith (2010b) describe a factual statement extractor system which identifies simplified statements within complex sentences. The system relies on the sentence's structure in order to identify splitting points and remove the least significant elements.<sup>7</sup> Simplification can also occur at the end of the generation process, to improve the quality of automatically generated questions (Gates, 2008).

Our simplification process relies on Tregex patterns associated with Tsurgeon operations. The sentences are first parsed using the Bonsai analyser. Then, several simplification rules are applied. Table 2 details the rules with some simplification examples in French.

In order to prevent information loss due to the simplification process, non-simplified sentences are also provided as input to the question generation system.

---

4. <http://fr.wiktionary.org>

5. <http://www.onisep.fr>

6. <http://nlp.stanford.edu/software/tregex.shtml>.

7. The system is downloadable online: <http://www.ark.cs.cmu.edu/mheilman/qg-2010-workshop/> [Visited August 23, 2011].

Table 2: Sentence simplification rules.

<b>Simplification</b>	<b>Before</b>	<b>After</b>
<i>Break up coordinated propositions</i>	Le trafic des trains à grande vitesse ICE a été provisoirement suspendu entre Francfort et Paris, <b>et</b> les trains ont été redirigés via Strasbourg.	Le trafic des trains à grande vitesse ICE a été provisoirement suspendu entre Francfort et Paris. Les trains ont été redirigés via Strasbourg.
<i>Break up colon-separated propositions</i>	Je reste chez moi : il pleut.	Je reste chez moi. Il pleut.
<i>Remove comma-separated sentence-initial NP</i>	<b>Texte de "compromis"</b> , cette nouvelle loi européenne a l'ambition de mettre l'UE "à la pointe dans la protection animale", selon les mots du commissaire européen John Dalli.	Cette nouvelle loi européenne a l'ambition de mettre l'UE "à la pointe dans la protection animale", selon les mots du commissaire européen John Dalli.
<i>Remove comma-separated sentence-final PP</i>	Texte de "compromis", cette nouvelle loi européenne a l'ambition de mettre l'UE "à la pointe dans la protection animale", <b>selon les mots du commissaire européen John Dalli</b> .	Texte de "compromis", cette nouvelle loi européenne a l'ambition de mettre l'UE "à la pointe dans la protection animale".
<i>Remove comma-separated sentence-initial adverb or PP</i>	<b>À ce stade</b> , l'enclos du Piton de la Fournaise demeure accessible au public.	L'enclos du Piton de la Fournaise demeure accessible au public.
<i>Remove PP surrounded by commas</i>	En France, <b>pour permettre une meilleure fréquentation de ses trains</b> , la SNCF a mis en service le TGV en 1981.	En France, la SNCF a mis en service le TGV en 1981.
<i>Remove a PP in a sequence of PPs</i>	La première locomotive à vapeur a été inventée <b>par Richard Trevithick en Angleterre en 1804</b> .	La première locomotive à vapeur a été inventée en Angleterre en 1804.
<i>Remove subordinate clauses</i>	Cette pratique fut énormément utilisée sur la Loire, pour la remontée de Nantes à Orléans, voire plus en amont <b>si les conditions le permettaient</b> .	Cette pratique fut énormément utilisée sur la Loire, pour la remontée de Nantes à Orléans, voire plus en amont.
<i>Remove parentheticals</i>	En Europe, il est cultivé dans la plaine du Pô <b>(Italie)</b> .	En Europe, il est cultivé dans la plaine du Pô.



### 4.3 Question Generation

The generation of questions is decomposed in two steps:

1. Identification of the target of the question, based on the question typology.
2. Transformation of the source sentence into a question.

The first step relies on manually defined Tregex expressions, which are detailed in Appendix B. It consists in verifying that the potential target constituents are in the source sentence. For instance, in order to generate a question about the subject, the system locates the subject in the source sentence and labels it with a special tag in the syntactic tree. Once the target constituents have been identified, Tsurgeon transformation rules are applied to the sentence. If no target constituent is found in the source sentence, then no question is generated.

Figure 2 shows how questions concerning locations are generated. The generation of other question types follows the same kind of procedure.

### 4.4 Challenges for French Question Generation

Generating questions in French implies two challenges that are not present in English:

- Subject-verb inversion
- Modifications in the verbal form, due to the syntactic transformations.

#### 4.4.1 SUBJECT-VERB INVERSION

Langacker (1965) describes two types of subject-verb inversions: simple and complex inversions.

**Simple inversion** The subject and the verb are simply inverted. This inversion is applied if the subject is a pronoun (1) and if the question is about the direct object (2) or the subject complement (3).

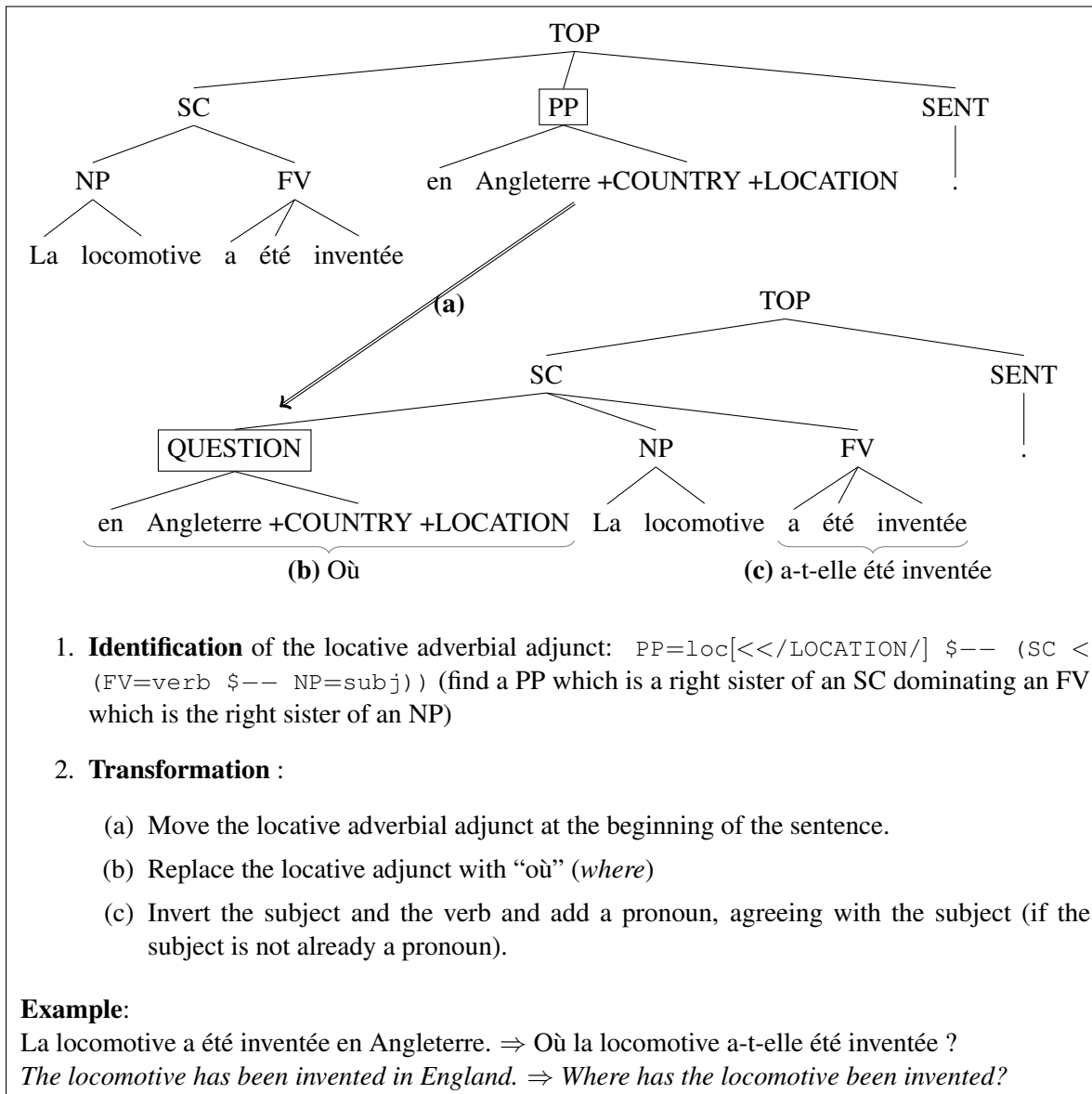
- (1) Il mange une pomme (eng: He eats an apple) → Que mange-t-il? (eng: What does he eat?)
- (2) Jean mange une pomme (eng: John eats an apple) → Que mange Jean? (eng: What does John eat?)
- (3) Jean est électricien (John is an electrician) → Qu'est Jean? (What is John?)

