Received January 10, 2020, accepted March 3, 2020, date of publication March 6, 2020, date of current version March 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2978977

Impact of Users' Beliefs in Text-Based Linguistic Interaction

VINCENZO CATANIA[1], (Member, IEEE), SALVATORE MONTELEONE[2], (Member, IEEE), MAURIZIO PALESI[1], (Senior Member, IEEE), AND DAVIDE PATTI[1], (Member, IEEE)

[1]Department of Electrical, Electronic, and Computer Engineering, University of Catania, 95100 Catania, Italy
[2]Institute of Advanced Studies and ETIS (UMR 8051—CNRS), University of Cergy-Pontoise, 95000 Cergy, France

Corresponding author: Davide Patti (davide.patti@dieei.unict.it)

This work was supported in part by the Institute of Advanced Studies of the University of Cergy-Pontoise through the Paris Seine Initiative for Excellence ''Investissements d'Avenir'' under Grant ANR-16-IDEX-0008.

ABSTRACT Linguistic interaction between humans and machines is one of the most challenging fields in the development of next-generation User Interfaces. In this work, we investigate the role of beliefs about the interlocutor in human-computer linguistic interaction. First, we introduced an experimental setup that makes use of filtered and post-processed web content to generate a realistic, generic linguistic interaction. Then, we collected dialogues from two different sets α and β, corresponding to users being unaware or aware of the artificial nature of the interlocutor, respectively. The results thus obtained, analyzed using a standard t-test procedure ($N = 30$), demonstrate a statistically significant difference between the two sets in some of the linguistic features selected, i.e., sentence length and the number of adjectives, providing further insights to expand some of the evidence previously found in the literature.

INDEX TERMS Human–computer interaction, linguistic behavior.

I. INTRODUCTION

Computer designers, cognitive scientists, psychologists, philosophers, and sometimes also artists have often found human-computer linguistic interaction an attractive topic to investigate [1]–[4]. While several other communication channels have been replicated with discrete success in human-machine interfaces [5]–[7], natural language interfaces are still at their infancy, with prominent applications focused on specific conceptual domains and predefined interaction templates [8]–[10]. As a consequence, the current hype is focused on improving the technology behind such interfaces, rather than studying the possible implications of dealing with those interactions. Nevertheless, with growth in complexity and diffusion of natural language-based interfaces, new models and methodologies oriented towards a human-centric analysis will be required.

In this work, we focus on a specific aspect of these kinds of interactions, that is, the role played by the beliefs about the interlocutor in human-computer dialogue. Specifically, we investigate how the beliefs about the human/artificial nature of the interlocutor influence the linguistic behavior in text-based interactions. Linguistic alignment between human and computer interlocutors has been investigated in literature from different perspectives, for both written and spoken dialogues.

The seminal work in [11] performs a comparison of keyboard conversations involving a computer and human partners, where the same human operator plays both kinds of partners employing a prefixed set of answering rules as defined in [12]. Results show how the utterances usage is affected by both the initial model assumed for the other partner and the partner's subsequent responses. Authors of [13] investigate the impact of beliefs in lexical alignment bringing evidence from human-computer dialogues. In particular, they introduce the notion of mediated alignment to denote a linguistic behavior that is affected by beliefs about the counterpart, in contrast to the unmediated alignment, which is a reaction to the interlocutor behavior (e.g., previous responses). Their study, although mainly finalized on gathering insights about human-human communications, seems to show that humans feel the need to adapt more towards the interlocutor if it is believed to be a computer. Authors of [14] analyze the role of preconceptions in talking to artificial entities such as computers and robots, identifying two prototypical preconceptions about the artificial communication

The associate editor coordinating the review of this manuscript and approving it for publication was Orazio Gambino.

TABLE 1. Some details of Consciousnet and other famous chatbots.

