Article

Yahoo! Answers as a Space for Informal Language Learning

Giuliana Dettori [1,*] **and Simone Torsani** [2]

[1] Institute for Educational Technology, Consiglio Nazionale delle Ricerche (CNR), Via De Marini 6, Genoa 16149, Italy

[2] Department of Modern Languages and Cultures, University of Genoa, Piazza Santa Sabina 2, Genoa 16124, Italy; E-Mail: simone.torsani@unige.it

* Author to whom correspondence should be addressed; E-Mail: dettori@itd.cnr.it; Tel.: +39-010-6475-683; Fax: +39-010-6475-300.

Abstract: Online social spaces, where users can exchange information, opinions and resources, have achieved wide popularity and are gaining attention in many research fields, including education. Their actual potential support to learning, however, still requires investigation, especially because portals can widely differ as concerns purpose and internal structure. This paper aims to contribute in this respect, by concentrating on question answering, a kind of social space not yet widely discussed in education. We analyzed a small corpus of posts from the Languages section of Yahoo! Answers Italy, checking if the questions reveal some inclination to learning or just the desire to obtain a service and if the answers provided by the community members can be considered as reliable sources of knowledge. Our analysis highlights the presence of a variety of question/answer types, from mere information exchange or help for task completion, up to language-related questions prompting valuable short lessons. The quality of answers may widely vary as concerns pertinence, correctness and richness of supporting elements. We found a high number of purely task-oriented questions and answers, but also a higher number of learning-oriented questions and correct, informative answers. This suggests that this kind of social space actually has valuable potential for informal learning.

Keywords: question answering; Yahoo! Answers; informal learning; language learning; online social space; peer learning; online communities

1. Introduction

Online social spaces have become increasingly popular in the past few years and are receiving the attention of researchers in the educational field as possible supports for formal learning or opportunities of informal learning [1–4]. Their learning potential, however, cannot be taken for granted just because learning has long been recognized as social in nature. Social spaces can widely differ as concerns aims, operation and internal structure, and all of these factors affect their usability and affordances. It is therefore necessary to investigate the potential of different types of social spaces in order to highlight their possible contribution to improve and innovate education.

In this paper, we take into consideration a kind of social space that has so far been scarcely considered in the educational field, that is, question answering (QA), with the aim to understand if informal learning opportunities are actually provided by such online environments.

Social QA services have been widely developing in the past decade, with the mission to be places where everybody can contribute what they know, because everybody, not only teachers and experts, has some knowledge to share [5]. Such services are also viewed as an expression of the collective intelligence of all of their users [6]. QA spaces, and in particular Yahoo! Answers (YA), which is currently one of the largest and most visited, are extremely popular, with hundreds of thousands of users and new questions every month [7]. They have become prominent places for online information seeking, especially since answered questions remain available in the website's database (for everybody, not only for community members) and can be retrieved also through search engines.

This exploratory study aims to shed light on the learning potential of QA spaces by building a descriptive picture of the kinds of information exchange (from the learning point of view) that actually take place in it and of the possible learning-oriented attitudes showed by its users. To this end, we selected and analyzed a small corpus of posts in the Languages section of the Italian chapter of YA. We choose to concentrate on one topic, because it is recognized [8] that there are wide differences among users' involvement and behavior in different content categories, and hence, the average outcomes of a transversal analysis would likely fail to faithfully mirror the real situation in any category. We chose the Languages section for our analysis, because languages are a study subject, but they are also used in everyday life to communicate in the current globalized world. This fact provides the opportunity for YA users to ask both academic and very practical questions, leaving aside the kind of vague, opinion-oriented questions (usually called "factoids" in the literature [9]) that are often diffused within other topics.

In the next section, we present a concise review of previous studies on QA spaces; then, we describe and discuss the organization and outcomes of our study.

2. QA Spaces in the Literature

A number of studies have been produced in the last decade, investigating QA systems from different points of view, analyzing a variety of aspects and characterizing them mainly as information seeking devices, social spaces and technological environments. Several interesting examples are reported below that can give an overall picture of the current research trends in this field.

Patterns of interaction are the focus of a study by Adamic, Zhang, Bakshy and Ackerman [5], who try to understand knowledge sharing activity within YA across its categories. Starting from the observation that some users focus only on specific categories, while others like to move across several ones, they map related categories, define an "entropy" of users' interests and combine these attributes to predict the choice of best answers.