For the compound tenses, the subject is placed between the auxiliary and the participle if it is a pronoun (4) and after the entire verbal form if it is not a pronoun (5).

- (4) Nous avons travaillé (eng: We have worked) → Avons-nous travaillé (eng: Have we worked?)
- (5) Jean a mangé une pomme (eng: John has eaten an apple) → Qu'a mangé Jean? (eng: What did John eat?)

In order to perform simple inversions, we use regular expressions which shift the verbal form before the subject. If the verb is a compound, the subject is inserted between the auxiliary and the rest of the form. Then, hyphens are added (see example 4) and elisions are performed if necessary (see example 5).

Figure 2: Generation of questions about locations.



**Complex inversion** The subject does not move, but is repeated after the verb, in a pronominal form having the same features as the subject. This inversion is applied in all cases not covered by the simple inversion.

- (6) Jean part à Londres (eng: John goes to London)  $\rightarrow$  Où Jean part-il? (eng: Where does Jean go?)

For the compound tenses, nominal subjects do not move and pronominals are placed between the auxiliary and the participle.

- (7) Jean est parti à Londres (eng: John went to London) → Où Jean est-il parti? (eng: Where did John go?)

In order to perform complex inversions, we begin by tagging the head of the subject and by recovering its features (gender and number). Then, we add the pronoun corresponding to those features after the verb if it is a simple tense, or between the auxiliary and the rest of the form if it is a compound one. Finally, we add the hyphens and make the elisions, as for simple inversions.

Sometimes, the two types of inversions are possible. In such case, our system applies the complex inversion.

- (8) Jean est parti à Londres (eng: John went to London) → Où Jean est-il parti? Où est parti Jean? (eng: Where did John go?)

#### 4.4.2 MODIFICATIONS OF INFLECTIONS

When a question is generated about the subject, hence introduced by *qui* or *qu'est-ce qui*, all the elements which are related to this subject have to be singular and masculine. Indeed, those interrogative pronouns are invariable. For example, we cannot generate the question (10) from the sentence (9). It is therefore necessary to transform the verb into a singular form.

- (9) Jean et Marie dorment. (eng: John and Mary are sleeping)

- (10) \*Qui dorment ? (eng: \*Who are sleeping?)

In order to perform this operation, we use a morphological inflection system that takes as input the lemma and the tense of the verb and gives as output the verb at the third person of the singular of the same tense. Furthermore, we put all the terms linked to the subject (attribute, adjective...) in their masculine-singular form, replacing the original forms by their lemma, given by the analyzer.

- (11) Jean et Marie sont malades? (eng: John and Mary are sick) → Qui est malade? (Who is sick?)

### 4.5 Post-Processing

The post-processing step yields a natural language question out of the syntactic tree obtained after transformation. To this aim, several regular expressions are applied to remove morphosyntactic, syntactic and semantic information from the parse tree, as well as information about lemmas. It also normalizes the case within the sentence by inserting an uppercase at the beginning of the question. More important to the French language, it manages elisions which are necessary for instance when the particle *est-ce que* is followed by a word beginning with a vowel. In this case, *que* is replaced with *qu'*, e.g. “Une phrase peut contenir plusieurs verbes conjugués. → Est-ce **qu'une** phrase peut contenir plusieurs verbes conjugués?” (eng: A sentence may contain several finite verbs. → May a sentence contain several finite verbs?).

## 5. Question Paraphrasing

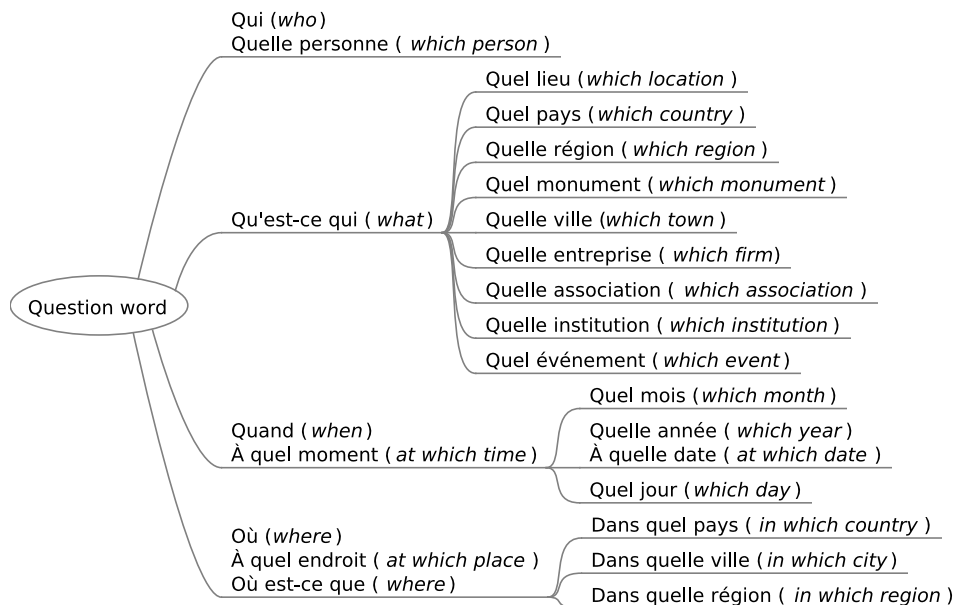
In order to further diversify the formulation of the automatically generated questions, we perform question paraphrasing, using two strategies: (i) Variations in the question words and (ii) Nominalisation of the main verb in the question.

## 5.1 Variations in Question Words

As explained in Section 2, varying question word aims to induce answers with different granularities (Garcia-Fernandez, 2010). A study by Tomuro (2003), based on manually paraphrased questions, showed that about half of the question reformulations consisted of variations in the interrogative part. In order to generate these variations, we make use of the named entity tag associated with the target answer. For instance, given the input sentence in Figure 2, we can either generate a generic question “*Où la locomotive a-t-elle été inventée ?*” (eng: *Where has the locomotive been invented?*), based on the named entity tag +LOCATION associated with the word *Angleterre* (eng: *England*) or a more specific question “*Dans quel pays la locomotive a-t-elle été inventée ?*” (eng: *In which country has the locomotive been invented?*), based on the specific tag +COUNTRY.

Figure 3 present the taxonomy of generic to specific question words used in our system. The structure of this taxonomy is close to the two-layer taxonomy proposed by Li and Roth (2002) for question classification.

Figure 3: Taxonomy of question word specialisations.



## 5.2 Question Nominalisation

Nominalisation consists in forming nouns out of a base word with a different part-of-speech (verb, adjective or noun). Here we focus on deverbal nominalisation, where a verbal construction is transformed into a corresponding nominal construction (Benetti and Corminboeuf, 2004). For instance, the following question “*When was Elizabeth II crowned?*” (fr: “*Quand Elizabeth II a-t-elle été couronnée ?*”) can be nominalised into “*When was the coronation of Elizabeth II?*” (fr: “*Quand a eu lieu le couronnement de Elizabeth II ?*”). Here, we focus specifically on deverbal event-denoting nouns which are morphologically derived from verbs, e.g. *destroy* – *destruction*. This kind of morphological process has been shown to be quite frequent in French question-answer pairs (Bernhard

et al., 2011). In other words, when an action verb is found in a question, then the answer often contains the corresponding deverbal noun. The ability to automatically generate question nominalisations could therefore help improving the results of automatic Question Answering systems.

The automatic generation of question nominalisation requires a resource which links verbs with morphologically derived nouns. Here, we make use of Verbaction, a lexicon of French action nouns linked to their corresponding verbs totalling 9,393 verb-noun pairs (Hathout et al., 2002; Hathout and Tanguy, 2002). Figure 4 displays an excerpt from Verbaction, which relates the noun *abandon* to the verb *abandonner* (eng: *to abandon*).

Figure 4: Excerpt from Verbaction.

```
<couple>
  <verb>
    <lemma>abandonner</lemma>
    <tag>Vmn-----</tag>
  </verb>
  <noun gender="masculine" number="singular">
    <lemma>abandon</lemma>
    <tag>Ncms</tag>
  </noun>
</couple>
```

The nominalisation process is based on Tregex patterns which identify specific question structures including action verbs. Overall, we defined 13 patterns. Once such a structure has been identified, the verb's lemma is looked up in a nominalisation resource and the question is subsequently transformed into its nominalised version.

