

# Open Research Online

---

The Open University's repository of research publications and other research outputs

## Varieties of Question Generation: introduction to this special issue

### Journal Item

How to cite:

Piwek, Paul and Boyer, Kristy Elizabeth (2012). Varieties of Question Generation: introduction to this special issue. *Dialogue & Discourse*, 3(2) pp. 1–9.

For guidance on citations see [FAQs](#).

© 2012 Paul Piwek and Kristy Boyer

Version: Version of Record

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.5087/dad>

<http://dad.uni-bielefeld.de/index.php/dad/article/view/2886>

---

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's [data policy](#) on reuse of materials please consult the policies page.

---

[oro.open.ac.uk](http://oro.open.ac.uk)

## Varieties of Question Generation: Introduction to this Special Issue

**Paul Piwek**

*Centre for Research in Computing  
The Open University  
Walton Hall  
Milton Keynes, MK7 6AA, United Kingdom*

P.PIWEK@OPEN.AC.UK

**Kristy Elizabeth Boyer**

*Department of Computer Science  
North Carolina State University  
890 Oval Dr.  
Campus Box 8206, Raleigh, NC 27695-8206, United States*

KEBOYER@NCSU.EDU

### Introduction

Questions are a topic that is ideally placed at the intersection of research on dialogue and discourse: in most dialogue, questions are the driving force, setting its direction, whereas answers to questions require discourse processing (e.g., contextual resolution in the case of short answers) or can be viewed as a piece of discourse themselves (e.g., an explanation in response to a why-question).

Current work on questions in both Formal and Computational Semantics can be traced back to early explorations into the application of modern logic to questions. In particular, Cohen (1929) proposed that the content of a question can be expressed as an open formula with one or more unbound variables. This idea reverberates in more recent work on the (computational) semantics of questions (e.g., Ginzburg & Sag, 2001; Piwek, 1998; Scha, 1983); at the same time, an alternative view of questions, originally proposed by C.L. Hamblin, has taken hold among formal semanticists; according to this view, the content of a question corresponds to the set of its answers (see Groenendijk & Stokhof, 1997). The past two decades have also seen a recognition of the centrality of questions in theories of dialogue – e.g., Ginzburg’s Questions Under Discussion (Ginzburg, 2012).

In the field of Natural Language Processing, questions have been extensively studied as part of the task of Question-Answering. Question-Answering research initially focused on answering questions from databases and knowledge representations (e.g., Green et al, 1961; Bronnenberg et al., 1979), but in the past two decades has refocused on retrieving answers from text – e.g., in 1999 the evaluation of question-answering systems became part of the Text Retrieval Conference (TREC) series. Simultaneously, there has been a strand of research on advisory dialogue systems – e.g., Winograd’s SHRDLU (Winograd, 1972) and more recently the DENK cooperative assistant (Ahn et al., 1995) – concentrating on theoretical issues by working with applications in restricted domains (*cf.* Hirschman and Gaizauskas, 2001). All the aforementioned systems were primarily aimed at *responding* to the user’s questions; even most advisory dialogue systems would only *ask* a question if there was some issue, which the user introduced, that needed clarification.<sup>1</sup>

<sup>1</sup> For an exception, see the seminal work of Power (1979) on machine-machine dialogue (rather than human-machine, dialogue), which involves dialogue agents asking questions as a part of joint plan construction. The questions are used by the agents to elicit answers that address gaps in their domain knowledge.

The rich tradition of research into questions in both Formal/Computational Semantics and Natural Language Processing has focused on the interpretation of questions and how to answer or respond to them. Until recently, there were few studies looking into the conditions under which a question gets asked, i.e., the principles and/or processes that underlie the generation of questions. In Natural Language Processing and more specifically the subfield of Natural Language Generation (McDonald, 1993; Reiter & Dale, 2000; Evans et al., 2002) research has been limited almost exclusively to generating text consisting of declarative sentences. An exception is work specifically on clarification questions in Natural Language Processing (e.g., Kievit et al., 2001; Purver, 2004) and also Stent's (2001) work on language generation for dialogue systems. In Formal Semantics, one of the few exceptions is Wisniewski's monograph 'The posing of questions' (Wisniewski, 1995). According to Wisniewski 'arriving at a question' is analogous to coming to a conclusion, i.e., as 'some premises [being] involved and some inferential thought processes [taking] place' (Wisniewski, 1995: xi).

