In Context: Integrating Domain- and Situation-specific Knowledge

Iryna Gurevych*, Robert Porzel, and Rainer Malaka

European Media Laboratory, GmbH
Schloss-Wolfsbrunnenweg 33
D-69118 Heidelberg, Germany
robert.porzel|rainer.malaka@eml-d.villa-bosch.de
iryna.gurevych@eml-r.villa-bosch.de
* Current affiliation EML Research, gGmbH.

Summary. We describe the role of context models in natural language processing systems and their implementation and evaluation in the SmartKom system. We show that contextual knowledge is needed for an ensemble of tasks, such as lexical and pragmatic disambiguation, decontextualizion of domain and common-sense knowledge that was left implicit by the user and for estimating an overall coherence score that is used in intention recognition. As the successful evaluations show, the implemented context model enables a multi-context system, such as *SmartKom*, to respond felicitously to contextually underspecified questions. This ability constitutes an important step towards making dialogue systems more intuitively usable and conversational without loosing their reliability and robustness.

1 Introduction

The human enterprise of answering or responding to conversational speech input in a suitable and felicitous manner is not imaginable without three essential features:

- the ability to recognize what was said by the questioner;
- the ability to infer information that is left implicit by the questioner;
- the ability to infer what constitutes a useful and felicitous answer.

The realization of such abilities poses a formidable challenge in the development of conversational intuitive dialogue systems with more than one domain, modality, or situational context. The SmartKom system has to deal with contextual dependencies as well as cross-modal references based on the system's symmetric multimodality [46], it can handle multiple requests in different domain contexts and features special scenario-specific situational contexts. Thus, *decontextualization* is needed to resolve the arising contextual ambiguities [29, 30]. In the case of restricted and controlled single domain systems, the problem of contextually implicit information can be solved by generating full paraphrases out of the underspecified user utterances [11]. In systems with multiple contexts additional knowledge sources and dynamic context modeling is needed. Herein we describe the central contextual processing unit which combines ontological and situative knowledge. In the SmartKom system, discourse contextual influences are handled by unification-based operations such as *overlay* [2, 27], which operate on the schemas automatically created from the SmartKom ontology [22, 39], and interact closely with the contextual processing module described herein, e.g. for the resolution of deictic expressions.

This integration of basic common-sense domain knowledge with situative context knowledge constitutes a necessary building block for scalable natural language understanding systems that facilitate felicitous cooperation and intuitive access to web-based, location-based and personal information. In single context systems, such as train schedule or help desk systems [5, 19], this does not constitute a problem, since conversational phenomena such as pragmatic and semantic ambiguities or indirect speech acts do not occur [36]. A multi-domain, multi-scenario and multi-modal system faces diverse usage contexts, (e.g., at home or in mobile scenarios), conversational phenomena (e.g., indirect speech acts and pragmatic ambiguities) and multiple cross-modal references (e.g., gestural and linguistic discourse objects). A comprehensive understanding of naturally occurring discourse and of the often implicit questions embedded therein still has many unsolved issues in computational linguistics. In this work, we describe research on context and knowledge modeling components that enable dialogue systems with multiple contexts to realize the needed capabilities outlined above.

For example, in most conversational settings passerby's responses to a question such as:

(1) Excuse me, how do I get to the castle?

will most likely not be followed by asking where and when the spatial instructions should start. More likely, directions will be given. But, as the collected field data (see Section 5.1) shows, the felicity of spatial instructions is also dependent on contextual factors such as distance, mobility of the questioner or weather. Information concerning time or place, for example, is rarely explicated when given *default* settings, based on *common ground* [26] hold. If not, however, such information is very likely to be expressed explicitely. In some cases, which are commonly labeled as *indirect speech acts* or *pragmatic ambiguities*, however, we are not only faced with implicit information, but also with implicit intentions.

It is, however, possible to resolve the ensuing ambiguities and determine appropriate default settings using additional context, dialogue and system knowledge. We will show how such knowledge can be based on collected data relevant to the domains and situations at hand. Next to the *Wizard-of-Oz* data collections and data collected in evaluations, based on the PROMISE framework [7], we included existing lexicographic and ontological analyses, e.g, a model of frame semantics [6] as well as an ontological top-level [44], and conducted additional experiments and data collections in *Hidden-Operator* experiments [42] and the newly developed *Wizard-and-Operator* paradigm [17, 35]. This collected, analyzed and modeled information, then, became part of the ontological domain knowledge, i.e., the hierarchies, relations and cardinalities modeled therein.

