

A Case Study on the Relevance of the Competence Assumption for Implicature Calculation in Dialogue Systems

Judith Victoria Fischer^(✉)

Sprachwissenschaftliches Institut, Ruhr-Universität Bochum, Bochum, Germany
fischer@linguistics.rub.de

Abstract. The *competence assumption* (CA) concerns the estimation of a user that an implicature, derived from an utterance generated in a dialogue or recommender system, reflects the epistemic state of the system about the validity of alternative expressions. The CA can be assigned globally or locally. In this paper, we present an experimental study on the effects of locally and globally assigned competence in a sales scenario. The results of this study suggest that dialogue systems should include means for modelling global competence and that assigning local competence does not improve the pragmatic competence of a dialogue system.

1 Introduction

In order to enhance user acceptance, systems that generate utterances in dialogical scenarios as, e. g., question-answering systems or recommender systems with natural language interfaces, should show some degree of pragmatic behavior. In particular, dialogue systems face the challenge of generating utterances with sufficient scope for the user's possible calculation of implicatures. In the following example, a virtual sales assistant S uses the scalar expression *good*, instead of the semantically stronger *excellent*.

S: *The HP Laptop has a good AMD Radeon graphics coprocessor.*

According to the standard view of quantity implicature calculation, a user, reading this statement, would reason as follows:

The system generated some ψ (*good*) instead of a semantically stronger alternative ϕ (*excellent*). There must be reasons for not generating ϕ : Either the system does not believe ϕ to be true or it believes ϕ not to be true:

$$\neg Bel_S(\phi) \vee Bel_S(\neg\phi)$$

Which interpretation holds, relies on the *competence* or *experthood assumption*: Only if it is shared knowledge of system and user that the system is competent in a way that allows them to predict the truth value of the semantically stronger ϕ , the user will eventually infer the stronger reading (Potts 2015). The inference can be outlined as follows:

1. ϕ is stronger than ψ , but S didn't claim ϕ .
2. Reason: $\neg Bel_S(\phi)$
3. CA: $Bel_S(\phi) \vee Bel_S(\neg\phi)$
4. 1 – 3 entail $Bel_S(\neg\phi)$

Two different inferences are involved in this overall picture. The first inference concerns the first epistemic assignment (i. e., 2) that just states that the speaker does not know whether or not ϕ holds. This so-called *weak implicature* was derived from the utterance ψ by following the Gricean quantity maxim. The *strong* or *secondary implicature* $Bel_S(\neg\phi)$ strengthens the weak implicature. It can be derived by considering the speaker's competence: Either the speaker believes that ϕ holds or that it doesn't hold. For further details on the relation between implicatures and the competence assumption, see Geurts (2010) and Sauerland (2004).

The competence assumption thus is a crucial component of implicature calculation, because it prevents scalar expressions to be inferred with the stronger implicature reading by default. For the competence assumption to work, certain requirements must be met. Consider the following example:

S: I've heard a lecture at ComputerCon about that generation of graphics coprocessors. The HP Laptop has a good AMD Radeon graphics coprocessor.

In this example, the statement that the agent had heard a lecture on graphics coprocessors suggests that S is at least competent with respect to these processors. Hence, the user will infer without doubt that the graphics coprocessor in question is not excellent, which might affect his purchase decision more strongly than the weak implicature that the system just doesn't know whether the coprocessor is excellent or not.

Thus, assuming competence for implicature calculation might have severe effects on the course of the conversation and its outcome. But does the additional information really change the user's assumptions about the system's reliability? Does the sentence *trigger* a competence assumption for the user and thus a stronger implicature reading?

The aim of this paper is to explore this issue in the context of sales and recommender systems: We are interested in positive, neutral, and negative competence triggers and their influence on the competence assumption. Furthermore, we want to know whether the CA should be triggered locally (i. e., for single assertions or themes) or globally (i. e., for the overall dialogue).

2 Related Work

Although there are a number of dialogue systems that deal with various aspects of implicature calculation, the role of the CA in legitimating implicature calculation has not been accounted for. Artificial agents have been utilized before,

for either examining various pragmatic reasoning phenomena or maximizing dialogue efficiency, or both. For example, Vogel et al. (2013) use artificial agents to show that they behave in a Gricean manner to maximize their joint utility when faced with reference games or other interactional scenarios. The agent's reasoning about the opponent's belief states, modeled as a variant of the partially observable Markov decision process (POMDPs), to maximize joint utility, results in implicature-rich readings, but the weak–strong distinction has not been accounted for.

Stevens et al. (2016) show that sales dialogue efficiency can be enhanced with pragmatic question answering with indirect answers and consideration of user's requirements by using a game-theoretic model of query answering for their agent: However, the implicature triggered by the indirect answer is, due to the probabilistic model of user types, a weak one only.

