

Génération automatique de questions à partir de textes en français *

Louis de Viron^{1,3} Delphine Bernhard¹ Véronique Moriceau^{1,2} Xavier Tannier^{1,2}

(1) LIMSI-CNRS, 91403 Orsay, France

(2) Université Paris Sud, 91405 Orsay, France

(3) Université Catholique de Louvain, Belgique

louis.devirion@student.uclouvain.be, {delphine.bernhard, moriceau, xtannier}@limsi.fr

Résumé. Nous présentons dans cet article un générateur automatique de questions pour le français. Le système de génération procède par transformation de phrases déclaratives en interrogatives et se base sur une analyse syntaxique préalable de la phrase de base. Nous détaillons les différents types de questions générées. Nous présentons également une évaluation de l'outil, qui démontre que 41 % des questions générées par le système sont parfaitement bien formées.

Abstract. In this article, we present an automatic question generation system for French. The system proceeds by transforming declarative sentences into interrogative sentences, based on a preliminary syntactic analysis of the base sentence. We detail the different types of questions generated. We also present an evaluation of the tool, which shows that 41 % of the questions generated by the system are perfectly well-formed.

Mots-clés : génération de questions, analyse syntaxique, transformation syntaxique.

Keywords : question generation, syntactic analysis, syntactic transformation.

1 Introduction

La génération automatique de questions à partir de textes consiste à transformer automatiquement une phrase déclarative en une phrase interrogative. Il s'agit d'une tâche complexe, qui mobilise nombre de ressources et outils du TAL tels que la détection d'entités nommées, l'analyse syntaxique, la résolution d'anaphores et la simplification de phrases. Les applications de la génération automatique de questions sont par ailleurs variées : création de tests et de questionnaires à choix multiples pour l'aide à l'apprentissage, systèmes de dialogue homme-machine ou de questions-réponses interactifs. Si le domaine a été largement traité dans le monde anglophone, et ce depuis longtemps (on retrouve un article de Wolfe (1976) qui évoque déjà le sujet), il n'existe pas, à notre connaissance, de travaux équivalents pour le français. Nous présentons dans cet article un système de génération automatique de questions pour le français qui procède par transformation d'arbre syntaxique à partir de l'analyse fournie par XIP (Ait-Mokhtar *et al.*, 2002) afin de produire à la fois des questions factuelles et des questions fermées.

2 État de l'art

La génération automatique de questions trouve son application dans deux domaines principaux :

- Les systèmes de dialogue à partir de textes expositifs (Prendinger *et al.*, 2007; Piwek & Stoyanchev, 2010). Ces dialogues peuvent être présentés comme des textes ou encore à l'aide de personnages virtuels.
- Les applications éducatives : génération de questions ouvertes pour la compréhension de texte (Wolfe, 1976; Gates, 2008) ou questions à choix multiple (Mitkov *et al.*, 2006).

Les systèmes de génération de questions développés pour l'anglais procèdent généralement selon le schéma suivant : (i) analyse morphosyntaxique et/ou syntaxique du texte source (ii) identification du syntagme cible sur lequel portera la question, (iii) déplacement du syntagme cible, (iv) remplacement du syntagme cible par le mot interrogatif approprié, (v) inversion et accord sujet-verbe, (vi) post-traitement de la question pour générer une

*. Ces travaux ont été partiellement financés par OSEO dans le cadre du programme QUAERO.

forme de surface grammaticale et bien formée (Wolfe, 1976; Gates, 2008; Heilman & Smith, 2009; Kalady *et al.*, 2010). Dans tous les cas, la génération nécessite au minimum une analyse morphosyntaxique de la phrase source. Il est également utile de disposer d'informations d'ordre sémantique telles que l'étiquetage en entités nommées (personne, organisation, pays, ville) ou la distinction animé-inanimé. Le système présenté dans cet article utilise la même méthode, appliquée au français.

La génération automatique de questions est généralement associée à une étape de simplification qui peut se situer en amont de la génération. Ainsi, Heilman & Smith (2010a) décrivent un algorithme d'extraction de propositions indépendantes à partir de phrases complexes. Pour ce faire, l'algorithme se base sur la structure de la phrase (propositions relatives et subordonnées, adverbes, appositions...) pour la découper et supprimer les éléments moins significatifs¹. La simplification peut également intervenir en fin de processus, afin d'améliorer les questions générées (Gates, 2008).

