A minimally supervised approach for question generation: what can we learn from a single seed?

Sérgio Curto, Ana Cristina Mendes and Luísa Coheur

Spoken Language Systems Laboratory - L^2F/INESC-ID
Instituto Superior Técnico, Technical University of Lisbon
R. Alves Redol, 9 - 2º– 1000-029 Lisboa, Portugal
{sergio.curto,ana.mendes,luisa.coheur}@l2f.inesc-id.pt

Abstract. In this paper, we investigate how many quality natural language questions can be generated from a single question/answer pair (a seed).
In our approach we learn patterns that relate the various levels of linguistic information in the question/answer seed with the same levels of information in text. These patterns contain lexical, syntactic and semantic information and when matched against a target document, new question/answer pairs can be generated. Here, we focus specifically on the task of generating questions.
Several works, for instance in Question Answering, explore the re-writing of questions to create (usually lexical) patterns; instead, we use several levels of linguistic information – lexical, syntactic and semantic (through the use of named entities). Also, the patterns are commonly hand-crafted, as opposed to our strategy where patterns are automatically learned, based on a single seed.
Preliminary results show that with the single question/answer seed pair – *"When was Leonardo da Vinci Born?"/1452* – we manage to generate several questions (from documents related with 25 personalities), from which 80% were evaluated as plausible.

1 Introduction

Question Generation (QG) has became an appealing line of research. Several workshops have been exclusively dedicated to this topic, including a shared evaluation challenge with the goal of generating questions from paragraphs and sentences [17].

The interest in QG has recently increased for several reasons. On one hand, generating questions (and answers) can be useful for Question-Answering (QA) or Dialogue Systems, as QG can act as a provider of questions to train a system to operate in a new domain. On the other hand, QG shows potential in tasks related with knowledge assessment, in two different perspectives: by reducing the amount of time allocated for the creation of tests by teachers which, if done manually, can be a time consuming trial and error process; by allowing the self evaluation of the knowledge acquired by learners.

Several works dedicated to QG rely in a set of hand-crafted rules, for instance [20, 4]. THE-MENTOR, a platform for the automatic generation of multiple-choice tests, is an exception to this trend: THE-MENTOR employs a minimally supervised approach to learn patterns from a set of seeds. These patterns will be later used to discover new questions and answers in text. The architecture of THE-MENTOR is composed by two main modules:

- The first module is dedicated to *learning patterns*. It receives as input a set of Question/Answer (Q/A) seeds and exploits the large amount of information in the Web to learn lexico-syntactic patterns that relate the question and answer of each seed.
- The second module is responsible for the *generation of new question/answer pairs*. It takes as input a text and the previously learned patterns and returns a new set of Q/A pairs that result from the match of the patterns against the text. Since the category of the seed question in the origin of the matched pattern is known, it also extracts distractors from the text. A multiple-choice test created by THE-MENTOR is, thus, composed by a question, one correct answer and a list of incorrect answers (distractors).

In this paper, we focus on the tasks of pattern learning and question generation[1] in THE-MENTOR. Our goal is to investigate how many quality natural language questions can be generated from a single seed. In particular, how many of those are semantically similar to the seed question. We also describe the advantages and problems that arise from using lexical and syntactic information in THE-MENTOR and detail how the system was adapted to add a semantic level to the learned patterns.

This paper is structured as follows: in Section 2 we present related work; in Section 3 we describe THE-MENTOR; in Section 4 we describe our experiments and in Section 5 we discuss the achieved results. We conclude in Section 6 where we also point to future work directions.

2 State of the Art

QG has recently called the attention of the research community; however, there is still a huge space for research in this topic. Some works in QG focus on the question taxonomies [7, 3]; others discuss the importance/impact of the nature of the information source [4], or point to ways of pre-processing the information sources [9, 11]. Although many QG systems use the Wikipedia as information

[1] Despite its importance to the generation of multiple-choice tests, we will not consider the generation of answers and distractors.

source [8], the QG system of [1] is executed over the Full Option Science System (FOSS). [20] refer the OpenLearn repository, which, besides covering a wide range of materials (in different formats), is also only authored by experts.