Detail name	Consciousnet	Eliza [17]	A.L.I.C.E.	CleverBot	PC Therapist III	A Neural Conversational Model [29]
Interaction Modality	Text	Text	Text	Text (Speech-To-Text web client)	Text	Text
AI features	Web mining	Simple keyword search	Keyword based. Heuristical pattern matching rules to the human's input	Big Data analysis, Machine Learning	keyword search	Sequence to sequence learning with Neural Networks
Application Domain	No specific domain	Psychologist emulation	Research and training	No specific domain	Entertainment	Research. Use cases: Helpdesk, Entertainment
Technologies	Google API, Perl, Regural Expressions	Script in several languages	Web client. XML Schema called AIML (Artificial Intelligence Markup Language) for specifying the heuristic conversation rules. AIML Interpreters in different languages (JAVA, PHP, LISP)	Web client, interfaces in many languages (eg., cleverbot.io)	Seven modules written in C, Turbo Prolog and QuickBasic	Long Short Term Memory Recurrent Neural Network (LSTM RNN)
Accessibility	Yes, Google API required	Yes	Yes	No. Only interfaces open (eg., cleverbot.io)	No	No
Source availability	Yes	Yes	Yes	—	—	—
Knowledge Base	Metaresponses, Web content	Database of answer patterns	Vocabulary, AIML	Past online conversations	Short-Term memory + Previous conversations	Conversations with an IT Helpdesk Troubleshooting and subtitles from OpenSubtitles
Estimated Knowledge Base size	Unlimited	KBs	About 150.000 words	More than 200 millions of conversations	Input content + Previous conversations + Half-MegaByte knowledgebase called KBASEK	IT Helpdesk Troubleshooting: 30M + 3M tokens for training and validation, respectively. Subtitles: 923M + 395M tokens for training and validation, respectively
Experience Support	Storage of Metaresponses	None	None	String metrics, contextual, and fuzzy responses	Analysis of past conversations	LSTM RNN
Language	Every present in the Web	Same as database	English	39 languages	Same as database and conversations	Same as database

partner, where it can be classified as a tool and as a social actor. In [15] authors analyze how expectations and beliefs about a system affect users' word choice. Specifically, they demonstrate that human adaptation in dialogue is dependent on the complexity level she/he assumes for the computer involved in the interaction. The study conducted in [16] further investigates the role of beliefs in lexical behavior, using pieces of evidence from a set of experiments in order to capture differences between humans and computers when choosing from a particular set of words. The results indicate that a modification in the lexical behavior of participants can be found when they believe to be interacting with a computer counterpart, aligning in linguistic choices more strongly when dealing with apparently less advanced computers. Assuming an even increasing generality and multi-domain applicability of the next-generation human-machine natural language interfaces, the contribution we intend to provide in this work is to investigate the role of beliefs about the interlocutor without assuming any specific conceptual domain a priori and not restricting the linguistic space to a reduced subset of word choices. First, we design and propose an experimental setup based on an artificial counterpart that uses filtered and post-processed web content to generate a realistic, generic linguistic interaction. Then, dialogues generated are collected into two different sets α and β, corresponding to users being unaware or aware of the artificial nature of the interlocutor before performing the interaction, respectively. Finally, the results thus obtained are analyzed using a standard t-test procedure ($N = 30$) to investigate a statistically significant difference between the two sets for a set of predefined metrics.

The remainder of this work is organized as follows: in Section III-A we describe the experimental setup consisting of the interaction platform and testing scenario specifically designed. Next, in Section IV, we show the methodology adopted to collect and process data, along with statistical analysis, to summarize the results. Finally, conclusions about results are drawn in Section V.

II. RELATED WORK

From a technical perspective, Consciousnet can be classified as a chatbot, i.e., a software simulating an artificial entity capable of textually interacting with human counterparts. This field, originated by the seminal work of Weizebaum [17], has been fueled by many contributions over the years [18]–[20], that differ from each other by two main aspects: (i) the complexity of model adopted (ii) the knowledge base used to generate content. Table 1 shows, at best of our knowledge, the most influential chatbots ever developed. For the reasons previously explained, only text-based interaction platforms have been taken into account. As shown, some essential features differentiate the proposed platform from other projects in Table 1, in which explicit knowledge bases are adopted. In particular, the application domain is not tied to any specific field a priori, being driven by the dynamics of the interaction, as it will be explained in Section III-A. Also, a potentially unlimited knowledge-base can be estimated since no explicit database is maintained and used as a source. Another feature is the multi-cultural language support: any language, slang, or even grammatical error, can emerge in the interaction. In other words, Consciousnet is simply in sync with the status of web content, thus being up to date with the cultural aspects the web reflects.

