# Acquisition and Assessment of Semantic Content for the Generation of Elaborateness and Indirectness in Spoken Dialogue Systems

**Louisa Pragst**[1]**, Koichiro Yoshino**[2]**, Wolfgang Minker**[1]**,**
**Satoshi Nakamura**[2] and **Stefan Ultes**[3]
[1]Ulm University, Ulm, Germany
[2] Nara Institute of Science and Technology (NAIST), Nara, Japan
[3] University of Cambridge, Cambridge, UK
{louisa.pragst,wolfgang.minker}@uni-ulm.de
{koichiro,s-nakamura}@is.naist.jp, su259@cam.ac.uk

## Abstract

In a dialogue system, the dialogue manager selects one of several system actions and thereby determines the system's behaviour. Defining all possible system actions in a dialogue system by hand is a tedious work. While efforts have been made to automatically generate such system actions, those approaches are mostly focused on providing functional system behaviour. Adapting the system behaviour to the user becomes a difficult task due to the limited amount of system actions available. We aim to increase the adaptability of a dialogue system by automatically generating variants of system actions. In this work, we introduce an approach to automatically generate action variants for elaborateness and indirectness. Our proposed algorithm extracts RDF triplets from a knowledge base and rates their relevance to the original system action to find suitable content. We show that the results of our algorithm are mostly perceived similarly to human generated elaborateness and indirectness and can be used to adapt a conversation to the current user and situation. We also discuss where the results of our algorithm are still lacking and how this could be improved: Taking into account the conversation topic as well as the culture of the user is likely to have beneficial effect on the user's perception.

## 1 Introduction

In a dialogue system (DS), the dialogue manager (DM) is responsible for choosing the system's contribution to a conversation. Several studies (e.g. (Ultes et al., 2015; Bertrand et al., 2011; Jaksic et al., 2006; Partala and Surakka, 2004)) show that adjusting the system's behaviour to the user can improve the user experience. To enable such adaptivity, the system needs several possible dialogue actions from which to choose. Often, those system actions are predefined manually. Hence, the amount of variants that the DM can choose from to adapt the system behaviour is limited by conversational skills, creativity and time of the person responsible for creating those actions. Foreseeing every possible situation the DS could find itself in and coming up with multiple viable system actions while considering possible types of users and their preferences in a conversation is demanding work. Approaches to the automatic generation of system actions, such as (Kadlec et al., 2015), have been presented to facilitate that process. However, those approaches often consider only system actions that are necessary from a functional point of view. There is no variety of system actions produced that would enable the DM to adapt to specific users characteristics or preferences. However, automatically generating variants of system actions can greatly increase the adaptability of a DM and thereby improve the user experience.

Studies (e.g. (Miehle et al., 2016; Pragst et al., 2017)) have shown that elaborateness and indirectness can be useful in adaptive DM. Here, elaborateness refers to the amount of additional information provided to the user and the level of indirectness describes how concretely information is addressed by a speaker. We have proposed the automatic generation of elaborateness and indirectness in (Pragst et al., 2016). In this work, we introduce an algorithm that, given a core statement on a semantic level, automatically creates more elaborated or indirect versions of that statement by retrieving semantic content from a knowledge base (KB) and assessing its relevance to the original

statement. Additionally, we ascertain that elaborateness and indirectness are suitable options for providing adaptability to the DM, and that our automatically generated system actions are mostly perceived similarly compared to human instances of elaborated and indirect statements. We further examine the circumstances under which the perception of automatically generated system actions deviates from human ones and discuss how to improve our algorithm based on those insights.

The remainder of the paper is structured as follows: In Section 2, we discuss related work. Section 3 gives an overview of the DS our approach is employed in. Our algorithm for the automatic generation of elaborateness and indirectness is presented in Section 4 and evaluated in Section 5. Finally, we draw a conclusion in Section 6.

