Automatic Question Generation using Discourse Cues

Manish Agarwal*, **Rakshit Shah*** **and Prashanth Mannem**
Language Technologies Research Center
International Institute of Information Technology
Hyderabad, AP, India - 500032
{manish.agarwal, rakshit.shah, prashanth}@research.iiit.ac.in

Abstract

In this paper, we present a system that automatically generates questions from natural language text using discourse connectives. We explore the usefulness of the discourse connectives for Question Generation (QG) that looks at the problem beyond sentence level. Our work divides the QG task into content selection and question formation. Content selection consists of finding the relevant part in text to frame question from while question formation involves sense disambiguation of the discourse connectives, identification of question type and applying syntactic transformations on the content. The system is evaluated manually for syntactic and semantic correctness.

1 Introduction

Automatic QG from sentences and paragraphs has caught the attention of the NLP community in the last few years through the question generation workshops and the shared task in 2010 (QGSTEC, 2010). Previous work in this area has concentrated on generating questions from individual sentences (Varga and Ha, 2010; Paland et al., 2010; Ali et al., 2010). Sneiders and E. (2002) used question templates and Heilman et al. (2009) used general-purpose rules to transform sentences into questions. A notable exception is Mannem et al. (2010) who generated questions of various scopes (general, medium and specific) * [1] from paragraphs instead of individual sentences. They boil down the *QG from paragraphs* task into first identifying the sentences in the paragraph with general, medium and specific scopes and then generating the corresponding questions from these sentences using semantic roles of predicates.

Discourse connectives play a vital role in making the text coherent. They connect two clauses or sentences exhibiting discourse relations such as $temporal$, $causal$, $elaboration$, $contrast$, $result$, etc. Discourse relations have been shown to be useful to generate questions (Prasad and Joshi, 2008) but identifying these relations in the text is a difficult task (Pitler et al., 2009). So in this work, instead of identifying discourse relations and generating questions using them, we explore the usefulness of discourse connectives for QG. We do this by analyzing the senses of the connectives that help in QG and propose a system that makes use of this analysis to generate questions of the type why, $when$, *give an example* and yes/no.

The two main problems in QG are identifying the content to ask a question on and finding the corresponding question type for that content. We analyze the connectives in terms of the content useful for question generation based on the senses they exhibit. We show that the senses of the connectives further help in choosing the relevant question type for the content.

In this paper, we present an end-to-end QG system that takes a document as input and outputs all the questions generated using the selected discourse connectives. The system has been evaluated manually by two evaluators for syntactic and semantic

*First two authors contributed equally to this work

[1] General scope - entire or almost entire paragraph, Medium scope - multiple clauses or sentences, and Specific scope - sentence or less

Proceedings of the Sixth Workshop on Innovative Use of NLP for Building Educational Applications, pages 1–9,
Portland, Oregon, 24 June 2011.

correctness of the generated questions. The overall system has been rated 6.3 out of 8 for QGSTEC development dataset and 5.8 out of 8 for Wikipedia dataset.

2 Overview

Question Generation involves two tasks, content selection (the text selected for question generation) and question formation (transformations on the content to get the question). Question formation further has the subtasks of (i) finding suitable question type (wh-word), (ii) auxiliary and main verb transformations and (iii) rearranging the phrases to get the final question.

There are 100 distinct types of discourse connectives listed in PDTB manual (PDTB, 2007). The most frequent connectives in PDTB are and, or, but, $when$, $because$, $since$, $also$, $although$, $for\ example$, $however$ and $as\ a\ result$. In this paper, we provide analysis for four subordinating conjunctions, *since, when, because* and *although*, and three adverbials, *for example, for instance* and *as a result*. Connectives such as *and, or* and *also* showing $conjunction$ relation have not been found to be good candidates for generating wh-type questions and hence have not been discussed in the paper. Leaving aside *and, or* and *also*, the selected connectives cover 52.05 per cent of the total number of the connectives in QGSTEC-2010 [2] dataset and 41.97 per cent in Wikipedia articles. Connective-wise coverage in both the datasets is shown in Table 1. Though *but* and *however* denoting $contrast$ relation occur frequently in the data, it has not been feasible to generate wh-questions using them.