Schlöder and Fernández (2015) develop a model for pragmatic rejection by means of implicature(s). Efstathiou and Lemon (2014) consider an account on non-cooperative dialogue in automated conversational systems and teach their agents to behave in this manner. It was shown that in a trading game, non-cooperative behaviour such as deception could increase the agents performance in comparison to a cooperative agent. The CA does not play a role in these models as well.

Insights on the CA originate from linguistic and philosophical analyses (e.g., Geurts (2010)), but these works do not consider requirements for developing computational systems with a generation component in a dialogue setting.

3 Testing the Competence Assumption Locally and Globally

Analogous to the situation in dialogue system research, the CA has not yet been a topic of empirical studies. The pragmatic approach in linguistics assumes that pragmatic reasoning is necessarily global (Sauerland 2004, p. 40), with which it refers to entire speech acts, not embedded sentences. In the context of this paper, *global* refers to an entire conversation, whereas *local* refers to a single speech act, primarily an assertion. It is not yet known on which level the CA is determined or whether it takes into account both levels of interpretation.

In our study, we confine ourselves to the question how the CA may be triggered and whether it will be established globally or locally. We consider the surface forms of the following aspects:

- politeness forms
- personally given indication of competence through additional information
- professionally induced indication of competence through additional information.

3.1 Participants

51 participants were consulted via <http://clickworker.com/> and the experiment was distributed by <https://www.testable.org/>, an open platform for web

experiments. Participation was limited to users from the US and UK. 15 participants were excluded due to failed attention checks, additional 9 participants were excluded, because they had response times below 2500 ms for more than one item. The threshold was set to 2500 ms as a result of a small pretest with three students, where response time minimum was 2900 ms. We assumed here that *clickworkers* are more familiar with the task at hand and that this would justify a lower response time minimum for them.

Thus included in the results are 24 participants.

3.2 Materials

The used items were priorly assigned to one of the categories “positive”, “neutral” or “negative”.

The examples were considered to be successful competence triggers if they achieved significant ratings within the spectrum of positive (100–70%), neutral (60–40%) or negative (30–0%) competence and in accordance to the priorly assigned categories. The examples were obtained through introspection, prior sales dialogue experiences – online and offline – and media research.

27 statements equally distributed within categories were tested. Of 27 statements, 9 were globally constructed competence triggers like short personal introductions and 18 statements were locally constructed competence triggers like mid-conversational sequences. For all statements, see the appendix; specific statements will hereafter be referred to with their item number (Table 1).

Table 1. Examples for global and local competence triggers

Global competence trigger:	
S:	<i>My name is Mr. Miller. My shift ends in 5 min, but I'm sure we'll find something for you.</i>
Local competence trigger:	
S:	<i>I have to admit I don't know a lot about that.</i>

The statements and two attention checks were randomized for each participant, which were then asked to “Rate the competence of the sales person on a scale from 0 (not competent) to 100 (very competent).” Answer was given for each statement with a slider from 0 to 100. Other specifics of the sliders grid were hidden from the participants.

3.3 Discussion and Results

First of all, we compared the mean ratings of all items from both groups (local and global) with their priorly assigned categories. As shown in Fig. 1, mean values of positive and neutral items met their categories prerequisites, whereas negative items did not.

We then proceeded to compare the participants ratings for global versus local items with their assigned categories:

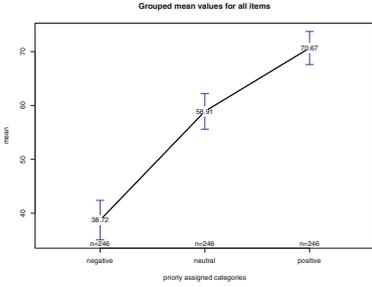


Fig. 1. Items by categories

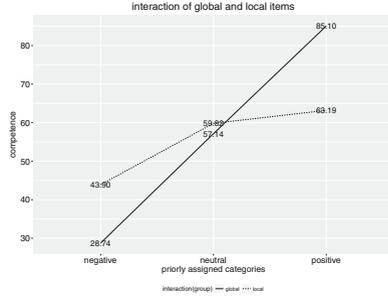


Fig. 2. Comparison of local and global triggers

As shown, global competence triggers are within their priorly assigned categories, whereas local competence triggers of the categories *positive* and *negative* are not. The results suggest that there might be a significant difference between positive-local and positive-global, negative-local and negative-global items.

The factorial ANOVA allows us to apprehend the effect of groups local and global and categories positive, neutral and negative on the competence ratings simultaneously.

The results of the ANOVA (Table 2) suggest that there is a significant difference within items and in the interaction between items and groups. No significant difference occurs within the groups local and global.

Table 2. Factorial ANOVA results

Analysis of variance table	
Response: competence	
	Pr(>F)
Item	<2.2e-16 ***
Group	0.4973
Item:group	3.856e-13 ***
Signif. codes: 0 *** 0.001 **	

This is also supported by the post-hoc Tukey analysis, which gives the analysis of significant differences of different groups and categories compared to each other (e.g., positive-local versus positive-global).