Overall, looking at the different features summarized in Table 1, Cleverbot is the work that can be compared the most to Consciousnet. It also uses a quite large database, while still supervised and based on previous online conversations performed by users. Nevertheless, considering it from an academic research perspective, it suffers from what we believe to be a severe flaw, i.e., the lack of freely available APIs and open-source code. Other works, not listed in Table 1, propose the mining of specific web resources to extract the knowledge base, including discussion forums [21], [22], logs from virtual environment games [23], ontology databases [24], semantic layers added to Wikipedia content [25], and TV/movie dialogue transcripts [26], [27]. However, they still may be considered as falling in the category of expert systems, and thus, not suitable for the kind of general-purpose experimental platform proposed in this work.

III. EXPERIMENTAL SETUP

A. WEB-BASED TEXT GENERATION PLATFORM

In this section, we describe the linguistic interaction platform used to carry out the experiments. In particular, we adopted three main design choices for the experimental setup:

- **Text-only interaction:** i.e., no added "realism" based on multi-modal techniques. This choice excludes speech synthesis/recognition, three- dimensional avatars, touch-based interactions, or robotics. The idea is to focus on the content of linguistic interaction, instead of introducing distracting elements that could reveal the artificiality of the interlocutor. For example, although speech synthesis is widely used to "humanize" user-experience, it makes the interlocutor more recognizable as non-human. Further, even when such artificiality is explicitly declared, several studies such as [28] have demonstrated that computers exposing human features may lead the user to have overly high expectations that eventually influence the resulting dialogue.
- **Generality:** not specialized or focused on a particular conceptual domain. This differentiates the platform used from expert systems and assistance/entertainment chat robots.
- **Unsupervised linguistic space:** neither explicit database of concepts nor responses to be maintained; the current status of web determines a "space" into which Consciousnet moves. This choice affects both semantic and syntactical aspects: e.g., slang, abbreviations, grammatical errors are part of linguistic space and can be generated by the environment.

Let us describe each architectural component and its role in the linguistic interaction process. Referring to the data flow shown in Figure 1, we consider a text input introduced by the user through the User Interface. This input is then is processed by Parser component, which produces an intermediate output, namely meta-response, following a set of assembling rules described in the Attitude configuration file (described below). It is fundamental to point out that a meta-response is not the final output of the artificial entity but a sort of "access key" that will be used by the next Net Inject component

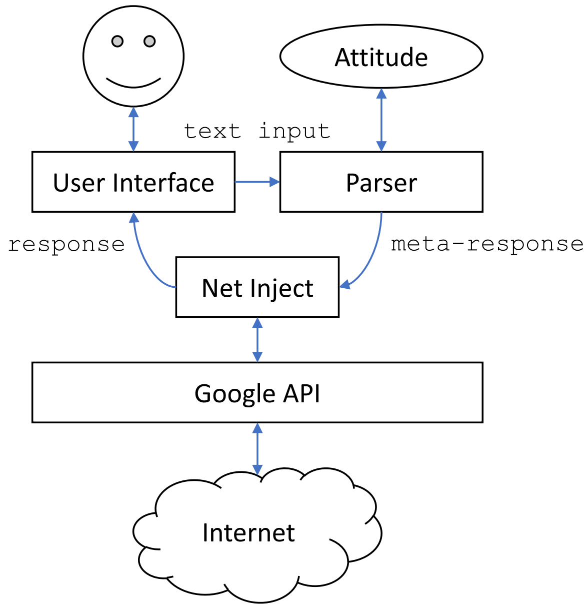

FIGURE 1. Consciousnet architectural components and data flow.

to extract web content. The production of a meta-response follows a set of rules (specified Attitude configuration) aimed at performing two main tasks on user's text: (i) detecting a potential set of keywords and selecting the one with the highest priority (ii) use the selected keywords to assemble an appropriate meta-response.

To give an example of how an Attitude data structure could work to generate a meta-response, let us consider the following simple Attitude configuration consisting in only three keywords rules:

```
<key "I" 10>
<pattern> * I *
<metaresponse> (1) people who (2)

<key "golf" 20>
<pattern> * golf *
<metaresponse>  why (1) golf (2)

<key default 0>
<pattern> *
<metaresponse> because *
<metaresponse> when they say *
```

Of course, this is an extremely trivial configuration, with only two regular keywords "I" and "golf" and a default keyword to use when no keyword is found. Now, let us consider as user's input the sentence "In the morning I talk to cats". Since the only keyword found is "I", the first rule is applied, and then a meta-response is assembled, using wildcards that replicate some parts of the sentence. The assembled result is: "In the morning people who talk to cats". Otherwise, if the input was "I play golf each weekend", even if both keywords are present, only the second with higher priority (20) would

TABLE 2. Entry-point list and abstraction levels.