3 Typologie des questions générées

En traitement automatique des langues, les travaux sur les phrases interrogatives ont surtout été effectués dans le cadre des systèmes de question-réponse. Les types de questions se focalisent généralement sur les questions factuelles (personne, temps, lieu, etc.), les questions de définition (*qu'est-ce que ? qui est ?*) et les questions de type liste (Dang *et al.*, 2006; Giampiccolo *et al.*, 2007). Dans notre étude, nous nous sommes limités aux questions factuelles et aux questions fermées. De plus, notre système ne traite pas les questions de quantité (*combien ?*) et de mesure (*quelle taille ? quelle longueur ?*).

Les questions sont générées par transformation syntaxique de la phrase de base. Nous avons donc identifié les constituants syntaxiques à propos desquels il est pertinent de poser des questions, sur la base de notre intuition et d'études linguistiques portant sur les phrases interrogatives (Langacker, 1965; Grevisse, 1975). De fait, ces constituants syntaxiques constituent les réponses attendues aux questions générées. Nous avons ensuite classé les constituants selon leur fonction grammaticale. Le tableau 1 détaille les résultats de cette analyse préalable. À cette typologie s'ajoutent les questions fermées (dont la réponse peut être *oui* ou *non*), qui ne portent pas sur un élément spécifique de la phrase.

Fonction	Type de question	Réalisation de la fonction
Sujet	Qu'est-ce qui / Qui	Groupe nominal / Pronom / Prop. infinitive / Prop. en « que » / Prop. relative
Complément d'objet direct	Que	Groupe nominal / Pronom / Prop. subordonnée
Complément d'objet indirect	À qui / À quoi / De qui / De quoi	Groupe nominal / Pronom
Attribut du sujet	Comment	Adjectif / Groupe nominal
Complément de lieu	Où	Groupe nominal / Pronom
Complément de temps	Quand	Groupe nominal / Pronom

TABLE 1 – Typologie des principales fonctions grammaticales et de leurs réalisations

4 Description du système

Nous détaillons maintenant notre système de génération de questions. Le système prend pour entrée des phrases en français. Les phrases passent tout d'abord par une étape de pré-traitement, comprenant toutes les opérations qui se déroulent entre la phrase brute fournie au système et celle qui sera effectivement traitée. Toutes les phrases sont à la base acceptées par le système et aucune sélection sémantique n'est donc opérée. La structure syntaxique des questions est ensuite générée à l'aide d'un ensemble de règles de transformation syntaxique. Enfin, la forme de surface des questions est obtenue à partir d'un ensemble de post-traitements opérant sur la structure profonde de la question.

1. Une démonstration de l'outil est accessible en ligne : <http://www.ark.cs.cmu.edu/mheilman/qg-2010-workshop/>

4.1 Pré-traitements

Simplification de phrases. La simplification de phrases est un enjeu crucial car elle permet de faciliter le processus de génération et, *in fine*, d'obtenir des questions de meilleure qualité. La simplification automatique constitue toutefois un domaine de recherche à part entière, nécessitant des techniques évoluées. Nous nous sommes donc limités à quelques règles de simplification basiques qui sont destinées à être améliorées dans le futur : suppression des portions de phrases entre parenthèses, virgules, tirets et crochets et scission des phrases au niveau des signes de ponctuation « ; » et « : ».

Construction de l'arbre syntaxique. Les phrases sont ensuite soumises à une analyse syntaxique à l'aide de XIP (Xerox Incremental Parser) (Ait-Mokhtar *et al.*, 2002), afin d'obtenir un arbre de constituants ainsi que des informations complémentaires utiles à la génération de questions telles que les informations morphologiques et l'étiquetage en entités nommées (villes, pays, dates, personnes, ...).

