A data-driven model of explanations for a chatbot that helps to practice conversation in a foreign language

Sviatlana Höhn
AI Minds
Vianden, Luxembourg
s.hoehn@aiminds.ai

Abstract

This article describes a model of other-initiated self-repair for a chatbot that helps to practice conversation in a foreign language. The model was developed using a corpus of instant messaging conversations between German native and non-native speakers. Conversation Analysis helped to create computational models from a small number of examples. The model has been validated in an AIML-based chatbot. Unlike typical retrieval-based dialogue systems, the explanations are generated at run-time from a linguistic database.

1 Introduction

Conversational agents tailored for communication with language learners are studied in the area of Communicative Intelligent Computer-Assisted Language Learning (CommICALL). Starting with the idea of creating a machine that behaves like a language expert in an informal chat, specific interactional practices need to be described where linguistic identities of interaction participants become visible. Such practices include repair with linguistic trouble source where non-native speakers address troubles in comprehension or production (Danilava et al., 2013).

Repair is a building block of conversation that helps to deal with troubles in understanding and production of talk. Depending on who produced a trouble source and who initiates a repair we distinguish between self-initiated and other-initiated repair. A repair can be carried out by the same speaker who produced the trouble source or by the other speaker (self-repair and other-repair).

Because there is a preference for self-repair, other-initiated self-repair is the most frequent repair type. It may become even more frequent in conversations where one of the speakers is more knowledgeable in some matters than the other, for instance in mastering professional terminology or communication in a second language not yet fully mastered. Therefore it is crucial for conversational agents acting in such environments to recognize and to handle repair initiations properly.

Repair sequences where the machine is the trouble-speaker are in focus of this article. The learner initiates a repair in response to something not (fully) understood, and the machine explains. This type of repair corresponds to *other-initiated self-repair with a linguistic trouble source where the language learner is the recipient of the trouble talk* ($OISR_L$).

CommICALL research is mainly grounded in Second Language Acquisition (SLA) theory (Petersen, 2010; Wilske, 2014). The model of explanation sequences, so called *negotiations of meaning* introduced by (Varonis and Gass, 1985) received a lot of attention and was highly re-used in subsequent CALL research (Fredriksson, 2012; Satomi Kawaguchi, 2012). The model includes a trigger, an indicator, a response and a reaction to response. However, this model has been criticized for its view on repair as something "marring the flow" of a conversation and for being inapplicable to non-institutional settings (Markee, 2000). Although repair in native/non-native speaker talk has been intensively studied in Conversation Analysis (CA) (Markee, 2000; Gardner and Wagner, 2004; Hosoda, 2006), the results have not been operationalized for an implementation in a CommICALL system. Therefore, this article has two objectives:

1. Identify typical interactional resources employed for initiation and carry-out of repair using methods of Conversation Analysis.

2. Create a computation models of the repair

of the type $OISR_L$ to be implemented in a CommICALL application.

We use a dataset of German native/non-native instant messaging conversations (Höhn, 2015) to analyze practices of repair in native/non-native speaker informal chat. All repair sequences have been annotated. Collections of similar cases have been built. Interactional resources used by language learners for repair initiations have been analyzed. Patterns of repair initiations have been obtained through generalization. In this way, rules for recognition of repair initiations have been created. An implementation case study was set up to validate the resulting computational models in an AIML-based chatbot.

2 Repair in Conversational Agents

Non-native speakers are usually not considered as the main user group of general-purpose dialogue systems. The assumption dominates that human users understand everything what an agent may say. This assumption is reflected in the two main problems addressed by research on repair for conversational agents: dealing with user's self-corrections which may make speech recognition difficult and managing system's lack of information in order to satisfy user's request.

These two research areas may be found under keywords *self-repairs*, sometimes *speech repairs* (Zwarts et al., 2010) or *disfluencies* (Shriberg, 1994; Martin and Jurafsky, 2009), and *clarification dialogues* or *clarification requests, CRs* in AI and NLP publications. What is referred to by the term *self-repair* in speech recognition domain corresponds to user's self-initiated self-repair in CA terminology.

Shriberg (1994) uses the term *reparandum* to refer to what is called *trouble source* in CA. The model considers pauses (moment of interruption) and lexicalised means to focus on the replacement (editing terms). These are interactional recourses used by speakers to signal trouble in production and to pre-announce a coming replacement.

The term *clarification dialogues* is mostly used to describe repairs dealing with insufficient information available for a system after speech recognition and language understanding (Kruijff et al., 2008; Jian et al., 2010; Buß and Schlangen, 2011). The term *miscommunication* was introduced to distinguish between *non-understandings* (the system could not match user's input to a representa-

tion) and *misunderstandings* (the system matched user's input to a wrong representation) (Dzikovska et al., 2009; Meena et al., 2015). These repair types correspond to other-initiated self-repair when the user is the trouble-speaker.