The main difficulty lies in selecting the correct nominalisation. Indeed, for some verbs, there are several nominalisations (Bauer, 2003). For instance, given the question “*En quelle année a été tourné le film "War Games" ?*” (eng: *In which year was the movie "War games" shot?*), there are several alternative deverbal nouns derived from the verb *tourner*: *tournage* (eng: *filming, shooting*), *tour* (eng: *turn*) and *tournée* (eng: *tour*). Only the first of these alternatives is correct. We therefore have to perform ambiguity resolution to select the best noun corresponding to the verb. For this, we use the Web, such as proposed by Sumita et al. (2005) for validating distractors in multiple-choice cloze items. For each candidate deverbal noun, we form several web queries including words situated in the immediate left and right context of the deverbal noun. We issue these queries to the Yahoo search engine and sum the number of hits. The deverbal noun which gets the largest amount of hits is eventually selected.

## 6. Evaluation

The evaluation of our question generation system aims at answering the following questions:

1. What is the quality of the automatically generated questions? What amount of post-editing is necessary for producing perfect questions out of them?
2. What is the recall of the system?
3. How well does question nominalisation perform?

4. How well does automatic question analysis for question answering perform for automatically generated questions?

We first present the evaluation datasets and then seek answers for all of the above questions.

## 6.1 Evaluation Datasets

For the evaluation of the QG system, we used two different datasets for the source sentences:

- 40 sentences from the CLEF French corpus, containing answers to questions for the 2005 and 2006 challenges. These sentences have been randomly selected among sentences containing answers to definition questions and factoid questions expecting an answer of type *person*, *organisation*, *date* and *place*. The CLEF French corpus is composed of about 177,000 well-formed news articles from *Le Monde* and *ATS* 1994 and 1995 (about 2 Gb). The sentences of these documents are supposed to be well-written and syntactically correct. CLEF 2005 questions are factoid and definition questions only; CLEF 2006 introduced a few list questions that we did not use in our evaluation.
- 40 sentences from Wikimini<sup>8</sup> and Vikidia,<sup>9</sup> which are simplified versions of the French Wikipedia targeted at children. We randomly selected the articles and we took the first sentence of each article, which we assume would contain the most important information.

**Sentence simplification** Prior to question generation, the input sentences were automatically simplified using the simplification method described in section 4.2. The simplification process was iteratively applied to simplified sentences in order to further simplify sentences. This iterative process is stopped when none of the sentences can be simplified anymore. For our corpora, the simplification process was terminated after two iterations (Simpl. 1 and Simpl. 2). Table 3 displays the statistics for the evaluation datasets before and after simplification. The figures indicate the number of sentences produced by each simplification step. For instance, for the CLEF dataset, 31 simplified sentences are produced out of the 40 source sentences. As expected, the simplification leads to shorter sentences. Moreover, sentences in the CLEF corpus are, on average, longer than those in the Vikidia-Wikimini dataset.

Table 3: Statistics for the evaluation datasets before and after simplification

	CLEF			Vikidia-Wikimini		
	Source	Simpl. 1	Simpl. 2	Source	Simpl. 1	Simpl. 2
<b># sentences</b>	40	31	8	40	19	3
<b># tokens</b>	978	665	167	567	222	20
<b>Avg. sentence length</b>	24.45	21.45	20.88	14.18	11.68	6.67

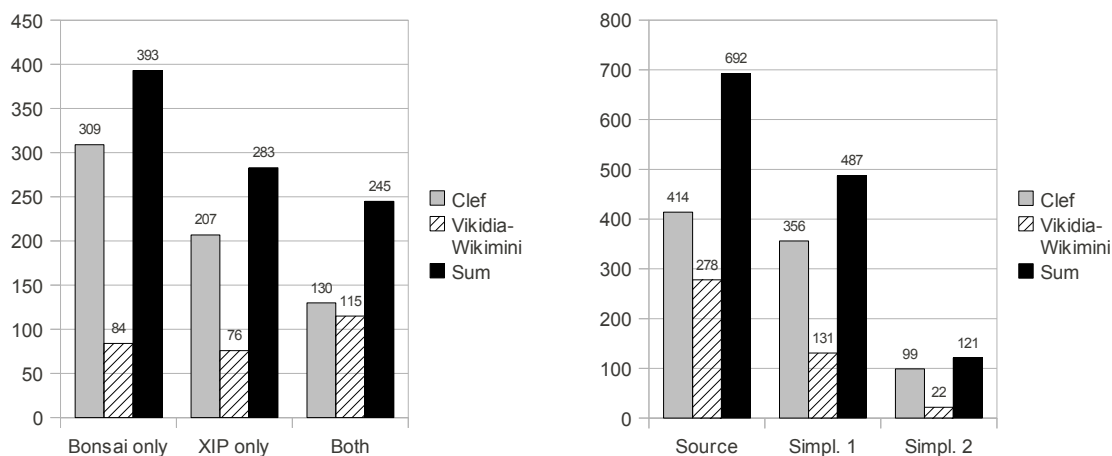
**Generated questions** Overall, 869 questions were generated for the CLEF dataset out of which 646 were unique. For the Vikidia-Wikimini dataset, we obtained 431 questions, out of which 275

8. <http://fr.wikimini.org>

9. <http://fr.vikidia.org>

were unique.<sup>10</sup> Figure 5 displays the number of questions generated (a) depending on the syntactic analyser used and (b) depending on the source sentence used (before and after simplification). The amount of questions generated is larger for sentences analysed with Bonsai. However, this discrepancy can be partly explained by the fact that the rules for generating acronym questions have been only implemented for Bonsai analyses. Acronym questions account for 45 questions in the CLEF dataset. Strikingly, the overlap between questions generated based on both syntactic analysers is rather low: the identical questions account for 130 questions in CLEF and 115 questions in Vikidia-Wikimini. This justifies the use of different syntactic parsers for question generation, as this makes it possible to obtain a larger variety of question types.

Figure 5: Number of questions generated.



(a) Depending on the syntactic analyser (unique questions)

(b) Depending on the simplification step

Figure 6 displays the repartition of question types for both corpora. Interestingly, no acronym or appositive questions were generated for the Vikidia-Wikimini dataset. This is due to the greater simplicity of the Vikidia-Wikimini sentences which rarely include appositive constructions or parenthesised acronyms. The proportion of locational and temporal questions is also larger for the CLEF dataset, here again due to the bigger complexity of the source sentences which more often contain prepositional phrases corresponding to adverbial adjuncts.

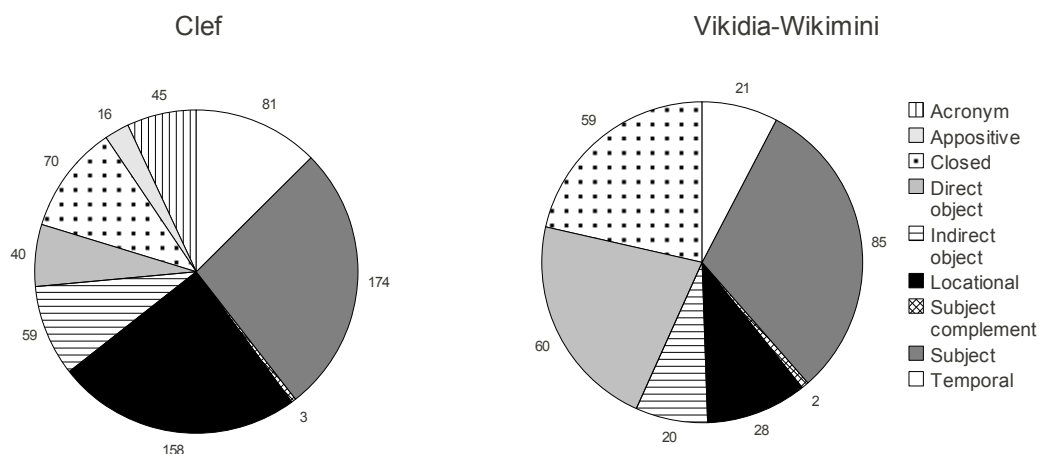
## 6.2 Intrinsic Evaluation of Question Quality

The acceptability of the generated questions was manually evaluated by native speakers. For each corpus, we randomly selected 100 automatically generated questions, including questions generated from simplified sentences. The questions were annotated by two independent annotators each. A 3-point scale was used for the annotation:

1. Unworthy questions, which were directly rejected because they were too inconsistent or ungrammatical to be redeemed.

<sup>10</sup>. Duplicates are due to identical questions generated at different simplification steps or based on the analysis of one of our two syntactic analysis tools.

Figure 6: Repartition of question types for the evaluation datasets.



2. Questions which were not perfect but could be improved, by performing some manual changes.
3. Perfect questions, which could be accepted without any change.