Construction de syntagmes complexes et deuxième phase de simplification. Les arbres de constituants produits par XIP ne donnent pas la structure syntaxique complète car cette information est par ailleurs disponible sous forme de relations de dépendance. Nous avons donc complexifié la structure de l'arbre de manière à ce qu'elle rende compte des éléments grammaticaux nécessaires à la génération de questions. En pratique, nous avons rassemblé sous un même nœud les éléments qui, ensemble, réalisent une seule fonction grammaticale. Parallèlement, une fois les syntagmes complexes repérés, nous avons supprimé les éléments superflus dans la phrase, tels certains adverbes ou syntagmes prépositionnels. En outre, quand plusieurs syntagmes nominaux sont réunis sous un même nœud (voir figure 1), la tête du syntagme est explicitement balisée dans l'arbre, afin d'utiliser les traits morpho-syntaxiques de la tête ultérieurement. Les nœuds sont regroupés à l'aide de 25 règles de transformation écrites dans le formalisme de Tregex et Tsurgeon et appliquées successivement à l'arbre². Tregex et Tsurgeon³ sont deux outils de manipulation d'arbres syntaxiques (Levy & Galen, 2006). Tregex permet d'explorer les arbres syntaxiques *via* un langage d'expressions régulières adapté et puissant. Ce formalisme fonctionne à partir d'une combinaison de patrons utilisant principalement des relations de fratrie et de filiation entre les nœuds dans l'arbre. Tsurgeon permet de transformer la structure des arbres sur la base des nœuds spécifiques identifiés par Tregex.

But : Réunir sous un même nœud NP un groupe nominal et un complément du nom
(ex : **Le chat de Jean** dort.)

- Tregex : NP=nom \$+ PP=cn [<</de/] (identification d'un NP suivi d'un PP dominant « de »)
- Tsurgeon : move cn > -1 nom (déplacement du groupe prépositionnel PP sous le nœud NP)

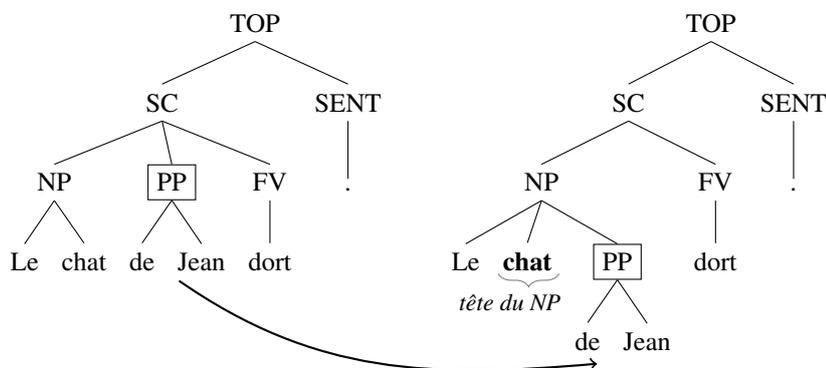


FIGURE 1 – Exemple de règle de regroupement de constituants

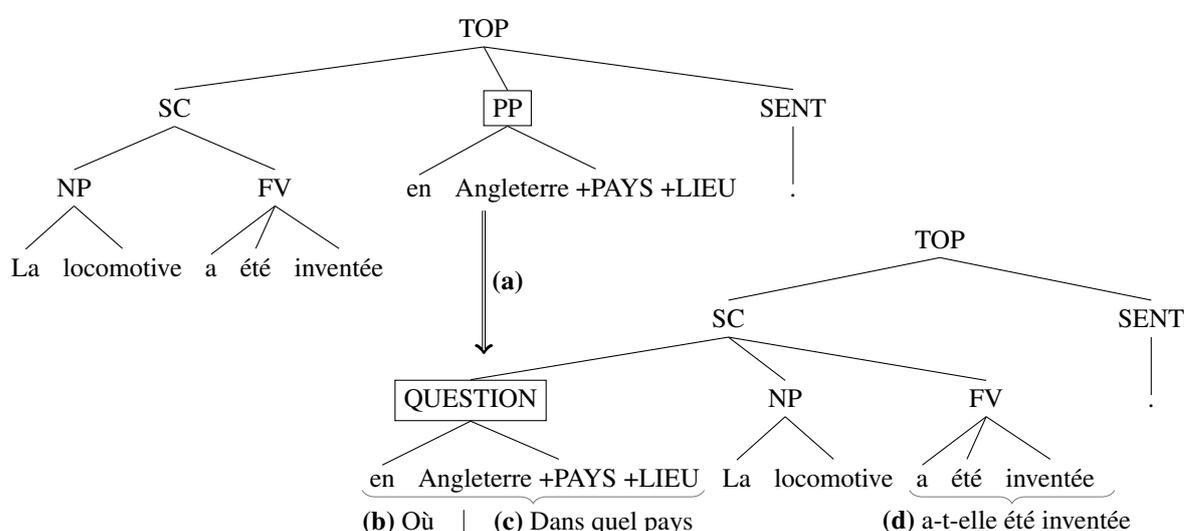
2. Ces regroupements pourraient également être réalisés à l'aide des relations de dépendance.

