Impact of Question Decomposition on the Quality of Answer Summaries

Finley Lacatusu, Andrew Hickl, Sanda Harabagiu

Language Computer Corporation
1701 N. Collins Blvd.
Richardson, TX 75080, USA
{finley, andy, sanda}@languagecomputer.com

Abstract

Generating answers to complex questions in the form of multi-document summaries requires access to question decomposition methods. In this paper we present three methods for decomposing complex questions and we evaluate their impact on the responsiveness of the answers they enable.

1. Introduction

Complex questions cannot be answered by a single entity or even a single sentence. Typically, complex questions address a topic that relates to many entities, events and complex relations between them. For example, when asking about *international organized crime* many concepts, such as criminals, gangs and their organized crimes come to mind. To convey information related to such complex topics, multiple sentences need to be used and organized in a summary if all relevant data is extracted from a document collection. Since such summaries need to be informative and to respond to complex questions, they constitute a good basis for evaluating the state-of-the-art of multi-document summarization. This was the goal of DUC-2005 [1].

The complex question-focused summarization task in DUC 2005 required systems to piece together information from multiple documents. The multi-document summaries (MDS) produced by such systems had to answer a question or a set of questions as posed by a DUC topic. NIST Assessors developed a total of 50 DUC topics to be used as test data. For each topic, the assessor selected 25-50 related documents from the Los Angeles Times and Financial Times of London and formulated a DUC topic statement, which was a request for information that could be answered using the selected documents. The topic statement could be in the form of a question or set of related questions and could include background information that the assessor thought would help clarify his/her information need. An example of a DUC 2005 topic is provided in Figure 1(a).

We argue that the quality of question-focused summaries like the ones evaluated in DUC 2005 depends in part on how complex questions are decomposed. In order to provide summary answers that are both accurate and responsive, we feel that complex questions need to be decomposed into a set of simpler questions that they entail. This can be accomplished in two ways. First, complex questions can be decomposed *syntactically* by extracting each of the overtly-mentioned subquestions included in a complex question. Since complex questions are used seek multiple types of information simultaneously, it is not uncommon to see complex questions that feature coordinated entities or clauses. For example, the complex question Q^c in Table 1 features two instances of syntactic coordination. There is on e coordination between two imperative verbs: V_1=[*identify*] and V_2=[*describe*]. There is also a coordination in the relative clause between the verb phrases. VP_1=[*crosses borders*] and VP_2=[*involves more than one country*]. The coordinated expressions are derived from the syntactic parse of the complex question, enabling us to generate four question decompositions, listed in Table 1. It is interesting to note that each question decomposition corresponds to a different dimension of the original question's information need.

International Organized Crime
Identify and describe types of organized crime that crosses borders or involves more than one country. Name the countries involved. Also identify the perpetrators involved with each type of crime, including both individuals and organizations if possible.

(a)

Semantic Signature

Types of Organized Crime
drug trafficking, money laundering, arms trading, smuggling of illegal immigrants

Perpetrator

Organizations	*Individuals*
Columbian drug cartels	Pablo Escobar
Italian Mafia	Manuel Noriega
Cosa Nostra	Pakistani tribal leaders
Camorra	Gen. Amaldo Ochoa Sanchez
Chinese Triads	Juan Sanchez-Perez
Japanese Yakuza	Leopoldo Piloto

Location
Panama, Mexico, Guatemala, Nigeria, Cuba, Pakistan, Russia, Peru

(b)

Figure 1: (a) Example of complex question evaluated in DUC 2005; (b) the Semantic Signature for the complex question.

Q^c: [Identify]$_{V_1}$ and$_{CC_1}$ [describe]$_{V_2}$ [types of organized crime] that [crosses borders]$_{C_1}$ or$_{CC_2}$ [involves more than one country]$_{C_2}$.
Q_1: $Identify_{V_1}$ types of organized crime that crosses borders.
Q_2: $Describe_{V_2}$ types of organized crime that crosses borders.
Q_3: $Identify_{V_1}$ types of organized crime that involves more than one country.
Q_4: $Describe_{V_2}$ types of organized crime that involves more than one country.

Table 1: Syntactic Decomposition of a Complex Question.