Regarding THE-MENTOR, it uses Li and Roth's two-layer taxonomy [13], that consists of a set of six coarse-grained categories (ABBREVIATION, DESCRIPTION, ENTITY, HUMAN, LOCATION and NUMERIC) and fifty fine-grained ones (*e.g.*, HUMAN:INDIVIDUAL, LOCATION:CITY, LOCATION:COUNTRY and NUMERIC:DATE). This taxonomy is widely used by the machine learning community [18], since their authors have published a set of nearly 6,000 labeled questions (the University of Illinois at Urbana-Champaign dataset), freely available on the Web. Concerning the information sources, THE-MENTOR uses the Web.

Different approaches to QG are described in the literature. Both QA and QG use patterns to bridge the gap between the question and the sentence in which the answer can be found. The idea is that the answer to a question will probably occur in sentences that contain a rewrite of the original question. The winning system of the QA track of TREC-10 [19] reports a list of such (surface) patterns. This line of work has been extensively followed, but these patterns are usually hand-made. [4], for instance, identifies key points, builds a situation model and uses question templates to generate questions from these; in order to generate questions, the Ceist system [20] uses Tregex [12], which receives as input a set of hand-crafted rules that extract questions from a given text.

THE-MENTOR also uses patterns to extract questions from the sources. However, contrary to other systems, the process of building these patterns is automatic, given a set of Q/A seeds. Our pattern learning algorithm was inspired in the one by [16], who use a question-answer seed pair to bootstrap new question rewrite patterns, which are afterwards validated. In our work, we do not validate the learned patterns. Moreover, the patterns learned by [16] only contain lexical information, for instance "<NAME> was born on <ANSWER>"; ours, however, are also based on syntactic and semantic information (like "NP VBD VBN in <ANSWER>") and, with this, we allow more flexibility in the matched text segments.

3 THE-MENTOR

In this section we present the overall architecture of THE-MENTOR. We detail how we extended the learned lexico-syntactic patterns with semantic information and describe the question generation step.

3.1 Overall Architecture

THE-MENTOR generates multi-choice tests about a free text document and it is composed by two main modules that perform the following tasks:

Learning lexico-syntactic patterns – A set of Q/A seeds is used to bootstrap patterns that relate questions with answers. Patterns are learned from the Web, and we exploit its linguistic variability to create plausible patterns. Moreover, we perform verb expansion and allow several types of patterns, according to the precision of the match against the original seeds.

Generation of test items – The learned patterns extract sentences where new Q/A pairs can be found: answers are explicitly present in the sentence and the respective questions can be built with information from the seed question. In order to discard low quality items, several filters are applied. Distractors are named entities that comply with the question category, extracted from the same text.

Afterwards, the user can evaluate the quality of the generated test items through a Web interface. The multiple-choice test will be composed by the test items the user considered as having quality.

3.2 Extending Patterns with Semantics

As previously said, our algorithm for learning patterns is based on the bootstrapping technique presented in [16], and involves the following two stages.

First, we use a Q/A seed pair – composed by a natural language question and its correct answer – to bootstrap lexico-syntactic patterns that relate questions and answers. For instance, given a seed pair *"Who painted the Birth of Venus ?"/Botticelli* and the syntactic structure of its question [$_{\text{WHNP}}$ Who] [$_{\text{VBD}}$ painted] [$_{\text{NP}}$ the Birth of Venus][2], the pattern "$\{$*ANSWER*$\}$ `has` *VBN*$_{\text{VBD}}$ *NP*"[3] is learned from the sentence *Botticelli has painted the Birth of Venus*. Note that there is a direct correspondence between the syntactic components of the seed question (except the Wh-phrase), the syntactic components of the sentence and the syntactic tags in the learned patterns. The lexical information that composes the patterns refers to the tokens found and caught in the sentence, but not present in the seed question.