Microcollaboration is investigated by Gazan [10], who shows how QA users sometimes engage in episodes of collaborative information seeking; this author spots social capital and affective factors as key elements apt at predicting the formation of such microcollaboration teams.

Knowledge sharing continuance is the focus of Jin, Zhou, Lee and Cheung [7], who propose a model to predict its occurrence, based on ex-post users' satisfaction and knowledge self-efficacy.

User satisfaction and effectiveness are also investigated by Shah [11], who concludes that YA is a very effective platform for QA, because posted questions receive a very fast answer, even though really satisfactory answers usually take longer to be posted, depending on the question's difficulty.

User motivations and expectations are investigated by Choi, Kitzi and Shah [12] based on gratification theory. Their findings highlight the importance of understanding the interrelationship among these aspects.

Sentiment analysis in questions and answers is carried out by Kucuktunc, Cambazoglu, Weber and Ferhatosmanoglu [13], who found that best answers differ from other answers, as concerns the sentiment they express, and predict the attitude that a question will provoke in answerers.

Best answer selection criteria are investigated by Kim, Oh and Oh [14], who detect seven value categories of various natures at the origin of users' choices.

Technical tools to automate some function within QA spaces include proposals like: identifying authoritative actors [15], learning to recognize reliable users and content [16], personalizing the interaction with the system based on user's interests [17] and a multi-channel recommender system for associating questions with potential answerers [18].

Two interesting reviews are also worth mentioning.

Shah, Oh and Oh [19], after reviewing the literature on online QA services, draw a research agenda for investigating information seeking behaviors in such settings; they identify three main areas of interest: users (including needs, tasks, expectations and motivation), information (including quality of questions and credibility of answers) and technology (including user interface, usability and business model), as well as some intermediate areas rising from the intersections of the main ones: collective knowledge, usage pattern/behavior and devices/policies related to user-generated content.

Gazan [20], after reviewing the current literature, identifies as major threads of QA research: question and answers classifications, quality assessment, user satisfaction, reward structures, motivation for participation, operationalization of trust and expertise. Directions for future research are also identified as extensions of these threads.

A point that emerges from this concise literature review is that the potential value of QA spaces in education is never addressed. Learning may be mentioned as a consequence of information seeking, but is never the focus of any of these studies. This is nevertheless an important issue to investigate, both to bring extra value to QA spaces and to foster innovation in the educational field by understanding, and exploiting, the learning potential of available resources. For this reason, we have

conceived of this exploratory study, with the aim to stimulate the development of research studies on the considered topic.

3. Methodology of Our Study

We extracted a corpus of 500 questions and related answers from the *Languages* section of YA Italy, posted and resolved within a few days in December 2013. We limited our choice to "closed" questions, that is, questions whose authors had already chosen the best answer, in order to work on a consolidated situation.

As pointed out in the previous section, current studies in this field concentrate on a variety of aspects apt at investigating social and information retrieval issues, without considering the learning potential. None of the classification approaches described in the literature, therefore, appeared suitable to support our analysis, and we needed to work out our own approach. Repeated explorations of YA *Languages*, carried out before the selection of the corpus to analyze, led us to spot a number of aspects as relevant to highlight users' learning orientation: types of questions, context/motivation and answers received as concerns the number, pertinence, correctness and richness from a language learning point of view. We defined each of them carefully (see the details in the next section), so as to avoid doubts and inconsistencies during post classification. Our choice to base this exploratory study on these aspects is due to four main reasons: (1) all of them contribute in some way to shed light on users' orientation to learning; (2) these pieces of information are usually found in the posts themselves and, hence, can be acquired straightforwardly without requiring inferences on our part; (3) these are the only learning-related clues that we have detected by analyzing the posts; and (4) they are sufficient to draw a meaningful picture of the language learning potential of the considered QA space.