Additional question decompositions can be identified by utilizing sources of semantic and pragmatic information. In this paper we present two novel methods used for decomposing complex questions that employ: (1) the expected an-

[1] The Document Understanding Conferences (DUC) are conference series run by the National Institute of Standards and Technology (NIST) to further progress in summarization and enable researchers to participate in large-scale experiments.

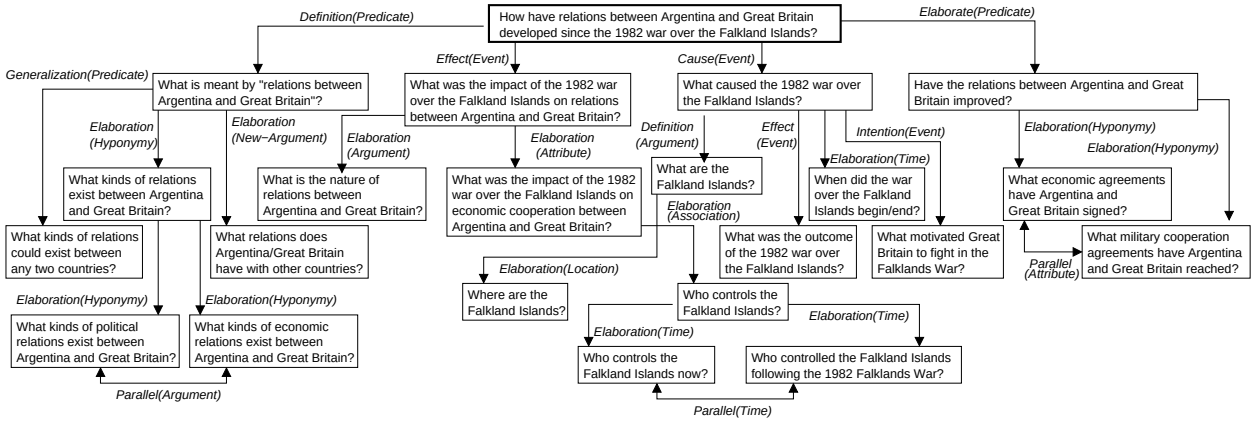

Figure 2: Top-Down Question Decomposition.

swer type (EAT), (2) possible discourse relations between questions; and (3) the semantic dependencies revealed by the predicate-argument structures discovered in questions. We define an expected answer type (EAT) as the set of concepts, events, and relations that represent the range of information sought by a question. In the case of question Q_1, the expected answer type consists of a list of events that pertain to the topic of international organized crime. The topic of a question is provided by the semantic signatures generated automatically with the method reported in (Harabagiu, 2004). (An example of one automatically-generated semantic signature is illustrated in Figure 1(b).)
The remainder of this paper is organized as follows. In Section 2 we describe three methods for automatically decomposing questions. In Section 3 we report the usage of question decompositions for generating answer summaries. Section 4 presents the results of our evaluations, whereas Section 5 summarizes the conclusions.

2. Question Decomposition

In this section, we describe three techniques for automatic question decomposition that we have implemented in order to create answer summaries for complex questions. In Subsection 2.1, we discuss how questions can be decomposed syntactically, while in Subsections 2.2 and 2.3, we introduce two complementary techniques for decomposing questions semantically.

2.1. Syntactic Question Decomposition

Complex questions often include multiple requests for information in the same sentence. In an analysis of 125 complex questions taken from the 2004 AQUAINT Relationship Q/A Pilot, the 2005 TREC Q/A Track Relationship Task and the 2005 DUC Question-Focused Summarization task, we found that 49 questions included more than one overt, simple question. We refer to complex questions that feature more than one overt question as *syntactically-complex questions*. We have identified three types of syntactically complex questions: (1) questions that feature coordination (of question stems, predicates, arguments, or whole sentences), (2) questions that feature lists of arguments or clauses, and (3) questions that feature embedded or indirect questions. Examples of each of these three types are provided in Table 2.

Coordination:	When and where did Fidel Castro meet the Pope?
Lists of Arguments:	What international aid organization operates in Afghanistan, Iraq, Somalia, and more than 100 other countries?
Embeded Questions:	The analyst would like to know of *any attempts by these governments to form trade or military alliances*.

Table 2: Syntactic Question Decompositions.

Questions that feature coordination of their stems are the easiest to decompose. For example, question Q_2^c: "*When and where did Fidel Castro meet the Pope?*" is decomposed into Q_2^1: "*When did Fidel Castro meet the Pope?*" and Q_2^2: "*Where did Fidel Castro meet the Pope?*" by simply selecting one of the question stems and following the surface realization of the dependencies in the original complex question.

