Procesamiento del Lenguaje Natural, Revista nº 51, septiembre de 2013, pp 101-108 recibido 30-04-2013 revisado 16-06-2013 aceptado 21-06-2013

Two Approaches to Generate Questions in Basque*

Dos aproximaciones para generar preguntas en euskera

Itziar Aldabe y **Itziar Gonzalez-Dios** y **Iñigo Lopez-Gazpio,**
Ion Madrazo y **Montse Maritxalar**
IXA NLP Group, University of the Basque Country UPV/EHU
Manuel Lardizabal Pasealekua 1 20018 Donostia
{itziar.aldabe, itziar.gonzalezd, ilopez077, jmadrazo003, montse.maritxalar}@ehu.es

Resumen: En este artículo se presenta un generador de preguntas basado en chunks y otro generador basado en dependencias sintácticas. Ambos generan preguntas en euskera a nivel de frase y utilizan el rasgo de animado/inanimado de los nombres, las entidades nombradas y los roles semánticos de los verbos, así como la morfología de los sintagmas nominales. Se describen dos experimentos de generación de preguntas basadas en textos didácticos, en los que una lingüista analiza la gramaticalidad y lo apropiado de las preguntas generadas a partir de frases simples, así como sus correspondientes pronombres interrogativos.
Palabras clave: generación de preguntas, recursos didácticos, rasgos semánticos, morfosintaxis

Abstract: This article presents a chunker-based question generator (QG) and a QG system based on syntactic dependencies. Both systems generate questions in Basque at the sentence level and make use of the animate/inanimate feature of the nouns, named entities, semantic roles of the verbs, and the morphology of the noun phrases. Two experiments to generate questions were carried out based on educational texts. Then, a linguist analysed the grammaticality and appropriateness of the questions generated from single sentences, as well as their interrogative pronoun.
Keywords: question generation, educational resources, semantic features, morphosyntax

1 Motivation

A new community of interdisciplinary researchers[1] have found a common interest in generating questions.[2] The first workshop on the question generation shared task and evaluation challenge (QGSTEC-2008) began the discussion on the fundamental aspects of question generation (QG) and set the stage for future developments in this emerging area. QG is defined (Rus and Graesser, 2009) as the task of automatically generating questions from some form of input, for which the input could vary from raw text to in-depth semantic representation.

Most of the current QG systems are mainly focused on the generation of questions based on single sentences (Rus and Graesser, 2009; Boyer and Piwek, 2010; Graesser et al., 2011). The generation task contains three steps (Rus and Graesser, 2009): content selection, question type selection, and question construction. The content identification and question type selection (i.e. interrogative pronoun) are usually carried out based on various linguistic information. This information is obtained by means of several natural language processing (NLP) tools: syntactic analysers, named entity recognisers, correference resolution systems and semantic role labellers. In contrast, the question construction is usually based on some transformation rules and patterns.

There is few research work on the generation of questions from paragraphs. An

* Acknowledgments: we thank the linguist Ainara Estarrona for her help in the definition of the wh-word lists related to the semantic roles. This research was partially funded by the IRAKURRI project (SPE12UN091), the Ber2Tek project (IE12-333), and the SKaTeR project (TIN2012-38584-C06-02)

[1] Researchers from various disciplines such as cognitive science, computational linguistics, computer science, discourse processing, educational technologies and language generation.

[2] http://www.questiongeneration.org/

ISSN 1135-5948

approach is presented in Mannem, Prasad, and Joshi (2010) where the generation is based on semantic roles of predicates. Agarwal, Shah, and Mannem (2011) also present a system which generates questions based on more than one sentence. For that task, they use discourse connectives. Other researchers address the task of generating the questions from a more pedagogical or psychological point of view. For instance, Mostow and Chen (2009) present a system based on a situation model which is based on characters' mental states. More recently, Olney, Graesser, and Person (2012) generate questions from concept maps based on psychological theories.

