

Understanding Texts in Attempto Controlled English

Norbert E. Fuchs
Institute of Computational Linguistics
University of Zurich

Abstract. Understanding texts in Attempto Controlled English (ACE) is considered undemanding, nonetheless hides some problems. To deal with these problems I propose an experiment based on Kuhn's ontographs that tests the understanding of simple ACE texts. Furthermore, I suggest to compare the relation between authors and readers with human verbal conversations. My conclusion is that the correct understanding of an ACE text is possible, but requires contributions from both authors and readers, quasi their cooperation.

Keywords. Attempto Controlled English, ACE, ontographs, human conversation, cooperative principle, Grice's maxims, repair, deduction, RACE, ACE comments

1. Can Texts in Attempto Controlled English Be Understood by Anybody?

We, the authors of Attempto Controlled English (ACE), claim that "once written, ACE texts can be read and understood by anybody"¹.

Is our claim really justified?

ACE was originally developed in the context of requirements engineering with the goal to make formal methods available to people who are not familiar with them. We have evidently achieved this goal as the current version of ACE is a rich and powerful knowledge representation language in the form of a subset of English that can automatically be translated into many formal, mostly logic, languages.

On top of that requirement engineering suffers from another problem that is related to our claim above. Requirement specifications, generally written in unconstrained natural language, are often not understood in the same way by the various people involved in the software development process [1,2]. This can lead – and has led – to misunderstandings with serious consequences for the software project. The misunderstandings are to a large part caused by the use of unconstrained natural language that is prone to ambiguity, vagueness, inconsistency, incompleteness, and hidden context dependencies. Other misunderstandings are due to human faults and to organisational shortcomings.

The question arises whether ACE – being a controlled language with a logical basis – can avoid, or at least alleviate, communication problems as those exhibited in requirements engineering. For a question so general, the answer must be no since ACE does not offer any assistance for organisational shortcomings. However for cases not involving organisational aspects, ACE seems to be the perfect answer since it eliminates most linguistic issues and provides means to curtail human faults.

¹ <http://attempto.ifi.uzh.ch/site/description/>

For his experiments Kuhn introduces ontographs, graphical notations of scenes called mini worlds together with a legend. The legend introduces types and (binary) relations, while the mini world shows individuals, their types and their relations. The prime quality of ontographs is that they provide concise, unambiguous and nonverbal descriptions of situations. Figure 1 shows an example.

In his first experiment Kuhn investigates the understandability of ACE by combining an ontograph with a set of 20 ACE sentences of which about half are true with respect to the depicted mini world, while the rest is false. Following is the set of sentences for the ontograph shown in figure 1.

John sees Tom. Lara sees Mary. Mary does not see Tom. Tom does not see Lara. Tom buys a picture. John buys a present. John sees no woman. Mary sees no man. Tom sees every woman. Lara sees every man. Tom sees nothing but women. John sees nothing but men. Lara buys nothing but presents. Lara buys nothing but pictures. No woman sees herself. No man sees himself. Every woman buys nothing but pictures. Every man buys nothing but presents. No man who buys a picture is seen by a woman. No woman who buys a picture is seen by a man.

The task of the 15 test persons consisted in identifying which of the 20 ACE sentences associated with the ontograph is true or is false. The test was performed with four ontographs of increasing complexity. The experiment showed that overall 83% of the ACE sentences were classified correctly. If three logically difficult sentences were not taken into account then the correct classification rose to almost 88%.

In a second experiment Kuhn compared the understandability of ACE with the understandability of a modified version of OWL's Manchester Syntax², a visibly formal language called MLL [3]. Each ontograph is accompanied by 20 ACE sentences or, alternatively, 20 logically equivalent MLL statements, both again about half correct and half incorrect with respect of the depicted mini world. The results with 64 test persons working with four diverse ontographs showed clearly, and statistically significantly, that ACE is better understood than MLL.

Kuhn concludes that his experiments clearly show that ACE is generally well understood by persons not familiar with ACE, and that ACE is understood significantly better than MLL.

3. Testing the Understanding of ACE Texts with Kuhn's Ontographs

As it happens, Kuhn's ontographs – by their very nature of being concise and unambiguous nonverbal descriptions of situations – can also be used to investigate the question whether a reader of an ACE text will understand the text as intended by its author.