The analyst is concerned with a possible relationship between the Cuban and Congolese governments. Specifically, the analyst would like to know of any attempts by these governments to form trade or military alliances.

Figure 3: Complex Question.

Syntactic question decomposition requires often the resolution of anaphora. We begin by identifying the antecedents of all referential expressions found in a complex question. For example, the complex question illustrated in Figure 3 requires the resolution of the NP "*these governments*" to the set: {*the Cuban government*; *the Congolese government*}. The coreference resolution algorithm we use was described in (Harabagiu et al., 2001). Once coreference is established, we process questions by using a set of syntactic patterns to extract embedded questions and to split questions that featured conjoined phrases or lists of terms into individual questions. For example, the complex question illustrated in Figure 3 is broken down in:
(1) "*What attempts have been made by [the Cuban and Congolese governments] to form trade alliances?*" and
(2) "*What attempts have been made by [the Cuban and Congolese governments to form military alliances?*".

2.2. Top-Down Question Decomposition

Even after syntactic question decomposition is performed, most complex questions still need to be be decomposed semantically before they can be submitted to a traditional Q/A system. In this subsection, we present a method for semantic question decomposition which utilizes relations that exist between questions in order to break down complex questions into the set of simpler questions that they entail.

Relation	Example
DEFINITION Relations	
DEFINITION(predicate)	What are the procedures for generating new drugs?
DEFINITION(argument)	What is a drug?
ELABORATION Relations	
ELABORATION(Hyponymy)	Which new drugs are being produced?
ELABORATION(Number)	How many new drugs are being produced?
ELABORATION(Time)	When are new drugs being produced?
ELABORATION(Location)	Where are new drugs being produced?
ELABORATION(Manner)	How do pharma companies produce new drugs?
ELABORATION(Quantity)	How many new drugs did pharma companies produce?
ELABORATION(Rate)	What was the greatest number of new drugs that pharma companies produced?
ELABORATION(Duration)	How long will pharma companies produce new drugs?
ELABORATION(Trend)	How much has pharma companies' production of new drugs increased/decreased?
ELABORATION(Inchoative)	When did pharma companies begin producing new drugs?
ELABORATION (Terminative)	When did pharma companies stop producing new drugs?
ELABORATION(Subjective)	How beneficial/detrimental was pharma companies' production of new drugs?
EVENT - EVENT Relations	
CAUSE(Event)	What steps did pharma companies take to produce new drugs?
INTENTION(Event)	Why did pharma companies produce new drugs?
EFFECT(Event)	What happened because pharma companies produced new drugs?
RESULT(Event)	What advantages resulted pharma companies producing new drugs?
OUTCOME(Event)	What profits did pharma companies take from producing new drugs?
TEMPORAL(Event)	What happened after/before pharma companies produced new drugs?
RELATIONSHIP(Event)	What is the connection between pharma companies producing new drugs and the higher incidence of autism in the US?

Relation	Example
GENERALIZATION / SPECIALIZATION Relations	
SPECIALIZATION (Predicate)	What kind of activities are involved in the creation of new drugs?
GENERALIZATION (Predicate)	What commercialization effects are typical in the drugs industry?
COUNTERFACTUAL Relations	
NEGATION(Predicate)	What pharma companies don't produce new drugs?
NEGATION(Argument)	What pharma companies have produced no (new) drugs?
NEGATION(Attribute)	What pharma companies produce only existing drugs?
EXCEPTIVE (Argument)	What pharma companies are producing new drugs other than MAOI inhibitors?
CONTRARY(Fact)	Despite the FDA's ban on new drug development, which pharma companies are producing new drugs?
ANSWER RESTRICTING Relations	
RESTRICT(Location)	What pharma companies are producing new drugs in the U.S.?
RESTRICT(Temporal)	What pharma companies are producing new drugs in 2006?
RESTRICT(Attribute)	What up-and-coming pharma companies are producing new drugs?
EPISTEMIC Relations	
EPISTEMIC(Event)	Are pharma companies producing new drugs?
EPISTEMIC-CONDITIONAL(Event)	Is it known if pharma companies are producing new drugs?
EPISTEMIC-EVIDENTIAL(Event)	Is there evidence that pharma companies are producing new drugs?
EPISTEMIC-REPORTED(Event)	Does anyone believe that pharma companies are producing new drugs?
EPISTEMIC-ALTERNATIVE(Event)	Do U.S. pharma companies produce new drugs or research new drugs?
EPISTEMIC-ELABORATION(Event)	Do pharma companies produce new drugs [with the help of foreign labs]?
PARALLEL Relations	
PARALLEL(Predicate)	What pharma companies work with infectious agents?
PARALLEL(Predicate)	What pharma companies research new drugs?
PARALLEL(Argument)	What pharma companies produce vaccines?
PARALLEL(Attribute)	What pharma companies produce affordable drugs?