Clarification requests in AI and NLP publications should not be confused with clarification requests in SLA publications where this term is used to refer to only a particular form of corrective feedback (Lyster et al., 2013), or to a dialogue move in meaning negotiations (Varonis and Gass, 1985).

Emphasising the importance of correct recognition of user's clarification requests, Purver (2004) provides a study of various types of clarification requests, see also follow-up publications (Purver, 2006; Ginzburg et al., 2007; Ginzburg, 2012). Purver (2004) uses the HPSG framework to cover the main classes of the identified classification scheme. Because different functions might be expressed by a clarification request of the same form, Purver (2004) analyses the *clarification readings* to cover the correspondence between the form and the meaning of the repair initiations. However, several points for critiques arise. For instance, some utterances may be formatted as repair initiations but have a different interactional function, such as expressing surprise and topicalization (not listed as possible readings). In addition, repair initiations designed to deal with troubles in understanding are put together with strategies for dealing with troubles in production (e.g. *gap fillers*). From the CA perspective, Purver (2004)'s *gap fillers* correspond to self-initiated other-repair, thus are sequentially completely different. Therefore, modifications in the classification proposed by (Purver, 2004) are needed in order to better comply with studies in CA, and therefore better reflect the state-of-the-art in CA-informed dialogue research.

Example 2.1. Different types of causes for clarification used in (Schlangen, 2004, Ex. (12)).

a.	A	I ate a Pizza with chopsticks the other day
	B	A Pizza with chopsticks on it?
b.	A	Please give me a double torx.
	B	What's a torx?
c.	A	Please give me a double torx.
	B	Which one?
d.	A	Every wire has to be connected to a power source.
	B	Each to a different one, or can it be the same for every wire?

Schlangen (2004) analyses communication problems leading to clarification requests focusing on

trouble source types (what caused the communication problem). Schlangen (2004) makes clear that a more fine-grained classification of causes for requesting clarification in dialogue may be needed, specifically, a model distinguishing between different cases in Example 2.1.

From the CA perspective, speakers' linguistic and professional identities and preferences play a role in speaker's selection of a specific format of a repair initiation. Speaker B in Example 2.1.b. positions herself as a novice in torx matters with her repair initiation, while speakers B in Examples 2.1.c. positions herself as knowledgeable in torx matters. In addition, utterances may be designed as repair initiations, but may in fact have a different function. For instance, the repair initiation produced by B in Example 2.1.a. may be analysed as a joke not requiring any explanation.

Other-initiated self-repair when the machine is the trouble-speaker is explored in (Gehle et al., 2014). Based on a corpus of video-recorded human-robot-interactions in a museum, the authors analyse interactional resources used by museum visitors to signal troubles in understanding robot's talk and dealing with misunderstandings. It was observed that people deal with different sorts of trouble similarly.

The potential user of a CommICALL system is a language learner who may have troubles in comprehension. While user-initiated repair has been subject of research of studies in human-robot interaction and general dialogue systems, not much attention has been paid to it in CommICALL. This article seeks to contribute to the research on repair in CommICALL by a microanalytic study of sequences of other-initiated self-repair when the native speaker is the trouble-speaker. Based on the results of the empirical study, the problem of computational modeling of system's reaction to the learner's repair initiation will be approached. The machine will need to recognize repair initiations, to extract the trouble source and to deliver an appropriate response. The the study contributes to language understanding for dialogue systems targeting language learners and has implications for user and expert models for CommICALL.

3 Practices of repair in chat

This section analyses interactional resources used by the non-native speakers in chat in order to other-initiate repair with a linguistic trouble source, that is to signal trouble and to reference the trouble source. Turn formats are specifically important for the future recognition of repair initiations by chatbots.

3.1 Repair initiations

Two abstract types of repair other-initiations were identified in the dataset: *statements of non-understanding* where a part of partner's utterance is marked as unclear, and *candidate understandings* where the own version of understanding of the problematic unit is provided. Non-understandings require an explanation of the trouble source in the repair while candidate understandings require a yes/no answer.

Repair other-initiations were found at two distinct types of position: *immediate* and *delayed*. The first type comes immediately after the trouble source turn. The second type comes later than the adjacent turn. Sequentially, both correspond to the next-turn repair initiation or second position repair described in CA literature as the first structurally specified place for other-initiated repair (Schegloff, 2000; Liddicoat, 2011). Delayed repair initiations occur because speakers in chat can produce turns simultaneously and follow distinct interleaved conversation threads. There is a dependency between the position of the repair initiation and the interactional recourses for repair initiation. Some resources are used exclusively in the immediate position.