Ontologies have traditionally been used to represent domain knowledge and are employed for various linguistic tasks, e.g., semantic interpretation, anaphora, or metonymy resolution. In our case, the aggregate model of situative and domain knowledge contains the SmartKom ontology [22, 39, 21]. As follows from interfacing with automatic speech, gesture and emotional recognition systems, a significant amount of uncertainty is involved, which is probably best reflected in the ensuing intention lattices and their confidence scores. Whether one looks at intention-, word lattices or n-best lists of hypotheses the problem of facing several different representations of a single utterance arises. This remains even though multi-modal systems can use the individual modalities to disambiguate each other. Different hypotheses of what the user actually might have said, of course, lead to a different understanding and in consequence to potentially different requests to the background system. The role of the context model in this light is to assist in evaluating the competing intention hypotheses against each other to find out what was said. Then, such contextual domain and situational knowledge can be used for augmenting such intention hypotheses with implicit information, to spell out their underlying intentions and, finally, to define a common background representation for the processed content, i.e., intention lattices in the case of the SmartKom system. Summarizing, a context model, therefore, can be employed in the following tasks:

- The explication of situationally implicit information.

This task can be further differentiated into two sub-tasks:

- provision of information that is indexical - such as time and place - based on common ground and -sense defaults and their dynamic instances, e.g. the current position of the user;
- provision of information that is pragmatic - such as speech acts and intentions and their dynamic instances, e.g. the actual open or closed state (accessibility) of the goal object.

- The scoring of individual interpretations in terms of their contextual coherence.

Again, this task can be further differentiated into two sub-tasks:

- using the ontological domain context to measure the semantic coherence of the individual interpretations, e.g. the ranking n-best lists or semantic interpretations thereof;
- using dynamic situational and discourse information, e.g. previous ontological contexts of prior turns.

Additionally, we can use the ontological knowledge modeled therein as the basis for defining the semantics and content of the information exchanged between the modules of multi-modal technology systems, as described by [22, 21].

After an overview of the state of the art of dialogue systems in the light of their domain- and context-specificity, we discuss the nature of domain- and situation models and their role in multi-domain, -scenario and -modal dialogue systems. Finally, we describe the architecture and processing of the context-model in the SmartKom system [46].

2 Contextual Processing in Dialogue Systems

Earlier approaches to handle conversational natural language input produced only toy systems. Their respective aims were to cope with special linguistic problems and/or to model particular cognitive capacities of language users. Broad coverage of constructions, lexical information sources and semantic/pragmatic behaviors was not the primary concern and also far outside the scope and capabilities of these systems. Today's linguistic development environments, representations and methodologies have shown that approximately complete coverage may be achieved in the areas of morphology and computational grammar. Furthermore, large lexical resources have been made available for linguistic applications in the area of parsing and also for natural language interfaces to application areas with restricted domains. Even though a broad coverage of frame semantic specification is still in the annotation progress [6], the handling of lexical semantics is still not set in stone [3, 18] and formal methods for dealing with pragmatic factors are in their beginning stages [8, 40, 36] - systems are in development that can offer suitable natural language interfaces both on the reception and the production side. These systems can be (and are) employed in domain-specific applications or demonstrators where they are commonly linked to non-linguistic applications (called the *background system*) such as databases [15], geographic information systems [28, 25], task planning systems [4, 13] or customer service systems [19].

Some open issues in handling the multi-domain problem are successfully beginning to be handled in the question-answering arena, by improved question parsing techniques coupled with more knowledge-based information understanding methods [24, 41, 31]. These information retrieval solutions assume more traditional desktop scenarios and more or less homogeneous content bases. While this is a reasonable thing to do for the type of information retrieval tasks with which the respective systems have to deal, conversational dialogue systems are faced with additional complications, that are added to the general open domain problem. Next to the spontaneous speech recognition input, additional factors are the changing context/situation of a mobile user and system on the one hand and the multitude of heterogeneous content bases that are needed to handle the topical informational need of a mobile user (e.g. a tourist) on the other. The content sources encompass, for example, rapidly changing online cinema information services, electronic program guides or hotel reservation systems, slower changing remote geographic information systems or relatively stable historical and architectural databases.

Natural language understanding in the area of parsing and pragmatically understanding questions as well as in terms of extracting their underlying *intentions* and finding suitable and felicitous answers is far from being solved. Still a variety of robust parsers can deliver valuable contributions beyond part of speech and treebank tags [34, 9, 16, 12].

The fact that multi-domain, -scenario and -modal conversational dialogue systems have so far been non-existent in the real word, is in part due to the fact that in all areas of NLP we face a mixture of context-variant and context-invariant factors that come into play at every level of natural language processing pipeline, e.g. speech recognition, semantic disambiguation, anaphoric resolution, parsing or generation. Ensembles of techniques and experiments are, therefore, needed to identify whether a factor is context-variant or not, and to identify specific types of contextual settings on which a given context-variant factor depends. Based on these findings and their application, decontextualization can be performed based on the contextual knowledge extracted, learned and modeled from the collected data. The results of such decontextualizations, e.g., for semantic disambiguation, then, in turn, can be evaluated using existing evaluation frameworks [47, 7].