Table 4 summarises the evaluation results for both corpora while Table 5 lists examples of generated questions with their corresponding score. The annotators agreed in 70% of the cases for the CLEF corpus and in 67% of the cases for the Vikidia-Wikimini corpus. The corresponding Kappa values (0.537 for CLEF and 0.484 for Vikidia-Wikimini) indicate a moderate level of agreement. However, disagreements mostly concern adjacent categories in our annotation and cases where one annotator considers a question as perfect while the other rejects it are rare (9 cases for CLEF and 4 cases for Vikidia-Wikimini).

The performance of the QG system is the lowest for direct object and indirect object questions. Closed questions and acronym questions perform best. This result is unsurprising as these questions are produced by rather simple generation rules. Question quality is also globally better for the simpler Vikidia-Wikimini corpus. If we compare both syntactic parsers, XIP obtains better scores for the CLEF corpus, but also leads to the generation of less questions than Bonsai. For the Vikidia-Wikimini corpus, on the contrary, Bonsai performs slightly better on average.

In order to further evaluate the quality of the generated questions, we asked the human annotators to post-edit the automatically generated questions that were neither perfect, nor rejected, which corresponds to a score of 2. The goal of the manual post-edition was to create targeted references for the questions, based on the questions generated by the system.

An example of an automatically generated question and the result after post-editing by both annotators is shown below:

**Question** *Qu'est-ce qui est une ville de Suisse et un port fluvial situé sur le Rhin, qui coupe la ville en deux ?*

**Post-edit 1** *Quelle ville de Suisse et port fluvial est coupé en deux par le Rhin ?*

**Post-edit 2** *Qu'est ce qui est une ville de Suisse et un port fluvial situé sur le Rhin ?*

Post-edition has been shown to be a valid evaluation method for natural language generation systems (Sripada et al., 2005) and the generation of multiple-choice test items (Mitkov et al., 2006).



Table 4: Question quality

	CLEF			Vikidia-Wikimini		
	# quest.	Annot. 1	Annot. 2	# quest.	Annot. 3	Annot. 4
<b>All</b>	100	2.20	2.05	100	2.27	2.38
<b>XIP</b>	43	2.37	2.19	64	2.28	2.41
<b>Bonsai</b>	73	2.15	2.05	79	2.41	2.61
<b>Acronym</b>	6	3.00	2.83	0	–	–
<b>Appositive</b>	3	2.33	2.33	0	–	–
<b>Closed</b>	6	2.33	2.33	22	2.73	2.77
<b>Direct object</b>	7	1.57	1.43	21	1.86	1.76
<b>Indirect object</b>	5	1.80	1.60	5	2.40	2.60
<b>Location</b>	34	2.23	2.09	10	2.20	2.60
<b>Subject</b>	26	2.11	1.88	31	2.35	2.71
<b>Subject complement</b>	0	–	–	1	1.00	1.00
<b>Temporal</b>	13	2.31	2.23	10	2.00	1.60

Table 5: Example questions with their scores

Score	Source sentence	Question
1	<p>Une fable est un court récit qui contient une morale, généralement à la fin.</p> <p><i>eng: A fable is a short story that contains a moral, usually at the end.</i></p>	<p>À quelle année Une fable est-elle un court récit qui contient une morale, généralement ?</p> <p><i>eng: At which year A fable is a short story that contains a moral, usually?</i></p>
2	<p>La circonférence d'un cercle est la longueur de sa ligne de contour.</p> <p><i>eng: The circumference of a circle is the length of its contour line.</i></p>	<p>Que la circonférence d'un cercle est-elle ?</p> <p><i>eng: What the circumference of a circle is it?</i></p>
3	<p>Le tabac a été découvert par Christophe Colomb en Amérique.</p> <p><i>eng: Tobacco was discovered by Christopher Columbus in America.</i></p>	<p>Qu'est-ce qui a été découvert par Christophe Colomb en Amérique ?</p> <p><i>eng: What was discovered by Christopher Columbus in America?</i></p>

The manually post-edited questions were used as targeted reference to compute Human-targeted Translation Edit Rate (HTER). The Translation Edit Rate measure has been developed to evaluate automatic machine translation (Snover et al., 2006). It corresponds to the number of edits needed to transform the output of a machine translation system into a reference translation, normalised by the number of words in the reference translation. The possible edit operations are the insertion, deletion and substitution of single words as well as changes in the position (shifts) of word sequences. We applied this measure to pairs made up of an automatically generated question (*hypothesis*) and the post-edited question (*reference*), using the TER-Plus evaluation tool.<sup>11</sup>

Table 6 displays the average HTER for both corpora, as well as the average number of inserts, deletes, substitutions and shifts. Average (Avg.) HTER corresponds to the average HTER values obtained while Global HTER corresponds to the total number of edits divided by the total numbers of words for all evaluated question pairs. Note that both values are multiplied by 100 here.

Table 6: HTER values.

	CLEF				Vikidia-Wikimini			
	Annot.1		Annot. 2		Annot. 3		Annot.4	
	Avg.	Sum	Avg.	Sum	Avg.	Sum	Avg.	Sum
<b>Ins</b>	3.83	153	3.70	174	1.08	56	1.04	25
<b>Del</b>	0.05	2	0.13	6	0.69	36	0.88	21
<b>Sub</b>	0.48	19	0.98	46	1.19	62	0.83	20
<b>Shift</b>	0.00	0	0.00	0	0.29	15	0.04	1
	<b>Avg.</b>	<b>Global</b>	<b>Avg.</b>	<b>Global</b>	<b>Avg.</b>	<b>Global</b>	<b>Avg.</b>	<b>Global</b>
<b>HTER</b>	24.64	22.74	30.32	25.11	30.81	26.67	29.64	23.51

The obtained average HTER values are roughly equivalent across both corpora, which indicates that the amount of post-edition work is about the same. In the CLEF corpus, the majority of the edits were due to insertions in the automatically generated sentences, which shows that the automatically generated questions tend to include extraneous material and are longer than what would be desirable. This is often due to indirect speech, when the reporting clause is kept in the question, e.g. “*Où l’écrivain américain Charles Bukowski a-t-il succombé mercredi à une pneumonie, à l’âge de 73 ans, a annoncé jeudi un de ses proches ?*” (eng: *Where did the American writer Charles Bukowski die on Wednesday from pneumonia, at the age of 73, did one his relatives announce on Thursday?*) In the Vikidia-Wikimini corpus, the amount of insertions and substitutions tend to be roughly equivalent. Moreover, the average number of insertions is lower than in the CLEF corpus, which shows that the generated questions do not include as much extraneous material, most certainly due to the shorter length of the source sentences.

### 6.3 Comparison with Original CLEF Questions

We also present a task-oriented experiment aimed at assessing the recall of the system. For this purpose, we automatically compare the questions generated by our system for the CLEF corpus

11. <http://www.umiacs.umd.edu/~snover/terp/>

with the corresponding questions of the CLEF campaign<sup>12</sup>. The goal of this study is to verify to what extent the system is able to produce questions identical or close to the meaning of the questions used in the CLEF evaluation campaign. For the CLEF 2005 subset, the system was able to generate 3 identical questions and 7 close questions out of 28. For the CLEF 2006 subset, the system was able to generate 2 identical questions and 1 close questions out of 12.

Two examples of automatically generated questions and the expected CLEF question are shown below:

<b>Source sentence</b>	<i>L'Organisation mondiale de la santé (OMS) a contredit mardi ses précédentes déclarations rassurantes sur l'épidémie d'Ebola au Zaïre.</i>
<b>CLEF question</b>	<i>Qu'est-ce que l'OMS ?</i>
<b>Generated question</b>	<i>Qu'est-ce que le OMS ?</i>
<b>Source sentence</b>	<i>Francesco De Lorenzo avait été ministre de la Santé de 1989 à 1993.</i>
<b>CLEF question</b>	<i>Quel politicien libéral était ministre de la santé en Italie de 1989 à 1993 ?</i>
<b>Generated question</b>	<i>Qui avait été ministre de la Santé de 1989 à 1993 ?</i>

Despite the large number of generated questions, the system still does not cover all the questions which could be generated from a sentence. This is due to the following problems:

- **Inferences and external knowledge.** In this case, the generation of the expected questions would require inferencing capabilities. For instance, given a sentence such as *Le Japon (...) estime que la création d'un sanctuaire pour les baleines est sans fondement scientifique.* (eng: *Japan believes that the creation of a sanctuary for whales is not scientifically based*), the generation of the question *Quel pays est contre la création d'un sanctuaire pour les baleines en Antarctique ?* (eng: *Which country is against the creation of a sanctuary for whales in Antarctica?*) requires some external knowledge about Japan's position on whaling. This is beyond the capabilities of our QG system in its current state.
- **Missing question generation rules.** Valuable information for generating questions is often contained in grammatical constituents which are not taken into account by our question typology, in particular noun postmodifiers:
  - prepositional phrases: *passage à la monnaie unique en 1999* → *Quelle est la date du passage à la monnaie unique ?*, (eng: *changeover to the single currency* → *What is the date of the changeover to the single currency?*)
  - relative clauses: *Le Dalaï Lama, qui a obtenu le prix Nobel de la paix en 1989* → *Quand le Dalaï Lama a-t-il été nommé Prix Nobel de la paix ?*, (eng: *The Dalai Lama, who obtained the Nobel Peace Prize in 1989* → *When was the Dalai Lama nominated for the Nobel Peace Prize?*)

This analysis provides important insights for the future developments of the question generation system in particular concerning the question types which should be tackled with the highest priority.