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

4.2 Génération de questions

Structure des règles de génération. Chaque règle de génération se décompose en une étape d'identification de la cible de la question et une étape de transformation. La phase d'identification utilise des expressions Tregex pour vérifier que les constituants nécessaires à la génération d'une question donnée soient bien présents dans la phrase. S'ils sont présents, on les marque comme tels à l'aide d'une étiquette spécifique. Une fois les constituants remarquables identifiés à l'aide de Tregex, la phrase est soumise à des transformations profondes à l'aide de Tsurgeon. Si aucune des expressions Tregex ne s'applique à la phrase, alors aucune question n'est générée.

Règles de génération. Le système comprend 9 règles de génération portant sur divers constituants de la phrase (sujet, complément d'objet direct, attribut du sujet, complément d'objet indirect, complément de lieu et complément de temps), ainsi qu'une règle de génération de questions fermées. La figure 2 illustre la génération d'une question de lieu. Les autres règles de génération fonctionnent sur le même principe.



1. Identification du complément de lieu :

PP=loc[<</LIEU/] \$-- (SC < (FV=verb \$-- NP=subj))

2. Transformation :

- (a) Déplacement du complément de lieu en tête de phrase
- (b) Remplacement du complément de lieu par « où »
- (c) Si le complément de lieu est un pays, une région, ou une ville, on peut le remplacer par, respectivement « dans quel pays », « dans quelle région » « dans quelle ville ». À terme, une procédure de génération de paraphrases plus complexe est envisagée mais le système actuel se limite à ces quelques règles.
- (d) Inversion Sujet-Verbe avec répétition d'un pronom, accordé avec le sujet, si celui-ci n'en est pas déjà un.

Exemple : La locomotive a été inventée en Angleterre. ⇒ Où la locomotive a-t-elle été inventée ? Dans quel pays la locomotive a-t-elle été inventée ?

FIGURE 2 – Génération de questions sur le complément de lieu.

4.3 Post-traitements

Les étapes de post-traitement permettent, à partir de la structure profonde de la question générée, de récupérer la question de surface, nettoyée à l'aide d'expressions régulières (suppression des éléments non-textuels, tels que les traits morfo-syntactiques, normalisation de la casse, suppression des espaces superflus, gestion des élisions...).

5 Évaluation

Nous avons procédé à une première évaluation manuelle des questions générées à partir de textes correspondant à des niveaux de difficulté langagière différents. Pour ce faire, trois corpus de 10 phrases chacun ont été utilisés. Le premier sous-corpus, appelé WIKI-SIMPLE, a été extrait de Wikimini et Vikidia, deux encyclopédies Web de type Wikipédia mais adaptées aux enfants. Le deuxième corpus, appelé WIKI, est quant à lui issu de Wikipédia et traite de sujets similaires au premier, tandis que le troisième corpus, appelé PRESSE, est tiré d'articles de presse en ligne. Les phrases ont été choisies aléatoirement dans les trois sources mentionnées⁴.

À partir de ces 30 phrases, 78 questions ont été automatiquement générées par le système (36 pour WIKI-SIMPLE, 21 pour WIKI et 21 pour PRESSE) et évaluées par 9 à 17 annotateurs humains différents⁵. Les évaluateurs devaient donner une mesure de la qualité des questions générées, sur une échelle allant de 1 à 4, la valeur 1 correspondant à une question irrecevable et grammaticalement incorrecte et la valeur 4 indiquant une question parfaite, grammaticale, bien formée et pertinente. Des explications et un exemple étaient également fournis.