(2) Where is the cinema Europa?

In real situations seemingly ‘simple’ questions such as (2) are difficult to understand in the absence of context. A felicitous answer often depends on the situational context in which the dialogue takes place. That is, as the data collected and analyzed shows [36], *where is* questions are either answered by localizations - if the reference object happens to be closed or with an instruction - if the reference object, e.g, a cinema or store, is open. In such cases of pragmatic ambiguity, the model resulting from the analyses of the corresponding data has to embed the utterance at hand into a greater situational context, e.g. by computing a contextual coherence score for the competing interpretations. The situation model consequently has to monitor the corresponding situational factors relevant to resolving such pragmatic ambiguities. Additionally, these situational observations are also needed for the resolution of indexicals, e.g. in the case of spatial- or temporal deixis [37].

The contextual coherence computations that are needed for decontextualization have to be able to deal with a variety of cases:

- For example, if decisions hinge on a number of contextual features, e.g. the situational accessibility of referenced objects [37], or domain-specific and pragmatic factors based on relations between referenced objects, as found in metonomyzation [32]. Here both ontological factors as well as situational factors come into play, e.g., semantic roles, weather, discourse factors, e.g., referential status, as well as user-related factors, e.g., tourists or business travelers as questioners and their time constraints.
- Additionally, if decisions hinge the contextual coherence of sets of concepts and their relations by applying both dialogical as well as semantic coherence measurements [20, 38], e.g. for ranking speech recognition hypotheses or semantic ambiguities.

3 Context Modeling

Utterances in dialogues, whether in human-human interaction or human-computer interaction, occur in a specific situation that is composed of different types of contexts. A broad categorization of the types of context relevant to spoken dialogue systems, their content and respective knowledge stores is given in Table 1.

types of context	content	knowledge store
dialogical context	what has been said by whom	dialogue model
ontological context	world/conceptual knowledge	domain model
situational context	time, place, etc	situation model
interlocutionary context	properties of the interlocutors	user model

Table 1. Contexts, content and knowledge sources

Following the common distinction between linguistic and extra-linguistic context[1] our first category, i.e. the dialogical context, constitutes the linguistic context, encompassing both co-text as well as intertext [8]. In linguistics the study of the relations between linguistic phenomena and aspects of the context of language use is called *pragmatics*. Any theoretical or computational model dealing with reference resolution, e.g. anaphora- or bridging resolution, spatial- or temporal deixis, or non-literal meanings, requires taking the properties of the context into account.

As knowledge sources in dialogue systems domain models are regarded to "hold knowledge of the world that is talked about" [14]. Following this general

[1] All extra-linguistic contexts are also often referred to as the *situational context* [10], however, we adopt a finer categorization thereof.

definition comes the observation that: "Information from the domain model is primarily used to guide the semantic interpretation of user's utterances; to find the relevant items and relations that are discussed, to supply default values, etc. The knowledge represented in a domain model is often coupled to the background system, e.g. a database system ... the domain knowledge is used to map information in the user's utterance to concepts suitable for database search" (ibid). We propose a different definition of the role of domain models in NLP systems, such as SmartKom. In our minds the knowledge contained in a domain model is to be modeled as an ontology proper, i.e., independent from the way an utterance or query is processed by the background system, e.g., the knowledge about (going to) cinemas, (seeing) movies and (getting) tickets and its representation is the same whether the background system is a specific database, a set of web-spidering agents or a combination thereof.

Statistical models based on specific corpora can serve to define *context groups* [48] and allow to differentiate between sets of distinct domain contexts that feature respective sense- and co-occurrence distributions. In our terminology, this formal context group function outputs a domain, i.e. the real world utterance-based linguistic target of our definition. It is important to note that despite the multitude of domains that are to be encompassed by the SmartKom system, the central aim is to create a kernel NLP system capable of dealing with multiple and extensible domains, which ultimately can be added to the system during runtime [43].

One of the central ideas embedded within the SmartKom research framework is to develop a kernel NLP system that can be used in a variety of situations, i.e. scenarios, domains and modalities, cf. [45], whereby:

- *scenarios* refer to different manifestations of the system, i.e. a home, office and public (booth) manifestation as well as a mobile one, and
- *modality* refers to speech, gesture, mimics, affectives and biometry
- *domain* refers to the general backdrop against which dialogues can be pitted, i.e., areas such as train schedules, movie information or hotel reservations.