THE-MENTOR learns three types of patterns, namely:

[2] Here, as well as throughout the entire paper, the Penn Treebank II Tags [2] are used.

[3] THE-MENTOR patterns are more complex than the ones presented, as they are explicitly linked to their seed question by indexes, mapping the position of each one of its components into the question components. However, for the sake of simplicity, these indexes are omitted.

1. STRONG patterns contain all phrases (and their contents) of the seed question. For instance, the STRONG pattern "{*ANSWER*} *VBD NP*" can be learned from the sentence *Botticelli painted the Birth of Venus*;
2. INFLECTED patterns are learned after the inflection of the main verb of the question in its various tenses. The auxiliary verb is not required to be in the pattern. For instance, the INFLECTED pattern "{*ANSWER*} `began` *VBG NP*" can be learned from the sentence *Botticelli began painting the Birth of Venus*;
3. WEAK patterns do not contain any verb phrase from the question. For instance, the WEAK pattern "{*ANSWER*}`'s` *NP*" can be learned from the sentence *Botticelli's Birth of Venus*.

The decision of learning different types of patterns resulted from the fact that, if patterns are too specific (STRONG patterns), they will not frequently match and not many questions will be generated; if patterns are too generic (WEAK patterns), the quality of the generated tests decreases. In [14] we concluded that the parameter that influenced the most the quality of the generated Q/A pair was the type of pattern on its origin. That is, STRONG patterns generate better Q/A pairs, however in a lower number, and WEAK patterns generate low quality Q/A pairs. Our goal is, thus, to learn the highest number possible of STRONG patterns, since they will lead to better quality questions and answers.

We noticed that the use of lexico-syntactic information alone was problematic when THE-MENTOR tries to match the patterns against the target documents. If a pattern like "*NP* `was` *VBN* `around 1486 by` {*ANSWER*}" is learned (and validated), it is unlikely that it will match a given text segment, since for that the text has to explicitly contain the expression "around 1486 by". The introduction of semantic information could solve such issues.

Indeed, a closer look to the pattern "*NP* `was` *VBN* `around 1486 by` {*ANSWER*}" can tell that this can be a more useful pattern if, instead of considering the instance of date (the year of 1486), we simply consider it as a named entity of type DATE. Being so, we extended the patterns with semantic information, to make such patterns valid and apt to find new Q/A pairs. For that purpose, we now include reference to the named entities that appear in the pattern. The previous pattern is now represented as follows: "*NP* `was` *VBN* `around 1486`*[Date]* `by` {*ANSWER*}"

We apply several strategies to recognize the named entities in the pattern, namely:

– A machine learning-based named entity recognizer (specifically, we used the Stanford's Conditional Random-Field-based named entity recognizer [6]) to detect entities of type HUMAN – including person's names.

– Regular expressions to detect NUMERIC and DATE type entities;
– Gazetteers to detect entities of type LOCATION, like cities and countries.

Note that the named entity recognition is performed uniquely on the lexical components of the pattern.

3.3 Question Generation

The question (and answer) generation step takes as input the previously learned set of patterns and a parsed text (a sentence, paragraph or a full document) from which the questions are to be generated. Afterwards, the patterns are matched against the given text.

The addition of a new level of linguistic information to the learned patterns implied the extension of our pattern matching algorithm in order to deal with the semantics in the pattern. The match is done at the lexical, syntactical and semantical levels and goes as follows:

1. Test if any segment in the sentence has the same syntactic structure as the learned pattern. This means that, in a depth-first search of the tree, the sequence of syntactic tags in the pattern has to occur. If so, collect the text segment and continue to the next step;
2. If the pattern does not contain any lexical information, continue to 6.;
3. If the pattern contains lexical information, test if the words in the pattern match with the tokens in the leaves of the parse tree of the text segment, in the same position;
4. If there is lexico-syntactic overlap between the pattern and the text segment, proceed to 6.;
5. If there is no lexico-syntactic overlap, extract all the named entities from the text segment of the same types as those in the pattern. If there is a lexico-semantic correspondence – the semantic categories of the named entities match and the remaining tokens also match – proceed to the next step;
6. Return the matched text segment.