Level	Type of Element	Examples
High	Commonly found, generic syntactical elements, *i.e.*, not useful for any restriction of the conceptual domains	I, me, are, you, sorry, yes, no
Mid-high	Verbs, nouns, elements denoting something less generic, such hypothesis, questions etc...	if, because, why, how, when, always
Mid	Terms introducing items that assume importance to the speaker	my X, your X
Mid-low	Terms introducing specific domains	music, cat, love, school, food, money
Low	Specific terms, referring to particular subjects	Schrodinger, Bowie, Berlin, spaghetti

be applied, leading to "why I play golf in the weekend". Thanks to the default keyword rule, even an input with no keywords would produce a meta-response. For example, e.g., "stop talking to me" would be assembled in "when they say stop talking to me" or "because stop talking to me". Table 2 shows an example of possible keywords, ordered by the abstraction level class. The approach chosen for experiments carried out in this work is to give the highest priority to keywords with the lowest level of generality (abstraction). This choice, while not mandatory, reflects the intuitive observation that the more abstract and general a word is, the less its semantic value is useful to understand the whole sentence.

The so assembled meta-response is then used by the Net Inject component to access the web content through a set Google APIs [30], freely available for non-commercial purposes. In particular, the Net Inject component performs a call of the Google's CustomSearch API method passing as argument the meta-response. The results obtained are a vector of strings selected according to Google's internal semantic algorithms. Please note that choosing this particular API is not relevant for the proposed text-generation approach, and any other equivalent library could be used instead. Finally, the Net Inject component processes the vector of strings, performing the following actions: (i) Filtering outputs in order to exclude some responses (e.g., too long, too short, with incomplete sentence, not allowed strings) (ii) Ordering on the basis of quality metrics (e.g., presence of particular keywords). The response string thus obtained is returned to the user, and then the interaction can repeat.

B. TESTING SCENARIO

In this Subsection, we describe the setup used to carry out the experimental tests using the interaction platform described in the previous Subsection. In order to understand the rationale behind the proposed testing setup, it might be useful to compare the proposed approach against the well-known test proposed by Turing. In a Turing test setup, shown in Figure 2(a), a human subject C interacts with two different counterparts A and B, which both asses to be "human". The purpose of the test is to check whether C can detect the artificial nature of one of the counterparts using a text-only interaction. The

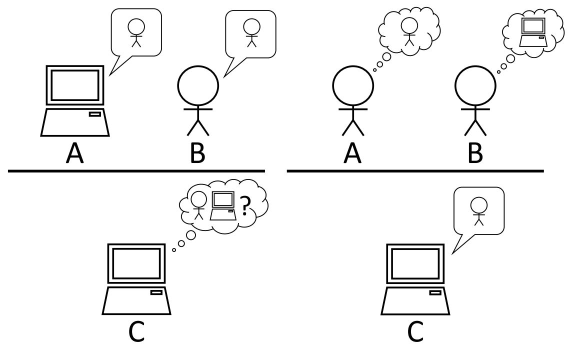

FIGURE 2. Comparison between the original Turing test setup (left) and the experimental setup (right) performed with Consciousnet.

experimental setup (Figure 2(b)) used in this work changes the perspective and the role of the three A, B, and C entities. First, both A and B assume an explicit belief about the nature of the counterpart. In other words, what changes between the experiments is not the actual nature of the counterpart C, but the assumption made by A and B. The comparison between the traditional Turing test and the proposed setup are summarized in Table 3, showing how the role of each of involved entities A, B, and C changes from both a functional and semantic perspective.

IV. EXPERIMENTAL RESULTS

A. DATA COLLECTION

A main set of 60 students, chosen from those attending the Computer Science for Psychology [31] course, held at the University of Catania, has been selected and then, each of them has been assigned to one the two sets α and β in order to have equal distribution of 30 male/female components with age ranging from 19 to 22 years. Students belonging to the set α were instructed that they were performing a generic text-only chat session with an American professor (Dr. Gioio), interested in testing a new form of interaction to be used in the future with his students. Each student entered (one at a time) in an isolated room, and, after a 10-minute chat session, moved to a separate room to avoid any influence on subsequent students. When all the students of the set α

TABLE 3. Comparison of roles and subjects involved in both traditional Turing test and adopted setup.