In contrast with Formal/Computational Semantics and Natural Language Processing, Question Generation has a substantial history in Education and Psychology, see Olney et al. (this volume) for an overview of work in this tradition which goes back at least as far as Piaget (1952). Particularly influential has been the work of Art Graesser and collaborators in which specific mechanisms for question generation are proposed. For example, Graesser et al. (1992) present 22 different mechanisms that are grouped into four categories: correction of knowledge deficit, monitoring common ground, social coordination of action, and control of conversation and attention.

Recently, there has been a broader uptake of research into Question Generation. Since 2008, researchers from different communities, ranging from, but not limited to, Discourse Analysis, Dialogue Modeling, Formal Semantics, Intelligent Tutoring Systems, Natural Language Generation, Natural Language Understanding, and Psycholinguistics, have met annually at the Question Generation workshop.<sup>2</sup> At the 3<sup>rd</sup> workshop in 2010, the first Question Generation Shared Task and Evaluation Campaign (see Rus et al., this volume) took place. With a significant body of work accumulating on Question Generation, it seems timely to collect a representative sample of high quality efforts in a special issue. The original call for the issue attracted 26 initial submissions. As a result of several selection stages, 7 papers were eventually accepted. Before briefly summarizing the contributions of these papers, the next section sets out to define the concept of Question Generation and to identify different types of Question Generation. Section 2 introduces the papers in this issue. We conclude with some remarks on the prospects and future direction of Question Generation as a topic of research.

## 1 Question Generation Tasks as Computational Problems

The emerging Question Generation (QG) community has adopted the following definition of Question Generation (cf. Rus et al., 2008):

“Question Generation is the task of automatically generating questions from various inputs such as raw text, database, or semantic representation. Question Generation is regarded as a discourse task involving the following four steps: (1) when to ask the question, (2) what the question is about, i.e. content selection, (3) question type identification, and (4) question construction.” (Rus, n.d.)

---

<sup>2</sup> The 1<sup>st</sup> workshop took place in Arlington (hosted by the NSF), the 2<sup>nd</sup> in Brighton (co-located with the 14<sup>th</sup> International Conference on Artificial Intelligence in Education), the 3<sup>rd</sup> in Pittsburgh (co-located with the Tenth International Conference on Intelligent Tutoring Systems) and the 4<sup>th</sup> event, in 2011, was organized as an AAAI symposium in Arlington.

This definition views QG as the task of constructing algorithms that transform inputs to certain outputs. In computer science, an algorithm is “a tool for solving a well-specified *computational problem*.” (Cormen et al., 2009:5) The definition of QG raises an entire family of computational problems, rather than a single one: it gives examples of different types of input, leaves the relation between the input and the output undetermined, and says of the output only that it should be a question. In other words, the definition of Question Generation gives rise to a large variety of question generation problems. In order to contextualise the papers in the current issue, we aim to chart some of this variety. Following Piwek et al. (2008), we characterize a specific question generation problem by the answers to the following three questions:

1. What is the *input*?
2. What is the *output*?
3. What is the *relation between* the input and the output?

Let us start with the *output*, since the constraints on this seem most stringent. According to the definition, the output should be a question. Even here, there is, however, scope for several interpretations. One possible interpretation of this is that the output should *syntactically* qualify as a question, i.e., an interrogative sentence. This seems, however, to be overly restrictive, and a more plausible interpretation is that the output should be pragmatically a question, i.e., a *request for information*. Requests for information can take many different forms, including the utterance of an interrogative sentence (“How late is it?”), the utterance of a declarative sentence with a question intonation (“Mary is at home?”) and the utterance of an imperative (“Tell me where I can find the off switch.”). Of course, the imperative “Tell me where I can find the off switch.” features an embedded question (“where I can find the off switch”), but not all embedded questions signal the presence of a pragmatic question. For example, the assertion “I know where the off switch is” is not a request for information, in contrast with “Tell me where I can find the off switch.” Additionally, not all pragmatic questions contain overt embedded questions; consider, e.g. “Tell me the time” as one way for asking what time it is.