12. Recall that the source sentences used for our evaluation correspond to sentences containing the answer to one question from either the CLEF 2005 or CLEF 2006 evaluation campaign.

#### 6.4 Evaluation of Question Nominalisation

We applied the question nominalisation rules presented in Section 5.2 to the list of French questions for the CLEF 2005 and CLEF 2006 evaluation campaigns, totalling 200 questions each. We did not use the questions generated by our QG system here since errors in the generated questions could have a detrimental effect on the nominalisation module. Overall, the system was able to generate 35 nominalisations for the CLEF 2005 dataset and 30 nominalisations for the CLEF 2006 dataset. About 1/3 of the nominalised questions were judged to be well formed. For the remainder of the questions, the nominalisation errors can be categorised in 6 different classes, exemplified in Table 7. Note that the percentages do not sum to 100 as several problems may be present in one and the same question.

Table 7: Question nominalisations and problem categories.

Category	Original question	Nominalised Question	Percent
<b>OK</b>	Quand le roi Talal de Jordanie <i>abdiqua</i> -t-il ?	Quelle est la date d' <i>abdication</i> du roi Talal de Jordanie ?	32
<b>Nominalisation is impossible</b>	Quelle multinationale française <i>a changé</i> son nom pour celui de groupe Danone ?	Quelle est la multinationale de <i>changement</i> de son nom pour celui de groupe Danone ?	40
<b>Incomplete question</b>	Quand l’Australie, la Suède et la Finlande <i>entreront</i> -elles dans l’Union Européenne ?	Quelle est la date d' <i>entrée</i> de l’Australie, la Suède et la Finlande ?	28
<b>Bad nominalisation</b>	Quelle église <i>a ordonné</i> des femmes prêtres en mars 1994 ?	Quelle est l’église d’ <i>ordre</i> de des femmes prêtres en mars 1994 ?	22
<b>Bad question word</b>	À qui <i>appartient</i> le Milan AC ?	Quelle est l’ <i>appartenance</i> du Milan AC ?	20
<b>No noun needed</b>	De combien d’étoiles <i>se compose</i> notre galaxie ?	Quelles sont les étoiles de <i>composition</i> de notre galaxie ?	20
<b>Pronominal verb</b>	Dans quel pays <i>se trouve</i> Euskirchen ?	Quel est le pays de <i>trouvaille</i> d’Euskirchen ?	18

These results might seem disappointing at first sight. However, some of the problems should be easy to solve by extending the nominalisation rules, in particular for solving the issue of incomplete questions, or by using additional constraints, for a better selection of the adequate question word or deverbal noun. Some of the problems would however require a more complete resource than Verbaction, which only relates deverbal action nouns to their verbs and does not provide additional information, e.g. the nominalisation’s argument structure. Such kind of information is included in the English NOMLEX lexicon (Macleod et al., 1998), which has been used to produce nominalisation patterns for information extraction (Meyers et al., 1998). It would also be helpful to possess resources about other kinds of deverbal nouns, such as deverbal agent nouns, e.g. *invent*, *inventor*. This would make it possible to further diversify the nominalisation transformation rules, in order to encompass questions such as the following: “*Who invented the television?*” → “*Who is the inventor of the television?*”.

## 6.5 Extrinsic Evaluation with a QA System

In the last part of the evaluation, we used the automatically generated questions as input for FIDJI, an open-domain QA system for French and English (Moriceau and Tannier, 2010). This system combines syntactic information with traditional QA techniques such as named entity recognition and term weighting in order to validate answers through different documents.

This particular evaluation consisted in estimating whether FIDJI was able to analyse the generated questions correctly (question type and expected answer type). We could have evaluated the full answer extraction process, but FIDJI uses XIP for answer extraction in a similar way that we used it for question generation. There would then be a bias since FIDJI would have found the sentences which have been used for question generation.

This experiment aims at investigating two possibilities:

1. An automatic evaluation of the quality of questions, if we can identify that the QA system fails at analysing bad questions but succeeds for good questions.
2. An evaluation of the robustness of the QA system, if it still manages to analyse correctly imperfect questions.

Question analysis in FIDJI turns the question into a declarative sentence where the answer is represented by the ‘ANSWER’ lemma. It also aims to identify:

- The syntactic dependencies,
- The expected type(s) of the answer (named entity type),
- The question type:
  - **factoid** questions are introduced by a specific *wh* word or by a specific trigger:
    - \* Questions in ‘*qui*’ (‘*who*’) expect a PERSON or ORGANIZATION named entity (NE).
    - \* Questions in ‘*quand*’ (‘*when*’) expect a DATE, as well as ‘*À quelle date*’ (‘*At which date*’), etc. Temporal NEs can be made more specific (a year, a day, etc.).
    - \* Questions in ‘*combien*’ (‘*how much/many*’) expect a NUMBER, as well as ‘*Quelle vitesse/température*’ (‘*what speed/temperature*’), etc., with more specific number NEs.
    - \* Questions in ‘*où*’ (‘*where*’) expect a LOCATION, as well as ‘*Dans quelle ville*’ (‘*in which city*’), etc.
    - \* Questions in ‘*quel*’ (‘*what/which*’) expect a specific answer type which may not have a corresponding NE type, as ‘*quelle déclaration*’ (‘*which declaration*’).
  - **definition** questions: ‘*Qu’est-ce que...*’ (‘*What is*’), ‘*Que signifie*’ (‘*What is the meaning of*’), ‘*Qui est*’ (‘*Who is*’), ...
  - **boolean** questions: ‘*Est-ce que...*’, ‘*... est-il*’ (yes/no question triggers), ...
  - **complex** questions are called ‘*comment*’ (‘*how*’) and ‘*pourquoi*’ (‘*why*’), but can be introduced by other means, such as ‘*pour quelle raison*’ (‘*for what reason*’), ‘*dans quel but*’ (‘*for what purpose*’), ‘*de quelle manière*’ (‘*in which manner*’)...

- **list** questions are those containing explicitly a plural answer type (e.g. *Which planets...?*, *Who are the...?*, etc.).

All the items above are determined by using the syntactic structure of the question. We added to XIP’s encrypted grammars some semantic lexical resources: about 200 nouns representing persons (such as teacher, minister or astronaut), organizations, locations, nationalities and numerical value units (currencies, physic units...). We also added to XIP’s grammars about 250 grammar rules to deal with the analysis of question syntactic constructions. Only those added rules are used for question analysis in FIDJI.

For example:

**Question:** Quel premier ministre s’est suicidé en 1993 ?

(Which Prime Minister committed suicide in 1993?)

**Dependencies:** DATE (1993)

PERSON (ANSWER)

SUBJ(se suicider, ANSWER)

attribut (ANSWER, ministre)

attribut (ministre, premier)

**Question type:** factoid

**Expected answer type:** person

These last two features (question type and expected answer type) are then compared with manually assessed types. We performed this evaluation on the same 100 generated questions coming from the CLEF corpus and evaluated in Section 6.2.

When both annotators did not agree on the quality of the question (see Table 4), we chose the “worst case” for this evaluation (*i.e.* the lowest value).

The first information is that FIDJI still finds question and answer types, even if the question is really bad (73% of type-1 questions, against 87% of types 2 and 3). Contrary to our initial hypothesis, an automatic evaluation of question quality based on FIDJI’s question analysis is thus impracticable. We then compared the system’s ability to analyse the question correctly, when the quality score was either 2 or 3.<sup>13</sup> The results are presented in Table 8. 61% of “medium” questions are still correctly analysed, against 73% for perfect questions. This means that a QA system like FIDJI is robust enough to deal with imperfect questions generated by an automatic system. As a next step, it would be interesting to check whether this finding holds for questions by humans which have been shown to be often ill-formed (Ignatova et al., 2008; Sitbon et al., 2008).