Posts' analysis was carried out by the two researchers involved in this study, who independently read and classified 20% of the selected posts, in order to check their inter-coder agreement before processing the whole corpus. Thanks to the preparatory work on aspect definition, five of the six classification aspects of our choice (types of questions, as well as the number, pertinence, correctness and richness of answers) were simply acquired from the posts, which easily led to complete coders' agreement on those variables. Some difficulties, on the other hand, arose as concerns context/motivation. For this variable, some explicit references present in most of the posts allowed us to split them into four groups (school, leisure, study and work); in a few cases, however, contexts or reasons motivating the questions were not explicitly mentioned, nor was it easy to determine, based on the overall content of such posts, if their authors were asking for suggestions and explanations while engaged in a formal learning context (which we classified as study) or were pursuing a free personal interest (which we classified as leisure). For this reason, we decided to introduce a fifth group, "study or leisure", including questions clearly showing a learning interest, but whose motivation for asking was not explicitly mentioned. Introducing such a mixed group does not bias the correctness of our approach, because our focus is on users' orientation to/interest in learning, and this was always evident for the posts included in this group, independently of the reasons that motivated the questions. With the introduction of this fifth group, it was possible to achieve complete agreement between the two coders also on this aspect. Once inter-coder agreement was verified, the analysis of the remaining posts was split between the two researchers.

The data collection was carried out by means of content analysis [21,22], a research methodology often applied to investigate aspects of learning of both a cognitive and affective nature, especially in online environments, which extensively use written communication. Post analysis was performed manually, because orientation to learning is a "latent" variable, *i.e.*, it cannot be associated with the use of particular expressions or constructs, hence excluding the possibility to apply automated procedures.

We also included in our analysis a list and the frequency of the addressed languages, even though these data do not directly contribute to determine users' orientation to learning, in order to draw a more expressive picture of the corpus analyzed. This information, moreover, contributes to highlighting what a variety of language-related questions can actually receive a (serious) answer in the considered QA space, hence widening its boundaries as a potential learning place.

On the other hand, it was not possible to analyze contributors' characteristics, because they use rather anonymous nicknames (as suggested by Yahoo! itself to protect privacy), mostly without explicit contextual features about themselves. We can only mention, therefore, that questions appeared to be mostly asked by different people, while only about 6% of answerers were frequent (and often appreciated) contributors. Finding a high number of different authors should not be surprising (even though the YA web site lists numerous answerers with a very high number of posts) being in line with the findings of other studies: Jin and colleagues [7], for instance, acknowledge a high turnover in online communities, pointing out that many people post a contribution just once, even though they continue for a while to visit the community's space, reading other members' posts.

4. Exploring the Languages Section of Yahoo! Answers Italy

4.1. Languages Addressed

Fourteen modern languages were addressed in the posts analyzed, plus three, classical, ancient languages (Latin, Sanskrit, Ancient Greek), a few special cases (Italian dialects, sign language, Elvish), five different combinations of two or three modern languages, together with general linguistic questions. We labelled as spam a small percentage of off-topic questions and messages posted twice. These data are summarized in Table 1 and Figure 1.

Table 1. Languages addressed in the considered posts.

Language	%	Language	%
English	45.0	Sanskrit	0.4
German	14.6	Romanian	0.4
Italian	10.8	Portuguese	0.4
Spanish	5.6	Ancient Greek	0.4
French	3.8	Swedish	0.2
Japanese	3.4	Hindi	0.2
Russian	1.2	more than 1 language	1.2
Latin	1.0	particular cases	1.2
Chinese	0.8	general/linguistics	5.4
Albanian	0.8		
Arabic	0.6	spam/double	2.6

Figure 1. Occurrences of languages in the considered posts.
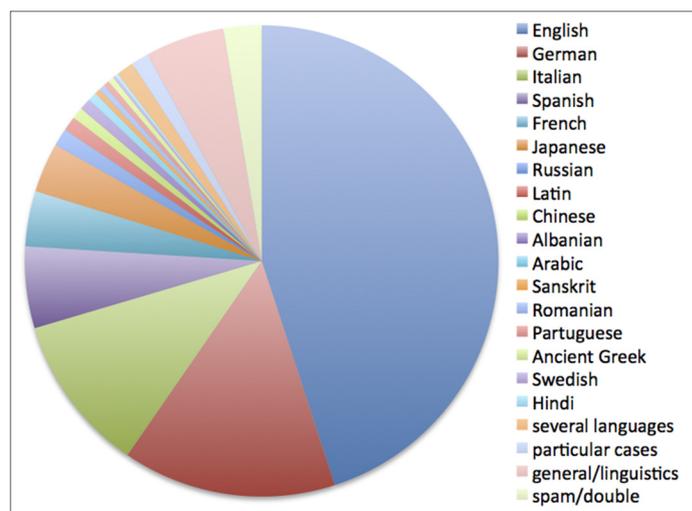
The most numerous group by far is represented by questions about the English language. This is likely due to the current wide diffusion of English, both as an object of study and a cultural means to communicate online or to understand cartoons, popular music and fiction.