Lastly, we assure that the tokens that represent the answer agree with the semantic category of the question. For that purpose, and since we have the mapping between question categories and types of named entities, we simply test if the answer is or contains a named entity of the expected type. Being so, the pattern "`NP was VBN around 1486[Date] by {ANSWER}`" matches the text segment *Guernica was painted around 1937 by Picasso*, given that 1937 is identified as a named entity of type DATE.

After the discovery of the text segments that match the learned patterns, the next step is the generation of new questions. This is done according to the type of matched pattern: STRONG, INFLECTED or WEAK. Since we keep track of the syntactic structure of the questions and the sentences on the origin of the patterns, the generation of new questions from the extracted text segments goes as follows:

- both STRONG and INFLECTED patterns result in a direct unification of all the components of the segment with the pattern's components. However, in the INFLECTED patterns, the verb is inflected with the tense and person existing in the question and the auxiliary in the question is also used;
- WEAK patterns lead to the unification of the segment components with the respective pattern components, and for all the components that do not appear in the segment, the components in the question are used.

4 Experiments

In this section we describe the experiments we conducted to understand how the automatic generation of questions is influenced by the use of only one Q/A seed pair.

4.1 Experimental Setup

Regarding the task of learning patterns, we used a single biographical question about Leonardo da Vinci and its respective answer as Q/A seed pair: *"When was Leonardo Da Vinci born?"/1452*.

The syntactic analysis of the question was done by the Berkeley Parser [15] trained on the QuestionBank [10], a treebank of 4,000 parse-annotated questions. In what concerns its classification, and as previously mentioned, we used Li and Roth's two-layer question taxonomy [13] and a machine learning-based classifier (SVM) fed with features derived from a rule-based classifier [18]. Being so, the syntactic structure of the question resulted to be `WHADVP VBD NP VBN` and the question category is NUMERIC:DATE.

The top 8 documents retrieved by the Web search engine Google were used to learn patterns.[4] Since THE-MENTOR sends multiple queries to the search engine, with the several permutations of the phrases that compose the seed question, the total number of retrieved documents considered in this step is, for each question, eight times the number of permutations.

For the generation of new questions, we use uniquely the STRONG and INFLECTED patterns found in the learning step. As target documents from which to

[4] Note that the number and nature of the documents used as source of information where the system finds the patterns is not fixed, being a topic for future work.

generate the questions, we collected the top 16 Web pages retrieved by Google for 25 different personalities belonging to 5 different domains, namely: writers (*e.g.*, Jane Austen), painters (*e.g.*, Pablo Picasso), politicians (*e.g.*, John F. Kennedy), musicians (*e.g.*, Jim Morrison) and actors (*e.g.*, Marlon Brando). We filtered out the html tags and images from the Web Pages, but no other preprocessing task was performed, like anaphora resolution or text simplification.

4.2 Pattern Learning

Regarding the task of learning patterns, THE-MENTOR was able to learn, from a single Q/A seed pair, a total of 24 patterns. Table 1 shows the distribution per type of the patterns learned, as well as some examples.

Type	# Learned Patterns	Examples
STRONG	10	"*NP VBD VBN* `in` {*ANSWER*}" "*NP VBD VBN* `in April`*[Date]* {*ANSWER*}"
INFLECTED	10	"*NP*, *VBN* `in` {*ANSWER*}" "{*ANSWER*} *NP* `is` *VBN*"
WEAK	4	"*NP* `lived from` {*ANSWER*}" "{*ANSWER*} `-- Renaissance polymath` *NP*"
Total	24	

Table 1. Learned patterns for the Q/A seed pair *"When was Leonardo Da Vinci born?"/1452*.

The learned patterns relate a noun phrase to a date. Recalling the Q/A seed pair, they associate the entity *"Leonardo da Vinci"* to his year of birth *1452*.