## 2 Related Work

Adaptive DMs can be beneficial to the user experience (Ultes et al., 2015; Bertrand et al., 2011; Jaksic et al., 2006; Partala and Surakka, 2004) and has been implemented in a number of DMs (e.g. (Gnjatović and Rösner, 2008; Ultes and Minker, 2014; Rieser and Lemon, 2011)). Often, adaptive DMs consider user characteristics such as culture (Aylett and Paiva, 2012; Mascarenhas et al., 2013) or emotion (André et al., 2004; Gnjatović and Rösner, 2008; Pittermann and Pittermann, 2007). Komatani et al. (2005) use the amount of information presented as adaptation mechanism to the user's knowledge and the degree of urgency. Such architectures provide the decision making process necessary for choosing the best suited system action. However, they depend on the availability of suitable system actions to perform optimally.

To facilitate the process of defining system actions, efforts have been made to model dialogues automatically, e.g (Beveridge and Fox, 2006; Kadlec et al., 2015; Zhai and Williams, 2014; Niraula et al., 2014). Those approaches are mostly focused on functional system behaviour. Only system actions that are necessary to solve a task are defined, limiting the possibilities for adaptation. Our goal is to generate variants of system actions that address the same functionality, and thereby increase the adaptability.

Our efforts to generate variants of system actions is paralleled by a number of tasks in the area of natural language genera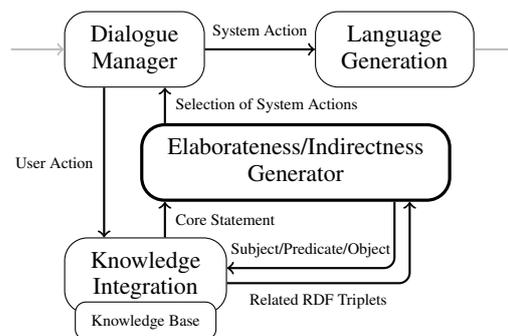tion. Natural language generators produce human-readable sentences from a more structured representation. With regard to surface realisation, one characteristic of good generators is their ability to provide variation in the generated sentences, which has been explored, among others, by Wen et al. (2015). With a similar goal, efforts towards the paraphrasing of sentences have been made (e.g. (Kozlowski et al., 2003; Langkilde and Knight, 1998)). Those approaches provide variation at the word level and preserve the semantic content of a sentence. They are complemented by our approach that focuses on variations of the semantic content of a system action. The content selection task is concerned with choosing relevant information that is to be communicated in the generated text, often with the goal of creating summaries (e.g. (Duboue and McKeown, 2003; Barzilay and Lapata, 2005)). While this research area certainly provides important insights to the generation of elaborateness, they need to be considered with respect to the peculiarities of dialogue. Instead of providing an overview over the most important information in a larger amount of data, the goal of our work is to augment an already determined piece of information with relevant further information. Hence, content selection is more concerned with filtering information, while our approach focuses on adding information.

## 3 System Architecture

We embed our approach to the generation of elaborateness and indirectness into an existing DS, the KRISTINA system (Wanner et al., 2016; Meditskos et al., 2016). It is employed in the health-care domain, with the overarching goal to support immigrants with health-care related issues in a socially competent manner. To enable a deeper understanding of the workings of the proposed algo-



Figure 1: Partial architecture of the KRISTINA system, enhanced by the proposed algorithm.

rithm, this section presents the system architecture of KRISTINA. A graphic representation of the relevant parts can be found in Figure 1.

The KRISTINA system does not rely on predefined system actions. Instead, a knowledge integration component (KI) (Meditskos et al., 2017) is responsible for interpreting the user request and searching a KB for the required information. The information retrieved from the KB is represented as RDF triplets. RDF triplets consist of subject, predicate and object and are used to describe the relationship between objects, e.g *(s:mother, p:is_from, o:Berlin)*. As a result of this representation, system actions are given as a set of RDF triplets that represent the semantic content to be conveyed to the user. A language generation component (Bouayad-Agha et al., 2012) transfers those RDF triplets into sentences.