Aldabe, Maritxalar, and Soraluze (2011) have probed the viability of the QG task for Basque language. They use a numerical entity recogniser and classifier to detect numerical entities and generate questions about them. Previous to the creation of the questions, the system automatically detects the clauses to be used for the generation. However, the present work deals with the automatic generation of Basque questions using single sentences as the source text for the generation.

This article presents two approaches to generate questions, a chunker-based generation and a generation based on syntactic dependencies. Both approaches created direct questions regarding the noun phrases of the sentences of the corpora. In general, we foresee a better performance when using syntactic dependencies. However, we expect that a chunker-based generation can also be suitable if we limit the source input to single sentences. As the final aim of the system is to be used in the education domain, authors evaluate both approaches whith texts prepared to work on science and technology at secondary school and texts prepared for a language learning scenario. The evaluation focused on how well each approach transforms a sentence into its corresponding interrogative form.

The paper is structured as follows. Section 2 presents the main features used for the generation process. Section 3 describes the implemented systems. Section 4 explains the results of the experiments. Finally, conclusions and future work are commented on section 5.

2 Question Generation

This work presents two question generation systems for Basque. The generation is based on the morphological information of the noun phrases. The systems also use semantic features during the generation process.

2.1 QG based on Noun Phrases

Basque is a Pre-Indo-European language and differs considerably in grammar from the languages spoken in surrounding regions. It is, indeed, an agglutinative head-final isolated language. The case system is ergative-absolutive. The inflections of determination, number and case appear only after the last element in the noun phrase. This last element can be the noun, but also typically an adjective or a determiner. Basque nouns belong to a single declension and its 18 case markers are invariant. Functions, normally fulfilled by prepositions, are realised by case suffixes inside wordforms.

In this work we intend to automatically generate questions about all the noun phrases appearing in each sentence. To that end, the detection of the case markers of the noun phrases is the starting point for the generation process. We report two QG systems in order to compare a chunker-based approach and a dependency-based approach.

We choose 5 case markers as the starting point for the experiments: absolutive (ABS), ergative (ERG), inessive (INE), allative (ALL) and ablative (ABL). As explained in section 4.1, all of them cover almost the 90% of all the noun phrases found in corpora when generating questions in our scenario. Absolutive and ergative cases accumulate the highest percentage of noun phrases in the corpus, as they are related to the subject and direct object syntactic functions. The inessive case is used in noun phrases with different adverbial functions (temporal, location, etc). And the ablative and allative cases give us the chance to work with the animate/inanimate features. The mentioned 5 cases need different wh-words (interrogative pronouns) depending on the features of the head in the noun phrase.

2.2 Semantic Features for QG

We have explored the animate/inanimate feature of the nouns, the use of named entities (person, location and organisation) and the semantic roles to deal with the generation of

questions.

- Animate/inanimate feature: the QG generators use the work done by Díaz de Ilarraza, Mayor, and Sarasola (2002), where semantic features of common nouns are extracted semi-automatically from a monolingual dictionary. Both systems consider the animate/inanimate feature of 15,000 nouns.
- Named entities: both QG systems include a named entity recogniser and classifier named *Eihera* (Alegria et al., 2003). They use this tool to identify person, place and organisation entities.
- Semantic roles: The QG generators take into account a corpus manually tagged at the predicate level with verb senses, argument structure and semantic roles (Aldezabal et al., 2010). This corpus is based on the work done in Aldezabal (2004), which includes an in-depth study of 100 verbs for Basque. Based on the occurrences of the 100 verbs, we have worked with the following roles from VerbNet (Kipper et al., 2006): actor, attribute, agent, beneficiary, cause, destination, direction, experiencer, extent, instrument, location, manner, patient, predicate, product, recipient, source, theme, temporal and topic. The mandatory roles of patterns with a probability higher than 75% are considered to be candidates.

