Using Grice Maxims In Ranking Community Question Answers

Abed Alhakim Freihat
University of Trento
Trento, Italy
Email: abed.freihat@unitn.it

Mohammed R H Qwaider
FBK, Trento
Trento, Italy
Email: qwaider@fbk.eu

Fausto Giunchiglia
University of Trento
Trento, Italy
Email: fausto@unitn.it

***Abstract*—Community question answering portals and forum Web sites are becoming prominent resources of knowledge and experience exchange and such platforms are becoming invaluable information mines. Getting to this information in such knowledge mines is not trivial and fraught with difficulties and challenges. One of these difficulties is to discover the relevant answers and/or to predict the best answer(s) among these. In this paper, we present a Grice cooperative maxims based approach for ranking community question answers.**

***Keywords*–Community Question Answering; Grice Maxims; Ranking Algorithms; Cooperative Principle.**

I. INTRODUCTION

Community question answering Web sites are growing rapidly. Web portals, such as Google Answers [1], Yahoo Answers [2], and other community forums are becoming rich resources for knowledge and experience exchange [3]. Despite the richness of these resources, benefiting from them - that depends on the ability of discovering relevant answers and/or ranking them - is still limited [4]. In the last decades, dozens of approaches to solve the ranking problem have been proposed. These approaches usually depend on feature extraction by using machine learning techniques [5].

In this paper, we address the problem of answer ranking in a different way. Our hypothesis is that linguistics offer us a good opportunity to predict relevancy of answers and rank them accordingly. In particular, we think that Grice maxims [6] give us a way to score answers and thus rank them accordingly.

Originally, Grice maxims were presented mainly from a pragmatic point of view as a way to explain how a listener perceives the utterances of a speaker so that he can understand the intention of the speaker of a sentence which seems extensionally unrelated to the conversation. For example, using these maxims explains that the speaker B understands the intention of the speaker A. The same holds for A who understands the indirect answer. An important point here is that the extensional logic relation between A and B is missing or implicit.

A *What is the time?*

B *The bus left five minutes ago.*

In this work, we use Grice maxims from engineering point of view, where we interpret and use them as a way for measuring the extensional relevancy of what a speaker says. For example, we are interested in: Does answer$_i$ contain more information than answer$_j$ and interpret it as answer$_i$ is better than answer$_j$ if it contains more information and vice versa. Thus, our approach does not consider the pragmatic (intentional) interpretation of Grice maxims as in the previous example. Instead, we are focusing on the extensional relation(s) between a question and an answer, and the relation(s) between answer$_i$ and answer$_j$.

The paper is organized as follows. Section II describes community question answering portals and illustrates the problem statement. Section III gives an overview of related works. In Section IV, Grice maxims are presented. In this section, we show how they can be interpreted and used as criteria for scoring answers in community question answering portals. In Section V, the implementation of our approach is described, where we depict the used resources, some of the approach experiments, and the proposed scoring algorithm for answers ranking in community question answering portals. The paper is concluded with future work discussion in Section VI.

II. PROBLEM STATEMENT

In the last two decades, several types of question answering Web sites have emerged. These sites offer usually the possibility to post a question and get several possible answers to the posted question. In general, the community question answering Web pages can be classified into two main categories [7].

- **Closed professional Web pages**. Such Web pages are usually specialized in one or more related domains. Answering the questions in these Web sites is restricted to trusted experts who work in these domains. The answers in such Web sites are written in well written and standard language. For example, medical consulting pages belong to this category.
- **Open non professional Web pages**. The questions in such pages usually belong to different domains and answering the questions is not restricted to specialized persons or experts. In contrary to the former type, the answers in such pages may contain malformed or not well written answers and may contain noisy punctuations, such as *:))*, *!!??,:((*, or non standard abbreviations such as *plz*, *thnx*, *u r*,...
  Community forums belong to this type of Web sites which are more likely to the social media platforms such as Facebook and Twitter in that they do not put any constraints on used language, punctuations, morphology, or orthography rules.

In this work, we focus in our research on open community forms. In particular, we are going to test our approach on Qatar Living forum [8], which is an open domain community forum. This forum is used mostly by expats who live and work in Qatar.