The national language—Italian—is the third most frequently mentioned in the posts analyzed; most questions concerned meaning of words or grammar rules and were submitted by native speakers, while only about 13% were submitted by foreigners trying to get in touch with native Italian speakers.

4.2. Types of Questions

We classified the posts analyzed into three groups, which we labelled suggestions/opinions, information and tasks. Requests of the type "suggestions/opinions" concern activities and strategies for improving language learning aspects (e.g., memorization or listening understanding) or help to choose a language to study ("which one is more useful or more interesting?"), references of books and websites to practice, as well as schools or websites for studying a (scarcely diffused) language (e.g., Hindi) or to earn a certification (e.g., Cambridge English language certification) and websites to meet native-speaker conversation mates or to find au-pair accommodations.

We classified as "information" questions concerning the meaning of single words or short sentences, correct spelling or pronunciation, form and use of idiomatic expressions, different shades of meaning between similar words or expressions and grammar explanations; all requests in which the emphasis is more on acquiring knowledge and building meanings than on obtaining material help with translations or other tasks (even though a translation or task may also be involved in the request).

Finally, we included among "tasks" all other questions, which essentially consisted of translations to or from languages, execution or correction of grammar and lexicon exercises and even the composition of letters or short essays.

Tasks constituted the largest group, followed by slightly less frequent information questions and a much lower amount of suggestions/opinions, as shown in Figure 2.

Figure 2. Types of questions in the analyzed set.
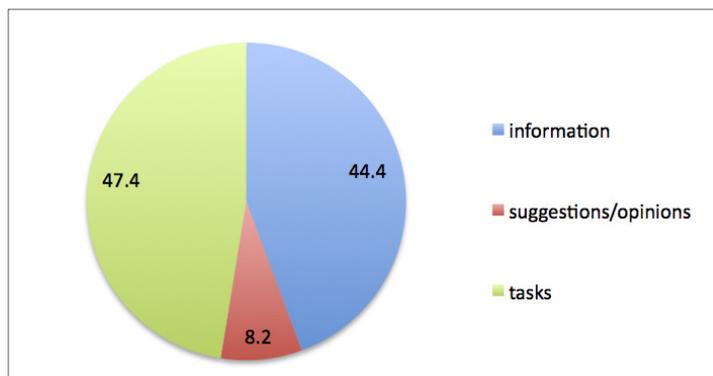
4.3. Contexts and Motivations of the Analyzed Questions

As already explained in the Methodology section, in relation to this aspect, we classified posts into five groups: school, leisure, study, study or leisure, work.

Questions in the school group explicitly refer to school-assigned tasks, such as sentences to be completed as grammar or lexicon exercises, translations and essays. A common characteristic of these questions is that their authors only ask for a product, not for explanations or suggestions.

The authors of leisure questions are often curious to know the meaning of a song, a cartoon, a sentence or an expression heard in a movie, or even of single words; in other cases, they are buying some item on the web and want to make sure to write the right message or interpret correctly an answer; or have a pen friend abroad and want to correctly describe (or interpret) a concept more complex than their usual conversational level. Sometimes, they also ask general questions on language use or structure. In all cases, these authors appear to be motivated by a personal interest not for the language addressed as an end in itself, but as a means to carry out some activity of their choice; in many cases, however, it is possible to identify an interest for acquiring some meaning (e.g., in song translations) or expressive form (e.g., for communicating with foreign friends).

Interest in learning, on the other hand, is evident in posts we grouped under the label study. Their authors' are sometimes clearly involved in a formal course and wish to improve their understanding of some topic; other times, they are following a personal learning interest. The explicit reference to learning of these questions made us decide to differentiate this group from school, because here, the focus is not on getting a product (e.g., an exercise or translation), but on receiving suggestions and explanations. The requested suggestions usually concern tools or methods for improving language learning; explanations mostly focus on complex grammar rules or the meaning of expressions that cannot be translated word by word. We included in the study group also requests of corrections (rather than completion) of essays and exercises, because their authors made the effort to work out the task themselves, and being helped to become aware of errors may certainly be a fruitful learning opportunity.