Table 8: Analysis of generated questions by FIDJI.

	<b>Perfect questions (“3”)</b>	<b>Medium questions (“2”)</b>	<b>Unworthy questions (“1”)</b>	<b>TOTAL</b>
<b>TOTAL</b>	26	41	33	100
<b>Correct analysis</b>	19 (73%)	25 (61%)	/	48
<b>Incorrect analysis</b>	7 (27%)	16 (39%)	/	19

13. Value 1 represents unworthy questions, for which finding the proper types makes no sense at all, since the questions are usually incomprehensible.

The reasons for FIDJI's incorrect analyses are quite different between types 2 and 3 questions. Failures on perfect questions are solely due to correct constructions that were unknown to FIDJI but frequently used as question word reformulations by the generation system, mainly '*À quel endroit...*' ('*In which place...*' instead of '*Where*') and '*Quelle personne...*' ('*Which person*' instead of '*Who*'). The automatic generation of variations in question words thus led to identifying gaps in FIDJI's question analysis module. For example:

**Question:** À quoi correspond le sigle OMS ?

(*What do initials OMS stand for?*)

is correctly analysed as a definition question with the relation `acronym(OMS, ANSWER)`, while the construction:

**Question:** Que représente le sigle OMS ?

(*What do initials OMS represent?*)

is not recognized and leads to an unknown type with dependency `OBJ(represent, ANSWER)`.

On the other hand, failures on type-2 questions come, in addition, from syntactic errors at the beginning of the generated question (as the use of '*Qu'est-ce qui*' ('*What*') when seeking for a person name or '*Pendant quand*' ('*During when*') for an event).

This is the case of the two following questions, where the type is not found:

**Question:** \*Qui est 46 ans ?

(*~ \*Who has 46 year-old?*)

**Question:** \*Pendant quand le président de l'ANC, a-t-il symbolisé la lutte pour le pouvoir de la majorité noire d'Afrique du Sud ?

(*~ \*During when the president of ANC has been the symbol of...?*)

## 7. Conclusion and Perspectives

We have presented a question generation system for the French language. The system proceeds by transforming declarative sentences into their interrogative counterparts, focusing on one grammatical constituent of the source sentence. This is, to our knowledge, the first available system of this type for French.

The evaluation shows that using two different syntactic analysis and named entity recognition tools leads to an increased variety in the questions generated, while preserving the average quality of the questions for different analyzers. We also proposed a question nominalisation procedure which generates nominal structures out of questions containing action verbs. Finally, we evaluated the quality of the questions generated by applying the question analysis module of a question answering system, FIDJI. This analysis showed that FIDJI is rather robust with respect to ill-formed questions.

While the system presented in this article is in principle not targeted at a single application, its expected uses are interactive Question Answering and educational assessment. We plan to integrate the question paraphrasing system into a QA system in order to improve the answer retrieval process. We are also interested in the automatic generation of educational assessment exercises, e.g. from Wikipedia.

We envision several directions for future work. First, a better automatic simplification of the source sentences would make it possible to improve the quality of the generated questions. Second, we also plan to develop a question classification module, similar to the one proposed by Heilman and Smith (2010a), in order to distinguish between well-formed and ill-formed questions.

Finally, we would like to stress one important limitation of question generation as implemented in our system, as well as many other similar tools. The input considered for generation is a single sentence, which may or may not have been simplified. We believe that an interesting direction for future work would be the ability to generate transversal questions, relying on pieces of information extracted from several sentences within the same paragraph or document. Consider for instance the following sentences:

Barack Obama, born August 4, 1961 Honolulu, Hawaii, is the 44th and current President of the United States. (...) He was named the 2009 Nobel Peace Prize laureate.

The following question contains information extracted from both sentences:

Which president of the United States was named the Nobel Peace Prize laureate in 2009?

This would require moving beyond sentence-level analysis towards discourse-level analysis and in particular the identification of co-reference chains.

## Acknowledgments

At the time of this work Louis de Viron was a student at the Université Catholique de Louvain, Belgium and Delphine Bernhard was at LIMSI-CNRS, Orsay, France. This research was partly supported by the Quaero Program, funded by OSEO. We would like to thank Anne Garcia-Fernandez, Anne-Laure Ligozat and Anne-Lyse Minard for their valuable help in evaluating the system.

## Appendix A. List of Tools and Resources

This appendix lists the tools and resources used in our question generation system. For the XIP tool, we did not use the Web service described here but our own copy of the tool which was kindly provided by the Xerox corporation.

Tool	Description	Availability
MELt	POS tagging	Free, downloadable from <a href="https://gforge.inria.fr/projects/lingwb/">https://gforge.inria.fr/projects/lingwb/</a>
Bonsai	Syntactic analysis	Free, downloadable from <a href="http://alpage.inria.fr/statgram/frdep/fr_stat_dep_bky.html">http://alpage.inria.fr/statgram/frdep/fr_stat_dep_bky.html</a>
XIP	Syntactic analysis and Named Entity Recognition	Web service, see <a href="http://open.xerox.com/Services/XIPParser">http://open.xerox.com/Services/XIPParser</a>
Tregex and Tsurgeon	Pattern matching in trees	Free, downloadable from <a href="http://nlp.stanford.edu/software/tregex.shtml">http://nlp.stanford.edu/software/tregex.shtml</a>

Table 9: List of tools.



Resource	Description	Availability
<i>Lefff</i>	Morposyntactic lexicon	Downloadable from <a href="http://atoll.inria.fr/~sagot/lefff.html">http://atoll.inria.fr/~sagot/lefff.html</a>
Lists of proper names	NE lexicon	Built from Prolexbase ( <a href="http://www.cnrtl.fr/lexiques/prolex/">http://www.cnrtl.fr/lexiques/prolex/</a> ) and in-house resources constituted for the system described in Elkateb-Gara (2005)
List of terms related to family	Lexicon used for person identification	Retrieved from <a href="http://fr.wiktionary.org">http://fr.wiktionary.org</a> (Category: “Lexique en français de la famille”)
Lists of jobs	Lexicon used for person identification	Retrieved from <a href="http://fr.wiktionary.org">http://fr.wiktionary.org</a> (Categories: “Métiers en français”, “Métiers du secteur primaire en français”, “Métiers du secteur secondaire en français”, “Métiers du secteur tertiaire en français”) and <a href="http://www.onisep.fr/">http://www.onisep.fr/</a>

Table 10: List of Resources.

## Appendix B. List of Tregex rules

### B.1 Questions about the subject

That/What clause:

```
/S(sub|C)/=subj [<<CS | <<CONJQUE] & [ $++ /VN|FV/=verb ] & [ >>, /SENT|TOP/ ]
```

Infinitive clause:

```
/VPinf|IV/=subj [ !<<, /PREP|PP/ ] & [ $++ /VN|FV/=verb ] & [ >>, /SENT|TOP/=root ]
```

Noun phrase:

```
NP=subj $++ (/VN|FV/=verb[ < /V|VERB/=verbalform & ?<-1 VPP=vpp1  
& ?< (VERB=vpp2 < /PARTPAS/ < /VMOD/) ] ) & [ >>, /SENT|TOP/ | $- (PONCT $- PP) ]
```

Clitic pronoun:

```
CLS=subj $++ (V=verb) & [ ! $++ CLO ] & [ >>, /VN/ ] & [ < /cIn|CLIT/ ]
```

### B.2 Questions about the direct object

Subordinate clause:

```
/SC|Ssub/=obj[ <</BG|CS/ ] & [ $- (SC[ <<(FV=verb[ <<VERB=verbalform ] & $ NP = subj)  
& > TOP=sent) ] | $- (VN=verb[ <<V=verbalform & <<CLS=subj & > SENT=sent) ] ]
```

Pronoun, rule 1:

```
/NP|CLS/=subj [ $++ (FV=verb < (PRON=cod $+ (PRON=coi[ <</lui|leur/ ] $+ VERB=verbalform))  
& >> TOP=sent) | > (VN=verb < (CLO=cod $+ (CLO=coi[ <</lui|leur/ ] $+ V=verbalform))  
& > SENT=sent) ]
```

Pronoun, rule 2:

```
/NP|CLS/=subj [$++ (FV=verb < (PRON=coi $+ (PRON=cod[<</le|la|les/] $+ VERB=verbalform))
& >> TOP=sent) | > (VN=verb < (CLO=coi $+ (CLO=cod[<</le|la|les/] $+ V=verbalform))
& > SENT=sent)]
```