 ISBN: 978-1-61208-620-0

Beside entertainment, this forum is a platform for knowledge and experience exchange about issues related to living and working in Qatar. This makes the forum besides its social side, a rich shared knowledge resource. Users who need advice or information about some issue related to living and working in Qatar, post their questions and they usually get several answers and comments from other registered users. The forum language is English though most of the forum members are non native speakers of English (which makes the task more challenging).

To summarize, given a question Q and a set of answers $\langle a_1, ..., a_n \rangle$, rank these answers according to their relevancy with respect to the question Q. The data sets, which we used for developing, and testing are taken from the SemEval 2016 Task 3 [9] competition.

III. Related Work

In this section, we give a brief review of community question answering approaches and a short overview of a similar approach that uses Grice maxims in computational approaches.

A. Community Question Answering Approaches

Question answers ranking is a task of great interest in both research and commerce. In the last couple of years, there were different shared tasks, in a wide contest in three SemEval editions [10], [9], and [11], also a more specific context of ranking a set of frequently asked questions for a given question [12].

In general, machine learning based approaches utilize a variety of features and techniques for solving the ranking problem, e.g., similarity features [12] such as cosine similarity applied to lexical, syntactic and semantic representations or distributed representations. In addition, machine learning approaches employ trigger words such as insulting, or degrading words, meta features such as user ID, or answer position in the list.

Other class of features is the class of automatically generated features, where these features are generated from syntactic structures using tree kernels [13]. The main classifiers used in these approaches are SVM (Support Vector Machine) classifiers [14] and Convolutional Neural Network [15].

The success of machine learning based ranking approaches depends on the learning platforms and the used set of features. For example, KeLP (Kernel-based Learning Platform) [16] had the best results on SemEval 2016 task 3, where they use KeLP machine learning platform [17] which learns the similarity of semantic representation between two given texts with the help of previously proposed features [18]. The second best results on SemEval 2016 task 3 belongs to ConvKN [18] which utilizes deep-learning techniques, by combining convolutional tree kernels and convolutional neural networks, together with text similarity and thread-specific features. The third best system is SemanticZ [19] that uses semantic similarity based on word embeddings and topics.

B. Grice Maxims Based Computational Approaches

Grice maxims have attracted many researchers who enriched the research community with variety of research proposals and articles about Grice theory. Most of these approaches are in the linguistics and pragmatics domains. In the following, we highlight few Grice maxims based computational approaches. We do not review Grice maxims theoretical approaches in the linguistics and pragmatics since they are worthy of dedicating one or more papers to review them.

In the current state of the art, we find interesting computational approaches that utilize Grice maxims for solving linguistic and other real world problems . For example, Vogel et al. [20] presented their approach that uses Grice maxims in multi-agent decision theory, where they suggest cognitively-inspired heuristics to reason about cooperative language resulting from Grice communication principle.

Another idea discussed a general game theoretic model of quantity implicature calculation [21], and proposed a procedure to construct interpretation games as models of the context of utterance from a set of alternative sentences, and a step-by-step reasoning process that selects the pragmatically feasible play in these games.

In another approach [22], Dale et al. used Grice maxims in generating referring expressions in natural language generating task. In another study, Kheirabadi et al. [23] consider news as a mutual conversational activity between the media and its audiences. Based on this observation, they introduce Grice pragmatic maxims as a set of linguistic criteria for news selectivity.

IV. Using Grice Maxims For Community Question Answers Ranking

Grice main idea is that communication between human beings is logic and rational. Following this idea, any conversation assumes cooperation between the conversation parties. This cooperation supposes in essence four maxims that usually hold in dialogues or conversations [6]. These maxims are:

1) **Quality**: Say only true things.
2) **Quantity**: Be informative as much as necessary.
3) **Relation**: Be relevant in your conversation.
4) **Manner**: Be direct and straightforward.

These maxims have been intensively researched in the domain of linguistics and pragmatics in the last decades, where the researchers focused on how to use Grice theory to explain speaker intention when he says some thing. In this work, we use these maxims partially to measure the appropriateness or relevancy of answer(s) of a given question. In this approach, we do not try to understand what the speaker (intentional) means. Instead, we try to understand if the speaker contribution contains (extensional) elements that comply with Grice maxims.