To enhance our system from a purely functionally oriented DS to a user oriented DS, our goal is to transform the RDF triplets retrieved by the KI in a way that makes them either more elaborated or more indirect, while preserving the original meaning. Thereby, the DM has more choices than just the functional answer to the user question. The KB employed by the KI is utilised to gather suitable RDF triplets for the new system action variants. Newly created system actions can be transformed to sentences by the language generation in the same manner as the output from the KI.

## 4 Generation of Elaborateness and Indirectness

The starting point for both the generation of elaborateness and indirectness is the set of RDF triplets that was selected to answer the user request by the KI. We call this set the *core statement*. To generate a more elaborated version of the core statement, further RDF triplets that are relevant to the core statement are added. To achieve a more indirect variant, the core statement is omitted from the system response and instead a set of RDF triplets that is closely related to it is used. This process is divided in two parts: the acquisition of relevant RDF triplets from the KB and the assessment of those triplets to find the ones most suitable with regard to the core statement. In the following, the procedure is described in more detail.

**Algorithm 1:** Pseudocode for the acquisition of RDF triplets from the KB.

**Data**: $coreStmt$, the set of RDF triplets selected by the KI
$spo$, a function that relates each triplet to the set of its subject, predicate and object
$maxDist$, the maximal number of iterations for the search of the KB
$minDist$, the minimal number of iterations after which triplets are included
$retrAll$, a function that retrieves all triplet containing the given resource or predicate from the KB
**Result**: $triplets$, the set of gathered triplets

$stmtSet \longleftarrow coreStmt$
$rmStmt \longleftarrow \emptyset$
**if** $minDist > 0$ **then**
    $rmStmt \longleftarrow coreStmt$

**for** $dist = 1$ **to** $maxdist$ **do**
    $stmtSet \longleftarrow \bigcup_{x \in stmtSet} \bigcup_{y \in spo(x)} retrAll(y)$
    **if** $dist < minDist$ **then**
        $rmStmt \longleftarrow stmtSet$

$triplets \longleftarrow stmtSet \setminus rmStmt$

### 4.1 Acquisition of Semantic Content

To avoid having to assess every RDF triplet stored in the KB with regard to its relevance to the core statement, triplets that are connected to the core statement are preselected. The pseudocode for this process is depicted in Algorithm 1. For every triplet in the core statement, the KB is searched for all further triplets that contain either its subject, predicate or object. Exemplary, if *(s:mother, p:is_from, o:Berlin)* is part of the core statement, the triplets *(s:Berlin, p:is_in, o:Germany)* and *(s:mother, p:has_age, o:42)* could be retrieved from the KB. This process is repeated for the newly gathered triplets to find further candidates. The number of iterations is determined by the desired level of elaborateness. The higher the targeted elaborateness, the more iterations are performed. A further parameter is used to adjust the level of indirectness. It determines the number of iterations that have to be performed before a triplet can be added to the final system action. If a triplet is encountered before sufficient iterations have been performed, it can be used to find further triplets, but is not allowed as part of the final system action. If an elaborated, but direct answer is desired, this parameter is set to 0. After gathering potential RDF triplets as candidates, the next step is to assess their relevance to the core statement.

**Algorithm 2:** Pseudocode for the assessment of semantic content.

> **Data**: $coreStmt$, the set of RDF triplets selected by the KI
> $triplets$, the set of gathered RDF triplets
> $nrTriplets$, the number of desired triplets, derived from the level of elaborateness
> $f$, a function to adjust the weight of the individual inputs to the rating
> **Result**: $addStmt$, the set of additional statements chosen
>
> $addStmt \longleftarrow \emptyset$
> **while** $|addStmt| < nrTriplets$ **do**
> $\quad$ $allStmts \longleftarrow coreStmt \cup addStmt$
> $\quad$ $p \longleftarrow getCondProb(t, allStmt)$
> $\quad$ $d \longleftarrow getDistance(t, allStmt)$
> $\quad$ $i \longleftarrow getInterrelation(t, allStmt)$
> $\quad$ $best \longleftarrow \text{argmax}_{t \in triplets} f(p, d, i)$
> $\quad$ $addStmt \longleftarrow addStmt \cup best$
> $\quad$ $triplets \longleftarrow triplets \setminus best$