Example 3.1. Open class repair initiation

615 L08 danke. good night)
          thank you. good night
617 N04 gn8 :-)
618 L08 ???
          ??? [repair initiation]
619 N04 gn8 ist ein zusammengeschrumpftes "gute Nacht" (lies: "g" = "gut" und "n8" = "N-Acht")
          gn8 is an abbreviation of "good night" (read: "g"="good" and "n8" = "n-ight")
620 N04 oder englisch, g=good, n-eight
          or English, g=good, n-eight
621 L08 aach sooo))
          I see

In Example 3.1, the learner initiates a repair by posting three question marks directly after the trouble source turn. The native speaker N04 is able to locate the trouble source, which is the abbreviation. In Example 3.1, the reference to the trouble source is realised by the immediate adjacent position, and signaling trouble with comprehension is realised by the questions marks.

Candidate understanding is another possibility

to mark a unit of an utterance as not (completely) clear. Example 3.2 shows a fragment of a chat where the native speaker N04 uses the word *über-fülltes* to describe an event in Munich (turn 222). The learner L08 checks her understanding of this term in turn 223 by copying the trouble source and providing her own understanding of the word. The trouble source is referenced through its repetition in the repair initiation. Signalling trouble is realised through the comparison token, the candidate understanding and the question mark.

Example 3.2. Many many people

221	L08	ja ich habe über Oktoberfest gehört, etwas lustiges und buntes))
		yes I have heard about Oktoberfest, something funny and colourful
222	N04	ja, und teures und überfülltes ;-)
		yes, and expensive and overfilled
223	L08	))überfülltes bedeutet "viele viele Leute"?
		overfilled means "many many people"?
224	N04	genau
		exactly

The repair initiations produced by the learners in the dataset always try to resolve problems with the meaning, none of them was concerned with the form by itself.

3.2 Repair carry-out

Repair carry-out strategies depend on the type of the trouble source and the repair initiation format and include confirmations / disconfirmations, definition work and paraphrasing of the trouble source. Direct definition work can be replaced or extended by a hyperlink to an example or a demonstration of an instance of the trouble source.

If the trouble source is an abbreviation, the definition work contained a full spelling of the abbreviated words and their explanation. For chat abbreviations, a full reading of the abbreviation was normally provided and enough for explanation, as Example 3.1 demonstrates. Problematic abbreviation were always repeated in the dataset, followed by the full spelling or reading.

If the trouble source is one semantic unit (one word or an idiomatic expression), a dictionary-like definition (synonyms + examples) is often selected to provide a repair. For longer messages or longer parts of longer messages, a strategy of splitting the message into smaller semantic units and a separate explanation of each unit can be chosen. Paraphrasing is also one of the strategies used by the native speakers to explain longer messages.

Example 3.3 shows how a machine translation

service can be used for definition work. Turn 376 contains an expression that the learner does not (fully) understand: "in sachen essen". This expression is being formally made to a trouble source in the repair initiation in turns 377 and 378. Turn 377 locates the trouble source and marks the expression as unclear. Turn 378 contains an instruction of what kind of explanation is desired.

Example 3.3. In Sachen Essen: repair is carried out with the help of machine translation.

376	N03	gibt es irgendwas moskau typisches in sachen essen?
		is there something of food which is typical for moscow?
377	L07	in sachen essen???
		in things food???
378	L07	übersetze bitte)))
		translate please [smile]
379	N03	какая пища является типичным Москве?
		which food is typical for Moscow?

4 Empirical findings

Regarding repair initiations, it was found that:

(1) Questioning is *the* practice to initiate repair in chat, confirming the results in the academic literature for oral interaction (Dingemanse et al., 2014). Other practices are declarations of lack of understanding such as *unklar* and *ich verstehe nicht*.

(2) Devices for signalling are question marks, dashes, explicit statements of non-understanding and presenting candidate understandings.

(3) References to trouble sources may be realised through the adjacent position, demonstrative expressions and full or partial repeats.

(4) Though all repair initiations were second-position initiations, they were not all immediate. Delayed repair initiation require more specific referencing to trouble source, open-class repair initiations cannot be used in a delayed second position.

(5) Repetition-based repair initiations may contain repetitions of one specific unit from the previous turn and contain a copy of the preceding turn regardless the unit boundaries. The latter may be placed between open class and restricted class repair initiations. Such types of repetitions have not been previously described in the academic literature and may be typical for non-native speakers.

(6) The communication medium influences repair initiation types and formats. In particular, repair initiations eliciting a repetition of the trouble source are uncommon in chat. Misreadings are

possible, but they are made visible through mis-productions in repetition-based repair initiations.

(7) The non-native speakers' identity influences the format of candidate understandings which differ from those in native speaker talk.

(8) Repair initiation is one option to deal with trouble in comprehension. Other options include dictionary look-up and the "let-it-pass" strategy.

Regarding repair carry-outs, it was found that:

(1) Explanations of the meaning through synonyms or paraphrases, translations and demonstrations are common forms of repair carry-outs.

(2) Repair design is linked to expectation of what is known to the repair recipient. Consequently, repairs are designed for the language learners targeting difficulties in linguistic matters.