In the following, we explain how we interpret the quantity, relation and manner maxims in our approach. We do not use the quality maxim and it is beyond the scope of our research.

A. Quantity Maxim

Grice summarizes this maxim as ”Speaker contribution is expected to be genuine and not spurious” and he gives criteria that indicate not violating the maxim.

1) Make your conversation as informative as required.
2) Avoid redundancy.

In our work, we use the first criterion in this maxim only. This means that we reward answers if they are informative and we do not penalize answers if they are redundant. In fact, we

do not have the mean to judge redundancy. We consider an answer as informative answer as follows.
Many of the Questions are usually inquiries about places, organizations, persons, or things. For example, the question *Is there any place where I can find scented massage oils in Qatar?* is asking about a place, the question *Does anyone have recommendations for which bank to use in Qatar?* is asking about an organization, the question *Can anybody give me details and information about where to find a very good dermatologist?* is asking about a person, and the question *What's the cheapest brand new car in Qatar?* is asking about thing.

Accordingly, answers for such questions are expected to contain information about the inquired entities or other entities that are in essence helpful for the user inquiry. For example, relevant answers for the question about the dermatologist may be names of hospitals rather than persons, such as the following answer *Try Apollo Clinic*.

In our approach, we interpret this maxim as How much an answer is informative as follows. Does the answer contain the following informative elements?

1) **Named entities**: A named entity here refers to person, organization, location, or product.
2) **References**: References here include Web urls, emails, and phone numbers.
3) **Currency**: We consider the presence of currency in an answer as informative element.
4) **Numbers**: In some cases, phone numbers, or currency are not recognized because they are implicit such as *20000 is a good salary*. For this reason, we consider the presence of numbers (2 digits or more) in answers as an informative element.

Of course, this list of informative elements is not exhaustive. However, these are the elements that we utilize in our approach.

B. Relation

Grice summarizes this maxim as "Speaker contribution is expected to be appropriate to immediate needs at each state of the transaction" and gives a very generic criterion to judge relevancy which is : Be relevant.
According to Grice himself and Grice theory researchers, this maxim is not well defined [24]

One of the problems related to defining answer relevancy in Grice theory is that while it can explain how the sentence B in the following conversation can be understood as the direct answer (*No, there is no milk left*), it does not give us a definition of what is a relevant answer, nor a way to compare the relevancy of direct answers such the answer shown above. This is very important since answers in community question answering are usually direct answers and according to our study, answers like B are extremely rare in community question answering portals.

A *Is there another pint of milk?*

B *Im going to the supermarket in five minutes.*

We think that defining what is a relevant contribution in the relation maxim and/or defining conversation relevancy in general is still an open issue that needs to be researched. At the same time, we try in this work to discover relevancy indicators and use them in our ranking algorithm. Accordingly, we can consider the following as relevancy indicators.

1) **Similarity**: Similarity between the question and the answer or at least overlapping between the question and the answer utterances.
2) **Imperatives**: Answers that contain imperative verbs such as *try*, *go to*, or *check* indicate that the answerer is explaining a way to solve a problem being discussed.
3) **Expression of politeness**: Expressions of politeness *I would*, *I suggest*, or *I recommend* are usually polite alternatives for imperatives. For example, *I suggest you to to do* is a polite way of saying *Do*.
   Although this indicator overlaps with the manner maxim, we think that using such expressions indicates that answerer is serious in his answer and hence indicates relevancy.
4) **Factoid answer particles**: For factoid questions *is/are*, *does/do* the answer particles *yes/no* indicate the relevancy of the answer.
5) **Domain specific terms**: Domain specific terms indicate relevancy. For example, terms such as *CV*, *NOC (National Occupational Classification)*, *torrent*, etc. are domain specific terms. Using such terms indicates also that the answerer is trying to help or is explaining how to solve the problem being discussed.

Again, this list of relevancy indicators is not exhaustive and it would be much better for our approach if could use concrete criteria that indicates the relation maxim. Nonetheless, these indicators are helpful in indicating relevancy and using them is better than not using this maxim at all.