3 QG-Malti and QG-Ixati

The article reports two question generation systems for Basque, QG-Malti and QG-Ixati. QG-Malti is a QG system based on *Maltixa* (Bengoetxea and Gojenola, 2010), a dependency parser for Basque. QG-Ixati is a QG system which uses *Ixati* (Aduriz et al., 2004), a chunker for Basque. Previous to the selection of the noun phrase (the content selection step), both systems perform a morphosyntactic analysis of the source texts.

As proposed in Rus and Graesser (2009), both QG systems can be described as a three-step process: content selection, question type selection and question construction. The main constraint on the present study is that the systems only select sentences which contain a single finite verb. Before the content selection process, QG-Malti also splits coordinate sentences into single sentences.

The goal of both approaches is to generate questions at sentence level. QG-Malti discards the sentences which have discourse elements whose function is to connect the sentence with other elements outside the sentence, but it rejects them only in case the discourse elements are not at the beginning of the sentence. QG-Ixati, however, can not discard this kind of sentences as the analyser Ixati does not detect this kind of discourse relations.

3.1 Target Selection

As mentioned, both systems generate questions related to all the noun phrases that occur in the sentences of the source text[3]. The generation process uses morphosyntactic features of the output of the corresponding analyser to select **the candidate target**. In the case of the QG-Ixati, the candidate target is the whole noun phrase (chunk). However, in the case of QG-Malti the candidate target is the word whose morphological analysis has the target case marker. And then, the dependency structure of the analysis is used to construct the corresponding noun phrase.

When there is more than one occurrence for the same case marker inside the same sentence, only one of those occurrences is used to generate a question. Based on the fact that in Basque the relevant information of a sentence is close to the verb, QG-Malti selects as the candidate target (word) the one which is closest to the verb. And, if there are two candidates at the same distance to the verb, it selects the one located on the left to the verb. The reason for this criterion is that in Basque the informationally relevant phrase of a sentence precedes immediately the verb.

QG-Ixati, however, establishes a preference criterion based on various semantic features of the candidate targets (noun phrases). It gives a higher priority to the animate/inanimate feature and named entity tag than to the semantic roles. The priority is obtained as follows:

1. If the head of the noun phrase is a named entity (person, place or organisation) or its animate/inanimate feature is known, the QG system establishes a weight of 2 for the noun phrase.
2. If the noun phrase fulfills one of the

[3]We use 5 declension cases in the experiments of the present work.

	Animate	Person	Inanimate	Place	Organisation	No semantic feature
ABS	*Nor*		*Zer*			*Nor/Zer*
ERG	*Nork*		*Zerk*			*Nork/Zerk*
INE	*Norengan*		*Non*			*Norengan/Non/Noiz*
ALL	*Norengandik*		*Nondik*			*Norengandik/Nondik*
ABL	*Norengana*		*Nora*			*Norengana/Nora*

Table 1: Question type based on named entities, animate/inanimate and case markers. Nor (Who-ABS); Zer (What-ABS); Nork (Who-ERG); Zerk (What-ERG); Norengan (To whom); Non (Where); Noiz (When); Norengandik (From whom); Nondik (From where); Norengana (To whom); Nora (To where)

mandatory roles of a particular verb subcategorization pattern, the QG system establishes a weight of 1 for the given noun phrase.

The system chooses the noun phrase with the highest priority. In the cases that the system still assigns the same weight to different noun phrases, the selected candidate is the one which is closest to the verb. And in case of still being a tie, the system chooses the phrase located on the left to the verb.

3.2 Question Type Selection

QG-Malti and QG-Ixati follow the same criteria when selecting the question type to be generated. The selection of the question type is based on the linguistic information of the corresponding candidate target. For each case marker and linguistic feature (animate/inanimate, named entity, semantic role and morphology), an expert in the field established the most probable question type (wh-word) based on linguistic studies, as well as on her experience.

Table 1 shows the question types selected by the QG systems related to the named entity, animate/inanimate feature and case marker of the candidate target. For example, if the head of the noun phrase is identified as a person named entity and its corresponding case marker is the absolutive, the *NOR* (Who-ABS[4]) wh-word is selected.