(3) Repair carry-outs may be immediate and delayed. Consequently, references to trouble source may be realised by the same resources as for repair initiations. However, there are dependencies between types of trouble source and participants' selection of resources for referencing the trouble source. For instance, abbreviations are usually repeated.

(4) *Split-repeat* is a type of a reference to the trouble source which did not appear in repair other-initiations but was found in the corresponding self-repair carry-outs. This way of referencing corresponds to self-repairs where native speakers only explained a few words from a longer turn or longer part of a turn marked as a trouble source. The trouble source was split in tokens, and only tokens that were supposed to cause the trouble were explained.

Repair carry-out is the preferred and the most frequent response to a repair initiation but other forms of responses are also possible, for instance a new repair initiation to deal with difficulties in identification of the trouble and responses which do not address the trouble. Finally, repair initiation and carry-out formats need to be "translated" into patterns and then into computational models of repair to make the findings applicable for computational purposes.

5 Computational model of $OISR_L$

In order to "serve computational interests" (Schegloff, 1996), the following needs to be taken into account for the purpose of modelling. Because repair initiations may occur *everywhere*, each user's utterance may be a repair initiation. Therefore, a repair initiation recognition routine needs to be activated after *every* user's turn. Two essential problems must be solved by a computer program in order to react to a repair initiation properly:
(1) Recognition of a repair initiation,
(2) Extraction of the trouble source.
A repair proper needs to be generated after that.

5.1 Recognition of repair initiations

Each class of repair initiations implies a specific form of referencing the trouble source. We consider the following types of referencing for modelling of the $OISR_L$-sequences:

1. Repeat-based initiations: *reuse* (a 1:1-copy of the trouble source), *recycle* the trouble source (rewriting it in a slightly different way),

2. Demonstratives-based initiations: using demonstrative determiners and pronouns.

3. Open-class initiations: referencing by a statement of non-understanding in the immediate position. The adjacent position of the repair initiation references the whole preceding turn as a trouble turn. Therefore we refer to this type of referencing as *reference by position*.

Each class of repair initiations references trouble of a particular size: either it is the whole preceding message (open-class and demonstratives-based repair initiations) or it is only a part of it (repeat-based and recycle-based initiations). Therefore, we consider three cases of trouble sources: single word (part of a longer message or a one-word message), part of a message (PoM) of two or more words and a whole message consisting of two or more words.

Signalling trouble involves symbolic and/or lexicalised means and a specific format designed either to mark something as unclear or to compare the trouble source with the own version of understanding. We call this *signalling format*.

The architecture of the repair initiation (RI) for $OISR_L$ can be formalised as follows. Depending on the time, different formats for the repair initiation may be used:

$RI = TIME \times RIFormat$

Time may be immediate or delayed: $TIME = \{immediate, \ delayed\}$. A repair initiation format is a combination of a reference to the trouble source and a selected signalling format:

$RIFormat = REF \times SignalFormat$

The referencing types are repeat-based $repeat(x)$, based on demonstratives Dem and reference by position AP. Signalling format may mark something in the trouble-turn as unclear $unclear(x)$ or present a candidate understanding $equals(x, y)$. The trouble source x and the candidate understanding y may be a single word, an idiomatic expression, part of a message or a complete turn (utterance).

$REF = \{repeat(x),\ AP,\ Dem\}$
$SignalFormat = \{unclear(x),\ equals(x,y)\}$
$x, y \in \{word,\ idiom,\ PoM,\ utterance\}$

This repair recognition procedure is also expected to differentiate between ordinary questions related to the subject of the ongoing talk and repair initiations. It works because ordinary questions are not formatted as $unclear(x)$ or $equals(x, y)$.

If a complete turn is recognised as a trouble source and this turn is a longer message, further filters may be applied to identify more precisely, which of the parts of the longer message may cause a problem with comprehension. This may be influenced by the learner model, but also by the system's capabilities to generate a repair proper. Section 5.3 will address this problem and provide examples of possible filters.

5.2 Generation of a repair carry-out

Repair carry-outs can contain a lexical reference to the trouble source, such as repeat-based and demonstratives-based references, or point to it just by the adjacent position to the repair initiation.

A confirmation or a disconfirmation is an appropriate type of self-repair carry-out after a repair other-initiation presenting candidate understandings $equals(x, y)$. All other self-repair carry-outs are expected to provide an explanation of the unit that is marked as problematic $explain(x)$. Because different options are available for referencing trouble source in immediate and delayed repair carry-outs, time needs to be taken into account in the abstract description:

$RCO = TIME \times RCOFormat$
$TIME = \{immediate,\ delayed\}$.

A self-repair carry-out is a product of a reference to the trouble source and the function $RCOF$, which it is expected to perform: confirming/disconfirming answer or an explanation.