C. Manner

Grice summarizes this maxim as "a speaker contribution is expected to be clear" and he gives four criteria that indicate not violating this maxim:

1) **Avoid obscurity of expressions**
2) **Avoid ambiguity**
3) **Be brief**
4) **Be orderly**

We think that these maxims need more research to define them and give us the possibility to implement them in a computational approach. In fact, we need concrete criteria that we can use to determine whether a speaker contribution is obscure, ambiguous, brief, or orderly. For example, an expression which is ambiguous or obscure in some context may be unambiguous and clear in other contexts. The same holds for brief, since to our knowledge, there is no approach that can classify answers in concise and redundant answers.
The last criterion also needs more explanation. In summary, these criteria are too generic and need to have more specific definitions.

In this work, we tried to give some criteria that can be used to judge that a speaker contribution complies with/ violates the manner maxim. These criteria are:

1) **Be positive**: By this criterion, we mean that the speaker contribution is expected to be tolerant and permissive.

2) **Avoid frustrating utterances**: Answers that contain such expressions are usually not useful in the conversation.
3) **Avoid ironic and humbling expressions**: We mean here that the answer tends to be formal and professional and that the answerer is aiming to give a direct useful contribution.
4) **Avoid insulting and degrading expressions**: Answers that contain such expressions are not expected to be be useful in any conversation.

We may also consider the grammatical and orthographic correctness as a criterion. We did not consider this because many of the members of Qatar Living are not native speakers of English.

V. Implementation

In the following, we present the ranking algorithm , where we start with explaining the used resources. Then, we illustrate some experiments that we have conducted in the framework of our approach, and finally we describe using Grice maxims in community question answers ranking.

A. Resources

In the following, we describe the resources that we used in our algorithm for each of Grice maxims.

Quality: No resources and this maxim was not used in the implementation.

Quantity: We have used an openNLP name finder [25] for Named Entity Recognition (NER). After testing state of the art name finders, we found that their performance is low in terms of precision and recall. This is due to the fact that the forum members usually do not follow English orthography in writing named entities. Named entity capitalization in the answers, which an important shape features for NER, is absent in many of the named entities in the answers. On the other hand, most of the used named entities in the forum are Arabic names (especially persons and locations), which makes the problem for the state of the art NER systems more difficult. To handle these two problems, we have trained the openNLP NER system on an annotated corpus which was taken from the training data set. The generated model reached, 91% precision, 83% recall, and 87% F measure. The annotated corpus and the model are available online [26] and can be freely used for both research and commercial purposes.

Relation: For the relation maxim, we have used four resources. These resources are:

a *Similarity*: For similarity, we used Word2Vec [27] and Brown and clark [28] embeddings.

b *Imperatives and Expression of politeness*: We have used an OpenNLP POS-tagger to detect imperatives and expressions of politeness. We reward answers that contains such expressions.

c *Domain specific terms*: Using the training data, a small dictionary that contains domain specific terms such as *router, CV, NOC, torrent...etc,* has been compiled. The terms in the dictionary are not classified and of course they are not exhaustive. Answers that contain such expressions are also rewarded.

Manner: We used here two resources for sentiment polarity lists [29], one positive sentiment word list and another negative sentiment words list.

a *Be positive*: For this criterion, we have used the positive sentiment list, which we use to reward answers that contain positive expressions.

b *Avoid frustrating expressions*: For this criterion, we used the negative sentiment list to penalize answers that contain frustrating expressions.

b *Avoid ironic and humbling expressions*: The negative sentiment list includes some of the ironic and humbling expressions. We have used the training data to extend the list with new ironic and humbling expressions that we found in the training data. Answers that contain such expressions are penalized.

c *Avoid insulting and degrading expressions*: The negative sentiment list includes some of the insulting and degrading expressions. We have extended the list with new expressions that we found in the training data. We penalize answers that contain such expressions.

B. Experiments

In the following, we describe some of the experiments that we conducted to compare their results with the results of our proposed algorithm which is described in the next section. We used the test data set taken from Semeval 2016 to evaluate the results of these experiments, where we used Mean Average Precision (MAP) as performance measure.