---

**Algorithm 3:** The function that estimates the conditioned probability of a triplet given the core statement.

> **Data**: $p$, the probability function
> $spo$, a function that relates each triplet to the set of its subject, predicate and object
>
> **Function** getCondProb($t$, *coreStmt*)
> $\quad$ **return**
> $\quad$ $mean_{s \in coreStmt} mean_{x \in spo(s)} mean_{y \in spo(t)} \frac{p(x,y)}{p(x)}$

---

**Algorithm 4:** The function that calculates the mean distance between a triplet and the core statement.

> **Data**: $spo$, a function that relates each triplet to the set of its subject, predicate and object
> $maxDist$, the maximal number of iteration for the search of the KB
> $retrAll$, a function that retrieves all triplet containing the given resource or predicate from the KB
>
> **Function** getDistance($t$, *coreStmt*)
> $\quad$ $accDist \longleftarrow 0$
> $\quad$ **for** $s \in coreStmt$ **do**
> $\quad\quad$ $dist \longleftarrow 0$
> $\quad\quad$ $stmtSet \longleftarrow \{s\}$
> $\quad\quad$ **while** $t \notin stmtSet \wedge dist \leq maxDist$ **do**
> $\quad\quad\quad$ $dist \longleftarrow dist + 1$
> $\quad\quad\quad$ $stmtSet \longleftarrow$
> $\quad\quad\quad$ $\bigcup_{x \in stmtSet} \bigcup_{y \in spo(x)} retrAll(y)$
> $\quad\quad$ **if** $t \notin stmtSet$ **then**
> $\quad\quad\quad$ $meanDist \longleftarrow meanDist + 100$
> $\quad\quad$ **else**
> $\quad\quad\quad$ $meanDist \longleftarrow meanDist + dist$
> $\quad$ **return** $\frac{accDist}{|coreStmt|}$

## 4.2 Assessment of Semantic Content

The overall process to choose triplets for the final system action is depicted in Algorithm 2, with Algorithms 3, 4 and 5 contributing necessary functions. All gathered RDF triplets are ranked with regard to the core statement and those with the highest rank are included in the final system action. After the inclusion of each new RDF triplet, the ranking is repeated. It takes as reference the newly added triplets as well as the core statement. This improves the overall consistency. The number of triplets in the final system action is restricted by the targeted level of elaborateness.

The ranking function $f$ takes into account several factors: $p$, the probability for the triplet to occur given the core statement, $d$, the mean distance between triplet and core statement in the KB, and $i$, the number of triplets in the core statement related to the triplet. It can be chosen freely to reflect the importance of the individual factors. In our experiments, we choose $f(p, d, i) = p + 2di$.

The probability for a triplet to occur given the core statement is derived from a corpus of dialogues between humans. An automated mapping of words to the semantic concepts that are used in the KB was performed and this data was used to calculate the probability that two concepts would appear in one dialogue turn as well as the overall probability that a concept would occur in a dialogue turn. From those probabilities, the conditioned probability that the concept of a new triplet will be in a turn if a concept of the core statement occurs in that turn can be calculated, as is shown in Algorithm 3. The mean of all conditioned probabilities between the concepts of the core statement and the triplet that is to be rated is used as input to the ranking.

Pseudocode for the calculation of the mean distance between a triplet and the core statement is given in Algorithm 4. The distance between two triplets in the KB refers to the number of iterations that have to be performed to find one triplet when starting from the other. If the triplet cannot be found due to the elaborateness restriction, a high number is assumed instead. The mean distance between a triplet and the core statement is used as a metric on how closely related they are.