**Pronoun, rule 3:**

```
/NP|CLS/=subj [$++ (FV=verb < (PRON=cod[<</le|la|les|en/] $+ VERB=verbalform)
& >> TOP=sent) | > (VN=verb < (CLO=cod[<</le|la|les|en/] $+ V=verbalform)
& > SENT=sent)]
```

**Pronoun, rule 4:**

```
SC[<< (NP=subj $++ (FV=verb < (PRON=cod[<</le|nous|vous|te|me/] $+ VERB=verbalform)))]
& $+ PP = coi
```

**Noun phrase**

```
NP=cod [$-- (SC < (FV=verb[<<VERB=verbalform] & $-- NP=subj)) & > TOP=sent
& ! << /COORDITEMS/ | $-- (VN=verb[<<V=verbalform
& $-- (NP=subj !<< PROREL) | < CLS=subj]) & > SENT=sent]
```

**B.3 Questions about the indirect object**

```
PP=coi [<< /PREP|P/=prep & < NP=obj] & [>+(SENT) (SENT=sent <
(VN=verb << V=verbalform $-- NP=subj $++ PP))
| $-- (SC=sent < (FV=verb << VERB=verbalform $-- NP=subj))]
```

**B.4 Questions about the subject complement**

```
AP=obj [$- (SC[<< (FV=verb[<< VERB=verbalform]
& $-- NP=subj) & > TOP=sent]) | $- (VN=verb[<< V=verbalform
& > SENT=sent] $-- NP=subj)]
```

**B.5 Questions about locative adverbial adjuncts**

```
PP=loc [<< /PREP|P/=prep & << /LIEU|LOCATION/]
& [>+(SENT) (SENT < (VN=verb $-- NP=subj ))
| $-- (SC < (FV=verb $-- NP=subj))]
```

**B.6 Questions about temporal adverbial adjuncts**

```
PP=time [<< /PREP|P/=prep & << /TIME|DATE/]
& [>+(SENT) (SENT < (VN=verb $-- NP=subj ))
| $-- (SC < (FV=verb $-- NP=subj))]
```

**B.7 Questions about appositives****People, rule 1:**

```
PUNCT < /,/ . (NP=person << /PERSON/ . (PUNCT < /,/))
```

**People, rule 2:**

```
NP . (PONCT < /,/ . (/NP|PP/=person << /PERSON/ . (PONCT < /,/)))
```

**People, rule 3:**

```
NP=person << /PERSON/ . (PUNCT < /,/ . (NP=full . /PUNCT|SENT/ ) )
```

**People, rule 4:**

```
/NP|PP/=person << /PERSON/ << /FIRSTNAME/ . (PONCT < /,/ . (/NP|PP/=full
!<< /PERSON/ . /PONCT|SENT/ ) )
```

**Dates:**

```
NP=event . (/PONCT|PUNCT/ < /,/ . (NP=date << /DATE/ . (/PONCT|PUNCT/ < /,/)))
```

**B.8 Questions about acronyms**

```
NP=full < (PONCT << /LRB/ << NP=abbrev << /RRB/)
```

**B.9 Closed questions**

```
NP=subj $+ /FV|VN/=verb > /SENT|SC/
```

**Appendix C. Details for resources**

Type	Description	Size
Association	Association name	81
Astronym	Name of astronomical object	27
Building	Famous building names	88
Celebrity	Famous people names	4,142
City	City names	41,992
Country	Country names	450
Demonym-association	Names of association members	8
Demonym-astronym	Names of astronomical objects inhabitants	8
Demonym-building	Names of building dwellers	8
Demonym-city	Names of city residents	63,602
Demonym-country	Names of country inhabitants	1,755
Demonym-ethnonym	Related to an ethnic group	5
Demonym-geonym	Names of geographical locations inhabitants	52
Demonym-hydronym	Names of body of water inhabitants	24
Demonym-region	Names of region inhabitants	2,985
Demonym-supranational	Names of supranational entity inhabitants	180
Dynasty	Dynasty names	57
Ensemble	Group names	14
Ethnonym	Ethnic group members	199
Family	Family member names	137
Feast	Celebration names	11
Firm	Firm names	23,290
Firstname	Given names	23,864

Table 11: Size of the resources used for list-based named entity tagging and person names tagging (part 1).

Type	Description	Size
Geonym	Geographical location names	217
History	Famous historical events	210
Hydronym	Body of water proper names	4,383
Institution	Institution names	68
Manifestation	Famous events	5
Meteorology	Meteorological phenomena	1
Object	Famous object names	1
Organization	Organization names	1,008
Prof	Job names	2,854
PseudoAnthroponym	e.g. C-3PO	3
Region	Region names	2,666
Supranational	Names of supranational entities	91
Vessel	Names of famous vessels	4
Way	Names of famous streets, roads and squares	16
Work	Names of famous books and texts	84

Table 12: Size of the resources used for list-based named entity tagging and person names tagging (part 2).

## References

- Salah Aït-Mokhtar, Jean-Pierre Chanod, and Claude Roux. Robustness beyond shallowness: incremental deep parsing. *Natural Language Engineering*, 8(3):121–144, 2002.
- Laurie Bauer. *Introducing Linguistic Morphology*. Georgetown University Press, 2003. 2nd edition.
- Laurence Benetti and Gilles Corminboeuf. Les nominalisations des prédicats d’action. *Cahiers de linguistique française*, 26:413–435, 2004.
- Delphine Bernhard and Iryna Gurevych. Answering Learners’ Questions by Retrieving Question Paraphrases from Social Q&A Sites. In *Proceedings of the 3rd Workshop on Innovative Use of NLP for Building Educational Applications held in conjunction with ACL 2008*, pages 44–52, Columbus, Ohio, USA, June 19 2008.
- Delphine Bernhard, Bruno Cartoni, and Delphine Tribout. A Task-Based Evaluation of French Morphological Resources and Tools, A Case Study for Question-Answer Pairs. *LILT (Linguistic Issues in Language Technology)*, 5(2), 2011.
- Béatrice Bouchou and Denis Maurel. Prolexbase et LMF: vers un standard pour les ressources lexicales sur les noms propres. *Traitement Automatique des Langues*, 49(1):61–88, 2008.
- Jonathan C. Brown, Gwen A. Frishkoff, and Maxine Eskenazi. Automatic question generation for vocabulary assessment. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 819–826, 2005.

- Marie Candito, Joakim Nivre, Pascal Denis, and Enrique Henestroza Anguiano. Benchmarking of Statistical Dependency Parsers for French. In *Coling 2010*, pages 108–116, Beijing, China, August 2010.
- Hoa Trang Dang, Jimmy Lin, and Diane Kelly. Overview of the TREC 2006 Question Answering Track. In *Proceedings of the Fifteenth Text REtrieval Conference (TREC 2006)*, 2006.
- Pascal Denis and Benoît Sagot. Coupling an annotated corpus and a morphosyntactic lexicon for state-of-the-art POS tagging with less human effort. In *Proceedings of the Pacific Asia Conference on Language, Information and Computation*, 2009.
- Anne R. Diekema, Ozgur Yilmazel, Jiangping Chen, Sarah Harwell, Elizabeth D. Liddy, and Lan He. What Do You Mean? Finding Answers to Complex Questions. In *Proceedings of the AAAI Spring Symposium: New Directions in Question Answering*, 2003.
- Pablo Ariel Duboue and Jennifer Chu-Carroll. Answering the question you wish they had asked: The impact of paraphrasing for Question Answering. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, pages 33–36, 2006.
- Faiza Elkateb-Gara. *L'organisation des connaissances, approches conceptuelles*, chapter Extraction d'entités nommées pour la recherche d'informations précises, pages 73–82. L'Harmattan, 2005.
- Olivier Ferret, Brigitte Grau, Martine Hurault-Plantet, Gabriel Illouz, Christian Jacquemin, Laura Monceaux, Isabelle Robba, and Anne Vilnat. How NLP can improve Question Answering. *Knowledge Organization*, 29:135–155, 2002.
- Anne Garcia-Fernandez. *Génération de réponses en langue naturelle orales et écrites pour les systèmes de question-réponse en domaine ouvert*. PhD thesis, Université Paris Sud 11 Orsay, 2010.
- Anne Garcia-Fernandez, Sophie Rosset, and Anne Vilnat. MACAQ : A Multi Annotated Corpus to Study how we Adapt Answers to Various Questions. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, 2010.
- Simon Garrod and Anthony Anderson. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27(2):181 – 218, 1987.
- Donna M. Gates. Automatically generating reading comprehension look-back strategy questions from expository texts. Master's thesis, Carnegie Mellon University, may 2008.
- Danilo Giampiccolo, Anselmo Peñas, Christelle Ayache, Dan Cristea, Pamela Forner, Valentin Jijkoun, Petya Osenova, Paulo Rocha, Bogdan Sacaleanu, and Richard Sutcliffe. Overview of the CLEF 2007 Multilingual Question Answering Track. In Alessandro Nardi and Carol Peters, editors, *Working Notes for the CLEF 2007 Workshop*, Budapest, Hungary, 19-21 September 2007.
- Arthur C. Graesser and Natalie K. Person. Question Asking During Tutoring. *American Educational Research Journal*, 31(1):104–137, 1994.
- Maurice Grevisse. *Le bon usage. Grammaire française avec des remarques sur la langue française d'aujourd'hui*. Duculot, Gembloux, 10e edition, 1975.