Experiment 1 (similarity run):

- **Method**: Rank the answers of a question using term frequencyinverse document frequency (Tf-IDF) [30] as a similarity function from the most similar answer to less relevant one.
- **Result**: The achieved result in this experiment was MAP=0.5839.

Experiment 2 (clusters / word representation 1):

- **Method**: We experimented mixing different combinations of word embeddings and similarity measure to rank the answers. We used Brown embedding with N-grams level, with a weight of 0.5 to embedding similarity and 0.5 to string similarity.
- **Result**: We got MAP=0.6089.

Experiment 3 (clusters / word representation 2):

- **Method**: Using Brown and Clark with weight of 0.3 to string similarity and 0.7 to cluster similarity.
- **Result**: we got MAP=0.5596.

Experiment 4 (clusters / word representation 3):

- **Method**: Including word2vec to Brown and clark, with a low-level features, like word shape with the same weight of 0.3 to string similarity and 0.7 to cluster similarity.
- **Result**: we got MAP=0.6422.

Experiment 5 (similarity rule based): In this experiment, we run the system in two phases:

1) Rank the comments depending on their token-based similarity score.

2) Re-rank it on background rules.
Having the first ranking clustered in three separated areas (good, potential useful, bad). Then we apply the following for each cluster. The answers of the same person were considered as duplicates.
Thus, we give priority to answers coming from different users. That means, we downgrade the answers of the same user if they are more than one answer.

In this experiment, the results were comparable with the previous experiments, where we got MAP=0.6403.

C. Grice Maxims Based Ranking Algorithm

In the following, we present our algorithm that uses Grice maxims to rank community question answers. The used abbreviations are explained as follows.

- SM: Similarity between question and answer.
- NE: Named entities.
- RE: Reference expressions.
- CN: Currency and numbers.
- IM: Imperative and polite expressions.
- DT: Domain specific terms.
- PS: Positive sentiment words.
- NS: Negative sentiment words.
- IR: Ironic and humbling words.
- ID: Insulting and degrading words.

Input:
Q: $\langle p, qText\rangle$, where p refers to the person who is asking, and $qText$ to the question text.
l: $\langle a_1, ..., a_n\rangle$, where $a_i = \langle p_i, aText_i, score_i\rangle$.
The variable p_i refers to the person who answered a_i, $qText_i$ to
the answer text, and $score_i$ to a number that represents the relevancy of a_i.
Output:
l: where l is the input list after sorting according to Grice Maxims.
algorithm GriceMaximxBasedRanking(q: $\langle p, qText\rangle$, l: $\langle ..., a_i = \langle p_i, aText_i, score_i\rangle, ...\rangle$)
begin
foreach answer a_i **in** l:
if $p_i = p$ **then** $score_i = i * -100$
else
$score_i = |SM_{qi}| + |NE_i| + |RE_i| + |CN_i| + |IM_i| + |DT_i| + |PS_i|$;
$score_i -= |NS_i| + |IR_i| + |ID_i|$;
sort l ;
return l;
end

The algorithm works in four steps as follows.

1) The algorithm checks whether the answerer is the same person who asked the question. The answers made by person who asked the question are downgraded such that they become the last answers in the list. Such answers according to our analysis are usually thanking messages or explanations of some aspects of their original question.
2) For the rest of the answers, the algorithm computes the similarity between the question Q and the answer a_i, where $0 \leq SM_{qi} \leq n$ ($n = |l|$).
3) Then, based on Grice maxims, the answers are rewarded or penalized as follows.
   a The answer a_i is rewarded according to the number of entities, reference expressions, currency and numbers, imperatives, domain specific terms, and positive sentiment words.
   b On the other hand, a_i is penalized according to the number of negative sentiment, ironic, and insulting words.
4) After rewarding and penalizing all answers, we then sort the list of answers according to their achieved scores in descending order. Best answer is the first answer in the list and so on.

TABLE I. RESULTS OF SOME COMMUNITY QUESTION ANSWER RANKING APPROACHES IN SEMEVAL 2017.

System	MAP
Baseline	0.623
Best System	0.884
Our System	0.785
Worst System	0.633

The proposed approach participated at SemEval 2017 task 3, where our system [31] achieved a MAP=0.785 as shown in Table I.