The question type is also selected based on the semantic role of the candidate target. In total, 11 different roles have been linked to targets with the absolutive case, 7 to the ergative, 8 to the inessive, 5 to the ablative, and 5 to the allative. Depending on the semantic role of the candidate target, the QG system establishes its corresponding wh-word. For each verb, mainly only one question type is linked to each role. But, there are some exceptions, for example, the verb *compare* can have an animate or inanimate *patient* that correspond to the *NOR* (Who-ABS) and *ZER* (What-ABS) question types respectively.

3.3 Question Construction

In this phase, each QG system applies its own strategy based on the information given by the corresponding analyser. QG-Malti constructs the questions using the dependency relation structure analysed in the source sentence. QG-Ixati uses the information of the chunks detected during the morphosyntactic analysis of the source sentence.

The question building is based on simple transformation rules defined in the system. The first element of the constructed question is the wh-word. Following the wh-word, the main verb is established. Then, the rest of the elements (dependency structures or chunks) that are to the left of the verb in the source sentence are added to the question. Finally, the elements that appeared on the right of the source sentence's verb are appended to the generated question.

During the development of the systems we realised that some discourse connectives (e.g. the connective *gainera* [5]) caused some noise to the generated questions. In most of the cases where the connective was at the beginning of the source sentence, such a noise could be avoided if the connective was deleted when constructing the question. That is why we decided to delete from the source sentences all the discourse connectives which appear at the beginning of the sentence.

[4]The ABS mark refers to the fact that the wh-word takes the absolutive case marker.

[5]Basque word for *in addition*

4 Evaluation

For the experiments, we chose texts about science and technology for secondary school learners and a specialised corpus in language learning because one of the final aims is to use QG systems into the education domain. In this work, as a first step, the evaluation focused on how well each QG system transforms a sentence into its corresponding interrogative form.

We focused on the evaluation of the syntactic correctness and fluency of the generated questions. To do so, a human judge followed the same classification as the one proposed in Boyer and Piwek (2010). We also studied the quality of the question types determining whether the generated wh-words asked about the source sentence. Finally, the expert also established whether the question was appropriate in relation to the source sentence.

4.1 Datasets

The science and technology (ST) dataset is composed of 5 texts about science and technology. One expert who works on the generation of learning materials defined the 5 texts as adequate for secondary school learners (Aldabe and Maritxalar, 2010). The main topics of these texts were: Continent; the Earth; Bats; the Arctic; and Computers respectively. All the texts have a similar length. In total, the dataset contains 176 sentences, being the average length of a sentence 13 words.

The language learning (LL) dataset focuses on a specialised corpus for Basque language learning, which is a collection of learning-oriented Basque written texts. The corpus is classified into different language levels[6] in accordance with the Common European Framework of Reference for Languages (Little, 2011). In the present work, the intermediate level of the corpus is the basis to generate the questions. The corpus is composed of near 80,000 sentences (over one million words), and the average length of a sentence is 13 words.

The ST dataset contains 646 noun phrases and the LL dataset has 200,000 noun phrases. Looking at the 5 case markers that are the starting point of the systems to generate the questions, almost 90% of the noun phrases are covered with the mentioned target case markers in both datasets. In the ST dataset, 55% of the noun phrases have the absolutive case marker, the 12% of the phrases have the ergative case marker, the 16% of phrases are inessive, the 3% of noun phrases have the allative case and the 2% of them the ablative case. In the LL dataset, the 60% have the absolutive case marker, the 11% of the phrases have the ergative case marker, and the 16% of phrases are inessive. Regarding the allative and ablative cases the percentage is near the 3%. The rest of case markers are under the 4% in both datasets.

4.2 Experiments

For each dataset, experiments with both QG systems were performed. The questions generated by QG-Malti and QG-Ixati were manually evaluated at different levels by one linguist.