- J. Gustafson, A. Larsson, R. Carlson, and K. Hellman. How do system questions influence lexical choices in user answers. In *Proceedings of Eurospeech '97*, pages 2275–2278, 1997.
- Sanda Harabagiu, Andrew Hickl, John Lehmann, and Dan Moldovan. Experiments with interactive question-answering. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pages 205–214, 2005.
- Nabil Hathout and Ludovic Tanguy. Webaffix: Discovering Morphological Links on the WWW. In *Proceedings of the Third International Conference on Language Resources and Evaluation*, pages 1799–1804, Las Palmas de Gran Canaria, Espagne, 2002. ELRA.
- Nabil Hathout, Fiammetta Namer, and Georgette Dal. *Many Morphologies*, chapter An Experimental Constructional Database : The MorTAL Project, pages 178–209. Cascadilla Press, 2002.
- Michael Heilman and Noah A. Smith. Question Generation via Overgenerating Transformations and Ranking. <http://www.cs.cmu.edu/~mheilman/papers/heilman-smith-qg-tech-report.pdf>, 2009.
- Michael Heilman and Noah A. Smith. Good Question! Statistical Ranking for Question Generation. In *Proceedings of NAACL/HLT*, 2010a.
- Michael Heilman and Noah A. Smith. Extracting simplified statements for factual question generation. In *Proceedings of QG2010: The third workshop of Question Generation*, Pittsburgh, 2010b.
- Kateryna Ignatova, Delphine Bernhard, and Iryna Gurevych. Generating High Quality Questions from Low Quality Questions. In *Proceedings of the Workshop on the Question Generation Shared Task and Evaluation Challenge*, Arlington, VA, September 2008.
- Saidalavi Kalady, Ajeesh Elikkotttil, and Rajarshi Das. Natural language question generation using syntax and keywords. In *Proceedings of QG2010: The third workshop of Question Generation*, Pittsburgh, 2010.
- Nikiforos Karamanis, Le An Ha, and Ruslan Mitkov. Generating Multiple-Choice Test Items from Medical Text: A Pilot Study. In *Proceedings of the Fourth International Natural Language Generation Conference*, pages 111–113, 2006.
- Ronald W. Langacker. French interrogatives: a transformational description. *Language*, 41(4): 587–600, october-december 1965.
- Wendy G. Lehnert. *The Process of Question Answering: A Computer Simulation of Cognition*. Lawrence Erlbaum Associates, Hillsdale, N.J., 1978.
- Roger Levy and Galen Andrew. Tregex and Tsurgeon: tools for querying and manipulating tree data structures. In *5th International Conference on Language Resources and Evaluation*, 2006.
- Xin Li and Dan Roth. Learning question classifiers. In *Proceedings of the 19th international conference on Computational linguistics*, pages 1–7, 2002.
- Catherine Macleod, Ralph Grishman, Adam Meyers, Leslie Barrett, and Ruth Reeves. NOMLEX: A Lexicon of Nominalizations. In *Proceedings of Euralex'98*, pages 187–193, 1998.

- Adam Meyers, Catherine Macleod, Roman Yangarber, Ralph Grishman, Leslie Barrett, and Ruth Reeves. Using NOMLEX to Produce Nominalization Patterns for Information Extraction. In *Proceedings of the Workshop on the Computational Treatment of Nominals*, pages 25–32, 1998.
- Ruslan Mitkov, Le An Ha, and Nikiforos Karamanis. A computer-aided environment for generating multiple-choice test items. *Natural Language Engineering*, 12(2):177–194, 2006.
- Véronique Moriceau and Xavier Tannier. FIDJI: Using Syntax for Validating Answers in Multiple Documents. *Information Retrieval, Special Issue on Focused Information Retrieval*, 13(5):507–533, 2010.
- Véronique Moriceau, Xavier Tannier, and Mathieu Falco. Une étude des questions "complexes" en question-réponse. In *Actes de la Conférence Traitement Automatique des Langues Naturelles (TALN 2010, article court)*, Montréal, Canada, July 2010.
- Gabriel Parent and Maxine Eskenazi. Lexical Entrainment of Real Users in the Let's Go Spoken Dialog System. In *INTERSPEECH-2010*, pages 3018–3021, 2010.
- Paul Piwek and Svetlana Stoyanchev. Generating Expository Dialogue from Monologue: Motivation, Corpus and Preliminary Rules. In *Proceedings of the 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 333–336, 2010.
- Jeffrey Pomerantz. A Linguistic Analysis of Question Taxonomies. *Journal of the American Society for Information Science and Technology*, 56(7):715–728, 2005.
- Helmut Prendinger, Paul Piwek, and Mitsuru Ishizuka. Automatic generation of multi-modal dialogue from text based on discourse structure analysis. In *Proceedings 1st IEEE International Conference on Semantic Computing (ICSC-07)*, pages 27–36, Irvine, CA, USA, September 2007.
- Fabio Rinaldi, James Dowdall, Kaarel Kaljurand, Michael Hess, and Diego Mollá. Exploiting paraphrases in a Question Answering system. In *Proceedings of the second international workshop on Paraphrasing*, pages 25–32, 2003.
- V. Rus, B. Wyse, P. Piwek, M. Lintean, Stoyanchev S., and C. Moldovan. The First Question Generation Shared Task Evaluation Challenge. In *Proceedings of the 6th International Natural Language Generation Conference (INLG 2010)*, Dublin, Ireland, 2010.
- Laurianne Sitbon, Patrice Bellot, and Philippe Blache. Evaluating Robustness Of A QA System Through A Corpus Of Real-Life Questions. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, 2008.
- Matthew Snover, Bonnie J. Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul. A Study of Translation Edit Rate with Targeted Human Annotation. In *Proceedings of AMTA*, 2006.
- Somayajulu G. Sripada, Ehud Reiter, and Lezan Hawizy. Evaluating an NLG System using Post-Editing. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 2005)*, pages 1700–1, 2005.

- Svetlana Stoyanchev and Amanda Stent. Lexical and Syntactic Priming and Their Impact in Deployed Spoken Dialog Systems. In *Proceedings of the 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers*, pages 189–192, 2009.
- Eiichiro Sumita, Fumiaki Sugaya, and Seiichi Yamamoto. Measuring Non-native Speakers' Proficiency of English by Using a Test with Automatically-Generated Fill-in-the-Blank Questions. In *Proceedings of the Second Workshop on Building Educational Applications Using NLP*, pages 61–68, June 2005.
- Noriko Tomuro. Interrogative reformulation patterns and acquisition of question paraphrases. In *Proceedings of the second international workshop on Paraphrasing*, pages 33–40, 2003.
- Noriko Tomuro and Steven Lytinen. Retrieval Models and Q&A Learning with FAQ Files. In Mark T. Maybury, editor, *New Directions in Question Answering*, chapter Retrieval Models and Q&A Learning with FAQ Files, pages 183–194. AAAI Press, 2004.
- Mickaël Tran and Denis Maurel. Prolexbase : un dictionnaire relationnel multilingue de noms propres. *Traitement Automatique des Langues*, 47(3):115–139, 2006.
- Weiming Wang, Tianyong Hao, and Wenyin Liu. Automatic Question Generation for Learning Evaluation in Medicine. In *Advances in Web Based Learning – ICWL 2007*, volume 4823 of *Lecture Notes in Computer Science*, pages 242–251. Springer Berlin / Heidelberg, 2008.
- John H. Wolfe. Automatic question generation from text - an aid to independent study. *SIGCUE Outlook, Proceedings of the SIGCSE-SIGCUE joint symposium on Computer science education*, 10(SI):104–112, 1976.
- Shiqi Zhao, Ming Zhou, and Ting Liu. Learning Question Paraphrases for QA from Encarta Logs. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pages 1795–1800, Hyderabad, India, January 6-12 2007.