	QGSTEC-2010 Dev. Data		Wikipedia Dataset	
Connective	**count**	**%**	**count**	**%**
because	20	16.53	36	10.28
since	9	7.44	18	5.14
when	23	19.00	35	10.00
although	4	3.30	22	6.28
as a result	5	4.13	6	1.71
for example	2	1.65	30	8.28
for instance	0	0.00	1	0.28
Total	121	52.05	350	41.97

Table 1: Coverage of the selected discourse connectives in the data

The system goes through the entire document and identifies the sentences containing at least one of the seven discourse connectives. In our approach, suitable *content* for each discourse connective which is referred to as *target argument* is decided based on the properties of discourse connective. The system finds the question type on the basis of discourse relation shown by discourse connective.

[2] QGSTEC 2010 data set involves Wikipedia, Yahoo Answers and OpenLearn articles.

3 Discourse connectives for QG

In this section, we provide an analysis of discourse connectives with respect to their target arguments and the question types they take.

3.1 Question type identification

The sense of the discourse connective influences the question-type (*Q-type*). Since few discourse connectives such as *when, since* and *although* among the selected ones can show multiple senses, the task of sense disambiguation of the connectives is essential for finding the question type.

Since: The connective can show *temporal, causal* or *temporal + causal* relation in a sentence. Sentence exhibits *temporal* relation in presence of keywords like time(7 am), year (1989 or 1980s), start, begin, end, date(9/11), month (January) etc. If the relation is *temporal* then the question-type is *when* whereas in case of *causal* relation it would be *why*.

1. *Single wicket has rarely been played **since** limited overs cricket **began**.*
   Q-type: **when**

2. *Half-court games require less cardiovascular stamina , **since** players need not run back and forth a full court.*
   Q-type: **why**

In examples 1 and 2, 1 is identified to show *temporal* relation because it has the keyword *began* whereas there is no keyword in the context of example 2 that gives the hint of *temporal* relation and so the relation here is identified as *causal*.

When: Consider the sentences with connective *when* in Figure 1. Although *when* shows multiple senses (*temporal, temporal+causal* and *conditional*), we can frame questions by a single question type, *when*. Given a new instance of the connective, finding the correct sense of *when*

Sentence: The San−Francisco earthquake hit when resources in the field already were stetched. (Temporal)
Question: When did San−Francisco earthquake hit ?

Sentence: Venice's long decline started in the 15th century, when it first made an unsuccessful attempt to hold Thessalonica against the Ottomans (1423−1430). (Temporal + Causal)
Question: When did Venice's long decline start in the 15th century ?

Sentence: Earthquake mainly occurs when the different blocks or plates that make up the Earth's surface move relative to each other, causing distortion in the rock. (Conditional)
Question: When do earthquake mainly occur ?

Figure 1: Questions for discourse connective ***when***

Discourse connectives	Sense	Q-type
because	causal	why
since	temporal causal	when why
when	causal + temporal temporal conditional	when
although	contrast concession	yes/ no
as a result	result	why
for example	instantiation	give an example where
for instance	instantiation	give an instance where

Table 2: Question type for discourse connectives

becomes unnecessary as a result of using discourse connectives.

Although: The connective can show *concession* or *contrast* discourse relations. It is difficult to frame a *wh*-question on *contrast* or *concession* relations. So, system generates a *yes/no* type question for *although*. Moreover, *yes/no* question-type adds to the variety of questions generated by the system.

3. *Greek colonies were not politically controlled by their founding cities , although they often retained religious and commercial links with them .*
   Q-type: **Yes/No**

A yes/no question could have been asked for connectives *but* and *however* denoting a *contrast* relation but it was not done to preserve the question-type variety in the final output of the QG system. Yes/no questions have been asked for occurrences of $although$ since they occur less frequently than but and $however$.