The process to determine the number of triplets a triplet is related to can be found in Algorithm 5. A triplet is related to another triplet of the core statement if the triplet could be reached from it during the acquisition. It can be assumed that a

**Algorithm 5:** The function that calculates the number of relations between a triplet and the core statement.

**Data**: $spo$, a function that relates each triplet to the set of its subject, predicate and object
$maxDist$, the maximal number of iteration for the search of the KB
$retrAll$, a function that retrieves all triplet containing the given resource or predicate from the KB

**Function** `getInterrelation`(*t, coreStmt*)

$\quad nrRel \longleftarrow 0$
$\quad$ **for** $s \in coreStmt$ **do**
$\quad\quad stmtSet \longleftarrow \{s\}$
$\quad\quad$ **for** $dist = 1$ **to** $maxdist$ **do**
$\quad\quad\quad stmtSet \longleftarrow$
$\quad\quad\quad \bigcup_{x \in stmtSet} \bigcup_{y \in spo(x)} retrAll(y)$
$\quad\quad$ **if** $t \in stmtSet$ **then**
$\quad\quad\quad nrRel \longleftarrow nrRel + 1$

$\quad$ **return** $nrRel$

---

**Statement 4**
KRISTINA: Your mother is not originally from here? Does she miss Germany sometimes?
TOM (ORIG.): Yes, my mother misses Germany.
TOM (HG): I think my mother misses Germany because most of her relatives and friends are there and when she is there she is able to communicate much better.
TOM (CG): My mother misses Germany. She and my father married there. They immigrated, but they visit Germany. My father is happy.

**Statement 6**
TOM: Do you know how the weather is going to be?
KRISTINA (ORIG.): It is going to rain this afternoon.
KRISTINA (HG): It is going to rain this afternoon, but it's not going to be cold, still 20°C. But I would take an umbrella.
KRISTINA (CG): It is going to rain this afternoon. It is not going to be cold in the afternoon, 20°C by then. The temperature tomorrow is also going to be 20°C.

Figure 2: Examples for human and computer generated elaborated statements.

---

triplet that is related to the whole core statement is more relevant to the situation than one that is related to part of it.

## 5 Evaluation

Our approach has been evaluated in an online user study. Participants were asked to rate both human generated (HG) and computer generated (CG) variants of dialogue contributions with regard to the original statement. The research question of the study was twofold: First, to compare the variants produced by our algorithm to HG ones. Second, to show that elaborateness and indirectness have the potential to be used in adaptive DM. In the following, an overview of the participants of the user study, the study design as well as the results are presented. Finally, the findings and their implications for the proposed algorithm and adaptive DM in general are discussed.

### 5.1 Participants

The study included 21 Japanese and 21 German participants, most of which were between the age of 20 and 30. The 26 male participants slightly outweigh the female participants. The language of the study was English, so to identify potential influences of the individual English reading skill, participants were asked to rate their English skill using either the Common European Framework of Reference for Languages (CEFR), which is often used in Germany to assess language skills, or the Test of English for International Communication (TOEIC), which is more common in Japan. All of the participants reported English reading skills above the beginner level (CEFR: A1/TOEIC reading: 115), with the majority even reporting skills at or above the upper intermediate level (CEFR: B2/TOEIC reading: 385 or better).

### 5.2 Study Design

The proposed algorithm was evaluated by comparing its results to actual human generated examples of elaborateness and indirectness. To this end, ten elaborated and ten indirect statements were extracted from natural conversations. The conversations take place between caregivers, caretakers and their relatives, with topics ranging from biographical information, eating preferences, health issues to recreational activities. For the extracted statements, the concise/direct version of the statement was determined manually, taking into consideration both previous and following parts of the conversation. Those concise/direct statements are referred to as the *original statements* for the remainder of the paper. The original statements were transformed into semantic representations of their content and the proposed algorithm was used to produce an altered semantic representation, either aiming to be more elaborated or more indirect. As the performance of our content acquisition and rating component was to be tested, not that of a language generation component, a human transformed the semantic representation of the content into sentences. They were instructed to create simple sentences and only include the information provided by the RDF triplets. Examples

**Statement 1**
TOM: My mother doesn't speak English very well.
KRISTINA (ORIG.): Can you translate for the nurses?
KRISTINA (HG): If the nurses need someone for translation can they contact you?
KRISTINA (CG): Can the nurses contact you to give them information about your mother?