	Turing Test	*test*
What is Fixed	*both A and B declare to be human*	*both A and B are actually human*
What changes	*the actual nature of A and B*	*the belief of A and B*
Who is the Liar	*A*	*C*
Who is the Tester	*C*	*A and B*
What is detected	*a difference in the interaction, due to the different nature of A and B*	*a difference in the interaction, due to the different beliefs of A and B*
Data analysed by Human	*C analyses the language used by A and B*	*A and B analyse the language used by C*

ended their sessions, the real experiment has been revealed, and permissions to access chat transcripts have been asked.

A total of 1116 and 943 sentences have been collected from participants of sets α and β, respectively, excluding those produced by the platform itself during the interaction. A quantitative and qualitative analysis of the transcripts was conducted to investigate the impact in terms of covered semantic domains, syntax-oriented metrics, and statistically relevant differences among the two sets α and β. In the pre-processing phase, some classes of words were excluded in order to have a less noisy input dataset to analyze. For example, words with a usage frequency under the threshold of 1% on the overall data. Another exclusion regarded those words commonly referred to as stop words, i.e., words that do not play any particular semantic role (for a complete list of the stop words adopted see also [32]). The remaining words were labeled using the Stanford POS Tagger [33] considering 4 categories (nouns, proper nouns, adjectives, adverbs, and verbs).

In order to perform a quantitative analysis of the samples belonging to set α and β, the following metrics have been chosen:

- **Sentence length:** the average length of each user sentence (in the number of characters).
- **Number of Sentences:** the number of sentences produced by each user during the interaction.
- **Nouns:** the number of nouns used by each user in the interaction.
- **Proper Nouns:** the number of proper nouns (e.g., names) used by each user in the interaction.
- **Adjectives:** the number of adjectives used by each user in the interaction.
- **Adverbs:** the number of adverbs used by each user in the interaction.
- **Verbs:** the number of verbs used by each user in the interaction.

While several further different choices could be possible, we judged the above metrics as an excellent trade-off to perform a quantitative analysis while still capturing some intuitive critical aspects of the produced linguistic interaction.

B. STATISTICAL ANALYSIS

This section describes the methodology adopted to evaluate any statistically significant difference between results coming from sets α and β. A key requirement of such analysis is that only a relatively limited amount of data is available for the evaluation. In order to overcome this limitation, a Unpaired Two-Sample t-test [34] procedure has been adopted to verify hypotheses regarding a statistically significant difference in the average values observed for the metrics defined above.

More formally, denoting with $\mu_{\alpha,i}$ and $\mu_{\beta,i}$ the true mean of metric i for sets α and β, respectively, we are interested in verifying whether there is enough statistical evidence to accept the null hypothesis $\mu_{\alpha,i} - \mu_{\beta,i} = 0$, i.e., no difference between the means of metric i when comparing the two sets.

The veracity of the null hypothesis will be verified together with a corresponding level of statistical reliability, represented by two error probabilities P_a and P_b, that are, the probability of having false positives and negatives, respectively. For the experiments presented in this section, we assume $P_b = 0.05$ and $P_a = 0.05$ targets, which are commonly accepted in literature as suitably small values, as suggested in [34].

Another important aspect is related to the size of samples collected in sets α and β, since only an appropriate number of samples would guarantee the capacity of identifying small differences while maintaining acceptable statistical significance levels. With this aim, we first chose, for each metric i, a threshold value $\delta_i = |\mu_{\alpha,i} - \mu_{\beta,i}|$ as critical difference, i.e., representing some discordance between the average values that we want to be sure to detect (if present). Next, we adopted the approach based on Cohen's d (know also as effect size) [35], using tables which express sample size in function of the desired P_a, P_b, and δ_i. With this regard, we chose a critical value of 8 characters for sentence length, 5 sentences for the total number of sentences, and 2 for nouns, proper nouns, adjectives, adverbs, and verbs. While these values are somewhat arbitrary, they do not affect the behavior of t-test, but only our choice of the sample size in order to capture differences that we judge to be significant. Using the above values for P_a, P_b, and δ_i, we found the number of sentences collected (about 1000 for each set) was

TABLE 4. Results of t-test performed on set α and β.