$RCOFormat = REF \times RCOF$
$REF = \{repeat(x),\ AP,\ Det,\ splitRepeat(x)\}$
$RCOF = \{explain(x),\ conf(equals(x,y))\}$

Delayed self-repairs need to update the focus of the talk, and therefore, a repeat-based reference makes more sense than other types of referencing.

In practice, the function $explain(x)$ needs to be implemented differently for different types of trouble source. The quality of the response is highly dependent on the linguistic resources available for the generation of the explanations. We discuss various practical issues in the next section.

5.3 Model validation

The purpose of this section is to validate the practical applicability of the abstract model described in the preceding section. Because language understanding and generation capabilities of each dialogue system determines the possibilities for implementation of the $OISR_L$ model, we took the simplest form of such a system, namely an AIML-based chatbot (Bush, 2006). AIML (Artificial Intelligence Markup Language) covers the language understanding and generation task (Droßmann, 2005) in form of pattern-template pairs shown below. If the chatbot finds an input that matches to WIE GEHTS, the utterance stored in the template tag will be delivered to the user as a response.

```
<category>
    <pattern>WIE GEHTS</pattern>
    <template>Gut, und selbst?
    Alles paletti?</template>
</category>
```

Example 5.1 illustrates how a chatbot can benefit from patterns extracted from the dataset to come closer to the behaviour of a language expert.

Example 5.1. A sub-dialogue with the chatbot: other-initiated self-repair where the chatbot is the trouble-speaker.

1	User	wie gehts?
		how are you?
2	Bot	Gut, und selbst? Alles paletti?
		I'm fine, and you? Everything okay?
3	User	paletti?
4	Bot	umgangssprachlich alles gut, alles in Ordnung, alles okay.
		colloquial everything good, everything fine, everything okay.

The bot uses a colloquial expression in turn 2 which is not clear for the user. The user initiates the repair in turn 3. The bot recognises turn 3 as a repair initiation and extracts the trouble source: the repeated word *paletti* and the corresponding idiomatic expression *alles paletti*. Bot's response in turn 4 is a repair carry-out generated from a linguistic database.

The work of the repair manager is organised in two steps determined by the model. Every user's

input that requires an explanation of a single entity (word, idiom) is redirected to the category that implements this function. The implementation of ProgramD includes so called `processors` to process specific AIML tags. A new AIML tag has been introduced for the purpose of this work: `<explanation>`. An additional processor named `explanation processor` has been implemented to generate a response.

The model for the recognition of repair initiations described in Section 5.1 is used for the implementation in form of the rules describing repair initiation formats. For instance, to recognise the repair initiation from Example 5.1, the chatbot matches the rule:

$$RI = immediate, repeat(x), unclear(x)$$

because the user repeats a part of bot's utterance placing a question mark after the repeated token and it happens immediately after the bot's turn.

In Example 5.1, the repair initiation contains only a part of an idiomatic expression and only the entire expression can be found in the linguistic database. Because all chatbot's utterances are known beforehand in AIML-based chatbots, it is possible to list all idioms to make their recognition easier. For this test implementation, a short list of idiomatic expressions and their parts was created. The explanation processor would first check, if the trouble source may be an idiom (comparing with the list and own preceding turns). If so, the entire expression will be set as the trouble source.

AIML provides a possibility to forward inputs with the same or similar meanings to a particular category handling responses to this meaning. Int this way, all recognised repair initiations with the meaning $unclear(x)$ are redirected to the category with the pattern:

```
<pattern>ICH VERSTEHE * NICHT</pattern>
```

where * is the matching token for the trouble source x.

The following template is responsible for the generation of repair carry-outs for all such trouble sources. The `<think>` tag allows processing of an input without without immediate output. The explanation processor searches for the trouble source in the linguistic database which contains only meanings, examples and notes about usage for German nouns, verbs, adjectives and adverbs. The database was automatically generated from Wiktionary. If the trouble source cannot be found in the linguistic database, the ex-

planation processor returns `<NOENTITY>` and the pre-stored *Response-1* is sent to the user. If the trouble source is found but its meaning is not stored in the database, the explanation processor returns `<ENTITY NOMEANING>`. A predefined *Response-2* is then sent to the user. Finally, if the explanation processor finds the trouble source in the database and at least one meaning of it is described, an explanation will be rendered. Five additional categories not shown here are responsible for rendering of the explanation and process meanings, examples and notes.

```
<template>
 <think>
  <set name="explanation-tmp">
   <explanation><star/></explanation>
  </set>
 </think>
 <condition name="explanation-tmp">
  <li value="NOENTITY">Response-1</li>
  <li value="ENTITY NOMEANING">
     Response-2</li>
  <li><srai>GETFIRSTMEANING
   <get name="explanation-tmp"/></srai>
  </li>
 </condition>
</template>
```

Every user's input that corresponds to an inquiry "does x mean y?" is redirected to the AIML category implementing meaning checks. An additional tag `<meaningcheck>` has been added to carry out the repair of this type. The handling of the meaning checks works in a similar way as the explanations described above. The program has been extended by a `meaning check processor` to process this tag in the following way. To generate a response to a candidate understanding, the chatbot needs to answer the question if x means the same as y? This is an instance of the textual entailment problem. If x is a single word, an idiom, a collocation or a proverb, the system can check the list of the synonyms of the corresponding entry in the linguistic database. If x and y are listed as synonyms, a confirming answer will be generated. Otherwise, the system will explain the meaning of x.