Figure 4: Relations between questions and examples. The examples were obtained when decomposing the complex question "*What pharmaceutical companies produce new drugs?*".

We believe that the relations that exist between a question and its decompositions may be of semantic nature (e.g. definitions, generalizations) or of discourse nature (e.g. elaborations, causes, effects, parallelisms). Furthermore, unlike discourse relations introduced by various coherence theories, the relations between questions have an argument. This argument may take any of the values: (1) PREDICATE, (2) EVENT, (3) ARGUMENT, (4) ATTRIBUTE, or (5) HYPERNYMY/HYPONYMY. The first four values refer to a predicate, event, argument or attribute detected in the mother-question, which will also be referred by the daughter-question. The last value (hypernymy/hyponymy) indicates that there is such a semantic relation (defined in WordNet) between a pair of concepts, one for the mother-question, one for the daughter-question. For example, in Figure 2 we illustrate several relations between questions that are labeled EFFECT(Event). This type of relation indicates (1) that the EAT of the decomposed question is a class of concepts that are caused by some event; and (2) the event is explicit in both questions. In Figure 2, the referred event is the 1982 war between Argentina and Great Britain. We have considered eight different classes of relations between questions that can be used to perform semantic question decompositions. The relations are illustrated in Figure 4.

When decomposing complex questions in a top-down manner, we need to have access to five forms of information:

◁1▷ the predicate-argument structures of current question;

◁2▷ the EAT of the current question;

◁3▷ the most relevant relations obtained from the topic signature of the text collection;

◁4▷ association-information connecting the question to the most likely decompositions.

Predicate-argument structures are provided by shallow semantic parsers trained on PropBank[2] and NomBank[3]. The EAT of the question is discovered with the technique reported in (Pasca and Harabagiu, 2001). The most relevant relations from the question topic are identified by the enhanced representations of topic signatures reported in (Harabagiu, 2004). For each EAT we have created a large set of 4-tuples (e_1, r, e_2, A), that we call *association information*. The association information consists of (1) e_1, the EAT of the mother-question; (2) r, the relation between the pair of questions; (3) e_2, the EAT of the daughter-question; and (4) A, the *alignment information* between the two questions. To produce the alignment information, we have used a lexical alignment module that was built for our Textual Entailment system called GROUNDHOG, which was described in (Hickl et al., 2006). The lexical alignment module, trained on 460,000 alignment pairs, uses a maximum entropy classifier that generates alignment information with a precision of 90.2% and recall of 94.4%. Figure 5 illustrates the association information relating two questions from Figure 2.

When generating a decomposition, we use the association information to build association rules similarly to the method introduced in (Nahm and Mooney, 2000). The association information is akin to the fillers of a template. Therefore, by representing it as binary features that are provided to a decision tree classifier (C5.0 (Quinlan, 1998)), we generate automatically association rules from the decision rules of the classifier. In order to find the new informa-