	95% c.i. of $\mu_{\alpha,i} - \mu_{\beta,i}$	P-value	mean alpha	mean beta	Null hypothesis
Sentence Length	8.16 ± 4.99	0.001393	40.16	32.00	Rejected
Number of Sentences	0.29 ± 5.50	0.917100	32.83	32.54	Accepted
Nouns	4.64 ± 5.56	0.100200	23.13	18.49	No evidence
Proper Nouns	-0.10 ± 1.97	0.917400	3.87	3.97	Accepted
Adjectives	2.81 ± 2.23	0.014230	8.88	6.07	Rejected
Adverbs	1.31 ± 2.47	0.293600	10.48	9.17	No evidence
Verbs	3.06 ± 4.87	0.213300	27.00	23.94	No evidence

more than sufficient to determine rejection/acceptance of the statistical hypothesis about the real mean difference between set α and β.

The results obtained for each metric are summarized in Table 4. In particular, columns report in order (from left to right): confidence interval of the mean difference $\mu_{\alpha,i} - \mu_{\beta,i}$, mean values for each of the two sets, and the corresponding $P - value$. The last column summarizes the conclusion that should be taken concerning the null hypothesis (i.e., no difference): accepted, rejected, or no evidence. Recalling the methodology described earlier, a sufficiently small $P - value$ (≤ 0.05) provide enough statistical evidence that the null hypothesis can be rejected, i.e., a statistically significant difference is present between α and β data sets. This is the case of the metrics representing the number of adjectives used and the sentence length. On the contrary, for the number of sentences and proper nouns, the null hypothesis is accepted, and we can assess that no significant difference has been detected.

While the obtained acceptance/rejection results of the null hypothesis for each metric can be a starting point for many different interpretations/considerations, but a few points can be assessed. First, unaware users of the group α tend to use (on average) longer sentences, at least eight characters per sentence. Please notice how the number of sentences for each interaction remains the same. Since the interaction time was fixed (10 minutes), and the sentence length of the artificial counterpart is statistically not dependent by the group being investigated, we can asses that the average time spent by users for producing each sentence was on average the same of both group α and β (32 sentences in 10 minutes, that is, 19 seconds per sentence). So, the eight chars difference in sentence length can be interpreted not only as ''more content'', but also as 25% quicker typing of the unaware group α users. A difference is also detected for the number of adjectives, which seems to be more frequently used by group α users. Finally, the acceptance of the null hypothesis for proper nouns is somewhat expected, usually being a poorly common lexical term (3-4 times per dialogue) appearing in a relatively predictable way, e.g., at the beginning of the session, when users write their names or cities.

A separate discussion is required for nouns, adverbs, and verbs. They show a third interesting possibility: some difference has been detected, but the $P - value$ is not sufficiently small to attribute any significance to the results. We can think of two possible reasons: (i) a difference in the actual means does not actually exist, and the non-zero values reported as difference are merely statistical fluctuations due to an insufficient amount of samples (ii) a difference exists, but it is smaller than the critical difference chosen when determining the sample size. For example, suppose that for some reason, users of set α tend to use a little more adverbs than users of set β. If such difference can be quantified on average as equal to 1, it will never be detected by our experimental setup, since the sample size is statistically capable of detecting only differences being at least equal to 2 (i.e., $\delta_i = 2$ for adverbs). Nevertheless, each critical δ_i has been chosen as representative of the minimal difference judged as significative, so increasing the sample size for detecting a smaller difference should be considered a lousy test design choice.

V. CONCLUSION

In this work, we investigated the impact of awareness about the nature of the counterpart in human-machine text-based interaction. Using a platform generation of a general-purpose linguistic interaction, we carried out experiments involving two different groups of users, representative of the two different beliefs about the interlocutor. A statistical analysis of data collected shows evidence of differences in terms of sentence length and number of adjectives used. In contrast, the total number of sentences and the use of proper nouns remain unaltered.

REFERENCES

[1] I. Goldstein and S. Papert, ''Artificial intelligence, language, and the study of knowledge,'' *Cogn. Sci.*, vol. 1, no. 1, pp. 84–123, 1977.