Only simple versions for each of paraphrasing and word-by-word explanation (split-reuse) were implemented. A word-by-word explanation only makes sense for words that could be difficult for the learner. We use a list of 100 and 1000 most frequently used German words[1] to filter those words that are supposed to be well known to everybody. The remaining words are explained separately.

---

[1] http://wortschatz.uni-leipzig.de/html/wliste.html

6 Results

The new model of other-initiated self-repair when the machine is the trouble-speaker allows recognising learner repair initiations and extracting the trouble source based on a description of language-specific and medium-specific resources for repair initiation. The model is created on a necessary level of abstraction to be applicable for text chat interaction in languages other than German. This assumption builds on (Dingemanse et al., 2014)'s finding that similar repair initiation formats exist across languages. Therefore, when provided a set of language-specific devices for repair initiation, it can be implemented for other languages. The extraction of the trouble source is based on abstract features like repetition of parts of the trouble-turn and adjacent position. These features are language independent.

The problem of the trouble source extraction is related to referring expression recognition or reference resolution described in NLP textbooks (Martin and Jurafsky, 2009, Ch. 21), which is addressed in a large number of scientific publications (Dahan et al., 2002; Iida et al., 2010). Usually only noun phrases or their pronominalised alternatives are considered for reference resolution in NLP. These are usually definite and indefinite noun phrases, pronouns, demonstratives and names. The analysis of repair initiations shows that verbs or parts of utterances may be used to refer to the trouble source. The presented model implicitly includes a local *discourse model* which "contains representations of entities which have been referred to in the discourse" (Martin and Jurafsky, 2009, p. 730). The local discourse model in repair sequences only concerns possible representations of the trouble source.

Compared to the model of clarification requests proposed in (Purver, 2004), the model introduced in this work has the following advantages. First, the inconsistencies form CA perspective found in (Purver, 2004)'s classification do not exist in the model presented in this work because of a close cross-disciplinary connection with CA. The model for repair initiations presented here strictly differentiates next-turn repair other-initiations from all other types of repair and describes only these repair initiations. Second, (Purver, 2004) introduced the model for clarification requests in a strong connection to the HPSG formalism. In contrast, the model presented in this work is already imple-

mentable with a simple language understanding technology. The separation between resources for signalling trouble and resources for referencing trouble source allows creating a rule-based grammar which can be implemented in dialogue systems with different levels of complexity.

With regard to the analysis of causes of troubles in understanding introduced in (Schlangen, 2004), mainly problems on the level of meaning and understanding were subject of learner's repair initiations. Consequently, the modelling was approached in this work with the assumption that the required kind of clarification is mainly determined by the user model targeting language learners. Similarly to the (Schlangen, 2004)'s approach to map the variance in form to a small number of readings, repair initiations in this work are mapped either to a content question *What does X mean?* or to a polar question *Does X mean Y?* where X is the trouble source and Y is the candidate understanding. In this way, the two approaches to modelling repair initiations are similar.

Models of repair covering repair initiations proposed in (Purver, 2004) and (Schlangen, 2004) and extended in follow-up work (Purver, 2006; Ginzburg et al., 2007; Ginzburg, 2012) were motivated by Conversation Analysis research. However, other approaches for modelling were preferred because of the insufficient operationalisation of CA findings for computational modelling. As an implication, the factors influencing the interaction that have been identified as important in CA studies and building a *system* did not become part of the baseline models in (Purver, 2004) and (Schlangen, 2004). Such factors include repair, turn taking, membership categorisation, adjacency pairs and preference organisation. In contrast to the previous models of repair (Purver, 2004; Schlangen, 2004) this work analyses repair initiations in a system of interconnected factors in conversation. More specifically, the proposed model of repair initiations takes turn taking and sequential organisation of interaction explicitly into account by distinguishing between immediate and delayed repair initiations and respective options for trouble source extraction. In addition, the new model takes virtual adjacency in chat into account. It explicitly differentiates repair initiated by the user from repair initiated by the system taking the sequential organisation into account. Finally, the preference organisation and recipient design were

taken into account by the user model. Based on the empirical findings, the user model assumes that language learners will request a special kind of clarification.

While recognition of repair initiations and trouble source extraction can be implemented using the simplest type of language understanding, namely, pattern-based language understanding, most repair carry-outs require more sophisticated linguistic capabilities.