**Statement 6**
KRISTINA: How much support does your father need? Can he walk on his own?
TOM (ORIG.): My father needs support walking.
TOM (HG): My father is unsteady and shaky if he has nothing to hold onto. He can do two, three steps if someone holds him.
TOM (CG): My father can do two, three steps if someone holds him.

Figure 3: Examples for human and computer generated indirect statements.

of the resulting sentence pairs can be found in Figure 2 and 3. A full list of the sentences pairs is provided in the additional material.

The study consisted of an online questionnaire, presenting pairs of an original statement and a HG or CG variant of it to the participants. The participants were not made aware that some of the statements were computer generated. Furthermore, the exchanges were presented as human-human instead of human-computer interaction. All participants assessed all HG and CG variants, resulting in 20 evaluated statements pairs for elaborateness and indirectness each. For elaborated variants, the participants were asked how relevant the additional information is. This question was rated on a five point scale from 1 - 'not at all' to 5 - 'very relevant'. For indirect statements, participants rated how easily they could derive the meaning of the original statement from the indirect statement on a five point scale from 1 - 'it is impossible' to 5 - 'it is obvious'. For all sentence pairs, participants were asked to rate which statement they preferred on a 5 point scale from 1 - 'the original one' to 5 - 'the elaborated/indirect variant'.

Apart from the comparison of the generation methods, differences between the nationalities and the individual original statements were also considered. Differing ratings in those areas suggest possible adaptations that may be employed by a DM to cater to different cultures or different situations.

The results for each research question were obtained using a three-way mixed ANOVA.
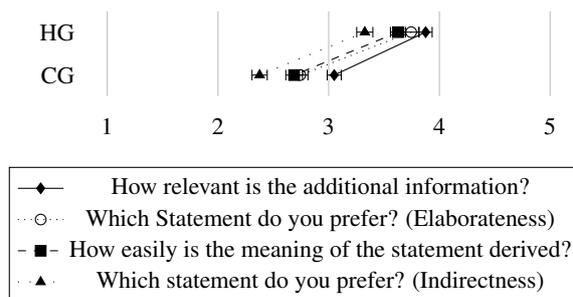


Figure 4: Comparison of the mean rating and standard error by the generation method used.

| Statement | Japanese | | German | |
|---|---|---|---|---|
| | HG | CG | HG | CG |
| 1 | 4.05 | 3.43 | 4.29 | 2.38 |
| 2 | 3.95 | 3.00 | 4.48 | 1.95 |
| 3 | 4.62 | 3.71 | 2.76 | 2.81 |
| 4 | 4.38 | 3.38 | 3.76 | 2.57 |
| 5 | 3.71 | 2.29 | 3.00 | 2.81 |
| 6 | 3.86 | 3.67 | 3.48 | 2.76 |
| 7 | 4.00 | 3.67 | 3.57 | 2.71 |
| 8 | 4.33 | 4.33 | 3.52 | 3.14 |
| 9 | 4.29 | 3.10 | 4.19 | 2.48 |
| 10 | 3.57 | 3.62 | 3.71 | 3.19 |

Table 1: Mean ratings for the question 'How relevant is the additional information?'.

### 5.3 Results

Figure 4 depicts the mean and standard error of the rating for each of our research questions. Statistical tests show that HG and CG statements yield mostly similar results. No significant differences between the generation methods can be found for a significance level of 0.05, except for the ease with which indirect statements can be interpreted. Here, CG statements are harder to understand than HG ones. However, this does not significantly influence the preference of participants for either direct or indirect statements. Additionally, we find that nationality and situation influence the rating, suggesting that adapting to them by changing the level of elaborateness and indirectness is viable. Apart from the main factors generation method, nationality and original statement, we also tested for influences of age, gender or proficiency in English on the results of our study but found no significant effects.