Identifying the question types for other selected discourse connectives is straight forward because they broadly show only one discourse relation (Pitler and Nenkova, 2009). Based on the relations exhibited by these connectives, Table 2 shows the question types for each discourse connective.

3.2 Target arguments for discourse connectives

A discourse connective can realize its two arguments, Arg1 and Arg2, structurally and anaphorically. Arg2 is always realized structurally whereas Arg1 can be either structural or anaphoric (PDTB, 2007; Prasad et al., 2010).

4. *[Arg1 Organisms inherit the characteristics of their parents]* ***because*** *[Arg2 the cells of the offspring contain copies of the genes in their parents' cells.](Intra-sentential connective because)*

5. *[Arg1* The scorers are directed by the hand signals of an umpire.*]* ***For example****, [Arg2 the umpire raises a forefinger to signal that the batsman is out (has been dismissed); he raises both arms above his head if the batsman has hit the ball for six runs.](Inter-sentential connective for example)*

Consider examples 4 and 5. In 4, Arg1 and Arg2 are the structural arguments of the connective *because* whereas in 5, Arg2 is the structural argument and Arg1 is realized anaphorically.

The task of content selection involves finding the *target argument* (either Arg1 or Arg2) of the discourse connective. Since both the arguments are potential candidates for QG, we analyze the data to identify which argument makes better content for each of the connectives. Our system selects one of the two arguments based on the properties of the discourse connectives. Table 3 shows the *target argu-*

Discourse connective	Target argument
because	Arg1
since	Arg1
when	Arg1
although	Arg1
as a result	Arg2
for example	Arg1
for instance	Arg1

Table 3: Target argument for discourse connectives

ment i.e. either Arg1 or Arg2, which is used as *content* for QG.

4 Target Argument Identification

Target argument for a discourse connective can be a clause(s) or a sentence(s). It could be one or more sentences in case of *inter-sentential*[3] discourse connectives, whereas one or more clauses in case of *intra-sentential*[4] connectives.

Discourse connectives *for example* and *for instance* can realize its Arg1 anywhere in the prior discourse (Elwell and Baldridge, 2008). So the system considers only those sentences in which the connectives occur at the beginning of the sentence and the immediate previous sentence is assumed to be the Arg1 of the connective (which is the *target argument* for QG).

In case of intra-sentential connectives (*because, since, although* and *when*) and *as a result* (*target argument* is Arg2 which would be a clause), identification of *target argument* is done in two steps. The system first locates the syntactic head or head verb of the *target argument* and then extracts it from the dependency tree of the sentence.

4.1 Locate syntactic head

Approach for locating the syntactic head of *target argument* is explained with the help of Figure 2 (generic dependency trees) and an example shown in Figure 3. Syntactic head of Arg2 is the first finite verb while percolating up in the dependency tree starting from the discourse connective. In case of intra-sentential connectives where Arg1 is the *target argument*, the system percolates up until it gets the second finite verb which is assumed to be target head

[3]Connectives that realize its Arg1 anaphorically and Arg2 structurally

[4]Connectives that realize both of its arguments structurally

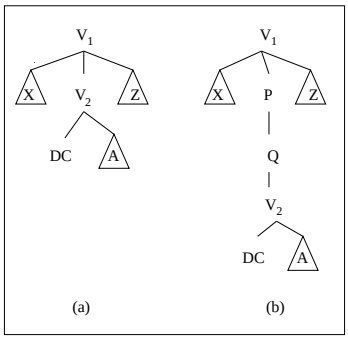

Figure 2: Head selection of the *target argument* for intra-sentential connectives (V_1,V_2: *finite verbs*; X,Z: *subtrees of* V_1; A: *subtree of* V_2; P,Q:*Not verbs*; DC:*discourse connective(child of* V_2*))*

of Arg1. Number of percolations entirely depend on structure and complexity of the sentence. Figure 2 shows two dependency trees (a) and (b). Starting from the discourse connective *DC* and percolating up, the system identifies that the head of Arg2 is V_2 and that of Arg1 is V_1.