Finally, we collected in the group work all questions that were clearly asked by people who need to produce or understand a sentence in a language they do not know, in the context of a work activity. The most notable example we found is a person likely working in a travel agency who needed to

translate into German excursion advertising leaflets organized in their region and who submitted those texts a piece at a time, over several posts (hence, constituting the only notable case of a multiple asker).

The percentage distribution in these five groups is shown in Figure 3. In all groups, the requests may concern both language understanding and production.

Figure 3. Contexts/motivations of non-spam question in the corpus analyzed.
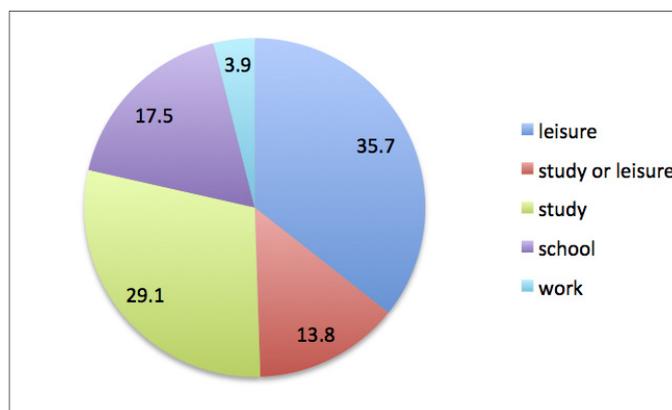

4.4. Answers: Number and Content

The questions analyzed received between one and 11 answers each, with a total of 1139 (average number of answers per question: 2.46; standard deviation: 1.68; median: 2; mode: 1).

We have been able to check the correctness of 90% of the questions (with a total of 1112 answers), excluding only translations and exercises for languages outside our competence fields; 95.1% of the questions we have been able to check received at least one correct answer. We grouped with incorrect answers also partially incorrect and incomplete ones, because, from the point of view of learning, they may anyway result in being misleading, even if not completely wrong.

We looked not only into the correctness, but also the richness of answers: do authors enrich their answers with links, references, citations, examples, alternative ways to express a concept or to translate an expression? How clear and concise are the explanations? Does the multiplicity of answers, when present, provide a multiplicity of points of view and help the asker to select the best answers? This led us to divide answers into four groups, corresponding to correct, correct and rich, incorrect or incomplete and spam answers, as shown in Figure 4.

We found that most of the answers could be considered satisfactory and reliable, with incorrect answers and spam limited to 21.2% in total. Moreover, 19% of the examined answers were not only correct, but also enriched by supporting elements. When links were provided, they aimed either to support key concept explanations (e.g., concept definition in Wikipedia) or to suggest extra resources (e.g., dictionaries, collections of exercises). Explanations, usually prompted by questions on grammar, lexicon and sentence constructions, were mostly correct, clear and concise; only in one case, a very formal explanation seemed copied from a (very formal) book; in two other cases, explanations were not completely correct.

Figure 4. Correct, rich, incorrect/incomplete and spam answers in the set analyzed.
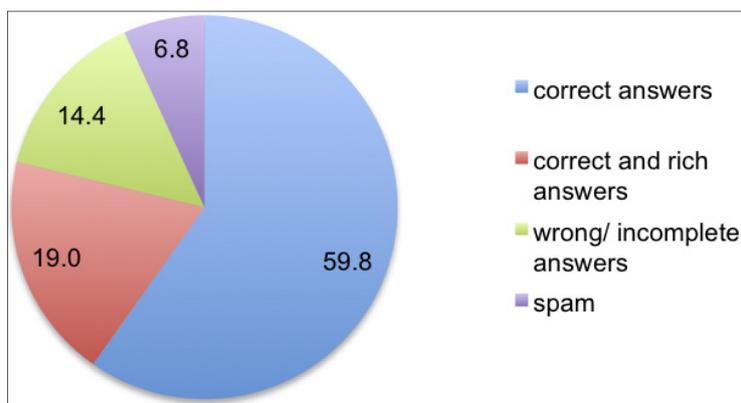
Not all explanations and references were explicitly requested in the questions: in about 30% of the cases of rich answers, the answers' richness went beyond the request. For instance, answering a question on books or links for individual study of Japanese, a user provided a reasoned list of references, explaining the focus of the various resources and suggesting how to combine them in a learning path. In another case, answering a request to correct a short English sentence, a user provided not only an explanation of the errors and one correct sentence, but also different possible ways to express the same concept. It was not rare, moreover, that people who answered a question on school homework not only did so, but also added some explanation on the rules to apply in the exercises.