As regards the question-types, a linguist judged whether the generated wh-words asked about the source sentence (yes/no). For that, the source sentence (input for the QG system) and the candidate target (answer to the generated question) were provided.

When checking the grammaticality, the linguist evaluated the syntactic correctness and fluency of the generated questions. For that, only the generated questions were provided. The questions were classified and differentiated among: i) correct questions; ii) questions which need minor changes (punctuation, capitalization, spelling or dialectical variants); iii) questions with major changes that are unnatural for native speakers even they are grammatically correct; and iv) incorrect questions due to the grammar, including oral speech style.

Finally, the judge established if the generated questions were appropriate (yes/no). For that, in addition to each question, the corresponding answer was also shown. When evaluating the appropriateness of the questions only correct questions and questions which nedded minor changes were considered.

4.2.1 ST dataset experiment

The experiment with the ST dataset reflects an educational scenario where the creation of updated material using texts from the web is crucial for the motivation of learners and

[6] Although the language level of a text can be a controversial aspect because it is difficult to define, in our source corpus, expert teachers classified the texts into specific levels.

	ST-common			ST-divergent		
	Wh-word	**Grammar**	**Appropr.**	**Wh-word**	**Grammar**	**Appropr.**
QG-Malti	76%	54%	46%	72.6%	59.7%	41.9%
QG-Ixati	88%	66%	64%	87.1%	48.4%	48.4%

Table 2: Results for the ST common and divergent inputs

teachers. Both systems generated questions for all the candidate targets of the 5 texts. Based on the 5 case markers, QG-Malti and QG-Ixati generated 112 and 81 questions respectively.

	Wh-word	**Grammar**	**Appropr.**
QG-Malti	75.0%	57.1%	44.6%
QG-Ixati	87.6%	59.3%	58.0%

Table 3: Percentage of correct questions of the QG systems for the ST dataset

Table 3 presents the evaluation results as regards wh-words, grammaticality and appropriateness in the ST dataset. The grammar column groups questions marked as correct and questions which need minor changes. In general, QG-Ixati obtains better results than QG-Malti, but, QG-Malti generates more questions. Thus, QG-Malti generates 64 grammatically correct questions while QG-Ixati generates 48.

The generation processes of QG-Malti and QG-Ixati differ mainly due to the analysers and the target selection criteria. However, both systems have in common some instances. We refer to common instances to those which have the same candidate target with the same case marker. Out of the 112 and 81 generated questions both systems have in common 50 instances. Thus, apart from the these common instances, QG-Malti selects 62 sentences to generate the questions, while QG-Ixati chooses other 31 different ones. Table 2 presents the manual evaluation results based on this distinction. As regards the common instances (ST-common column), QG-Ixati obtains better results in terms of wh-words (88%), grammaticality (66%) and appropriateness (64%). The comparison of the divergent samples (ST-divergent column) with the common instances of each system shows different results. On the one hand, it is remarkable the improvement of the grammaticality of QG-Malti (59.7%) compared to its common instances (54%). On the other hand, QG-Ixati obtains worse results in terms of grammaticality (48.4%) and appropriateness (48.4%), compared to the common instances (66% and 64% respectively).

Thus, even the overall results are better for QG-Ixati, the number of grammatically correct questions of the divergent dataset is higher in the case of QG-Malti. These results must be analysed deeply as we foresee that the target case markers and the used analysers can have an influence on the results.

4.2.2 LL dataset experiment

The aim of the LL dataset experiment is to analyse the influence of the case marker of the noun phrase chosen as the answer to the generated question. This is why the sample contains 20 questions per case marker for each QG system selected at random[7]. In this experiment, a total of 100 generated questions for each system are evaluated.