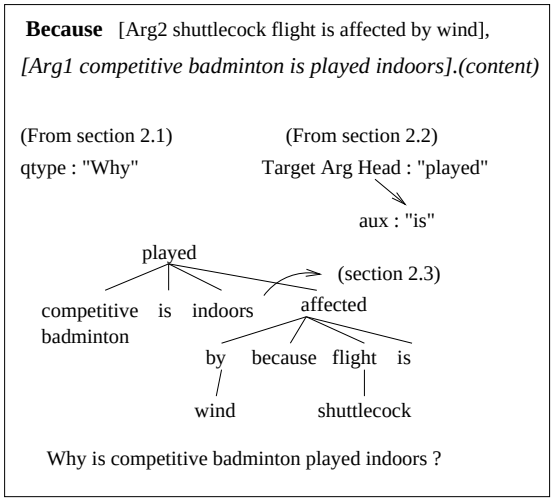

Figure 3: Question Generation process

Since the discourse connective in the example of Figure 3 is *because*, the *target argument* is Arg1 (from Table 2). By percolating up the tree starting from *because*, the head of Arg2 is *affected* and that of Arg1 is *played*. Once we locate the head of the *target argument*, we find the auxiliary as Mannem et al. (2010) does. For the example in Figure 3, the auxiliary for question generation is *is*.

4.2 Target Argument Extraction

The extraction of the *target argument* is done after identifying its syntactic head. For *as a result*, the *target argument*, Arg2, is the subtree with head

Score	Description	Example
4	The question is grammatically correct and idiomatic/natural.	In which type of animals are phagocytes highly developed?
3	The question is grammatically correct but does not read as fluently as we would like.	In which type of animals are phagocytes, which are important throughout the animal kingdom, highly developed?
2	There are some grammatical errors in the question.	In which type of animals is phagocytes, which are important throughout the animal kingdom, highly developed?
1	The question is grammatically unacceptable.	On which type of animals is phagocytes, which are important throughout the animal kingdom, developed?

Table 4: Evaluation guidelines for syntactic correctness measure

as the head of the connective. For intra-sentential connectives, the *target argument*, Arg1, is the tree remaining after removing the subtree that contains Arg2.

In Figures 2 (a) and (b) both, a tree with head V_1 and its children, X and Z, is left after removing Arg2 from dependency trees, which is the *content* required for generating the question. Note that in the tree of Figure 2(b), the child P of the head verb V_1 is removed with its entire subtree that contains Arg2. Thus, subtree with head V_2 is the unwanted part for the tree in Figure 2(a) whereas subtree with head P is the unwanted part for the tree in Figure 2(b) when the target argument is Arg1.

In Figure 3, after removing the unwanted argument Arg2 (subtree with head *affected*), the system gets *competitive badminton is played indoors* which is the required clause (*content*) for question generation. The next section describes how the *content* is transformed into a question.

5 Syntactic Transformations and Question Generation

The syntactic transformations used in this work are similar to those by Mannem et al. (2010). At this stage, the system has the question type, auxiliary and the *content*. The following set of transformations are applied on the *content* to get the final question. (1) If the auxiliary is present in the sentence itself then it is moved to the beginning of the sentence; otherwise auxiliary is added at the beginning of the sentence. (2) If a wh-question is to be formed, the question word is added just before the auxiliary. In case of Yes/No questions, the question starts with the auxiliary itself as no question word is needed. (3) A question-mark(*?*) is added at the end to complete the question.

Consider the example in Figure 3. Here the *content* is *competitive badminton is played indoors*. Applying the transformations, the auxiliary is first moved at the start of the sentence to get *is competitive badminton played indoors*. Then the question type *Why* is added just before the auxiliary *is*, and a question-mark is added at the end to get the final question, *Why is competitive badminton played indoors ?*

Scope: In QGSTEC 2010 the question had to be assigned a scope, specific, medium or general. The scope is defined as: *general* - entire input paragraph, *medium* - one or more clauses or sentences and *specific* - phrase or less. Questions generated using discourse connectives are usually of the scope specific or medium. Mannem et al. (2010) assigned medium scope to the questions generated using the semantic roles such as ARGM-DIS (result), ARGM-CAU (causal) and ARGM-PNC (purpose) given by the SRL. However, most of the times, the scope of the answer to these questions is just a clause or a sentence and should have been assigned specific scope instead of medium.