Variants were not only highlighted within some posts, but also emerged from the whole set of answers given to a same question; in about 10% of the cases of multiple answers, answerers made explicit reference to other answers, sometimes even commenting on their being right or wrong or correcting them. The presence of several answers to the same question, however, was not very frequent in the corpus analyzed, as highlighted by the data presented at the beginning of this section.

Multiple answers did not seem to help questions' authors to pick the best answers, not even in the cases in which there was some discussion among answerers or correct answers were nicely commented on: in about 10% of these cases, authors selected as best answers an incorrect, or incomplete, or scarcely informative one. Unfortunately, YA does not require users to justify their votes, and most authors in our sample missed doing so; hence, it was not possible to determine the reasons for such non-optimal choices, nor to find out if the askers who had received multiple answers had compared them before choosing their preferred one. Neither did the limited number of multiple answers, and the brevity of most posts, allow us to try to apply any of the research results on the best answer selection mentioned in Section 2; such a fine-tuned analysis, anyway, would be beyond the aims of the current study.

5. Discussion

Let us now consider these data from the point of view of learning opportunities. This entails understanding if questions' authors seem to have some interest in learning and if answerers actually show competence, as well as the availability and ability to share it.

As concerns questions, two groups (school and work) do not show any learning orientation, but only the wish to obtain a product. While this is quite understandable as regards the work group, it is

much less so as concerns the school one: it is fine for students to ask for help outside class, but to have somebody else carry out their duties can hardly be seen as a way to learn. In this respect, a QA space may even appear to hinder learning rather than support it, giving lazy students an opportunity to avoid putting time and effort into learning tasks. Such use of QA communities, anyway, is not a surprise, but also acknowledged in the literature [23]. School group's size, however, is not very large, leaving much space to more fruitful orientations and uses of the portal. In fact, the joint percentages of the two non-learning-oriented groups amount only to 21.4%, while the two groups with clear learning orientation (study and study or leisure) reach twice as much with 42.9%. This last percentage is also higher than that of the leisure group (35.7%), whose intention to learn is not confirmed nor disconfirmed by elements of their posts; hence possible, and perhaps incidental, but not evident.

This good enough orientation to learning is confirmed by the data concerning question types (Figure 2), which show that task questions amount to less than half the total requests; it is important to remember, moreover, that tasks do not always correspond to a non-learning orientation, because they include correction requests, which we classified as study rather than school, as explained above. All of these figures suggest a strong (even though not very strong) awareness that a social space of this kind can represent a useful resource for learning. This is very important from the point of view of learners' maturity, especially in a life-long learning perspective, since awareness of, and ability to exploit, learning resources is a characterizing feature of self-regulated learners [24,25].

Answers' analysis shows an even higher learning potential. The low occurrence of spam suggests that the members of this online community take the task of answering their fellow members' questions rather seriously. We note that the amount of rich and correct answers is almost three times that of spam. This, together with the fact that the total amount of correct answers reaches almost 80% of the examined sample, suggests that this online community has a good level of content and resource knowledge and is willing to share it. Willingness is underlined by the presence of careful explanations beyond users' requests.

The high number of correct answers and limited spam we found is not in line with other studies in the literature that found among answers numerous low quality ones [26] and much noise [17]. This disagreement may be due to the limited size of our sample or to the fact that our study focused on a single topic, which, by its nature (as noted in the Introduction), lends itself to the formulation of practical questions, more than of generic ones.

The 14.4% wrong answers is certainly noticeable, especially since several users seemed not to recognize them as such, as indicated by the fact that in a few cases they voted one of them as the best answer, even when correct answers to their questions had been given by other answerers. A way to avoid such incongruous selections would be to let the best answers emerge from the votes of all readers rather than of just one person (the asker), as is the case in other social spaces (e.g., bookmark collections). Voting for incorrect answers, however, is not always dependent on the asker's poor judgment, but may also be due to content-unrelated factors, such as timing: one study [27] points out that the sooner an answer to a question appears, the higher its chances to be selected as the best answer by the question's author, while another study [11] highlights that the most satisfactory answers are usually not the first ones to be posted. Moreover, askers receive YA points (that they can then spend to ask other questions on YA) when they select the best answer to their questions, and this may likely

encourage a timely selection of a "best answer" even without real satisfaction. Finally, it must be remembered that recognizing reliable content is a non-trivial skill, as highlighted in the literature [16].