These additional scenario-specific contexts feature:

- dynamic mobility of the user - where traditional input modalities, such as keyboard- and mouse-based input, are highly unsuitable;
- prolonged dialogues throughout sometime hour-long spatial navigation task; and
- context-dependent intentions.

Therefore, dynamic, e.g. situational context information has to be integrated together with the domain knowledge.[2]

[2] As noted in [36] current natural language understanding systems need a systematic way of including situational factors, e.g., the actual accessibility of goal objects has been shown to be a deciding contextual factor determining whether a

Speakers may not always be aware of the potential ambiguities inherent in their utterances. They leave it to the context to disambiguate and specify the message. Furthermore, they trust in the addressee's ability to extract that meaning from the utterance that they wanted to convey. In order to interpret the utterance correctly, the addressee must employ several context-dependent resources. Speakers in turn anticipate the employment of these interpretative resources by the hearer and construct the utterance knowing that certain underspecifications are possible since the hearer can infer the missing information. In the same way certain ambiguities become permissible due to shared common ground [26]. The role of the interlocutionary context is, therefore, also of importance in this process.[3]

4 Decontextualization in SmartKom

In line with our proposal stated above to separate domain- and application knowledge the implementation within the SmartKom system exhibits a clear distinction between domain-specific knowledge and application-specific knowledge. This is consequently mirrored by respective modules: the domain and situation model (each can be addressed via separate blackboards/communication pools) implemented in a module called *modeler.knowledge* and the function model implemented in a module called *modeler.function*[4].

4.1 The Modeler Knowledge Module

The running module for situational and ontological knowledge receives dynamic spatio-temporal information, e.g., GPS coordinates and local times as well as (multiple) representations of user utterances in form of *intention hypothesis* as input. It converts the incoming documents into document object models, on which it operates[5]). After processing, a decontextualized *intention hypothesis* document is returned as output.[6]

given `Where interrogative` at hand is *construed* as an instructional or a descriptive request.

[3] Since it is assumed in the SmartKom context, that general user model information is supplied via external sources, e.g., via a user's *SmartCard*, only the interaction preferences of the users are monitored actively by the system.

[4] This module can be described as the module that contains the knowledge of how specific plans are realized (given the actual software agents, databases and hardware devices). It therefore can be regarded as a translator between representations coming from the NLP and knowledge system and those of the background system.

[5] See www.w3c.org/DOM for the specification.

[6] The modeler.knowledge module features additional task- and domain-independent functionalities to probe and manipulate and compute on the ontology (www.w3c.org/OWL) as well as on the schema hierarchy (www.w3c.org/XMLS), the dynamic respective situational data and the static database information.

The context-dependent tasks performed by the context module implemented in SmartKom are:

- to know which information is ontologically required and provide the adequate situational and ontological information;
- to detect situationally appropriate readings; and
- to compute contextual coherence scores for alternative intention hypotheses.

4.2 Modeler Knowledge at Work

The first and foremost function is to add situation-specific discourse and dialog knowledge. For example, no agent can check room vacancies without knowing the arrival date and duration of the intended stay, neither can a theater agent reserve tickets without knowing the seat numbers, etc. In human-human dialogues this knowledge is responsible for determining relevant answers to given questions. Consider the following exchange given in Examples (3) and (4), where additional turns, asking the user to specify time and place, are avoided by decontextualizing the question and providing corresponding answers.

(3)	*Was*	*läuft*	*im*	*Kino*
	What	runs	in the	cinema

(4)	*Hier*	*sehen*	*Sie*	*was*	*heute*	*in*	*den*	*Heidelberger*	*Kinos*	*läuft*
	Here	see	you	what	today	in	the	Heidelberg	Cinema	runs

The SmartKom context model enables the system to act analogously, i.e. to provide - hitherto implicit - knowledge concerning what is talked about. The simplified structures given in Table 4.2 show insertions (in bold face) into an intention hypothesis made by the model in the case of a question such as given in Example (3). In this case the insertions made via contextual knowledge are threefold:

- For the cardinally required time and place slot in the performance object respective defaults are inserted.
- These indexical defaults are contextually resolved[7], by means of accessing a global position system (GPS). This information that can be used to resolve `here` with an appropriate level of granularity, e.g. town or spatial points, by means of a geographic information system in much the same way as `today` will also be replaced with granularity-specific temporal information, e.g. year, date or time.