The analysis shows that the following categories are significantly different ($p < .05$):

- neutral-negative
- positive-negative
- positive-neutral.

Furthermore, the following groups and categories are not significantly different:

- neutral:local-neutral:global ($p = 0.972$)
- positive:local-neutral:global ($p = 0.50$)
- positive:local-neutral:local ($p = 0.848$).

This confirms what we have seen before in Fig. 2. Ignoring the groups local and global the categories show a significant difference towards each other. When comparing the interaction between groups and categories, it seems that above all positive-local items and neutral items struggled to differentiate themselves. Also, the findings confirm that negative-local and negative-global items as well as positive-local and positive-global items are significantly different.

4 Implications for Future Work

The results from this study on the role of the CA in dialogue systems suggest that the global competence assumption can be established fairly well. This holds especially true for the competence triggers of the category *positive* and *negative*, which had the most extreme competence ratings. In these categories, polite behaviour and indication of competence with mention of a certain position received the best mean ratings. The negative competence trigger with the lowest mean item number (8) worked with both attributes as well, whereas items number (7) and (9) – direct admission of incompetence and lack of time – didn't compete. Future work thus should pay attention to the factor of polite behaviour and indication of competence. In terms of negative competence triggers, competence should be indicated in an indirect manner, as direct admission of incompetence did not score as well in the mean values.

Local competence triggers of positive-local and neutral-local, as well as neutral-global items failed to distinguish themselves. It should be considered whether neutral competence triggers are beneficial in the first place, or if they can more easily be computed by a lack of competence triggers. Furthermore, local competence triggers did not score as well as their global counterparts. Hence, for computing local competence a more blunt use of polite behaviours or competence indicators should be reviewed.

The global items of the categories *positive* and *negative* were the most promising and therefore suggest that it would be beneficial to integrate a model of CA into the information states of a dialogue system. These first results of our ongoing study on competence triggers will be used for subsequent studies on the role of the CA for deriving weak and strong implicatures. On the whole, they will provide the empirical grounding for a probabilistic model of content determination in a question-answering system in a sales scenario.

Appendix

global-positive: (1) My name is Mr. Miller, I am the assistant Head of the Computer Department at this store. (2) My name is Mr. Miller, I am the designated laptop expert of this store. (3) My name is Mr. Miller and I am here to help you with whatever question you might have.

global-neutral: (4) My name is Mr. Miller. (5) My name is Mrs. Miller. (6) My name is Mrs. Kowalsky.

global-negative: (7) My name is Mr. Miller. Unfortunately I have to inform you, that this is not my department, I am covering this shift for a colleague. (8) My name is Mr. Miller. This is my first day at this store. So please; cut me some slack. (9) My name is Mr. Miller. My shift ends in 5 min, but I'm sure we'll find something for you.

local-positive: (10) Oh, that is easy: (11) I heard a lecture about this at ComputerCon. (12) I'm glad you ask. (13) When I was upgrading my home PC, I came across some information about this: (14) I recently read an article about that. (15) Ah, that is a fascinating piece of hardware.

local-negative: (16) Let me look that up for a second. (17) I have to admit I don't know a lot about that. (18) Sadly, this is not my field of expertise. (19) Well, I don't know much about those to be honest, but what I can tell you is that: (20) I don't think this makes much of a difference, but: (21) I'll have to check in with my colleague about that.

local-neutral: (22) Interesting question. (23) Of course. (24) Well, in regards to this question, I'd have to say: (25) The specifics to that are: (26) As far as I am concerned: (27) On behalf of that, it turns out that.

References

- Efstathiou, I., Lemon, O.: Learning non-cooperative dialogue behaviours. In: Proceedings of the SIGDIAL 2014 Conference, pp. 60–68. Association for Computational Linguistics (2014)
- Geurts, B.: Quantity Implicatures. Cambridge University Press, New York, NY (2010)
- Potts, C.: Presupposition and implicature. In: Lappin, S., Fox, C. (eds.) *The Handbook of Contemporary Semantic Theory*, 2nd edn, pp. 168–202. Wiley-Blackwell, Oxford (2015)
- Sauerland, U.: Scalar implicatures in complex sentences. *Linguist. Philos.* **27**(3), 367–391 (2004)
- Schlöder, J.J., Fernández, R.: Pragmatic rejection. In: Proceedings of the 11th International Conference on Computational Semantics, pp. 250–260. Association for Computational Linguistics (2015)

- Stevens, J.S., Benz, A., Reuße, S., Klabunde, R.: Pragmatic question answering: a game-theoretic approach. *Data Knowl. Eng.* **106**, 52–69 (2016)
- Vogel, A., Potts, C., Jurafsky, D.: Implicatures and nested beliefs in approximate decentralized-pomps. In: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pp. 74–80. Association for Computational Linguistics (2013)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