Even though we believe that the appropriate concept of a question for the purpose of the Question Generation community is that of a pragmatic question, i.e., a request for information, most work so far has focused on the more narrow problem of generating interrogative sentences. This holds true, for example, for the contributions to this issue. Additionally, the work reported here concentrates on written questions in a Natural Language (specifically, English and French). Many alternative question generation problems can, however, be envisaged if we change the output medium (auditory, visual, tactile, ...) or modality (Natural Language, Gesture, Knowledge Representation, ...). For example, a spoken question (medium: auditory; modality: Natural Language) could consist of a declarative sentence that is uttered with question intonation or which is uttered in a context which leads the addressee to infer that it is a question rather than an assertion (see Beun 1990 on ‘declarative questions’). Requests for information can also be multimodal (modality: Natural Language & Gesture) as in ‘Who is that?’ (accompanied by a pointing act). It may even be possible to ask a question using facial expressions only (e.g., raising an eyebrow at the right point in a conversation).

Turning our attention to the *input* side, we have the same wealth of possibilities. Again, the ground covered so far in QG research seems to be limited. The work in this issue focuses on input in the form of written text and knowledge representations. By varying medium and modality we arrive at many further possibilities: spoken language input, diagrams, presentations by virtual agents, etc. The input also need not consist of declarative, asserted, information; for example, the input could itself be a question.

Input and output do not need to share medium and modality: the input could be diagram and the output a written (or spoken) question about that diagram. Additionally, even when the

medium or modality is constant across input and output there may be significant differences. For example, input and output could be in Natural Language, but with the input language German and the output language French. This would give rise to a *cross-language* question generation problem.

In third place, and perhaps most importantly, there is the *relation* between the input and the output. Here again, research on QG has focused mainly on one particular relation: the situation where the output question is answered by the input (e.g., input = “John bought five cakes”, and output = “How many cakes did John buy?”) There has, however, also been some interest in question reformulation, where the output question is a clearer, better or different formulation of the input (which is a question or a search query), see Marciniak (2008). Wisniewski (1995) pioneered the notion that an output question can be raised by an input, as in the following example. The input ‘If John tried to check in after 12:00, he missed the flight. Did John miss the flight?’ raises the question ‘Did John try to check in after 12:00?’ Finally, one further relation between input and output is that of clarification: the output question requests information about the form, content, or intended use of the input (e.g., input = “Remove that block” and output = “The green one?”). There are many more relations, some of which are implicit in the QG Mechanisms that are described in Graesser et al. (1992).

The papers in this issue deal with declarative (asserted) input, either in Natural Language (French or English) or a Knowledge Representation formalism (concept maps), and output consisting of interrogative sentences, which are answered by the input. This significantly narrows down the set of QG problems that are addressed. The algorithms for *solving* these problems vary, however, in a number of ways. We conclude this section, by focusing on one particular dimension along which they differ, which is loosely based on the Vauquois triangle, or pyramid, from research in Machine Translation (Vauquois, 1968). Machine Translation (MT) algorithms take a source language text and map it to a target language text. Three approaches can be distinguished known as direct, transfer and interlingua MT. In direct MT, the transformation rules operate directly on the strings of the source text to obtain the target text. For transfer, an intermediate (syntactic or semantic) structure is constructed for the input text; this is mapped to the corresponding structure in the target language; the structure for the target text is then mapped to the target language text itself. In contrast with the transfer approach, the interlingua approach is based on a *single* (conceptual) structure to which the source text is mapped and from which, in turn, the target text is generated.

In Text-to-Text QG, the input is a declarative sentence. Usually, a pre-processing step is involved which divides this sentence up into sentences of a size that is appropriate for generating questions. Similar to the hierarchy in MT (direct, transfer and conceptual), there is a hierarchy of QG approaches. In principle, it is possible to carry out QG strictly on the string level (Figure 1.A). To our knowledge, there are, however, no systems that directly implement this approach. Rather, most systems carry out syntactic processing and some semantic processing to arrive at an intermediate representation, which may or may not include the original source text. The amount of semantic processing varies from only partial analysis (e.g. recognition of named entities) to full processing into a semantic representation language. The approaches share the use of transformation rules. There are, however, two different kinds of transformation rule: 1) rules that map the representation of the input text, which includes the input text itself, directly to the target question without any further intermediate representation (Figure 1.B), and 2) rules that map the representation of the input text, which may or may not include the input text itself, to a representation of the target question (Figure 1.C and D). The transformation rules of the second kind require a further step that consists of generating the target question text from the target question representation. At the top of the pyramid in Figure 1 we have transformations from a semantic representation of input text to a semantic representation of the target question. Note the absence of an equivalent to interlingua MT: in MT, the conceptual representation for the content of the target and source language text is the same (it resides at the peak of the pyramid); in

contrast, there is no shared conceptual representation (i.e., the pyramid has no peak) that captures both the content of the declarative source text and the interrogative target question.