[2] *http://www.cis.upenn.edu/∼mpalmer/project_pages/ACE.htm*

[3] *http://nlp.cs.nyu.edu/meyers/NomBank.html*

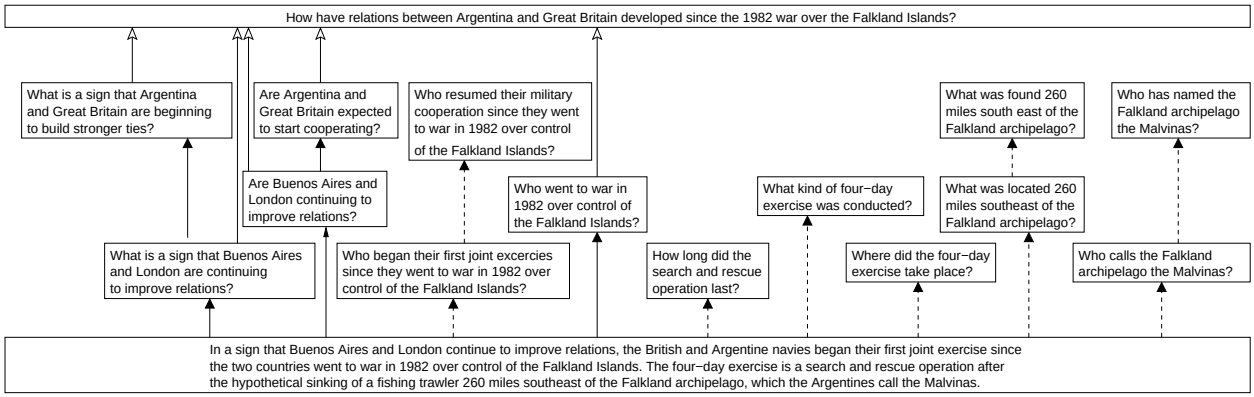

Figure 6: Bottom-Up Question Decomposition.

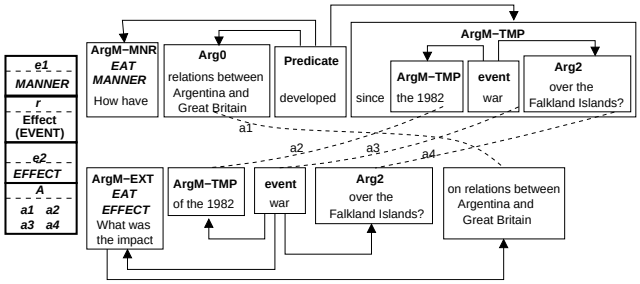

Figure 5: Example of Question Patterns.

tion, that specializes the decomposed question, we select the topic relation that (a) fits best the predicate-argument structure of the decomposed question, and (b) produces similar lexical alignment while preserving grammaticality. These last two conditions must be met by the question surface realization function. The Top-Down Question Decomposition Procedure is illustrated in Figure 7.

Step 1:	Generate predicate-argument structures for complex question.
Step 2:	Find the EAT of the question.
Step 3:	Find association rules from the classification of association information.
Step 4:	Use the association rules to have access to most likely (a) discourse relation to decomposed question (b) EAT of decomposed question (c) lexical alignment between complex question and decomposed question
Step 5:	Select most likely topic relations that fit the association information.
Step 6:	Produce surface realization of decomposed question.

Figure 7: Top-Down Question Decomposition Procedure.

2.3. Bottom-Up Question Decomposition

In contrast to the Top-Down question decomposition described in Subsection 2.2, complex questions can also be semantically decomposed in a Bottom-Up fashion by identifying the potential decomposition relations that may exist between sets of factoid questions related to the same topic. In previous work (Harabagiu et al., 2005b) we have described an approach that used syntactic patterns – in conjunction with semantic dependency and named entity information – to generate factoid questions automatically from large text corpora. This approach is illustrated in Figure 8. Figure 6 illustrates a bottom-up decomposition produced by the procedure from Figure 8. In Figure 6, all dashed arrows correspond to further produced decompositions that are not distancing themselves semantically from the complex question (Step 10 in Figure 8).