[2] D. Isles, ''Artificial intelligence as a possible tool for discovering laws of logic,'' *Cognit. Sci.*, vol. 2, no. 4, pp. 329–360, Feb. 1978.

[3] J. McCarthy and P. Hayes, *Some Philosophical Problems From the Standpoint of Artificial Intelligence*. Stanford, CA, USA: Stanford Univ., 1968.

[4] A. C. Clarke, *2001: A Space Odyssey*. New York, NY, USA: New American Library, 1968.

[5] H. S. Hasan and S. A. Kareem, ''Human computer interaction for vision based hand gesture recognition: A survey,'' in *Proc. Int. Conf. Adv. Comput. Sci. Appl. Technol. (ACSAT)*, Nov. 2012, pp. 55–60.

[6] P. Xia, ''Haptics for product design and manufacturing simulation,'' *IEEE Trans. Haptics*, vol. 9, no. 3, pp. 358–375, Jul. 2016.

[7] S. C. Eastwood, V. P. Shmerko, S. N. Yanushkevich, M. Drahansky, and D. O. Gorodnichy, "Biometric-enabled authentication machines: A survey of open-set real-world applications," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 2, pp. 231–242, Apr. 2016.

[8] S. Cascianelli, G. Costante, T. A. Ciarfuglia, P. Valigi, and M. L. Fravolini, "Full-GRU natural language video description for service robotics applications," *IEEE Robot. Autom. Lett.*, vol. 3, no. 2, pp. 841–848, Apr. 2018.

[9] F. Lamberti, F. Manuri, G. Paravati, G. Piumatti, and A. Sanna, "Using semantics to automatically generate speech interfaces for wearable virtual and augmented reality applications," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 1, pp. 152–164, Feb. 2017.

[10] R. Pringle, "Love, philosophy, and processors: Interview with a robot [Interview]," *IEEE Technol. Soc. Mag.*, vol. 35, no. 4, pp. 52–55, Dec. 2016.

[11] S. Brennan, "Conversation with and through computers," *User Model. User-Adapted Interact.*, vol. 1, no. 1, pp. 67–86, 1991.

[12] J. F. Kelley, "An iterative design methodology for user-friendly natural language office information applications," *ACM Trans. Inf. Syst.*, vol. 2, no. 1, pp. 26–41, Jan. 1984.

[13] H. P. Branigan, M. J. Pickering, J. Pearson, J. F., McLean, and C. Nass, "Syntactic alignment between computers and people: The role of belief about mental states,"in *Proc. 25th Annu. Conf. Cogn. Sci. Soc.*, New York, NJ, USA: Lawrence Erlbaum Associates, 2013, pp. 186–191.

[14] K. Fischer, "The role of users' preconceptions in talking to computers and robots," in *Proc. Workshop How People Talk to Comput., Robots, Artif. Commun. Partners*, vol. 2, 2006, pp. 112–130.

[15] J. Pearson, J. Hu, H. P. Branigan, M. J. Pickering, and C. I. Nass, "Adaptive language behavior in HCI: How expectations and beliefs about a system affect users' word choice," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. (CHI)*, vol. 2, 2006, pp. 1177–1180.

[16] H. P. Branigan, M. J. Pickering, J. Pearson, J. F. McLean, and A. Brown, "The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers," *Cognition*, vol. 121, no. 1, pp. 41–57, Oct. 2011.

[17] J. Weizenbaum, "ELIZA—A computer program for the study of natural language communication between man and machine," *Commun. ACM*, vol. 9, no. 1, pp. 36–45, Jan. 1966.

[18] K. M. Colby, "Simulations of belief systems," in *Computer Models of Thought and Language*, R. C. Schank and K. M. Colby, Eds. San Francisco, CA, USA: Freeman, 1973, ch. 6, pp. 251–286.

[19] B. A. Shawar and E. S. Atwell, "Using corpora in machine-learning chatbot systems," *Int. J. Corpus Linguistics*, vol. 10, no. 4, pp. 489–516, 2005.

[20] P. Tarau and E. Figa, "Knowledge-based conversational agents and virtual storytelling," in *Proc. ACM Symp. Appl. Comput. (SAC)*, New York, NY, USA, 2004, pp. 39–44.

[21] J. Huang, M. Zhou, and D. Yang, "Extracting chatbot knowledge from online discussion forums," in *Proc. 20th Int. Joint Conf. Artif. Intell. (IJCAI)*, San Francisco, CA, USA: Morgan Kaufmann, 2007, pp. 423–428.