We should add that it is possible to use transformation rules that overgenerate and are followed by a post-processing step that weeds out ill-formed outputs using an independent measure, e.g. by ranking generated question with an appropriate language model (see Heilman & Smith, 2009; 2010).

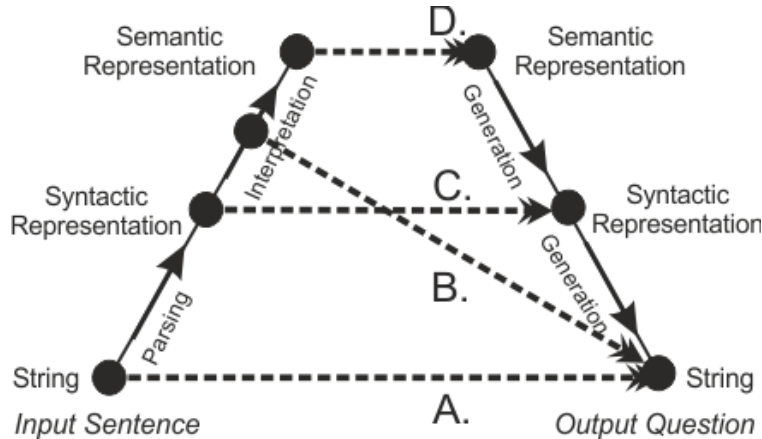


Figure 1 The Question Generation Pyramid

## 2 Overview of Papers in this Issue

Rather than try to replicate the excellent abstracts that precede each of the papers, in this section we contrast and compare the papers using some of the distinctions that were introduced in the previous section. We group the seven papers into four broad categories, primarily based on domain of application: 1) application-neutral generation of questions from sentences, 2) question generation in educational settings, 3) question generation for virtual agents, and 4) question generation shared tasks and evaluation.

### 2.1 Application-neutral Generation of Questions from Sentences

The paper by Yao, Bouma & Zhang entitled “Semantics-based Question Generation and Implementation” falls firmly in the category *question generation from declarative sentences*. The main contribution of this paper is to demonstrate the feasibility of semantics-based QG. The open source MrsQG system is the first of its kind, implementing a semantic rewriting approach (see Figure 1.D). It relies on parsing the English input sentence (after some preprocessing) to a structure in Minimal Recursion Semantics (MRS; Copestake et al., 2005). A rewriting step then maps the MRS for the declarative sentence into an MRS for a corresponding question. Language generation is deployed to turn this MRS into an English output question. Both parsing and language generation are based on existing open source tools for MRS. The rewriting step is in principle language independent: it maps the language neutral MRS for the input to an MRS for the output question. The system performed very well in the first Question Generation Shared Task and Evaluation Campaign (QGSTEC; Rus et al., this volume). The QGSTEC was set up to be application neutral, using texts on a variety of topics and evaluation measures that ranked approaches based on generic criteria (such as syntactic and semantic correctness of the generated questions).

The paper by Bernhard, De Viron, Moriceau & Tannier (“Question Generation for French: Collating Parsers and Paraphrasing Questions”) relies less on semantics and more on syntax: in particular, it draws on the outputs of two different syntactic parsers in combination with named entity recognition. In terms of Figure 1, the approach falls somewhere between C. and D. In this system, generation concerns removing information from the output question’s syntax tree and application of elisions (e.g., from *que* to *qu’*). Additionally, the system can generate reformulations based on substitution of different question words. The paper is a first in that it introduces an application-neutral questions-from-sentences generator for French. The authors describe a range of ways in which their system was evaluated. These include human judgments on acceptability of generated question, use of post-edited questions as a target reference against which generated questions are compared using the Human-targeted Translation Edit Rate (HTER; Snover et al., 2006), recall in terms of the CLEF Question Answering campaign corpus, and extrinsic evaluation in terms of how well generated questions can be parsed correctly automatically.