Step 1:	Text passages corresponding to candidate answers to factoid questions are identified and extracted from text, using techniques first developed for answer type detection for factoid Q/A (Harabagiu et al., 2005a)
Step 2:	Once these answers were identified, we used a pattern specification language to generate a factoid-style natural language generator from each answer's sentence. Examples of these automatically-generated questions are presented in Figure 6.
Step 3:	In order to measure the coverage of the set of questions generated from a text, we used a paraphrase acquisition system (similar to the method proposed in (Shinyama et al., 2002)) to generate additional questions that could be associated with each identified answer. Under this approach, each of the generated questions were parsed using a semantic parser trained on PropBank. Pairs of entities assigned a semantic role by the same predicate were then selected and used to generate a web query that returned the top 500 documents containing both entities. Sentences containing both terms were then extracted, and a method described in (Clifton and Teahan, 2004) was used to extract the intervening text (or paraphrase) that occurred between the terms. In order to ensure that only semantically equivalent paraphrases were used to create new questions, the set of extracted paraphrases were clustered (using a complete-link clustering algorithm introduced in (Barzilay and Lee, 2003); only clusters containing text passages extracted from the original question were considered to be viable paraphrases.
Step 4:	Once a set of new questions are generated, a concept similarity function = (2× *Nr of Alignments*) / (*Nr of Predicates and Arguments in both Questions*) was used to calculate the similarity between each pair of questions in the collection. This score was then used in a KNN clustering algorithm to cluster questions into sets which were assumed to seek similar types of information.
Step 5:	For each question cluster, select it's centroid question.
Step 6:	Find association rules from the classification of association information when considering the centroid question.
Step 7:	Use association rules to have access to the most likely (a) discourse relation to a more complex question, (b) EAT of the more complex question, (c) lexical alignment between the complex and the more complex question that is proposed.
Step 8:	Use the alignment information to select the most likely topic relations that fit the association information.
Step 9:	Produce surface realization of the more complex question.
Step 10:	Measure the distance to the original question, by using the concept similarity described in Step 4. If the distance increased, STOP.
Step 11:	If more than one complex question was produced, GO TO Step 4.
Step 12:	When the conceptual similarity is above a threshold, STOP the decomposition.

Figure 8: Bottom-Up Question Decomposition Procedure.

3. Using Question Decompositions for MDS

Multiple Document Summaries (MDS) that meet the information needs of a complex question can be created when we have access to question decompositions. We have devised three methods for selecting sentences that are incorporated in the MDS. The first method extracts keywords from the question decompositions in order to rank the sentences. This method was employed in our DUC-2005 evaluation system (Lacatusu et al., 2005). The second method has used the decomposed questions for finding their answers with a Q/A system which ranked the answer passages. The third Method used textual entailment to select the sentences for MDS.

3.1. Method 1

Each complex question is associated with a cluster of documents from which the MDS needs to be produced. For each cluster of documents, two topic signatures TS_1 and TS_2 are automatically devised. TS_1 were introduced in (Lin and Hovy, 2000) as a list of ranked terms that are relevant to the topic discussed in the document cluster. Each relevant term received a weight, given by likelihood ratios. TS_2, introduced in (Harabagiu, 2004), represent a topic by its relevant binary relations (between concepts). Each relation is also weighted.

In this method, keywords were extracted automatically from each subquestion, and stopwords were removed. These keywords were then associated with sets of alternations originally developed for the automatic Q/A system (Harabagiu et al., 2005a). A sample of these alternations for two different terms is provided in Figure 9. All the keywords extracted from all decomposed questions and their alternations are collected in a set K_a. Each sentence S from the document cluster receives a score which is equal to the sum of (a) the number of topic-relevant terms from TS_1 encountered in S; (b) the number of topic-relevant relations from TS_2 encountered in S; and (c) the number of keywords from K_a which are recognized in S.

benefit: advance, advantage, aid, ameliorate, assist, avail, better, build, contribute to, favor, further, improve, make it, pay, pay off, profi t, promote, relieve, serve, succor, work for, acquire, derive, come by, receive, fi nd, collect, obtain, help, payment

problem: state, resolve, diffi culty, condition, affairs, effort, overcome, grapple, bear, bitch, predicament, quandry, plight, extricate, diffi cult, unplesant, trying, awkward, entangle, pinch, fi x, holem, jam, mess, muddle, pickle, situation, hard, rough, stress, strain, job, trouble, hindrance, wrinkle, interfere, question, matter, issue

Figure 9: Alternations for the question keywords *benefit* and *problem*

Summary answers were generated by merging the top-ranked sentences selected from each subquestion into a single paragraph. Two simple optimizations were then performed to improve the overall quality and legibility of summaries. In order to reduce redundancy, we used a semantic parser to create predicate-argument structures for each sentence included in the summary; lower-ranked sentences that featured the same predicate-argument structure as higher-ranked sentences were dropped from the summary. In addition, we attempted to resolve pronouns in summary sentences by including the immediately preceding sentence from the sentence's original document; if this immediately preceding sentence also contained a pronoun, both sentences were dropped from the summary. This process was repeated until the summary reached a total of 250 words.