6 Evaluation and Results

Automatic evaluation of any natural language generated text is difficult. So, our system is evaluated manually. The evaluation was performed by two graduate students with good English proficiency. Evaluators were asked to rate the questions on the scale of 1 to 4 (4 being the best score) on syntactic and semantic correctness (Evalguide, 2010) of the question and an overall rating on the scale of 8 (4+4) is assigned to each question.

The syntactic correctness is rated to ensure that the system can generate grammatical output. In addition, those questions which read fluently are given greater score. The syntactic correctness and fluency is evaluated using the following scores: 4 - gram-

Discourse Connective	Example
because	One-handed backhand players move to the net with greater ease than two-handed players **because** the shot permits greater forward momentum and has greater similarities in muscle memory to the preferred type of backhand volley (one-handed, for greater reach). *Why do one-handed backhand players move to the net with greater ease than two-handed players ?* (**Causal**)
since	Half-court games require less cardiovascular stamina, **since** players need not run back and forth a full court. *Why do half-court games require less cardiovascular stamina ?* (**Causal**) Single wicket has rarely been played **since** limited overs cricket began. *Since when has single wicket rarely been played ?* (**Temporal**)
when	A one-point shot can be earned **when** shooting from the foul line after a foul is made. *When can a one-point shot be earned ?* (**Conditional**)
although	A bowler cannot bowl two successive overs, **although** a bowler can bowl unchanged at end for several overs. *Can a bowler bowl unchanged at the same end for several overs?* (**Contrast, concession**)
as a result	In the United States sleep deprivation is common with students because almost all schools begin early in the morning and many of these students either choose to stay up awake late into the night or cannot do otherwise due to delayed sleep phase syndrome. **As a result**, students that should be getting between 8.5 and 9.25 hours of sleep are getting only 7 hours. *Why are students that should be getting between 8.5 and 9.25 hours of sleep getting only 7 hours?* (**Result**) **As a result** of studies showing the effects of sleep-deprivation on grades , and the different sleep patterns for teenagers , a school in New Zealand , changed its start time to 10:30, in 2006, to allow students to keep to a schedule that allowed more sleep. *Why did a school in New Zealand change its start time ?* (**Result**)
for example	Slicing also causes the shuttlecock to travel much slower than the arm movement suggests. **For example**, a good cross court sliced drop shot will use a hitting action that suggests a straight clear or smash, deceiving the opponent about both the power and direction of the shuttlecock. *Give an example where slicing also causes the shuttlecock to travel much slower than the arm movement suggests.* (**Instantiation**)
for instance	If the team that bats last scores enough runs to win, it is said to have ”won by n wickets”, where n is the number of wickets left to fall. **For instance** a team that passes its opponents’ score having only lost six wickets would have won ”by four wickets”. *Give an instance where if the team that bats last scores enough runs to win, it is said to have ”won by n wickets”,where n is the number of wickets left to fall.* (**Instantiation**)

Table 5: Examples

matically correct and idiomatic/natural, 3 - grammatically correct, 2 - some grammar problems, 1 - grammatically unacceptable. Table 4 shows syntactic correctness measure with examples.

The semantic correctness is evaluated using the following scores: 4 - semantically correct and idiomatic/natural, 3 - semantically correct and close to the text or other questions, 2 - some semantic issues, 1 - semantically unacceptable.