## 2.2 Question Generation in Educational Settings

Olney, Graesser & Person (“Question generation from concept maps”) present a method for generating questions for tutorial dialogue. This involves automatically extracting concept maps from textbooks. In contrast with the aforementioned QG from sentences approaches, this approach does not deal with the input text on a sentence-by-sentence basis only. Rather, various global measures (based on frequency measures and comparison with external ontologies) are applied to extract an optimal concept map from the textbook. The template-based generation of questions from the concept maps allows for questions at different levels of specificity to enable various tutorial strategies, from asking more specific questions (which indicated precisely to the student which information they are expected to provide) to a student who is struggling, to the use of less specific questions to stimulate extended discussion. Evaluation of the approach follows the QGSTEC methodology (Rus et al., this volume), enriched with an evaluation of the pedagogical value of questions.

Liu, Calvo & Rus (“G-Asks: An Intelligent Automatic Question Generation System for Academic Writing Support”) introduce a system which helps students to develop their writing skills. It focuses on writing around citations. The approach is template-based, and takes as input individual sentences. In contrast with most other approaches in this issue, the relation between the input sentence and the question is not one of answerhood. Rather, the questions generated by the system typically ask for evidence or support for the claim that is made by the input sentence (e.g., input = “Cannon (1927) challenged this view mentioning that physiological changes were not sufficient to discriminate emotions.”; output = “Why did Cannon challenge this view mentioning that physiological changes were not sufficient to discriminate emotions?”).

## 2.3 Question Generation for Virtual Agents

The papers in this section are concerned with the generation of questions for the benefit of dialogue agents, and more specifically virtual agents, i.e., dialogue agents that have a computer-animating embodiment. Work on such agents faces the challenges of portability and scalability: how to create agents that answer and/or ask questions in new or very large domains.

Yao, Tosch, Chen, Nouri, Artstein, Leuski, Sagae & Traum (“Creating Conversational Characters Using Question Generation Tools”) use question generation technology to automatically construct a repository of question-answer pairs from text. A virtual agent can use such a repository to respond to user questions by retrieving an answer from the repository that answers a question that is similar to the user’s question. Yao et al. have built a tool that incorporates the Question Transducer of Heilman and Smith (2009) and perform three experiments to gauge the prospects

of this approach for automatically extracting question-answer repositories. Their experiments address the issue whether the approach is feasible at all for creating new dialogue characters, and also whether it can be used to enrich manually created repositories (without degrading their quality too much).

In contrast with Yao et al., Mendes, Curto & Coheur (“Question Generation based on Lexico-Syntactic Patterns Learned from the Web”) describe work on a virtual agent which primarily *asks* rather than *answers* questions. Their work is about automatically harvesting suitable questions for such an agent. The THE-MENTOR platform starts from a set of seed question/answer pairs. It learns patterns between questions and answers based on information obtained from querying the Web with queries that have been automatically constructed from the seed questions. The patterns are used to generate fresh questions from new input sentences. Evaluation of the QG patterns was performed on the QGSTEC 2010 development set and a Wikipedia page, rating grammatical and semantic correctness. In terms of the QG Pyramid (Fig. 1), the approach is closest to B.

#### 2.4 Question Generation Shared Tasks and Evaluation

The final paper by Rus, Wyse, Piwek, Lintean, Stoyanchev & Moldovan provides “A Detailed Account of The First Question Generation Shared Task Evaluation Challenge”. The QGSTEC was held in 2010 as part of the 3<sup>rd</sup> Workshop on Question Generation (Boyer & Piwek, 2010) and involved two application-neutral tasks: A. Question Generation from Paragraphs and B. Question Generation from Sentences. Five teams participated in the QGSTEC, with approaches ranging from those based on shallow surface features and syntax to deep semantics. Human judges evaluated generated questions on several criteria: Relevance, Question Type, Syntactic Correctness and Fluency, Ambiguity and Variety. Additionally, for the QG from paragraphs task, the task was to generate questions with different scope (from broad, paragraph level, to phrase level or less).

### 3 Concluding Remarks

This special issue presents a range of Question Generation problems and approaches that are reasonably representative of the current direction of QG research. This research is somewhat biased towards textual input, possibly partly as a result of the recent Question Generation Shared Task and Evaluation Campaign (QGSTEC; Rus et al., this volume). For the QGSTEC, raw text was chosen as the input format to avoid ruling out participants due to specific theoretical commitments that other input formats (e.g., knowledge representation formalisms) might entail. With Semantic Web notations (such as OWL) gaining widespread acceptance, there is, however, good reason to revisit this decision and investigate whether future challenges could also include tasks based on non-linguistic input. An attractive aspect of this is that genuine knowledge representation inputs allow deeper reasoning about the content of questions and, consequently, also about decisions regarding which question to ask. This, in turn, would make it possible to stimulate a strand of deep Question Generation in contrast with, for example, the current ‘Generating Questions from Sentences’ task which stimulates primarily research that focuses on direct shallow relations between the input text and the output question.