3.2. Method 2

Top-down question decompositions do not have any sentences associated with them as answer. Bottom-up question decompositions have such sentence-answers known, due to the process in which they were created. However, it is not clear whether the answers selected from the bottom-up question decomposition are the most informative answers for the complex question. This observation is important, since, as detailed in Section 4, less than 25% of the questions produced by the two decomposition methods are identical or overlap. Therefore, we chose to use all question decompositions and to find their answers with a Q/A system that was tuned such that it ranks on the first place the known answers of the questions devied by the bottom-up process. To achieve this goal, we tuned the parameters of the density score for the answer extraction detailed in (Harabagiu et al., 2005a). The rank of each sentence is given by the $Density\_Value(sentence) = Density(keywords) \times Match\_Quality(keywords) \times Match_Proximity(keywords)$. $Density(keywords)$ is the ratio $\frac{k_S}{k_Q}$, where k_S counts the number of keywords and alternations extracted from any question decomposition that are matched in the sentence, whereas k_Q is the cardinal of the set of keywords. The match-quality score uses three values for each keyword matching: (a) 1.0 for perfect match, (b) 0.8 for morphological variation, and (c) 0.6 for semantic synonyms. The match-quality score adds these corresponding values for every keyword matching. Finally, the keyword-proximity score takes pairs of keywords and it favors keywords matched in the same clause (value = 1.0) over keywords matched in the same sentence (value = 0.7), or keywords matched in the same paragraph (value = 0.5). Sentences ranked with this method were sent to the summary generator that proceeded similarly to Method 1.

3.3. Method 3

Textual Entailment evaluated in the PASCAL RTE challenges was used as another method for selecting sentences for MDS. We have used the textual entailment system described in (Hickl et al., 2006) to decide whether a sentence can be entailed or not from a decomposed question. The TE system generates a YES/NO answer as well as a confidence score in the entailment decision. We have ranked each sentence from the document cluster that was entailed by at least one question decomposition by the entailment confidence. After this ranking was produced, sentences were sent to the summary generator that continued the processing similarly to Method 1.

4. Evaluation Results

Multi-Document Summarization can be evaluated by several scores: (1) the ROUGE metric (Lin, 2004), (2) the Pyramid score (Nenkova and Passonneau, 2004), and (3) the responsiveness score. We selected to evaluate the impact of question decompositions by using the responsiveness score. The responsiveness score, is given by a computational linguist who selects an integer value between 1 and 5 to assess his/her satisfaction with the information contained in the summary as an answer to the question. A score of 1 represents the least responsive summary and 5 is given to the most responsive summary.

Experiment	Summarization Method	Question Decomposition
E_1	Method 1	Syntactic
E_2	Method 1	Top-Down
E_3	Method 1	Bottom-Up
E_4	Method 1	Top-Down and Bottom-Up
E_5	Method 2	Syntactic
E_6	Method 2	Top-Down
E_7	Method 2	Bottom-Up
E_8	Method 2	Top-Down and Bottom-Up
E_9	Method 3	Syntactic
E_{10}	Method 3	Top-Down
E_{11}	Method 3	Bottom-Up
E_{12}	Method 3	Top-Down and Bottom-Up

Table 3: Description of Experiments.

To be able to evaluate the impact of question decomposition on multi-document summarization, we have performed

twelve different experiments, which are listed in Table 3. The results of the experiments for the 50 topics from DUC-2005 are illustrated in Figure 10. The best results were obtained for 14 topics.

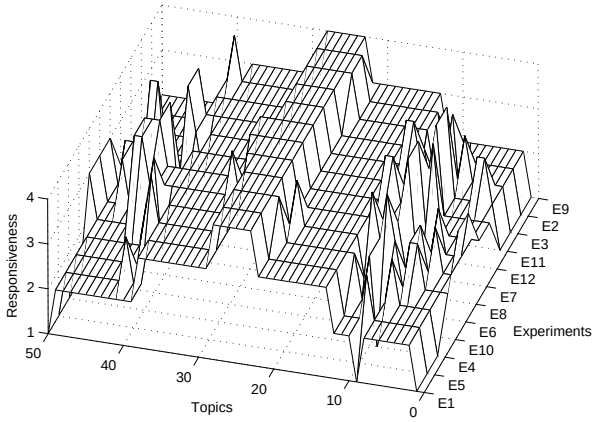

Figure 10: The responsiveness score for the 50 topics, using the twelve summarization methods.