VI. CONCLUSION

In this paper, we have presented a community question answers ranking approach based on Grice Maxims. In this approach, we gave extensional interpretation of Grice maxims rather than the intentional interpretation in pragmatics. We have demonstrated that Grice maxims indeed offer an effective method for solving challenging linguistic problems.

Although our approach did not reach the performance of machine learning based approaches, it gave a linguistic motivated solution which can be improved so that it reaches the performance of machine learning methods. We hope that the presented work will attract researchers to pay more attention to bridge the gaps in Grice maxims by defining solid criteria for these maxims. In particular, defining the criteria for what is informative, brief, redundant, obscure, ambiguous, or concise speaker contribution are very important for Grice based computational approaches such as the one presented in this paper.

We believe that more effort in this direction will offer us new powerful solutions that can achieve high quality results with significant performance.
In our planned future work, we plan to do more research on defining concrete criteria for the relation maxim. We think that defining the relation maxim can enhance the achieved results in the current work.

Another important concept that we plan to work on, is to explore the role of domain specific terms in community answers to classify questions and answers in domains. Our hypothesis is that domain specific classification of questions and answers improves the results of our current approach.

REFERENCES

[1] Google, "Google answers." https://answers.google.com/, 2018.

[2] Yahoo, "Yahoo answers." https://answers.yahoo.com/, 2018.

[3] M. S. Pera and Y.-K. Ng, "A community question-answering refinement system," in *Proceedings of the 22Nd ACM Conference on Hypertext and Hypermedia*, HT '11, (New York, NY, USA), pp. 251–260, ACM, 2011.

[4] X. J. Wang, X. Tu, D. Feng, and L. Zhang, "Ranking community answers by modeling question-answer relationships via analogical reasoning," in *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '09, (New York, NY, USA), pp. 179–186, ACM, 2009.

[5] M. Nguyen, V. Phan, T. Nguyen, and M. Nguyen, "Learning to rank questions for community question answering with ranking SVM," *CoRR*, vol. abs/1608.04185, 2016.

[6] H. P. Grice, "Logic and conversation," in *Syntax and Semantics: Vol. 3: Speech Acts* (P. Cole and J. L. Morgan, eds.), pp. 41–58, San Diego, CA: Academic Press, 1975.

[7] D. Mollá and J. L. Vicedo, "Question answering in restricted domains: An overview," *Comput. Linguist.*, vol. 33, pp. 41–61, Mar. 2007.

[8] Q. Living, "Qatar living: best place for cars - properties - buying - selling - renting and Qatar news and events." http://www.qatarliving.com/, 2018.

[9] P. Nakov, L. Màrquez, A. Moschitti, W. Magdy, H. Mubarak, A. A. Freihat, J. Glass, and B. Randeree, "Semeval-2016 task 3: Community question answering," in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, SemEval '16, (San Diego, California), pp. 525–545, Association for Computational Linguistics, June 2016.

[10] P. Nakov, T. Zesch, D. Cer, and D. Jurgens, eds., *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*. SemEval '15, Denver, Colorado: Association for Computational Linguistics, June 2015.

[11] P. Nakov, D. Hoogeveen, L. Màrquez, A. Moschitti, H. Mubarak, T. Baldwin, and K. Verspoor, "SemEval-2017 task 3: Community question answering," in *Proceedings of the 11th International Workshop on Semantic Evaluation*, SemEval '17, (Vancouver, Canada), Association for Computational Linguistics, August 2017.

[12] E. R. Fonseca, S. Magnolini, A. Feltracco, M. R. H. Qwaider, and B. Magnini, "Tweaking word embeddings for FAQ ranking," in *Proceedings of Third Italian Conference on Computational Linguistics (CLiC-it 2016) & Fifth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2016), Napoli, Italy, December 5-7, 2016.*, 2016.

[13] A. Moschitti, "Making tree kernels practical for natural language learning," in *11th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 113–120, 2006.