[7]For example, the topical resolution of *here* and *now* - enable the system to produce a suitable response, such as retrieving a map of the cinemas of Heidelberg and the specific performances. Therefore `here` and `today` constitute *placeholders* for defaults that are replaced almost immediately with actual values by the situation model or discourse model.

```
<informationSearchProcess>
  <entertainment>
    <performance>
      <cinema>
        <contact>
          <address>
            <town>
              here
            </town>
          </address>
        </contact>
      </cinema>
      <time>
        <beginTime>
          <at>
            now
          </at>
        </beginTime>
      </time>
    </performance>
  </entertainment>
</informationSearchProcess>
```

```
<contact>
  <x> 70.345 < /x>
  <y> 49.822 < /y>
  <town>
    Heidelberg
  </town>
</contact>
<time>
  <at> 19:00:00T26:08:03 </at>
</time>
```

```
<scores>
  <contextualCoherence>
    0.46
  </contextualCoherence>
</scores>
```

Table 2. Contex-specific insertions into a sample intention hypothesis resulting from the intepretation of a speech recognition hypothesis

- A contextual score for each hypothesis is computed indicating the contextually most adequate reading, as SmartKom processes *intention lattices* consisting of several intention hypotheses.

By means of explicating such information and providing topical and contextually adequate values, the system can retrieve appropriate information from web sites or databases on what is currently playing in town, produce maps featuring cinema locations and then offer further assistance in navigation or reservation for example.

We have, therefore, linked the context model to interfaces providing contextual information. For example within both the SmartKom and the Deep Map framework [28], a database called the *Tourist-Heidelberg-Content Base* supplies information about individual objects including their opening and closing times[8]. By default, objects with no opening times, e.g. streets, can be considered always to be open. A global positioning system built into the mobile device supplies the current location of the user which is handed to the geo-

[8]Additional information extraction agents are able to gather data and information from the web, using ontological translators and updating the local database.

graphic information system that computes among other things the respective distances and routes to the specific objects. It is important to note that this type of context monitoring is a necessary prerequisite for context-dependent analysis.

5 Application and Evaluation in SmartKom

5.1 Data and Annotations

For demonstrating and realizing context-dependent effects in the SmartKom scenarios we collected two types of data. Firstly, we collected field data, by asking SmartKom-specific questions to pedestrians on the street and tracking the situational context factors and responses. The logged and classified field data was then used to train classifiers for recognizing specific intentions based on contextual factors. In a previous study another corpus of questions was collected and annotated in terms of their underlying intentions and turned into a gold standard [36]. Secondly, we collected laboratory data, i.e. dialogues in *Hidden-Operator*- and *Wizard-and-Operator* experiments [42, 17, 35]. All utterances were transcribed. Then specific sets of the audio files were sent to the speech recognizer. We logged the speech recognition hypothesis (SRH), n-best lists of SRHs and the module's in- and output for all utterances.

Using the laboratory data we created specific corpora for annotation experiments. In a first set of annotation experiments on a corpus of 1300 SRHs the SRHs were annotated within the discourse context, i.e. the SRHs were presented in their original dialogue order. For each SRH, a decision had to be made whether it is semantically coherent or incoherent with respect to the best SRH representing the previous user utterance. In a second experiment the annotators saw the SRHs together with the transcribed user utterances. The task of annotators was to determine the best SRH from the n-best list of SRHs corresponding to a single user utterance. The decision had to be made on how well the SRH expressed the intentional content of the utterance [38]. In the first experiment the inter-annotator agreement was 80% and in the second 90%. Lastly, the annotators had to create corresponding gold standards by means of discussing the cases of disagreement until an optimal solution was found.

5.2 The Evaluation in SmartKom

For evaluating the performance of the model describes above we computed the task-specific accuracies as compared to the gold standards described above. The situational models trained on the field data of 366 subjects using a c4.5 machine learning algorithm [49] achieved an intention recognition accuracy of 88% as compared to baseline achieved by a context-insensitve model of 59%

evaluated against the annotated gold standard of a corpus of dialogues with 50 subjects featuring various kinds of spatial interrogatives [36].

For evaluating the contextual coherence scores of the model we logged the scores of all scoring modules (speech recognizer, parser and discourse model) that rank n-best lists of speech recognition hypotheses produced out of word graphs [33] and those that rank the representations produced by the parser [12]. As described above these speech recognition hypotheses were annotated in terms of their coherence, correctness and *best-ness* and turned into corresponding gold standards [23, 38].

For computing contextual coherence in the evaluation the module employed three knowledge sources, an ontology of about 730 concepts and 200 relations and a lexicon (3.600 words) with word to concept mappings, covering the respective domains of the system and a conceptual dialogue history including the concepts and relations of the previous best hypothesis. The final evaluation of was carried out on a set of 95 dialogues. The resulting dataset contained 552 utterances resulting in 1.375 SRHs, corresponding to an average of 2.49 SRHs per user utterance.

The task of hypothesis verification, i.e., finding out what was said, in our multi-modal dialogue system is to determine the best SRH from the n-best list of SRHs corresponding to a given user utterance. The baseline for this evaluation was the overall chance likelihood of guessing the best one, i.e. 63.91%.