Definitions provide an explanation of the trouble source. Existing online dictionaries such as Wiktionary or Wikipedia may be used to create linguistic knowledge bases. Because one term may have multiple meanings, a linking to the correct meaning may be required. This problem is related to lexical ambiguity resolution also known as meaning resolution (Small et al., 1987) and is part of a larger area of computational lexical semantics (Martin and Jurafsky, 2009, Ch. 20).

Paraphrases provide a reformulation of the trouble source. A lot of efforts have been put in automatic paraphrase generation and recognition. Several recent publications are (Metzler et al., 2011; Regneri and Wang, 2012; Marton, 2013).

Synonyms provide usually a short reformulation of the trouble source. Existing language resources such as WordNet (Fellbaum, 2010) and GermaNet (Hamp et al., 1997) can be used for finding synonyms. Multiple meanings of a word may need to be resolved.

Translations may be generated by using existing machine translation systems (Avramidis et al., 2015; Burchardt et al., 2014). Open source statistical machine translation systems such as Moses[2] make experimental implementations feasible. Commercial machine translation API can be integrated into the dialogue manager, for instance Google Translate API[3].

Demonstrations include hyperlinks to websites containing relevant information examples of an object referenced by the trouble source. For semi-automatically created databases of linguistic knowledge, such information may be included into examples. Wikipedia articles sometimes also contain links to example websites and pictures, which may be used as examples of concepts described in the article.

Explicit handling of repairs targeted for lan-

guage learners allows an implementation in a CommICALL system that helps to practice conversation. In this way, this research advances state-of-the-art in ICALL and strengthens multidisciplinary connections to related disciplines, such as Conversation Analysis and NLP. Other types of tutorial dialogues where a clarification of the terminology may be necessary would also benefit from the presented model.

7 Conclusions

This article describes typical interactional resources employed for repair in native/non-native speaker chat with the purpose of computation modelling of repair for a conversational agent in a CommICALL application. The study shows that CA methods provide a valuable set of tools for computational modelling of rare phenomena in talk from a small number of examples. To be successful, such approaches require datasets replicating the speech exchange systems that are envisioned in the communication with the agent. In particular, this research showed that native/non-native speaker chat data can be used for computational models of dialogues in a CommICALL application.

Acknowledgements

References

Eleftherios Avramidis, Maja Popovic, and Aljoscha Burchardt. 2015. DFKI's experimental hybrid MT system for WMT 2015. In *Proceedings of the 10th Workshop on Statistical Machine Translation*. ACL, pages 66–73.

Aljoscha Burchardt, Arle Richard Lommel, Georg Rehm, Felix Sasaki, Josef van Genabith, and Hans Uszkoreit. 2014. Language technology drives quality translation. *MultiLingual* 143:33–39.

Noel Bush. 2006. Program D. http://www.aitools.org/Program_D.

Okko Buß and David Schlangen. 2011. DIUM – An Incremental Dialogue Manager That Can Produce Self-Corrections. In *SEMDIAL*.

_____

[2]http://www.statmt.org/moses/
[3]https://cloud.google.com/translate/docs

Delphine Dahan, Michael K Tanenhaus, and Craig G Chambers. 2002. Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language* 47(2):292–314.

Sviatlana Danilava, Stephan Busemann, Christoph Schommer, and Gudrun Ziegler. 2013. Towards Computational Models for a Long-term Interaction with an Artificial Conversational Companion. In *Proc. of ICAART'13*.

Mark Dingemanse, Joe Blythe, and Tyko Dirksmeyer. 2014. Formats for other-initiation of repair across languages: An exercise in pragmatic typology. *Studies in Languag* 3(81):5–43.

Christian Droßmann. 2005. German AIML set. http://www.drossmann.de/wordpress/alicebot/.

Myroslava O Dzikovska, Charles B Callaway, Elaine Farrow, Johanna D Moore, Natalie Steinhauser, and Gwendolyn Campbell. 2009. Dealing with interpretation errors in tutorial dialogue. In *Proceedings of the SIGDIAL 2009 Conference: The 10th annual meeting of the special interest group on discourse and dialogue*. Association for Computational Linguistics, pages 38–45.

Christiane Fellbaum. 2010. Princeton university: About WordNet. http://wordnet.princeton.edu. WordNet.

Christine Fredriksson. 2012. About collaboration, interaction, and the negotiation of meaning in synchronous written chats in l2-german. In Linda Bradley and Sylvie Thouësny, editors, *CALL: Using, Learning, Knowing, EUROCALL Conference, Gothenburg, Sweden, 22-25 August 2012, Proceedings*. Research-publishing.net, pages 88–92.

Rod Gardner and Johannes Wagner. 2004. *Second Language Conversations: Studies of Communication in Everyday Settings*. A&C Black - Verlag.

Raphaela Gehle, Karola Pitsch, and Sebastian Wrede. 2014. Signaling trouble in robot-to-group interaction. emerging visitor dynamics with a museum guide robot. In *Proceedings of the second international conference on Human-agent interaction*. ACM, pages 361–368.