Here is the outline of an experiment that is based on Kuhn's first experiment and that distinguishes between the role of the author and the role of the reader. The author takes an ontograph and generates a set of true ACE sentences describing the ontograph. For concreteness, let us assume that the author chooses the ontograph of figure 1 and the "true" subset of its associated ACE sentences, namely

John sees Tom. Mary does not see Tom. John buys a present. Mary sees no man. Tom sees every woman. Tom sees nothing but women. Lara buys nothing but presents. Lara

² <https://www.w3.org/TR/owl2-manchester-syntax/>

buys nothing but pictures. No man sees himself. Every woman buys nothing but pictures. No man who buys a picture is seen by a woman.

The reader – who acts as test person – is now shown the above ACE sentences together with a small set of ontographs including the one that is described by the ACE sentences. The reader's task is then to identify the correct ontograph.

For lack of resources I did not do the experiment, but – given the excellent results of Kuhn's two experiments – I am highly confident that it would show that the test persons can successfully relate a set of true ACE sentences to the correct ontograph.

However, a positive outcome of the experiment should only be considered as a supportive argument, but not as a proof, that the reader of an ACE text will understand it in the way that the author intended. I have two reasons for this reservation. First and most importantly, the ACE texts associated with ontographs consist only of individual, unrelated sentences, they do not constitute one coherent text. Second, Kuhn's texts use only a subset of ACE, namely those ACE constructs that can be mapped to the semantic web language OWL. Some commonly occurring constructs are missing, for example explicit if-then sentences, anaphoric references interrelating sentences, intransitive and ditransitive verbs, modifiers of nouns and verbs.

When present, these ACE constructs lead to more structured, more connected and more involved texts that go beyond what can be expressed by ontographs. Alternative approaches are needed to investigate the understanding of these texts. One such alternative is offered by human conversation.

4. Understanding ACE Texts in the Context of Human Conversations

Conversations are interactive verbal communications between two or more people that underlie explicit and implicit social rules, linguistic rules and timing constraints [4]. Most important in our context is the mutually expected and required cooperation between speakers and listeners to achieve a successful communication, i.e. a clear and immediate understanding of the information exchanged. Clark [5] and Brennan [6] introduced the concept of grounding in conversations that comprises the collection of "mutual knowledge, mutual beliefs, and mutual assumptions" that is essential for the communication between people. Successful grounding in communication requires all parties "to coordinate both the content and process". Conversations usually consist of a series of utterances and possibly clarifying questions, where speakers and listeners take turns in order to gradually complete the communication, but it is also possible that a single exchange takes place between one speaker and one or more listeners.

I contend that a form of cooperation, a form of grounding, is also required between the author and the reader of an ACE text to make sure that the reader understands the intentions of the author. This cooperation, too, requires explicit and implicit rules to be successful, details of which we will see in the following. This motivates me to consider the relation between author and reader as a form of conversation. The analogy turns out to be fruitful with respect to the question whether and how far a reader understands the authors intentions, though it has its limitations and peculiarities: the communication is not verbally, but in written form; there is no taking turns of author and reader, but only one exchange of a complete text; a reader usually cannot ask an author clarifying questions that are often an integral part of a verbal conversation.

Continuing with the analogy I define a conversation of this kind as successful if the reader understands the ACE text as its author intended.

Next we will have a closer look at the necessary contributions of the author and the reader to achieve this goal.

5. The Contributions of the Author

The task of authors to write clearly and to keep their readers in mind has been spelled out since ancient times by many teachers of good writing style. Here I will focus on those aspects that are related to controlled languages, specifically to ACE. I will assume that the author is sufficiently familiar with ACE.

The ACE Trouble Shooting Guide³ gives some general hints on the construction and interpretation of an ACE text that can be summarised as "Though the ACE parser will unravel every syntactically correct sentence, however complex, you [and your readers] may have problems to do so. Thus keep your sentences short and simple."

These hints are similar to the writing rules defined for Simplified Technical English STE⁴, but are not as detailed and do not radically restrict the length of certain language constructs. The decisive difference between ACE and STE is that the grammatical correctness of ACE texts can be checked by parsers while there are no checkers for STE. The hints of the ACE Trouble Shooting Guide also recall the Guidelines on Transparency defined for the EU General Data Protection Regulation (GDPR).