The context- and knowledge-based system [20, 38] achieves an accurracy of 88%[9] The knowledge-based system without the dialogical context features already exceeds that of the acoustic and language model scores produced by the automatic speech recognizer reaching 84.06% on the same task.

The evaluation of the contextual coherence scoring in terms of its disambiguation performance meant to calculate how often contextual coherence picks the appropriate reading - given an ambiguous lexicon entry such as *kommen* associated in the lexicon with both WatchPerceptualProcess and MotionDirectedTransliterated. For this evaluation we tagged 323 lemma with their contextually appropriate concept mappings and achieved an accuracy of 85% given an aggregate majority class baseline averaged over the majority class baselines of each individual lemma of 42%.

6 Conclusion

The basic intuition behind explicating contextual dependencies originally proposed by McCarthy [29] was that any given axiomatization of a state of affairs, meanings or relations presupposes an implicit context. Any explicit context model employed in processing information, therefore, needs to provide the

[9]This means that in 88% of all cases the best SRH defined by the human *gold standard* is among the best scored by the module.

information why specific meaning can be assigned to the underspecified information and, thusly, applied to its processing. This has often been called *fleshing out* and was considered impossible in its maximal form, e.g. Akman and Surav [1] state that:

> It is seen that for natural languages a fleshing-out strategy – converting everything into decontextualized eternal sentences – cannot be employed since we do not always have full and precise information about the relevant circumstances. (ibid:60)

Herein, we have presented a context model that performs a set of *fleshing out* tasks, which, as the successful evaluations show, suffice to enable a multi-context system, such as *SmartKom*, to respond felicitously to contextually underspecified questions. We have developed a corresponding system that integrates domain, dialogue and situative context in a multi-domain, -scenario and -modal dialogue system. We have shown how that:

- this knowledge can be used for improving the speech recognition reliability in the case of hypothesis verification, i.e., for finding out what was said;
- this knowledge can be used to explicate contextually implicit information, i.e., for resolving indexical expressions;
- this knowledge can be used to resolve context-dependent ambiguities, i.e. for lexical and pragmatic disambiguation.

We have, therefore, demonstrated that the inclusion of such contextual interpretation in natural language processing can enable natural language understanding systems to become more conversational without loosing the reliability of restricted dialogue systems. Given the challenge to extract the underlying intentions from conversational utterances such as "*is there a bakery close by*" or "*I don't see any bus stops*", we presented the necessary knowledge stores and inferential capabilites necessary for their decontextualization, which is a prerequisite for understanding utterances and responding felicitously. This enables us to restate McCarthy's original claim to say that, for natural languages a fleshing-out strategy can be employed if we have sufficient and precise knowledge about the relevant contextual circumstances.

Acknowledgments

This work was partially funded by the German Federal Ministry of Education, Science, Research and Technology (BMBF) in the framework of the SmartKom project under Grant 01 IL 905 K7 and by the Klaus Tschira Foundation. The authors would like to thank Ralf Panse, Christof Müller and Hans-Peter Zorn for their valuable implementation and evaluation work.