When we wanted to measure the impact of question decomposition on multi-document summarization, we compared the results of the experiments listed in Table 3 against three baseline experiments in which no question decomposition is available. BE_1 is the baseline experiment in which Method 1 uses only the keywords extracted from the complex question. BE_2 is the baseline experiment in which the density value of sentences is measured by using only keywords extracted from the complex question. BE_3 is the baseline experiment in which textual entailment is performed between the sentences and the complex question. By computing the difference in responsiveness score between the results obtained in the experiments listed in Table 3, and the baseline experiments, we have found that the largest impact (25%) of question decomposition for MDS was obtained in experiment E_8 for the seven topics that included topics D385 – "*Electric Automobile Development*", and D401 – "*Foreign minorities in Germany*". The least impact (8%) was obtained in experiment E_1, which however produced competitive scores in the DUC-2005 evaluations.

5. Conclusions

In this paper we presented three methods for question decomposition that enable improved results for question-driven multi-document summarization. The first method decomposes questions based on syntactic information, whereas the other two use semantic and coherence information for question decomposition. One of the semantic question decomposition methods operated in a top-down manner, whereas the other operates in a bottom-up manner. In experimental results, we have found that by combining the two semantic-based question decomposition methods we achieved the highest responsiveness scores, which were improved by 25% from results that are produced by baseline methods that have no access to question decompositions.

6. Acknowledgments

This material is based upon work funded in whole or in part by the U.S. Government and any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the U.S. Government.

7. References

Regina Barzilay and Lillian Lee. 2003. Learning to paraphrase: An unsupervised approach using multiple-sequence alignment. In *HLT-NAACL*.

T. Clifton and W. Teahan. 2004. Bangor at TREC 2004: Question Answering Track. In *Proceedings of the Thirteenth Text Retrieval Conference*.

S. Harabagiu, D. Moldovan, M. Pasca, M. Surdeanu, R. Mihalcea, R. Girju, V. Rus, F. Lacatusu, P. Morarescu, and R. Bunescu. 2001. Answering Complex, List and Context Questions with LCC's Question-Answering Server. In *Proceedings of the Tenth Text REtrieval Conference*.

S. Harabagiu, D. Moldovan, C. Clark, M. Bowden, A. Hickl, and P. Wang. 2005a. Employing Two Question Answering Systems in TREC 2005. In *Proceedings of the Fourteenth Text REtrieval Conference*.

Sanda Harabagiu, Andrew Hickl, John Lehmann, and Dan Moldovan. 2005b. Experiments with Interactive Question-Answering. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*.

Sanda Harabagiu. 2004. Incremental Topic Representations. In *Proceedings of the 20th COLING Conference*, Geneva, Switzerland.

Andrew Hickl, John Williams, Jeremy Bensley, Kirk Roberts, Bryan Rink, and Ying Shi. 2006. Recognizing Textual Entailment with LCC's Groundhog System. In *Proceedings of the Second PASCAL Challenges Workshop (to appear)*.

F. Lacatusu, A. Hickl, P. Aarseth, and L. Taylor. 2005. Lite-GISTexter at DUC 2005. In *Proceedings of the Document Understanding Workshop (DUC-2005) Presented at the HLT/EMNLP Annual Meeting*.

Chin-Yew Lin and Eduard Hovy. 2000. The automated acquisition of topic signatures for text summarization. In *Proceedings of the 18th COLING Conference*, Saarbrücken, Germany.

Chin-Yew Lin. 2004. ROUGE: a Package for Automatic Evaluation of Summaries. In *Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004)*, Barcelona, Spain.

Un Yong Nahm and Raymond J. Mooney. 2000. A Mutually Beneficial Integration of Data Mining and Information Extraction. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence*.

Ani Nenkova and Rebecca Passonneau. 2004. Evaluating Content Selection in Summarization: the Pyramid Method. In *HLT-NAACL 2004*, Boston, MA.

Marius Pasca and Sanda Harabagiu. 2001. High Performance Question/Answering. In *Proceedings of the 24th Annual International ACM SIGIR Conference*.

R. Quinlan. 1998. C5.0: An Informal Tutorial. RuleQuest.

Y. Shinyama, S. Sekine, K. Sudo, and R. Grishman. 2002. Automatic Paraphrase Acquisition from News Articles. In *Proceedings of Human Language Technology Conference*, San Diego, CA.