[22] Y. Cao, W.-Y. Yang, C.-Y. Lin, and Y. Yu, "A structural support vector method for extracting contexts and answers of questions from online forums," *Inf. Process. Manage.*, vol. 47, no. 6, pp. 886–898, Nov. 2011.

[23] J. Orkin and D. Roy, "Semi-automated dialogue act classification for situated social agents in games," in *Agents for Games and Simulations II*. Springer, 2011, pp. 148–162.

[24] J. Bang, S. Han, K. Lee, and G. G. Lee, "Open-domain personalized dialog system using user-interested topics in system responses," in *Proc. IEEE Workshop Autom. Speech Recognit. Understand. (ASRU)*, Dec. 2015, pp. 771–776.

[25] A. Augello, G. Pilato, G. Vassallo, and S. Gaglio, "A semantic layer on semi-structured data sources for intuitive chatbots," in *Proc. Int. Conf. Complex, Intell. Softw. Intensive Syst.*, Mar. 2009, pp. 760–765.

[26] L. Nio, S. Sakti, G. Neubig, T. Toda, M. Adriani, and S. Nakamura, *Developing Non-Goal Dialog System Based on Examples of Drama Television*. New York, NY: Springer, 2014, pp. 355–361.

[27] R. E. Banchs and H. Li, "Iris: A chat-oriented dialogue system based on the vector space model," in *Proc. ACL Syst. Demonstrations (ACL)*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2012, pp. 37–42.

[28] C. Nass, *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. Cambridge, MA, USA: MIT Press, 2005.

[29] O. Vinyals and Q. Le, "A neural conversational model," 2015, *arXiv:1506.05869*. [Online]. Available: http://arxiv.org/abs/1506.05869

[30] Google Inc. (2016). *Google CustomSearch*. [Online]. Available: https://developers.google.com/custom-search/

[31] University of Catania. (2018). *Computer Science for Psychology*. [Online]. Available: http://syllabus.unict.it/insegnamento.php?id=7215&eng

[32] Ranks.nl. (2016). *English Stopwords*. [Online]. Available: http://www.ranks.nl/resources/stopwords.html

[33] The Stanford Natural Language Processing Group. (2016). *Stanford Log-Linear Part-of-Speech Tagger*. [Online]. Available: http://nlp.stanford.edu/downloads/tagger.shtml

[34] D. C. Montgomery, *Design and Analysis of Experiments*, 5th ed. Hoboken, NJ, USA: Wiley, 2000.

[35] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. New York, NJ, USA: Lawrence Erlbaum Associates, 1977.

VINCENZO CATANIA (Member, IEEE) received the Laurea degree in electrical engineering from the University of Catania, Catania, Italy, in 1982. He was with STMicroelectronics, Catania, till 1984, responsible for testing microprocessor system. He is currently a Full Professor of computer engineering with the University of Catania. His current research interests include performance and reliability assessment in parallel and distributed systems, the Internet of Things, and pervasive mobile computing.

SALVATORE MONTELEONE (Member, IEEE) received the B.Sc. and M.Sc. degrees in computer engineering and the Ph.D. degree in communications and computer engineering from the University of Catania, in 2007, 2010, and 2014, respectively. He is currently a Research Fellow at the Institute of Advanced Studies and at ETIS (UMR 8051—CNRS), University of Cergy-Pontoise, France. His research interests include comprehend cooperative systems, user-centric systems, and network-on-chip architectures.

MAURIZIO PALESI (Senior Member, IEEE) received the Ph.D. degree in computer engineering from the University of Catania, Italy, in 2003. He is currently an Associate Professor of computer engineering at the University of Catania. His main research interests focuses on on-chip communication system architectures with particular emphasis on energy efficient implementations. He collaborates with several international research institutes, serves as a TPC Member and in the editorial board for several international conferences and journals. He is also a HiPEAC Member.

DAVIDE PATTI (Member, IEEE) received the Laurea and Ph.D. degrees in computer engineering from the University of Catania, Catania, Italy, in 2003 and 2007, respectively. He is currently an Assistant Professor with the University of Catania. He has published several works on platform-based system design, design space exploration, and low-power techniques for embedded systems and network-on-chip architectures.