References

1. Varol Akman and Mehmet Surav. Steps toward formalizing context. *AI Magazine*, 17(3):55–72, 1996.
2. Jan Alexandersson and Tilman Becker. The Formal Foundations Underlying Overlay. In *Proceedings of the Fifth International Workshop on Computational Semantics (IWCS-5)*, pages 22–36, Tilburg, The Netherlands, February 2003.
3. James F. Allen. *Natural Language Understanding.* Menlo Park, Cal.: Benjamin Cummings, 1987.
4. James F. Allen, Lenhart K. Schubert, George Ferguson, Peter Heeman, Chung Hee Hwang, Tsuneaki Kato, Marc Light, Nathaniel Martin, Bradford Miller, Massimo Poesio, and David Traum. The TRAINS project: A case study in building a conversational agent. *Journal of Experimental and Theoretical AI*, 7:7–48, 1995.
5. Harald Aust, Martin Oerder, Frank Seide, and Volker Steinbiss. The Philips automatic train timetable information system. *Speech Communication*, 17:249–262, 1995.
6. Collin F. Baker, Charles J. Fillmore, and John B. Lowe. The Berkeley FrameNet Project. In *Proceedings of COLING-ACL*, pages 86–90, Montreal, Canada, 1998.
7. Nicole Beringer, Ute Kartal, Katerina Louka, Florian Schiel, and Uli Türk. PROMISE: A Procedure for Multimodal Interactive System Evaluation. In *Proceedings of the Workshop 'Multimodal Resources and Multimodal Systems Evaluation*, pages 77–80, Las Palmas, Spain, 2002.
8. Harry Bunt. Dialogue pragmatics and context specification. In H.C. Bunt and W.J. Black, editors, *Computational Pragmatics, Abduction, Belief and Context; Studies in Computational Pragmatics*, pages 81–150. John Benjamins, Amsterdam, 2000.
9. Michael J. Collins. A new statistical parser based on bigram lexical dependencies. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics,* Santa Cruz, Cal., pages 184–191, 1996.
10. John H. Connolly. Context in the study of human languages and computer programming languages: A comparison. *Modeling and Using Context*, 2116:116–128, 2001.
11. Christian Ebert, Shalom Lappin, Howard Gregory, and Nicolas Nicolov. Generating full paraphrases out of fragments in a dialogue interpretation system. In *Proceeding of the 2nd SIGdial Workshop on Discourse and Dialogue*, pages 58–67, Aalborg, Denmark, 2001.
12. Ralf Engel. SPIN: Language understanding for spoken dialogue systems using a production system approach. In *Proceedings of the International Conference on Speech and Language Processing 2002*, pages 2717–2720, Denver, USA, 2002.
13. George Ferguson and James F. Allen. TRIPS: An intelligent integrated problem-solving assistant. In *Proceedings of the 15th National Conference on Artificial Intelligence & 10th Conference on Innovative Applications of Artificial Intelligence,* Madison, Wisc., pages 567–573, 1998.
14. Annika Flycht-Erriksson. A survey of knowledge sources in dialogue systems. In *Proceedings of the IJCAI-99 Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, pages 41–48, 1999.
15. F. Gallwitz, M. Aretoulaki, M. Boros, J. Haas, S. Harbeck, R. Huber, H. Niemann, and E. Nöth. The Erlangen Spoken Dialogue System EVAR: A State–of–the–Art Information Retrieval System. In *Proceedings of 1998 International*

Symposium on Spoken Dialogue (ISSD 98), pages 19–26, Sydney, Australia, 1998.

16. Marsal Gavaldá and Alex Waibel. Growing semantic grammars. In *Proceedings of the 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics,* Montréal, Québec, Canada, 10–14 August 1998, pages 451–456, 1998.
17. David Gelbhart, John Bryant, Andreas Stolcke, Robert Porzel, Manja Baudis, and Nelson Morgan. SmartKom English: From robust recognition to felicitous interaction. In W. Wahlster, editor, *SmartKom - Foundations of Multimodal Dialogue Systems*. Springer, Berlin, 2004.
18. Yolanda Gil and Varun Ratnakar. A comparison of (semantic) markup languages. In *Proc.of the 15th Int.FLAIRS Conference*, pages 408–412, Florida, USA, 2002.
19. A. L. Gorin, G. Riccardi, and J. H. Wright. How may I help you? *Speech Communication*, 23(1/2):113–127, 1997.
20. Iryna Gurevych, Rainer Malaka, Robert Porzel, and Hans-Peter Zorn. Semantic coherence scoring using an ontology. In *Proceedings of HLT/NAACL*, pages 88–95, Edmonton, CN, 2003.
21. Iryna Gurevych and Robert Porzel. The SmartKom Ontology. In W. Wahlster, editor, *SmartKom - Foundations of Multimodal Dialogue Systems*. Springer, Berlin, 2004.
22. Iryna Gurevych, Robert Porzel, and Stefan Merten. Less is more: Using a single knowledge representation in dialogue systems. In *Proceedings of the HLT/NAACL Text Meaning Workshop*, pages 14–21, Edmonton, Canada, 2003.
23. Iryna Gurevych, Robert Porzel, and Michael Strube. Annotating the semantic consistency of speech recognition hypotheses. In *Proceedings of the Third SIGdial Workshop on Discourse and Dialogue*, pages 46–49, Philadelphia, USA, July 2002.
24. E. Hovy, L. Gerber, U. Hermjakob, C. Lin, and D. Ravichandran. Toward semantics-based answer pinpointing. In *Proceedings of the First International Conference on Human Language Technology*, 2001.
25. Michael Johnston, Srinivas Bangalore, Gunaranjan Vasireddy, Amanda Stent, Patrick Ehlen, Marilyn Walker, Steve Whittaker, and Preetam Maloor. Match: An architecture for multimodal dialogue systems. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 376–383, Philadelphia, Germany, 2002.
26. Robert Krauss. The role of the listener: Addressee influences on message formulation. *Journal of Language and Social Psychology*, 6:91–98, 1987.
27. Markus Loeckelt, Tilman Becker, Norbert Pfleger, and Jan Alexandersson. Making sense of partial. In *Proceedings of the 6th workshop on the semantics and pragmatics of dialogue*, pages 101–107, Edinburgh, Scotland, 2002.
28. Rainer Malaka and Alexander Zipf. Deep Map: Challenging IT research in the framework of a tourist information system. In *Proceedings of the International Conference on Information and Communication Technologies in Tourism, Barcelona, Spain*, pages 15–27, Berlin, Germany, 2000. Springer.
29. John McCarthy. Notes on formalizing contexts. In Tom Kehler and Stan Rosenschein, editors, *Proceedings of the Fifth National Conference on Artificial Intelligence*, pages 555–560, Los Altos, California, 1986. Morgan Kaufmann.