Table 5 shows questions generated by the system for each connective. The results of our system on QGSTEC-2010 development dataset are shown in Table 6. The overall system is rated 6.3 out of 8 on this dataset and the total number of questions generated for this dataset is 61. The instances of the connectives were less in the QGSTEC-2010 development dataset. So, the system is further tested on five Wikipedia articles (football, cricket, basketball, badminton and tennis) for effective evaluation. Results on this dataset are presented in Table 7. Overall rating of the system is 5.8 out of 8 for this dataset and 150 are the total number of questions generated for this dataset. The ratings presented in the Tables 6 and 7 are the average of the ratings given by both the evaluators. The inter-evaluator agreement (Cohen’s kappa coefficient) for the QGSTEC-2010 develop-

ment dataset for syntactic correctness measure is 0.6 and is 0.5 for semantic correctness measure, and in case of Wikipedia articles the agreement is 0.7 and 0.6 for syntactic and semantic correctness measures respectively.

Discourse connective	No. of questions	Syntactic Correctness(4)	Semantic Correctness(4)	Overall Rating(8)
because	20	3.6	3.6	7.2
since	9	3.8	3.2	7
when	23	2.3	2.2	4.5
although	4	4	3.8	7.8
as a result	5	4	4	8
Overall	61	3.2	3.1	6.3

Table 6: Results on QGSTEC-2010 development dataset

Discourse connective	No. of questions	Syntactic Correctness(4)	Semantic Correctness(4)	Overall Rating(8)
because	36	3.3	3.2	6.5
since	18	3.1	3	6.1
when	35	2.4	2.0	4.4
although	22	3.1	2.8	5.9
as a result	6	3.6	3.2	6.8
for example	16	3.1	2.9	6.0
for instance	2	4	3	7
Overall	135	3.0	2.8	5.8

Table 7: Results on the Wikipedia data(cricket, football, basketball, badminton, tennis)

On analyzing the data, we found that the Wikipedia articles have more complex sentences (with unusual structure as well as more number of clauses) than QGSTEC-2010 development dataset. As a result, the system's performance consistently drops for all the connectives in case of Wikipedia dataset.

No comparable evaluation was done as none of the earlier works in QG exploited the discourse connectives in text to generate questions.

7 Error Analysis

An error analysis was carried out on the system's output and the four most frequent types of errors are discussed in this section.

7.1 Coreference resolution

The system doesn't handle coreference resolution and as a result of this, many questions have been rated low for semantic correctness by the evaluators. Greater the number of pronouns in the question, lesser is the semantic rating of the question.

6. *They grow in height **when** they reach shallower water, in a wave shoaling process.*
   Question: *When do **they** grow in height?*

Although the above example 6 is syntactically correct, such questions are rated semantically low because the context is not sufficient to answer the question due to the pronouns in it. 13.54% of the generated questions on the Wikipedia dataset have pronouns without their antecedents, making the questions semantically insufficient.

7.2 Parsing Errors

Sometimes the parser fails to give a correct parse for the sentences with complex structure. In such cases, the system generates a question that is unacceptable. Consider the examples below.

7. *In a family who know that both parents are carriers of CF , **either because** they already have a CF child **or as a result** of carrier testing , PND allows the conversion of a probable risk of the disease affecting an unborn child to nearer a certainty that it will or will not be affected.*
   Question: *Why do in a family who know that both parents are carriers of CF , either or will not be affected ?*

In example 7 above, the sentence has a complex structure containing paired connective, either-or, where the argument of *either* has *because* and that of *or* has *as a result* in it. Here the question is formed using *because* which is correct neither syntactically nor semantically due to the complex nature of the sentence. 9.38% sentences in the datasets are complex with either three or more discourse connectives.

7.3 Errors due to the inter-sentential connectives

For inter-sentential connectives, system considers only those sentences in which the connectives occur at the beginning of the sentence and the immediate previous sentence is assumed to be the Arg1 of the connective (which is the target argument for QG). But this assumption is not always true. Of the total number of instances of these connectives, 52.94% (for Wikipedia dataset) connectives occur at the beginning of the sentences. Consider the paragraph below.