Though all the approaches in this volume deal with Question Generation as a computational problem, they draw on other disciplines, notably linguistics, computational semantics, psychology and education. There is, however, less evidence that current work builds on the extensive literature in the field of Formal and Computational Semantics that deals with the semantics and pragmatics of questions (Groenendijk & Stokhof, 1997; Ginzburg & Sag, 2001). Here, there seems to be a further opportunity, for example, to draw on notions such as partial and conditional



answerhood, to gain a deeper and richer understanding of the relation between questions and their answers.

### Acknowledgements

The call for this special issue followed on from the 3<sup>rd</sup> workshop on Question Generation. In response to the call for statements of intent, we received 26 abstracts. Some of the submissions, but not all, were based on papers presented at the workshop. In a first round, the special issue editors provided authors with feedback on the suitability of their proposed topic for the special issue. In response, 14 full papers were submitted that then went through a two-stage reviewing process. Eventually, 7 papers were accepted for the special issue. Each paper was reviewed by three reviewers. Paper authors were asked to review one other submission. Additionally, we are grateful to the following reviewers: Greg Aist, Lee Becker, Nadjet Bouayad-Agha, Marc Cavazza, Mark Core, Barbara di Eugenio, Aldabe Itziar, Rodger Kibble, James Lester, François Mairesse, Smaranda Muresan, Rodney Nielsen, Rashmi Prasad, and Mariët Theune. Also, we acknowledge the help of Anton Dil, Chris Fox, Julius Goth, David King, Chris Mitchell, and Alistair Willis with proofreading the contributions to this issue. Finally, we would like to thank the managing editors of *Dialogue & Discourse*, in particular, Jonathan Ginzburg (Editor-in-Chief) and Amanda Stent for their support throughout the preparation of this issue.

### References

- Ahn, R.M.C., Beun, R.J., Borghuis, T., Bunt, H.C. & van Overveld, C.W.A.M. (1995). The DENK-architecture: a Fundamental Approach to User-Interfaces. *Artificial Intelligence Review Journal*, 8(2): 431-445.
- Beun, R.J. (1990). The recognition of Dutch Declarative Questions. *Journal of Pragmatics*, 14:39-56.
- Boyer, K. and P. Piwek (eds.) (2010). *Proceedings of the 3<sup>rd</sup> Workshop on Question Generation, Part of the 10<sup>th</sup> Conference on Intelligent Tutoring Systems (ITS 2010)*, Carnegie Mellon University, Pittsburgh PA., June 18, 2010.
- Bronnenberg, W.J., H.C. Bunt, J. Landsbergen, P. Medema, R. Scha, W.J Schoenmakers and E. van Utteren (1979). The question-answering system PHLIQA 1. In: L. Bolc (Ed.), *Natural communication with computers*. MacMillan.
- Cohen, F.S. (1929). What is a Question? *The Monist* 39: 350-364.
- Copestake, A., D. Flickinger, C. Pollard, and I. Sag (2005). Minimal Recursion Semantics: An Introduction. *Research on Language & Computation*, 3(4): 281-332.
- Cormen, T., C. Leiserson, R. Rivest and C. Stein (2009). *Introduction to Algorithms*. MIT Press.
- Evans, R., P. Piwek and L. Cahill (2002). What is NLG? *Proceedings of International Natural Language Generation Conference INLG02*, New York, USA, 1-3 July 2002.
- Ginzburg, J. & I. Sag (2001). *Interrogative Investigations*. CSLI Publications/University of Chicago Press.
- Ginzburg, J. (2012). *The Interactive Stance: Meaning for Conversation*. Oxford University Press.
- Graesser, A. C., Person, N. K., & Huber, J. D. (1992). Mechanisms that generate questions. In: T. E. Lauer, E. Peacock, & A. C. Graesser (Eds.), *Questions and information systems* (pp. 67-187). Erlbaum.
- Green, B. F., Wolf, A. K., Chomsky, C. and Laughery, K. (1961). BASEBALL: An automatic question answerer. In: *Proceedings Western Joint Computer Conference 19*, pp. 219-224.