In the WEAK patterns, the relation between the noun phrase and the date can only be discovered if we analyze the context of the matched text segment. For instance, the pattern "{*ANSWER*} `-- Renaissance polymath` *NP*" does not explicitly indicate that the relation between the noun phrase and the answer is of birth date. Indeed, WEAK patterns have the disadvantage of matching sentences where the semantics is completely different from that on the question seed, and the relation between the noun phrase and the date has to be discovered afterwards.

In the STRONG and INFLECTED patterns, the relation between the noun phrase and the date is expressed by the verb phrase. This allows THE-MENTOR to match text segments and generate questions with similar semantics to that of the seed question (*e.g.*, *Mozart was born in 1756*) or with different semantics (*e.g.*, *Mein Kampf was published in 1925.*).

At this stage, the usefulness and applicability of these patterns is merely indicative, since their quality in the context of THE-MENTOR can be only assured if they allow to generate new quality questions.

4.3 Generation of Questions

Here we detail the evaluation of the questions generated by THE-MENTOR.

Evaluation Guidelines We follow the guidelines of [4], who classify questions as being *plausible* – if they are grammatically correct and if they make sense regarding the text from where they were extracted – and *implausible* (otherwise). However, we refine each of the previous definitions and consider that plausible questions can be split into three different types:

- PL-A for plausible, pronominal anaphoric questions. That is, if the question is well formulated at the lexical, syntactical and semantic levels and makes sense in the context of the sentence that originated it, but contains a pronoun. For instance *When was she awarded?* is marked as PL-A;
- PL-C for plausible questions that need a context to be answered. That is, if the question is well formulated at the lexical, syntactical and semantic levels and makes sense in the context of the sentence that originate it, but can only be understood if the user is aware of that context. For instance *When was the manuscript published?* is marked as PL-C; [5]
- PL, a plausible question, without pronouns and self contained in the sense that it is clear what is being asked, regardless of the context. For instance *When was the Andy Warhol Museum announced?* is marked as PL.

  By the same token, we split implausible questions in:

- IMPL-I: incomplete questions. As example, question *When was he invited* is marked as IMPL-I (the snippet that originated it was *In early 1980 he was invited to perform at the official ceremony celebrating the transformation of the former British colony Rhodesia into the independent state of Zimbabwe*).
- IMPL: a question that makes no sense or is not supported by the context from which it were extracted. For instance, the question *When was December 1926 Agatha identified* makes no sense.

Results THE-MENTOR could generate 34 questions from 20 learned patterns (of types INFLECTED and STRONG), meaning that 4 of the learned patterns were completely useless. The evaluation results, according to the evaluation guidelines, are summarized in Table 2.

Concerning the total generated questions, most of them are considered plausible. While 20% of the questions (7 in 34) are considered implausible, and

[5] The PL-C questions are a general case of the PL-A questions. We have decided to clearly separate these two situations.

PL	PL-A	PL-C	IMPL	IMPL-I	Total
10	6	11	2	5	34

Total PL	Total IMPL
27	7

Table 2. Evaluation of the generated questions.

should be almost totally reformulated or discarded, nearly 30% of the questions (10 in 34) can be used without any further modifications. The remaining 50% require anaphora resolution or need context from the surrounding sentences to be fully understood.

Regarding the nature of the generated questions, only 6 of the 34 generated questions are biographical questions about the birth date of a personality. This result is somewhat understandable because we do not impose any constraint to the target documents as they should contain the birth date of the personality and, even in the cases where the birth year is in the document, it is not sure that this information is stated in a way that the learned patterns can capture them. Nevertheless, all these questions were marked as PL. For instance, *"When was Francisco Goya born?"* was a question correctly generated by THE-MENTOR.

All the other 24 questions are semantically different from the seed, like *"When was Jackson arrested?"* or *"When was a common-law marriage abolished?"*. Indeed, the system's performance decreased in the generation of these questions. However, note that here lies the potential of using syntactic information in the patterns: the patterns are flexible enough that allow to generate questions that are not semantically related with the Q/A seed pair.