30. John McCarthy. Generality in artificial intelligence. In Vladimir Lifschitz, editor, *Formalizing Common Sense: Papers by John McCarthy*, pages 226–236. Ablex Publishing Corporation, Norwood, New Jersey, 1990.
31. Dan Moldovan, Marius Pasca, Sanda Harabagiu, and Mihai Surdeanu. Performance issues and error analysis in an open-domain question answering system. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 33–40, 2002.
32. Geoffrey Nunberg. *The Pragmatics of Reference*. Indiana Linguistics University Club, 1987.
33. Martin Oerder and Hermann Ney. Word graphs: An efficient interface between continuous-speech recognition and language understanding. In *ICASSP Volume 2*, pages 119–122, 1993.
34. R. Pieraccini, E. Tzoukermann, Z. Gorelov, J. Gauvainand E. Levin, C. Lee, and J. Wilpon. A speech understanding system based on statistical representation of semantics. In *Proceedings of the International Conference on Acoustics Speech Signal Processing*, San Francisco, USA, 1992.
35. Robert Porzel and Manja Baudis. The Tao of CHI: Towards fefilicitous human-computer interaction. In *Proceedings of HLT/NAACL*, page (in press), Boston, USE, 2004.
36. Robert Porzel and Iryna Gurevych. Towards context-adaptive utterance interpretation. In *Proceedings of the 3rd SIGdial Workshop on Discourse and Dialogue,* Philadelphia, USA, 2002.
37. Robert Porzel and Iryna Gurevych. Contextual coherence in natural language processing. In P. Blackburn, C. Ghidini, R. Turner, and F. Giunchiglia, editors, *Modeling and Using Context*. LNAI 2680, Springer, Berlin, 2003.
38. Robert Porzel and Iryna Gurevych. Ontology-based contextual coherence scoring. In *Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue,* Saporo, Japan, July 2003, 2003.
39. Robert Porzel, Norbert Pfleger, Stefan Merten, Markus Löckelt, Ralf Engel, Iryna Gurevych, and Jan Alexandersson. More on less: Further applications of ontologies in multi-modal dialogue systems. In *Proceedings of the 3rd IJCAI 2003 Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, Acapulco, Mexico, 2003.
40. Robert Porzel and Michael Strube. Towards context adaptive natrual language processing systems. In Manfred Klenner and Henriette Visser, editors, *Proceedings of the International Symposium: Computional Linguistics for the New Millenium*, pages 141–156. Peter Lang, Berlin, 2002.
41. J. Prager, D. Radev, and K. Czuba. Answering what-is questions by virtual annotations. In *Proceedings of the First International Conference on Human Language Technology*, 2001.
42. Stefan Rapp and Michael Strube. An iterative data collection approach for multimodal dialogue systems. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation*, Las Palmas, Spain, 2002.
43. Stefan Rapp, Sunna Torge, Silke Goronzy, and Ralf Kompe. Dynamic speech interfaces. In *Proceedings of the ECAI Workshop on Artificial Intelligence in Mobile Systems*, Berlin, Germany, 2000.
44. Stuart J. Russell and Peter Norvig. *Artificial Intelligence. A Modern Approach.* Prentice Hall, Englewood Cliffs, N.J., 1995.

45. Wolfgang Wahlster. SmartKom: Fusion and fission of speech, gerstures and facial expressions. In *Proceedings of the Firsat International Workshop on Man-Machine Symbiotic Systems*, Kyoto, Japan, 2002.
46. Wolfgang Wahlster. SmartKom: Symmetric multimodality in an adaptive an reusable dialog shell. In *Proceedings of the Human Computer Interaction Status Conference*, Berlin, Germany, 2003.
47. Marilyn A. Walker, Candace A. Kamm, and Diane J. Litman. Towards developing general model of usability with PARADISE. *Natural Language Engeneering*, 6, 2000.
48. Domonique Widdows. A mathematical model of context. In P. Blackburn, C. Ghidini, R. Turner, and F. Giunchiglia, editors, *Modeling and Using Context*. LNAI 2680, Springer, Berlin, 2003.
49. Patrick Henry Winston. *Artificial Intelligence*. Addison-Wesley, 1992.