Table 4 shows the evaluation results per case marker. In general, both systems obtain grammatically better questions when the generation is based on noun phrases with absolutive or inessive case markers. QG-Ixati obtains better overall results compared to QG-Malti. It is noticeable the difference on the grammaticality of the absolutive (QG-Ixati, 85% and QG-Malti, 60%) and ergative (QG-Ixati, 65% and QG-Malti, 45%) case markers. In contrast, QG-Malti performs better in terms of grammaticality and appropriateness of the allative and ablative case markers, and the wh-word of the ergative.

Although the source sentences are the same for both systems, the systems sometimes differ in the source candidate targets for the generation process. Out of the 100 questions, both systems have in common 47 questions. Table 5 presents the results of the 100 questions (LL-overall column) and the 47 common questions (LL-common column) in terms of wh-word, grammaticality and appropriateness.

The grammatically is better for QG-Malti

[7]The source sentences were the same for both QG systems.

	QG-Malti			QG-Ixati		
	Wh-word	**Grammar**	**Appropr.**	**Wh-word**	**Grammar**	**Appropr.**
ABS	70%	60%	55%	85%	85%	60%
ERG	95%	45%	40%	80%	65%	60%
INE	60%	85%	60%	70%	85%	60%
ALL	70%	45%	35%	85%	35%	30%
ABL	50%	50%	45%	55%	40%	35%

Table 4: Percentages per case markers (20 questions per case marker)

	LL-overall			LL-common		
	Wh-word	**Grammar**	**Appropr.**	**Wh-word**	**Grammar**	**Appropr.**
QG-Malti	69%	57%	47%	70.2%	63.8%	57.4%
QG-Ixati	75%	62%	50%	68.1%	59.6%	48.9%

Table 5: Results for the LL overall and LL common inputs

when looking at the common instances (from 57% to 63.8%) and it is lower for QG-Ixati (from 62% to 59.6%). Looking at the case markers of the 47 questions, just 3 out of the 47 questions correspond to the absolutive noun phrases and this is the main reason for getting worst results when using QG-Ixati.

4.3 Preliminary Error Analysis

The analysis of the results as well as the subsequent meetings with the expert allowed us to carry out a preliminary error analysis of the systems.

As regards the grammatical correctness of the questions, we have classified the erroneous questions in different categories: (i) questions which are grammatically correct but unnatural as regards the speakers; (ii) questions which contain orthographic errors; (iii) questions which are incorrectly generated in terms of morphology; (iv) questions which refer to oral speech; (v) problems with punctuation marks; and (vi) questions with an incorrect word order.

One of the reasons to generate ungrammatical questions is due to the type of the source input. In the analysis of the results we detected: i) some source input that correspond to subordinate clauses; ii) some source input that correspond to relative clauses and iii) some typos or spelling errors at word level. When analysing the results without taking into account the mentioned questions, the grammaticality and appropriateness measures of both QG systems improve 6 points for the ST dataset, and more than 10 points for the the LL dataset. In contrast, the number of correct wh-words hardly varies.

5 Conclusions and Future Work

Our QG systems created questions in order to ask about noun phrases at sentence level. With that end, a chunker-based QG system, QG-Ixati, and a QG system based on dependency structures, QG-Malti, have been implemented. Both systems deal with Basque language and make use of the animate/inanimate feature of the nouns, named entities (person, location and place), the semantic roles of the verbs, as well as the morphology of the noun phrases.

The results of the experiments show that QG-Malti generates a higher number of questions in a real scenario (ST dataset), however its general performance is slightly worse than QG-Ixati. The results for the LL dataset show a noticeable difference in grammaticality between both systems when generating questions about noun phrases with the absolutive and ergative case markers.

Future work will focus on the improvement of the systems. Once the roles are detected automatically, the semantic role approach would cover more verbs. Thus, we plan to focus on the analysis and integration of new Basque NLP tools or knowledge representations in order to generate questions that require deeper understanding. In addition, in the case of QG-Malti we want to improve the system using the information about syntactic dependencies, to discard ungrammatical source input for the generator, and to improve the results of the identification of the question type.