6. Conclusions

This exploratory study aimed to break new ground and stimulate the development of research studies on the learning potential of QA spaces. To this end, we built a descriptive picture of what kind of information exchange (from a learning perspective) is taking place in the *Languages* section of the Italian chapter of YA.

The data presented suggest that there is a real potential for language learning in this social space. This cannot be compared to the kind of learning that can take place by making use of courses, lectures, books, journals and the like, because questions and answers necessarily cover limited content. As properly remarked in the literature, "the knowledge shared in YA is very broad…but generally not very deep" [5]. Learning taking place here is necessarily fragmentary, related to limited and focused chunks of knowledge, but is not for this less valuable, especially in an informal learning perspective.

QA online communities have several benefits to offer: the possibility to receive answers tailored to one's questions [12], *i.e.,* to one's own needs and wishes, the possibility to ask questions on any language and on any aspect of language use and acquisition, because questions are answered not by a single person or small group, but by a large community with diverse competence and skills; and the possibility to easily reach native speakers of many languages. These possibilities appear to be actually exploited by YA users in the Languages section, considering how many languages and linguistic aspects were addressed in the relatively small sample analyzed.

The difficulty met by some users in discriminating reliable and unreliable content in the presence of multiple answers of diverse value stresses the (widely recognized) need to guide students to develop critical skills during formal education, in order to help them become able to suitably exploit potentially valuable opportunities of informal learning (like QA), which are brought about by the wide diffusion of online social spaces and resources.

Further studies will be necessary to explore more in depth and, in various directions, the learning potential detected. Possible developments could concern users' learning orientation in other sections of YA or deepen the reasons for non-optimal best answer choices. Keeping a focus on language learning, a relevant extension of the current study should investigate the nature of learning taking place in QA spaces, analyzing answers' linguistic content in detail and spotting what kinds of answer are given to a variety of specific linguistic questions. These learning-oriented developments add to the various research threads on the nature and functioning of QA spaces that have been highlighted in the literature review.

Author Contributions

Both authors jointly designed the study. The bibliographical research and data analysis was split in equal parts among them. The paper was jointly written in all its parts.

Abbreviations

QA: question answering;
YA: Yahoo! Answers

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Anne Bartlett-Bragg. "Reframing practice: Integrating social software to enable informal learning." *Training and Development in Australia* 35 (2008): 6–21.
2. Neil Selwyn, and Lyndsay Grant. "Researching the realities of social software use." *Learning, Media and Technology* 34 (2009): 79–86.
3. Catherine McLoghlin, and Mark J. W. Lee. "Social software and participatory learning: Pedagogical choices with technology affordances in the Web 2.0 era." Paper presented at the Ascilite 2007 conference, Singapore, 2–5 December 2007.
4. Giuliana Dettori, and Simone Torsani. "Enriching Formal Language Learning with an Informal Social Component." *Educational Technology & Society* 16 (2013): 93–103.
5. Lada A. Adamic, Jun Zhang, Eytan Bakshy, and Mark S. Ackerman. "Knowledge Sharing and Yahoo! Answers: Everyone knows something." Paper presented at ACM WWW Conference, Bejing, China, 21–25 April 2008.
6. Ze Li, Haiying Shen, and Joseph E. Grant. "Collective intelligence in the online social network of Yahoo! Answers and its implications." Paper presented at the ACM CIKM Conference, Mauii, HI, USA, 29 October–2 November 2012.
7. Xiao-Ling Jin, Zhongyun Zhou, Matthew K. O. Lee, and Christy M. K. Cheung. "Why Users Keep Answering Questions in Online Question Answering Communities: A Theoretical and Empirical Investigation." *International Journal of Information Management* 33 (2013): 93–104.
8. Amit Rechavi, and Sheizaf Rafaeli. "Knowledge and social Networks in Yahoo! Answers." Paper presented at the 45th Hawaii International Conference on System Sciences, Maui, HI, USA, 4–7 January 2012.
9. Yandong Liu, and Eugene Agichtein. "On the evolution of the Yahoo! Answers QA Community." Paper presented at the International Conference SIGIR'08, Singapore, 20–24 July 2008.
10. Rich Gazan. "Microcollaborations in a social Q&A community." *Information Processing and Management* 46 (2010): 693–702.
11. Chirag Shah. "Measuring effectiveness and user satisfaction in Yahoo! Answers." *First Monday*, 16 February 2011. Available online: http://www.firstmonday.dk/ojs/index.php (accessed on 16 October 2014).
12. Erik Choi, Vanessa Kitzie, and Chirag Shah. "Investigating motivations and expectations of asking a question in social Q&A." *First Monday*, 19 March 2014. Available online: http://www.firstmonday.dk/ojs/index.php (accessed on 16 October 2014).