Inter-annotator agreement The Cohen's kappa coefficient [5] was used to calculate the inter-annotator agreement. Two annotators evaluated the questions generated from the Web pages retrieved for the 25 personalities.

Annotators agreed in 33 of the 34 questions. The only difference lied in a question classified as IMPL-I by one annotator and as IMPL by the other. Therefore, the relative observed agreement among raters, Pr(a), is 0.97, the probability of chance agreement, Pr(e), is 0.24 and, finally, the kappa coefficient (K)[6] is 0.96, which is, as considered by some authors an almost perfect agreement.

5 Discussion

Although we only used one seed Q/A pair, our approach for learning patterns based on the linguistic variability of large-scale information sources, resulted

[6] K = (Pr(a)-Pr(e))/(1-Pr(e)).

in a total of 24 linguistically rich patterns. Most of these patterns are of types INFLECTED and STRONG, which we have found, in a previous work, to generate the best questions. These learned patterns lead to a set of 34 new questions, from which nearly 80% are considered plausible, requiring little or no modification. Some of these questions are directly related with the semantics of the seed, but others focus on different topics.

The possibility of generating questions semantically different from the seed question is due to the use of syntactic information in the learned patterns. Particularly the verb phrase that we include in the INFLECTED and STRONG patterns allow us to generate questions that are mostly considered as grammatically correct, making sense in the context from where they were extracted. However, the best results of THE-MENTOR are achieved for the questions that are semantically equivalent to the seed question, that is, that ask for the birth date of the personality under consideration.

The question chosen as seed of extreme importance, as it happens in approaches based in minimally supervised learning. Indeed, THE-MENTOR is highly dependent on the nature of these seed questions. In this work, we have chosen a simple factoid question, well studied by the QA community, classified as NUMERIC:DATE. Being so, all the generated questions belong to this category. Moreover, the entity in the seed question (*Leonardo da Vinci*) is a fairly known and acknowledged personality, with many references to it in the Web; thus, the probability of finding documents containing the birth date of Leonardo da Vinci is also high. This is, obviously, a constraint to the learning task process: if a question that will match a few set of texts is given as seed, the number of learned patterns will be much less.

Regarding the target documents against which the patterns were matched and from which the questions were generated, we used a set of documents retrieved by the search engine for several personalities. We had no control on these documents. One of the consequences of this decision was visible in a later analysis of this corpus: some documents were not directly related with the personality we asked for. For instance, a page about the JFK airport was retrieved for the query *"John F. Kennedy"*.

6 Conclusions and Future Work

In this paper we investigated the number and quality of the questions generated from a unique seed. We presented our approach to generate questions based on patterns learned using a minimally supervised learning method. We described THE-MENTOR and our efforts to add a semantic level to the used patterns, and how it influenced the tasks of pattern learning and Q/A pairs generation.

With a single seed, composed by a natural language question and its respective correct answer, THE-MENTOR was able to learn several patterns that lead to a set of questions, most of which were considered as having quality. Regarding the semantics of the generated questions, all those that are related with the seed question were considered plausible; nevertheless, several questions were also generated that were not semantically related with the seed question, and most of these were also plausible.

Currently, we are working on the actual bootstrapping technique within THE-MENTOR. Given that the newly generated questions are of quality, we are using them as seed to learn new patterns which will afterwards give rise to new generated questions.

A line for future work direction is the application of algorithms for co-reference resolution, both to texts where the patterns are learned and those where the questions (and answers) are generated. We believe that, when we start dealing with this linguistic phenomenon, THE-MENTOR's results will largely improve. Also regarding the pre-processing of texts, we intend to apply normalization in order to reduce the linguistic variations of the same entity.

Acknowledgments

This work was supported by FCT (INESC-ID multiannual funding) through the PIDDAC Program funds, and also through the project FALACOMIGO (ProjectoVII em co-promoção, QREN n 13449).

Ana Cristina Mendes is supported by a PhD fellowship from Fundação para a Ciência e a Tecnologia (SFRH/BD/43487/2008).