Bibliography

Aduriz, Itziar, María Jesús Aranzabe, Joxe Mari Arriola, Arantza Díaz de Ilarraza, Koldo Gojenola, Maite Oronoz, and Larraitz Uria. 2004. A cascaded syntactic analyser for Basque. *Computational Linguistics and Intelligent Text Processing*, pages 124–134.

Agarwal, Manish, Rakshit Shah, and Prashanth Mannem. 2011. Automatic question generation using discourse cues. In *Proceedings of the 6th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 1–9. Association for Computational Linguistics.

Aldabe, Itziar and Montse Maritxalar. 2010. Automatic distractor generation for domain specific texts. In *Proceedings of the 7th International Conference on NLP, IceTAL 2010*. Springer, pages 27–38.

Aldabe, Itziar, Montse Maritxalar, and Ander Soraluze. 2011. Question generation based on numerical entities in Basque. In *Proceedings of AAAI Symposium on Question Generation*, pages 2–8.

Aldezabal, Izaskun. 2004. *Aditazpikategorizazioaren Azterketa Sintaxi Partzialetik Sintaxi Osorako bidean. 100 aditzen azterketa, Levin-en (1993) lana oinarri hartuta eta metodo automatikoak baliatuz.* Ph.D. thesis, Euskal Filologia Saila. UPV/EHU.

Aldezabal, Izaskun, María Jesús Aranzabe, Arantza Díaz de Ilarraza, Ainara Estarrona, and Larraitz Uria. 2010. EusPropBank: Integrating semantic information in the Basque dependency treebank. In *Computational Linguistics and Intelligent Text Processing*. Springer, pages 60–73.

Alegria, Iñaki, Nerea Ezeiza, Izaskun Fernandez, and Ruben Urizar. 2003. Named Entity Recognition and Classification for texts in Basque. In *II Jornadas de Tratamiento y Recuperación de Información, JOTRI, Madrid.*

Bengoetxea, Kepa and Koldo Gojenola. 2010. Application of different techniques to dependency parsing of Basque. In *Proceedings of the NAACL HLT 2010 First Workshop on Statistical Parsing of Morphologically-Rich Languages*, pages 31–39. Association for Computational Linguistics.

Boyer, Kristy Elizabeth and Paul Piwek, editors. 2010. *Proceedings of QG2010: The Third Workshop on Question Generation*. Pittsburgh: questiongeneration.org.

Díaz de Ilarraza, Arantza, Aingeru Mayor, and Kepa Sarasola. 2002. Semiautomatic labelling of semantic features. In *Proceedings of the 19th international conference on Computational linguistics-Volume 1*, pages 1–7. Association for Computational Linguistics.

Graesser, Arthur, James Lester, Jack Mostow, Rashmi Prasad, and Svetlana Stoyanchev. 2011. Question generation papers from the AAAI fall symposium. Technical report, FS-11-04.

Kipper, Karin, Anna Korhonen, Neville Ryant, and Martha Palmer. 2006. Extending VerbNet with novel verb classes. In *In Proceedings of 5th international conference on Language Resources and Evaluation*.

Little, David. 2011. The common european framework of reference for languages: A research agenda. *Language Teaching*, 44(03):381–393.

Mannem, Prashanth, Rashmi Prasad, and Aravind Joshi. 2010. Question generation from paragraphs at UPenn: QGSTEC system description. In *Proceedings of QG2010: The Third Workshop on Question Generation*, pages 84–91.

Mostow, Jack and Wei Chen. 2009. Generating instruction automatically for the reading strategy of self-questioning. In *Proceedings of the 2009 conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling*, pages 465–472. IOS Press.

Olney, Andrew M, Arthur C Graesser, and Natalie K Person. 2012. Question generation from concept maps. *Dialogue & Discourse*, 3(2):75–99.

Rus, Vasile and Arthur C. Graesser, editors. 2009. *The Question Generation Shared Task and Evaluation Challenge*.