The major part of the ACE Trouble Shooting Guide is dedicated to trouble shooting per se. The use of each ACE construct is examined in great detail, possible pitfalls and their avoidance are considered, and solutions for often occurring problems suggested.

Though hints, guidelines and writing rules contribute to the understanding of the author's text they do not address the essential and most critical point, namely the content that is to be communicated to the reader.

Effective communication requires – as I stated above – a cooperation between the author and the reader. In the context of verbal conversations Grice introduced the cooperative principle⁵ that quite naturally can also be applied to the author-reader case. In the author-reader case, the cooperative principle binds foremost the author as the one who produces the text, but also the reader who relies on the author to adhere to the principle. From the cooperative principle Grice derived four maxims that substantiate it. I rephrase Grice's maxims to be suitable for the author-reader case and for ACE, and add some clarifying comments.

- Maxim of Quality

The text should be true. Though using an English syntax, ACE is a formal logic language. Thus an ACE texts can be given a model-theoretic or a proof-theoretic meaning [7]. Considered model-theoretically, an ACE text is the description of a modelled domain in which case – as Kuhn's ontograph experiment shows – some sentences can describe the domain correctly, while others do not. This approach is also called truth-conditional since the correct sentences are labelled true, the others false. Considered proof-theoretically, the sentences of an ACE text are not

³ <http://attempto.ifi.uzh.ch/site/resources/>

⁴ <http://www.asd-ste100.org>

⁵ https://en.wikipedia.org/wiki/Cooperative_principle

interpreted with respect to a modelled domain, but are assumed to describe a true state of affairs. The meaning of a sentence is accordingly no longer described in terms of truth-conditions but in terms of the logical inferences that can be drawn from that sentence. It remains the responsibility of the writer to avoid contradictions. For computational purposes – for instance to answer questions from the ACE text – the proof-theoretic approach is preferred since it supports logical deduction on a syntactic not on a semantic basis. The Attempto reasoner RACE that will be presented later as a tool for the reader interprets an ACE text proof-theoretically.

- Maxim of Quantity

The text should be complete, but should not contain additional information. As the experience with requirement specifications and instruction manuals shows this maxim is hard to fulfil, and often it is not. This maxim makes the most exacting demand on the author of an ACE text.

- Maxim of Relevance

The text should be relevant to the topic described. For ACE texts this maxim seems to coincide with the maxim of quantity.

- Maxim of Manner

The text should be perspicuous. This maxim should restrain the author from creating convoluted texts and from showing off. Following this maxim is made easier by the unambiguousness of ACE and by the guidelines of ACE's Trouble Shooting Guide.

There are potential pitfalls in ACE relevant for the above maxims. ACE's handling of anaphoric references and ambiguities can lead even people familiar with ACE into a conflict between the reading enforced by ACE's interpretation rules and a reading suggested by common sense. The author of an ACE text must be aware of these pitfalls that, if not avoided, could also negatively affect the reader of the ACE text.

Here is an – admittedly contrived – case of misleading resolutions of anaphoric references. The author wants to express that a manager calls a subordinate to give the subordinate some order, and writes

A manager calls a subordinate. He gives him some orders.

In standard English this would be understood as intended since we use context and common sense to correctly resolve the anaphoric references *he* and *him*. However, according to the ACE interpretation rules the pronoun *he* of the second sentence refers to the *subordinate* and the pronoun *him* to the *manager*, thus incorrectly stating that the subordinate gives orders to his superior. The author can easily avoid this unintended meaning using more explicit anaphoric references, for example

A manager calls a subordinate. The manager gives the subordinate some orders.

Next a potentially misleading case of ambiguity and anaphora resolution – even more contrived.

A girl owns a telescope. A girl owns a camera. A man sees the girl with the telescope.

Reading this in standard English we most probably would interpret the third sentence to mean that the man sees the girl that owns the telescope. However, according to the ACE interpretation rules the third sentence has quite another meaning: the anaphoric reference *the girl* refers to the textually closer *a girl* in the second sentence, the one that owns a camera, the prepositional phrase *with the telescope* modifies the verb *see* not *the girl*, and *the telescope* refers to *a telescope* of the first sentence. To express the "common sense meaning" the author could, for instance, write

A girl owns a telescope. A girl owns a camera. A man sees the girl that owns the telescope.