Le système de génération de questions obtient un score moyen de 2,87 pour toutes les questions confondues, ce qui est supérieur au score de 2,5 marquant le milieu de notre échelle de valeurs allant de 1 à 4. De plus, 41 % des questions se voient attribuer un score majoritaire de 4 et 29,5 % un score majoritaire de 3. L'accord inter-annotateurs, mesuré à l'aide du score kappa de Fleiss (Fleiss, 1971) sur chaque sous-corpus, varie de 0,3 à 0,381. Ceci correspond à un accord faible, mais l'accord est nettement supérieur pour les valeurs extrêmes (1 et 4) que pour les valeurs centrales, avec un kappa maximum de 0,622. Ceci indique que la différence entre les valeurs centrales était sans doute moins bien perçue par les évaluateurs et pose la question de la pertinence d'une échelle aussi précise. Nous indiquons ici pour chaque score un exemple de question qui l'a obtenu majoritairement :

1. Le riz est la céréale la plus utilisée comme aliment par l'homme. → Que Le riz est plus utilisée comme aliment par l'homme ?
2. Un train est composé de plusieurs voitures (pour transporter des personnes) et/ou de plusieurs wagons (pour transporter des marchandises), et peut être tracté par une locomotive. → De quoi Un train est-il composé et peut être tracté par une locomotive ?
3. En France, pour permettre une meilleure fréquentation de ses trains, la SNCF a mis en service le TGV en 1981. → En quelle année la SNCF a-t-il mis en service le TGV ?
4. L'observatoire volcanologique du Piton de la Fournaise enregistre depuis le 14 août une augmentation lente mais continue du nombre de séismes. → Qu'enregistre L'observatoire volcanologique du Piton de la Fournaise depuis le 14 août ?

On obtient globalement de meilleurs résultats sur les questions générées à partir de phrases soit très courtes, soit très longues (ce qui confirme l'intérêt d'un algorithme de simplification de phrases tel que nous l'avons développé). Mais les phrases longues conduisent à la génération d'un nombre plus restreint de questions. Quant au type de questions, on retrouve les scores les plus élevés sur les questions fermées, de type « est-ce-que » (moyenne : 3,25), puis sur les questions sur le sujet (moyenne : 2,92). Ces types sont aussi proportionnellement les plus fréquents puisqu'ils rassemblent respectivement 22 et 24 questions (environ 60% des questions générées). Les autres types obtiennent des résultats légèrement inférieurs (2,34 de moyenne pour les 16 questions sur le complément direct, 2,35 pour 7 questions sur le complément indirect et 2,57 pour 9 questions sur les compléments de lieu et de temps).

Quelques problèmes récurrents ont pu être identifiés suite à une analyse détaillée des questions générées, tels que des majuscules intempestives⁶, des problèmes d'accord ou des erreurs dues aux règles implémentées qui donnaient lieu à l'extraction d'un syntagme cible incorrect pour la question. Parmi ceux-ci, certains étaient liés à l'analyseur, conduisant par exemple à un mauvais accord entre le sujet et le pronom (exemple : *Quand la SNCF a-t-il mis en service le TGV ?* où *la SNCF* est considéré comme masculin par l'analyseur). D'autres problèmes pourraient être résolus par l'amélioration de différents aspects du système de génération, en particulier par une utilisation plus systématique des informations sur les relations de dépendance. Notons aussi qu'une prise en compte plus approfondie de la « sémantique » du texte permettrait de préciser davantage les questions, notamment en résolvant les anaphores, qui forment des constituants non pertinents pour la réponse aux questions.

4. <http://fr.wikimini.org>, <http://fr.vikidia.org>, <http://www.wikipedia.org>

5. Pour des raisons pratiques, les questions générées ont été divisées en quatre parties différentes, évaluées chacune par des annotateurs différents. Par ailleurs, le nombre d'annotateurs pour chaque sous-partie n'était pas forcément le même.

6. Ces problèmes devraient pouvoir être résolus assez facilement, par l'utilisation de listes de noms propres qui permettraient de déterminer les termes qui doivent conserver leur majuscule.