8. *A game point occurs in tennis whenever the*

player who is in the lead in the game needs only one more point to win the game. The terminology is extended to sets (set point), matches (match point), and even championships (championship point). *For example, if the player who is serving has a score of 40-love, the player has a triple game point (triple set point, etc.) as the player has three consecutive chances to win the game.*

Here in example 8, the third sentence in which the example is specified is related to the first sentence but not the immediately previous sentence. For these connectives, the assumption that immediate previous sentence is Arg1 is false 14.29% of the times.

7.4 Fluency issues

The system does not handle the removal of predicative adjuncts. So the questions with optional phrases in it are rated low for syntactic correctness measure.

8 Conclusions and Future Work

Our QG system generates questions using discourse connectives for different question types. In this work, we present an end-to-end system that takes a document as input and outputs all the questions for selected discourse connectives. The system has been evaluated for syntactic and semantic soundness of the question by two evaluators. We have shown that some specific discourse relations are important such as $causal$, $temporal$ and $result$ than others from the QG point of view. This work also shows that discourse connectives are good enough for QG and that there is no need for full fledged discourse parsing. In the near future, we plan to implement coreference resolution and sentences with more than two connectives. We aim to improve the system with respect to the sentence complexity and also incorporate other discourse connectives.

Acknowledgements

We would like to thank Suman Yelati and Sudheer Kolachina from LTRC, IIIT-Hyderabad for their helpful discussions and pointers during the course of this work. Thanks also to the anonymous reviewers for useful feedback.

References

2010 *Question generation shared task and evaluation challenge*, http://questiongeneration.org/QG2010

Andrea Varga and Le An Ha 2010 *WLV: A Question Generation System for the QGSTEC 2010 Task B*, Proceedings of QG2010: The Third Workshop on Question Generation

Santanu Paland Tapabrata Mondal, Partha Pakray, Dipankar Das and Sivaji Bandyopadhyay 2010 *QGSTEC System Description JUQGG: A Rule based approach* , Proceedings of QG2010: The Third Workshop on Question Generation

Husam Ali, Yllias Chali, and Sadid A. Hasan 2010 *Automation of Question Generation From Sentences*, Proceedings of QG2010: The Third Workshop on Question Generation

Eriks Sneiders 2002. *Automated question answering using question templates that cover the conceptual model of the database.* In Proceedings of the 6th International Conference on Applications of Natural Language to Information Systems (pp. 235-239).

Michael Heilman, Noah A. Smith. 2009 *Question generation via overgenerating transformations and ranking* Technical Report CMU-LTI-09-013, Carnegie Mellon University.

Prashanth Mannem, Rashmi Prasad and Aravind Joshi 2010 *Question Generation from Paragraphs at UPenn: QGSTEC System Discription*, Proceedings of QG2010: The Third Workshop on Question Generation

Rashmi Prasad and Aravind Joshi 2008 *A Discourse-based Approach to Generating why-Questions from text*, Proceedings of the Workshop on the Question Generation Shared Task and Evaluation Challenge Arlington, VA, September 2008

Emily Pitler, Annie Louis and Ani Nenkova 2009 *Automatic sense prediction for implicit discourse relations in text* , ACL '09 Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - Volume 2

2007 *PDTB 2.0 Annotation Manual*, http://www.seas.upenn.edu/ pdtb/PDTBAPI/pdtb-annotation-manual.pdf

Emily Pitler and Ani Nenkova 2009 *Using syntax to Disambiguate Explicit Discourse Connectives in Text*, ACLShort '09 Proceedings of the ACL-IJCNLP 2009 Conference Short Papers

Rashmi Prasad and Aravind Joshi and Bonnie Webber 2010 *Exploiting Scope for Shallow discourse Parsing*, LREC2010

Robert Elwell and Jason Baldridge 2008 *Discourse connective argument identification with connective specific rankers,* In Proceedings of ICSC-2008

Evaluation guidelines 2010 *In QGSTEC-2010 Task B evaluation guidelines*, http://www.question generation.org/QGSTEC2010/ uploads/QG-fromSentences-v2.doc