References

1. Becker, L., Nielsen, R.D., Ward, W.H.: What a pilot study says about running a question generation challenge. In: The 2nd Workshop on Question Generation. Brighton, England (July 2009)
2. Bies, A., Ferguson, M., Katz, K., Macintyre, R., Contributors, M., Tredinnick, V., Kim, G., Marcinkiewicz, M.A., Schasberger, B.: Bracketing guidelines for treebank ii style penn treebank project (1995)
3. Boyer, K.E., Lahti, W., Phillips, R., Wallis, M., Vouk, M., Lester, J.: An empirically derived question taxonomy for task oriented tutorial dialogue. In: The 2nd Workshop on Question Generation. Brighton, England (July 2009)
4. Chen, W., Aist, G., , Mostow, J.: Generating questions automatically from informational text. In: The 2nd Workshop on Question Generation. Brighton, England (July 2009)
5. Cohen, J.: A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 20(1), 37 (1960)

6. Finkel, J.R., Grenager, T., Manning, C.: Incorporating non-local information into information extraction systems by gibbs sampling. In: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. pp. 363–370. ACL '05, Association for Computational Linguistics, Stroudsburg, PA, USA (2005), http://dx.doi.org/10.3115/1219840.1219885
7. Forăscu, C., Drăghici, I.: Question generation: Taxonomies and data. In: The 2nd Workshop on Question Generation. Brighton, England (July 2009)
8. Heilman, M., Smith, N.: Ranking automatically generated questions as a shared task. In: The 2nd Workshop on Question Generation. Brighton, England (July 2009)
9. Heilman, M., Smith, N.: Extracting simplified statements for factual question generation. In: The 3rd Workshop on Question Generation. Pittsburgh (June 2010)
10. Judge, J., Cahill, A., van Genabith, J.: Questionbank: creating a corpus of parse-annotated questions. In: ACL-44: Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics. pp. 497–504. Association for Computational Linguistics, Morristown, NJ, USA (2006)
11. Kalady, S., Elikkottil, A., Das, R.: Natural language question generation using syntax and keywords. In: The 3rd Workshop on Question Generation. Pittsburgh (June 2010)
12. Levy, R., Andrew, G.: Tregex and tsurgeon: tools for querying and manipulating tree data structures. In: In LREC 2006 (2006)
13. Li, X., Roth, D.: Learning question classifiers. In: Proceedings of the 19th international conference on Computational linguistics. pp. 1–7. Association for Computational Linguistics, Morristown, NJ, USA (2002)
14. Mendes, A.C., Curto, S., Coheur, L.: Bootstrapping multiple-choice tests with the-mentor. In: CICling 12th International Conference on Intelligent Text Processing and Computational Linguistics (2011)
15. Petrov, S., Klein, D.: Improved inference for unlexicalized parsing. In: Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference. pp. 404–411. Association for Computational Linguistics, Rochester, New York (April 2007), http://www.aclweb.org/anthology/N/N07/N07-1051
16. Ravichandran, D., Hovy, E.: Learning surface text patterns for a question answering system. In: ACL '02: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. pp. 41–47. Association for Computational Linguistics, Morristown, NJ, USA (2002)
17. Rus, V., Wyse, B., Piwek, P., Lintean, M., Soyanchev, S., Moldovan, C.: The first question generation shared task evaluation challenge. In: The sixth international Natural Language Generation Conference. Dublin, Ireland (July 2010)
18. Silva, J., Coheur, L., Mendes, A., Wichert, A.: From symbolic to sub-symbolic information in question classification. Artificial Intelligence Review 35, 137–154 (2011), http://dx.doi.org/10.1007/s10462-010-9188-4, 10.1007/s10462-010-9188-4
19. Soubbotin, M.M.: Patterns of potential answer expressions as clues to the right answers. In: TREC (2001)
20. Wyse, B., Piwek, P.: Generating questions from openlearn study units. In: The 2nd Workshop on Question Generation. Brighton, England (July 2009)