Of course, contriving cases where the "standard English meaning" and the "ACE meaning" clash does not properly depict the reality of using ACE where such clashes are rare and – as demonstrated – easily avoided.

6. The Contributions of the Reader

I assume that also the reader is familiar with ACE. Furthermore, I assume that the reader cannot contact the author, but has only the author's complete ACE text available.

An important concept in verbal conversation is repair [4], elsewhere called "establishing a common ground" [5,6]. Repair can be performed by the speakers who correct themselves while speaking, but in most cases it is the responsibility of the listeners to contribute to the success of the conversation by asking clarifying questions or by showing signs of not understanding.

Defining the relation between the author of an ACE text and its reader as conversation, I must also redefine the concept "repair" given that the reader cannot ask clarifying questions or express lack of understanding. How can repair – understood as an attempt to better understand the ACE text – be defined in these circumstances?

My answer is logical deduction for which ACE is supported by several automated tools. Best suited as a tool for repair is RACE [8,9], a first-order reasoner with equality that can show the (in-) consistency of an ACE text, deduce one ACE text from another one and answer ACE questions on the basis of an ACE text. All three capabilities of RACE – consistency checking, deduction, question answering – can help the reader to better understand the given ACE text, to "repair" the author-reader conversation in the newly defined sense.

To begin with, I will focus on RACE questions that can extract explicit and implicit information from the ACE text. RACE offers altogether 11 forms of questions that we will use to investigate the following example ACE text.

John's red cat catches a mouse in a garden. Some black cats of John sleep silently. No red cat is a black cat. Every cat is an animal. Every mouse is an animal.

Most general are *yes/no* questions that just ask for the existence or non-existence of facts, for instance

Is there a black cat that sleeps? Does a mouse sleep? Is a cat a mouse?

If a *yes/no* question can be answered then RACE will show all answers and for each answer all sentences of the ACE text needed for that answer. For lack of space, I will omit here and in the following the – mostly obvious – answers generated.

Detailed information can be gained by *wh*-questions, i.e. questions with query words. If a *wh*-question can be answered then RACE will show all ACE sentences needed for the answer and, most importantly, the substitutions for the query words. Again, there can be more than one answer.

- Asking for subjects and objects of sentences is done by *who*-questions, *whose*-questions, *what*-questions and *which*-questions.
Who catches what in a garden? Whose cats sleep? Which cat sleeps?
- Asking for adverbs or prepositional phrases that modify verbs is done via *how*-questions, *where*-questions and *when*-questions.
Where does a cat catch a mouse? How does a black cat sleep?
- Asking for the cardinality of countable objects or for the amounts of mass

objects is performed by *how-many*-questions and *how-much*-questions.

There are how many cats? There are how many animals? How many cats sleep?

- Finally, asking for verbs can be done via *do*-questions.

What does a red cat do? What do some black cats do? What does a cat do?

In human conversations presuppositions⁶ – implicit assumptions based on shared common knowledge – and implicatures⁷ – information that is only suggested by a conversation – play an important role. Thus also the reader of an ACE text may be tempted to interpret the text beyond what it concretely states, respectively seems to logically imply.

Actually, this behaviour is justified in a few cases. RACE relies for its deductions on built-in auxiliary axioms that express domain-independent common knowledge, for example the relation between singular and plural nouns, or the laws of arithmetic. Some of the deductions using auxiliary axioms have the quality of presuppositions or implicatures since they add information that is not explicit in the ACE text. Using the above example text RACE can, for instance, derive

- for a noun its existence expressed by *there is* or *there are*
There is a red cat. There are some cats. There is a mouse. There is a garden.
- for a noun its replacement by *somebody* or *something*
There is somebody. Somebody's cats sleep. Somebody catches something.
- from possessives the ownership expressed by the verb *have*
John has some black cats.
- from *some X* either *one X* or *at least one X* or *more than one X* or *two X* or *at least two X*
At least one black cat sleeps. More than one black cat sleeps.

Other presuppositions or implicatures are not supported by RACE because they are either domain-dependent or too specific.