Jonathan Ginzburg. 2012. *The interactive stance*. Oxford University Press.

Jonathan Ginzburg, Raquel Fernández, and David Schlangen. 2007. Unifying self-and other-repair. In *Proceeding of DECALOG, the 11th International Workshop on the Semantics and Pragmatics of Dialogue (SemDial07)*.

Birgit Hamp, Helmut Feldweg, et al. 1997. Germanet-a lexical-semantic net for german. In *Proceedings of ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*. Citeseer, pages 9–15.

Sviatlana Höhn. 2015. deL1L2IM: Corpus of long-term instant messaging NS-NNS conversations. ELRA http://islrn.org/resources/339-799-085-669-8/.

Yuri Hosoda. 2006. Repair and relevance of differential language expertise in second language conversations. *Applied Linguistics* 27(1):25–50.

Ryu Iida, Shumpei Kobayashi, and Takenobu Tokunaga. 2010. Incorporating extra-linguistic information into reference resolution in collaborative task dialogue. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, pages 1259–1267.

Cui Jian, Desislava Zhekova, Hui Shi, and John Bateman. 2010. Deep Reasoning in Clarification Dialogues with Mobile Robots. In *Proceedings of the 19th European Conference on Artificial Intelligence*. IOS Press, pages 177–182.

Geert-Jan M. Kruijff, Michael Brenner, and Nick Hawes. 2008. Continual planning for cross-modal situated clarification in human-robot interaction. In *The 17th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, pages 592–597.

Anthony J. Liddicoat. 2011. *An Introduction to Conversation Analysis*. Continuum.

Roy Lyster, Kazuya Saito, and Masatoshi Sato. 2013. Language teaching. *Oral corrective feedback in second language classrooms*. 46.

Numa Markee. 2000. *Conversation Analysis*. Mahwah, N.J.: Lawrence Erlbaum.

James H Martin and Daniel Jurafsky. 2009. *Speech and language processing: an introduction to natural language processing, computational linguistics and speech resognition*. Pearson International Edition, second edition edition.

Yuval Marton. 2013. Distributional phrasal paraphrase generation for statistical machine translation. *ACM Transactions on Intelligent Systems and Technology (TIST)* 4(3):39.

Raveesh Meena, José Lopes Gabriel Skantze, and Joakim Gustafson. 2015. Automatic detection of miscommunication in spoken dialogue systems. In *16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. pages 354 – 363.

Donald Metzler, Eduard Hovy, and Chunliang Zhang. 2011. An empirical evaluation of data-driven paraphrase generation techniques. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*. Association for Computational Linguistics, pages 546–551.

Kenneth A Petersen. 2010. *Implicit Corrective Feedback in Computer-Guided Interaction: Does Mode Matter?*. Ph.D. thesis, Georgetown University.

Matthew Purver. 2004. *The theory and use of clarification requests in dialogue*. Ph.D. thesis, King's College, University of London.

Matthew Purver. 2006. CLARIE: Handling clarification requests in a dialogue system. *Research on Language and Computation* 4(2-3):259–288.

Michaela Regneri and Rui Wang. 2012. Using discourse information for paraphrase extraction. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. Association for Computational Linguistics, pages 916–927.

Yuan Ma Satomi Kawaguchi. 2012. Corrective feedback, negotiation of meaning and grammar development: Learner-learner and learner-native speaker interaction in esl. *Open Journal of Modern Linguistics* 2:57–70.

Emanuel A Schegloff. 1996. Issues of relevance for discourse analysis: Contingency in action, interaction and co-participant context. In *Computational and conversational discourse: Burning Issues – An Interdisciplinary Account*, Springer-Verlag Berlin Heidelberg, pages 3–35.

Emmanuel A. Schegloff. 2000. When 'others' initiate repair. *Applied Linguistic* 21(2):205–243.

David Schlangen. 2004. Causes and Strategies for Requesting Clarification in Dialogue. In *5th Workshop of the ACL SIG on Discourse and Dialogue*.

Elizabeth Ellen Shriberg. 1994. *Preliminaries to a Theory of Speech Disfluencies*. Phd, University of California at Berkeley.

Steven Small, Garrison Cottrell, and Michael Tanenhaus. 1987. *Lexical ambiguity resolution*. Morgan Kaufman Publishers, Inc., Los Altos, CA.

Evangeline Marlos Varonis and Susan Gass. 1985. Non-native/non-native conversations: A model for negotiation of meaning. *Applied Linguistics* 6(1):71–90.

Sabrina Wilske. 2014. *Form and Meaning in Dialog-Based Computer-Assisted Language Learning*. Ph.D. thesis, University of Saarland.

Simon Zwarts, Mark Johnson, and Robert Dale. 2010. Detecting Speech Repairs Incrementally Using a Noisy Channel Approach. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*. August, pages 1371–1378.