6 Conclusion et perspectives

Nous avons présenté un système de génération automatique de questions pour le français. Le système de génération procède par transformation de phrases déclaratives en interrogatives et se base sur une analyse syntaxique préalable de la phrase de base. Il s'agit, à notre connaissance, du premier système de ce type pour le français. Une évaluation préliminaire a montré que les questions générées par le système sont par ailleurs de bonne qualité : 41 % d'entre elles sont considérées comme parfaites par des évaluateurs humains. Les perspectives d'amélioration de l'outil sont nombreuses : meilleure simplification automatique des phrases de base, utilisation plus systématique des relations de dépendance, reformulation des questions via la génération automatique de paraphrases (Tomuro, 2003), implémentation d'un système de classification automatique des questions à l'aide d'un algorithme d'apprentissage supervisé (Heilman & Smith, 2010b) ou encore génération de questions transversales au sein d'un paragraphe. Les applications potentielles du système sont également variées : génération de questions pour systèmes de question-réponse interactifs et génération automatique de questionnaires pour des applications éducatives.

Références

- AIT-MOKHTAR S., CHANOD J.-P. & ROUX C. (2002). Robustness beyond shallowness : incremental deep parsing. *Natural Language Engineering*, **8**(3), 121–144.
- DANG H. T., LIN J. & KELLY D. (2006). Overview of the TREC 2006 Question Answering Track. In *Proceedings of the Fifteenth Text REtrieval Conference (TREC 2006)*.
- FLEISS J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, **76**(5), 378–382.
- GATES D. M. (2008). Automatically generating reading comprehension look-back strategy questions from expository texts. Master's thesis, Carnegie Mellon University.
- GIAMPICCOLO D., FORNER P., PEÑAS A., AYACHE C., CRISTEA D., JIKKOUN V., OSENOVA P., ROCHA P., SACALEANU B. & SUTCLIFFE R. (2007). Overview of the CLEF 2007 Multilingual Question Answering Track. In A. NARDI & C. PETERS, Eds., *Working Notes for the CLEF 2007 Workshop*, Budapest, Hungary.
- GREVISSE M. (1975). *Le bon usage. Grammaire française avec des remarques sur la langue française d'aujourd'hui*. Gembloux : Duculot, 10e édition.
- HEILMAN M. & SMITH N. A. (2009). *Question Generation via Overgenerating Transformations and Ranking*. Technical Report CMU-LTI-09-013, Language Technologies Institute, Carnegie Mellon University.
- HEILMAN M. & SMITH N. A. (2010a). Extracting Simplified Statements for Factual Question Generation. In *Proceedings of the 3rd Workshop on Question Generation*, Pittsburgh.
- HEILMAN M. & SMITH N. A. (2010b). Good Question! Statistical Ranking for Question Generation. In *Proceedings of the NAACL-HLT 2010*, p. 609–617.
- KALADY S., ELIKKOTTIL A. & RAJARSHI D. (2010). Natural language question generation using syntax and keywords. In *Proceedings of the 3rd Workshop on Question Generation*, Pittsburgh.
- LANGACKER R. W. (1965). French interrogatives : a transformational description. *Language*, **41**(4), 587–600.
- LEVY R. & GALEN A. (2006). Tregex and Tsurgeon : tools for querying and manipulating tree data structures. In *5th International Conference on Language Resources and Evaluation*.
- MITKOV R., HA L. A. & KARAMANIS N. (2006). A computer-aided environment for generating multiple-choice test items. *Natural Language Engineering*, **12**(2), 177–194.
- PIWEK P. & STOYANCHEV S. (2010). Generating Expository Dialogue from Monologue : Motivation, Corpus and Preliminary Rules. In *Proceedings of NAACL-HLT 2010*, p. 333–336.
- PRENDINGER H., PIWEK P. & ISHIZUKA M. (2007). Automatic generation of multi-modal dialogue from text based on discourse structure analysis. In *Proceedings of ICSC-07*, p. 27–34, Irvine, CA, USA.
- TOMURO N. (2003). Interrogative reformulation patterns and acquisition of question paraphrases. In *Proceedings of the second international workshop on Paraphrasing*, p. 33–40.
- WOLFE J. H. (1976). 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.