To sum up, RACE – offering a method of "repair" of the author-reader relation – provides the reader with powerful methods to investigate the ACE text at hand, and ultimately to understand it better.

7. The Role of ACE Comments

RACE offers many powerful forms of questions, but leaves out one highly important for the understanding of the ACE text, namely the question "why?" that would reveal the intentions and decisions of the author.

The question "why?", and similar ones, cannot be answered by the ACE text since they would have to provide information about the text itself, that is meta-information for which there are no – and there cannot be – ACE language constructs.

Being privy to authors meta-information must be provided by them. One way to do so is offered in the form of ACE comments that so far have not found their proper attention as an important means of information transfer between authors and readers.

⁶ <https://en.wikipedia.org/wiki/Presupposition>

⁷ <https://en.wikipedia.org/wiki/Implicature>

ACE comments have the advantage that they can contain any information using the less constrained standard English – which, alas, could also be the cause of further misunderstandings.

To the above example ACE text the author could add a comment to explain why the sentence *No red cat is a black cat.* is necessary to correctly count the cats.

```
...  
# Red cats and black cats must be declared distinct to correctly count all cats.  
No red cat is a black cat.  
...
```

Comments can also be used by authors for other forms of meta-information that need to be transmitted to the reader, for instance clarifying examples. Furthermore, comments can provide helpful structuring information of the ACE text, like titles of chapters and sections.

8. The Cooperation between Authors and Readers Ensures Understanding

The question whether and to which extent the reader of an ACE text understands the text as the author intended is a difficult one.

To find an answer to this question I first suggest an experiment based on Kuhn's ontographs. Users have to identify the ontograph described by a given ACE text. This experiment is instructive, but – using a subset of ACE in a restricted way – has only limited evidence with respect to the question of understanding.

My second approach is based on the analogy of the author-reader relation to human verbal conversations. This analogy turns out to be fruitful since it allows me to show that the understanding of an ACE text does not concern the reader alone, but necessarily requires contributions from both the author and the reader. For authors I derive guide-lines for the presentation and for the content of the ACE text. I also offer solutions to possible pitfalls. For readers I suggest logic deduction as a powerful tool to increase the understanding of the ACE text.

Furthermore, I advocate ACE comments for the transmission of important meta-information between author and reader.

To summarise: My investigations validate our claim that a given ACE text can be understood by any reader – however only provided that both the author and the reader of the ACE text cooperate to ensure a successful communication.

Acknowledgements

I would like to thank the Institute of Computational Linguistics, University of Zurich, for its hospitality. Furthermore, I very much appreciate the critical and constructive comments of the four anonymous reviewers of a previous version of this paper.

References

1. A. Al-Rawas, S. Easterbrook, Communication Problems in Requirement Engineering: A Field Study, Proc. First Westminster Conference on Professional Awareness in Software Engineering, London, 1996.
2. L. Li, *Desiree: a Refinement Calculus for Requirements Engineering*, PhD Thesis, DISI, Università degli Studi di Trento, 2016.
3. T. Kuhn, *Controlled English for Knowledge Representation*, Doctoral thesis, Faculty of Economics, Business Administration and Information Technology of the University of Zurich, 2010.
4. N. J. Enfield, *How we talk. The inner workings of conversation*, Basic Books, 2017.
5. H. H. Clark, S. E. Brennan, Grounding in communication. In L. B. Resnick et al (eds.), *Perspectives on socially shared cognition*, Washington, DC: American Psychological Association (1991) 127–149.
6. S. E. Brennan, The Grounding Problem in Conversations With and Though Computers. In S. R. Fussell, R. J. Kreuz (eds.), *Social and cognitive psychological approaches to interpersonal communication*, (1998), 201-225.
7. U. Schwertel. *Plural Semantics for Natural Language Understanding – A Computational Proof-Theoretic Approach*, PhD thesis, University of Zurich, 2004.
8. N. E. Fuchs, *First-Order Reasoning for Attempto Controlled English*, Proceedings of the Second International Workshop on Controlled Natural Language (CNL 2010), Springer, 2012.
9. N. E. Fuchs, *Reasoning in Attempto Controlled English: Non-Monotonicity*, Proceedings of the Fifth International Workshop on Controlled Natural Language (CNL 2016), Springer, 2016.