- Groenendijk, J. and M. Stokhof (1997). Questions. In: Van Benthem, J. and A. ter Meulen (Eds.), *Handbook of Logic & Language*. North-Holland.
- Heilman, M. and N. Smith (2009). Question generation via overgenerating transformations and ranking. Technical Report CMU-LTI-09-013, Carnegie Mellon University Language Technologies Institute.
- Heilman, M. and N. Smith (2010). Good Question! Statistical Ranking for Question Generation. In: Proc. NAACL/HLT 2010, Los Angeles.
- Hirschman, L. and R. Gaizauskas (2001). Natural language question answering: the view from here. *Natural Language Engineering* 7(4): 275-300.
- Kievit, L., P. Piwek, R.J. Beun and H. Bunt (2001). Multimodal Cooperative Resolution of Referential Expressions in the DENK system. In: Bunt, H. and R.J. Beun, *Cooperative Multimodal Communication, Lecture Notes in Artificial Intelligence Series 2155*, Springer, Berlin/Heidelberg, pp.197-214.
- Marciniak, T. (2008). Language generation in the context of yahoo! Answers. In: V. Rus and A. Graesser (eds.), *online Proceedings of 1<sup>st</sup> Question Generation Workshop, September 25-26, 2008*, NSF, Arlington, VA.
- McDonald, D. (1993). Issues in the choice of a source for natural language generation. *Computational Linguistics*, 19(1):191–197.
- Piaget, J. (1952). *The origins of intelligence*. International University Press.
- Piwek, P. (1998). *Logic, Information & Conversation*. PhD Thesis, Eindhoven University of Technology. URL: <http://alexandria.tue.nl/extra2/9803039.pdf>
- Piwek, P., H. Prendinger, H. Hernault, and M. Ishizuka (2008). Generating Questions: An Inclusive Characterization and a Dialogue-based Application. In: Rus, V. and A. Graesser (eds.), *online Proceedings of 1<sup>st</sup> Question Generation Workshop, September 25-26, 2008*, NSF, Arlington, VA.
- Power, R. (1979). The organisation of purposeful dialogues. *Linguistics*, 17, 107–152.
- Purver, M. (2004). *The Theory and Use of Clarification Requests in Dialogue*. PhD Thesis, University of London, 2004.
- Reiter, E. and R. Dale. (2000). *Building Natural Language Generation Systems*. Cambridge University Press, Cambridge, UK.
- Rus, V. (n.d.) Question Generation Main Website. [questiongeneration.org](http://questiongeneration.org) (Accessed 19 January 2012).
- Rus, V., A. Graesser, A. Stent, M. Walker & M. White (2007). Text-to-Text Generation. In: Dale, R. & M. White (Eds.), *Shared Task and Comparative Evaluation in Natural Language Generation, Workshop Report*, Arlington, VA.
- Rus, V., Z. Cai and A. Graesser (2008). Question Generation: Example of A Multi-year Evaluation Campaign. In: Rus, V. and A. Graesser (eds.), *online Proceedings of 1<sup>st</sup> Question Generation Workshop, September 25-26, 2008*, NSF, Arlington, VA.
- Scha, R. (1983). *Logical Foundations for Question Answering*. Ph.D. Dissertation, University of Groningen, M.S. 12.331, Philips Research Laboratories, Eindhoven, The Netherlands.
- Snover, M., B. Dorr, R. Schwartz, L. Micciulla, and J. Makhoul (2006). A Study of Translation Edit Rate with Targeted Human Annotation. In: *Proceedings of Association for Machine Translation in the Americas, 2006*.
- Stent, A. (2001). *Dialogue Systems as Conversational Partners: Applying conversation acts theory to natural language generation for task-oriented mixed-initiative spoken dialogue*. Unpublished PhD Thesis. University of Rochester.
- Vauquois, B. (1968). A survey of formal grammars and algorithms for recognition and transformation in machine translation. In: *IFIP Congress-68*, pp. 254-260.
- Winograd, T. (1972) *Understanding Natural Language*. Academic Press, New York.
- Wisniewski, A. (1995). *The Posing of Questions: Logical Foundations of Erotetic Inferences*. Synthese Library / Volume 252. Kluwer Academic Publishers.