13. Onur Kucuktunc, B. Barla Cambazoglu, Ingmar Weber, and Hakan Ferhatosmanoglu. "A large-scale sentiment analysis for Yahoo! Answers." Paper presented at the ACM WSDM'12 Conference, Seattle, WA, USA, 8–12 February 2012.

14. Soojung Kim, Jung Sun Oh, and Sanghee Oh. "Best-Answer selection criteria in a social Q&A site from user-oriented relevance perspective." *Proceedings of the American Society for Information Science and Technology* 44 (2007): 1–15.

15. Mohamed Bouguessa, Benoît Dumoulin, and Shengrui Wang. "Identifying authoritative actors in Question-Answering forums—The case of Yahoo! Answers." Paper presented at the ACM KDD'8 Conference, Las Vegas, NV, USA, 24–27 August 2008.

16. Jiang Bian, Yandong Liu, Ding Zhou, Eugene Agichtein, and Hongyan Zha. "Learning to Recognize Reliable Users and Content in social Media." Paper Presented at the ACM WWW Conference, Madrid, Spain, 20–24 April 2009.

17. Zoltán Gyöngyi, Georgia Koutrika, Jan Pedersen, and Hector García-Molina. "Questioning Yahoo! Answers." *Technical Report Stanford InfoLab*. Available online: http://ilpubs.stanford.edu: 8090/819/1/2007–35.pdf (accessed on 16 October 2014).

18. Gideon Dror, Yehuda Koren, Yoelle Maarek, and Idan Szpektor. "I want to answer, who has a question? Yahoo! Answer recommender system." Paper presented at the ACMKDD'11 Conference, San Diego, CA, USA, 21–24 August 2011.

19. Chirag Shah, Sanghee Oh, and Jung Sun Oh. "Research agenda for social Q&A." *Library and Information Science Research* 31 (2009): 205–09.

20. Rich Gazan. "Social Q&A." *Journal of the American Society for Information Science and Technology* 62 (2011): 2301–12.

21. Liam Rourke, Terry Anderson, Randy D. Garrison, and Walter Archer. "Methodological issues in the content analysis of computer conference transcripts." *International Journal of Artificial Intelligence in Education* 12 (2001): 8–22.

22. James W. Potter, and Deborah Levine-Donnersteirn. "Rethinking validity and reliability in content analysis." *Journal of Applied Communication Research* 27 (1999): 258–84.

23. Rich Gazan. "Seekers, sloths, and social reference: Homework questions submitted to a question-answering community." *New Review of Hypermedia and Multimedia* 13 (2007): 239–48.

24. Ivo Roll, Vincent Aleven, Bruce McLaren, and Kenneth Koedinger. "Designing for metacognition—Applying cognitive tutor principles to the tutoring of help seeking." *Metacognition and Learning* 2 (2007): 125–40.

25. Giuliana Dettori. "Demystifying Self-Regulated Learning." *Future Learning* 2 (2013): 55–59.

26. Yunjie Xu, Cheng Yian Tan, and Li Yang. "Who will you ask? An Empirical Study of Interpersonal Task Information seeking." *Journal of the American Society for Information Science and Technology* 57 (2006): 1666–77.

27. Amit Rechavi, and Sheizaf Rafaeli. "Not all is gold that glitters: Response time and satisfaction rates in Yahoo! Answers." Paper presented at the IEEE International Conference on Social Computing, Boston, MA, USA, 9–11 October 2011.