[14] M. Dinarelli, A. Moschitti, and G. Riccardi, "Re-ranking models based-on small training data for spoken language understanding," in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3 - Volume 3*, EMNLP '09, (Stroudsburg, PA, USA), pp. 1076–1085, Association for Computational Linguistics, 2009.

[15] K. Tymoshenko, D. Bonadiman, and A. Moschitti, "Convolutional neural networks vs. convolution kernels: Feature engineering for answer sentence reranking," in *NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016*, pp. 1268–1278.

[16] S. Filice, D. Croce, A. Moschitti, and R. Basili, "Kelp at semeval-2016 task 3: Learning semantic relations between questions and answers," in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, (San Diego, California), pp. 1116–1123, Association for Computational Linguistics, June 2016.

[17] M. Nicosia, S. Filice, A. Barrón-Cedeño, I. Saleh, H. Mubarak, W. Gao, P. Nakov, G. Da San Martino, A. Moschitti, K. Darwish, L. Màrquez, S. Joty, and W. Magdy, "Qcri: Answer selection for community question answering - experiments for arabic and english," in *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, (Denver, Colorado), pp. 203–209, Association for Computational Linguistics, June 2015.

[18] A. Barrón-Cedeño, G. Da San Martino, S. Joty, A. Moschitti, F. Al-Obaidli, S. Romeo, K. Tymoshenko, and A. Uva, "Convkn at semeval-2016 task 3: Answer and question selection for question answering on arabic and english fora," in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, (San Diego, California), pp. 896–903, Association for Computational Linguistics, June 2016.

[19] T. Mihaylov and P. Nakov, "Semanticz at semeval-2016 task 3: Ranking relevant answers in community question answering using semantic similarity based on fine-tuned word embeddings," in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, (San Diego, California), pp. 879–886, Association for Computational Linguistics, June 2016.

[20] A. Vogel, M. Bodoia, C. Potts, and D. Jurafsky, "Emergence of gricean maxims from multi-agent decision theory," in *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 9-14, 2013, Westin Peachtree Plaza Hotel, Atlanta, Georgia, USA*, pp. 1072–1081, 2013.

[21] M. Franke, "Quantity implicatures, exhaustive interpretation, and rational conversation," *Semantics and Pragmatics*, vol. 4, pp. 1–82, June 2011.

[22] R. Dale and E. Reiter, "Computational interpretations of the gricean maxims in the generation of referring expressions," *CoRR*, vol. cmp-lg/9504020, 1995.

[23] R. Kheirabadi and F. Aghagolzadeh, "Grice's cooperative maxims as linguistic criteria for news selectivity," *Theory and Practice in Language Studies*, vol. 2, no. 3, p. 547, 2012.

[24] R. Frederking, "Grice's maxims: do the right thing," *Proc. of AAAI SpringSymp. on Compl. Implicature: Computational Approaches to Interpreting and Generating Conversational Implicature*, 1996.

[25] Appache, "Apache opennlp." http://opennlp.apache.org/, 2018.

[26] A. A. Freihat, "Named entity recognizer for Qatar." https://www.researchgate.net/project/Named-Entity-Recognizer-For-Qatar, 2018.

[27] J. Turian, L.-A. Ratinov, and Y. Bengio, "Word representations: A simple and general method for semi-supervised learning," in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, (Uppsala, Sweden), pp. 384–394, Association for Computational Linguistics, July 2010.

[28] R. Agerri and G. Rigau, "Robust multilingual named entity recognition with shallow semi-supervised features," *Artificial Intelligence*, vol. 238, pp. 63 – 82, 2016.

[29] J. Breen, "twitter-sentiment-analysis-tutorial-201107." https://github.com/jeffreybreen/twitter-sentiment-analysis-tutorial-201107, 2017.

[30] A. Rajaraman and J. D. Ullman, *Mining of Massive Datasets*. New York, NY, USA: Cambridge University Press, 2011.

[31] M. R. H. Qwaider, A. A. Freihat, and F. Giunchiglia, "Trentoteam at semeval-2017 task 3: An application of grice maxims in ranking community question answers," in *Proceedings of the 11th International Workshop on Semantic Evaluation, SemEval@ACL 2017, Vancouver, Canada, August 3-4, 2017*, pp. 271–274, 2017.