In the following, a more detailed description of the results is presented. A complete list of mean ratings for each research question can be found in Tables 1, 2, 3 and 4.

| Statement | Japanese | | German | |
|---|---|---|---|---|
| | HG | CG | HG | CG |
| 1 | 3.75 | 2.95 | 4.14 | 2.24 |
| 2 | 3.76 | 2.86 | 4.48 | 1.90 |
| 3 | 3.95 | 3.57 | 2.86 | 2.57 |
| 4 | 4.29 | 2.81 | 3.95 | 2.62 |
| 5 | 3.29 | 2.00 | 2.57 | 2.81 |
| 6 | 2.86 | 2.62 | 3.62 | 2.57 |
| 7 | 3.24 | 1.95 | 3.76 | 2.43 |
| 8 | 3.90 | 3.90 | 3.76 | 3.19 |
| 9 | 4.33 | 3.10 | 4.24 | 2.52 |
| 10 | 4.10 | 3.10 | 4.05 | 3.14 |

Table 2: Mean ratings for the question 'Which statement do you prefer?' (Elaborateness).

| Statement | Japanese | | German | |
|---|---|---|---|---|
| | HG | CG | HG | CG |
| 1 | 4.19 | 3.24 | 4.81 | 3.00 |
| 2 | 3.33 | 3.14 | 4.24 | 2.86 |
| 3 | 1.67 | 1.19 | 2.29 | 1.00 |
| 4 | 2.67 | 2.19 | 2.48 | 1.71 |
| 5 | 4.00 | 3.48 | 4.76 | 4.38 |
| 6 | 3.86 | 3.81 | 4.62 | 4.24 |
| 7 | 2.81 | 1.81 | 4.14 | 1.71 |
| 8 | 3.29 | 2.71 | 3.95 | 2.38 |
| 9 | 3.00 | 1.38 | 2.90 | 1.05 |
| 10 | 4.52 | 4.10 | 5.00 | 4.38 |

Table 3: Mean ratings for the question 'How easy is it to derive the original meaning?'.

### 5.3.1 Impact of Generation Method

The relevance of additional information as well as the user preference for either the original or the elaborated/indirect statement do not show significant differences regarding the generation method with a significance level of $0.05$. Only the ease with which the meaning of an indirect statement can be derived is significantly influenced by the generation method ($F(1, 37) = 5.401$, $p = .026$). This indicates that overall participants perceived HG and CG statements to be similar, but had problems to interpret CG indirect statements. In addition to those results, several significant interaction effects can be found. Those interaction effects offer valuable information about potential improvements that can be made to the proposed algorithm. Hence, they are examined in closer detail in the following.

Significant interaction effects between generation method, nationality and original statement exist for both the relevance of additional information ($F(9, 333) = 2.731$, $p = .004$) as well as the user preference for either the original or the elaborated statement ($F(9, 333) = 2.486$, $p = .009$). For both question, several interaction patterns can be observed, depending on the original statement:

| Statement | Japanese | | German | |
|---|---|---|---|---|
| | HG | CG | HG | CG |
| 1 | 3.57 | 3.05 | 4.19 | 2.57 |
| 2 | 3.05 | 2.38 | 2.95 | 1.43 |
| 3 | 2.62 | 2.14 | 2.24 | 1.38 |
| 4 | 1.95 | 1.52 | 2.00 | 1.38 |
| 5 | 3.90 | 2.57 | 3.90 | 2.57 |
| 6 | 3.86 | 3.29 | 3.76 | 3.19 |
| 7 | 4.33 | 3.00 | 3.10 | 2.38 |
| 8 | 3.81 | 3.67 | 3.71 | 2.76 |
| 9 | 3.95 | 2.29 | 2.71 | 1.52 |
| 10 | 3.29 | 3.10 | 3.62 | 1.33 |

Table 4: Mean ratings for the question 'Which statement do you prefer?' (Indirectness).



(a) Statement 3

(b) Statement 4

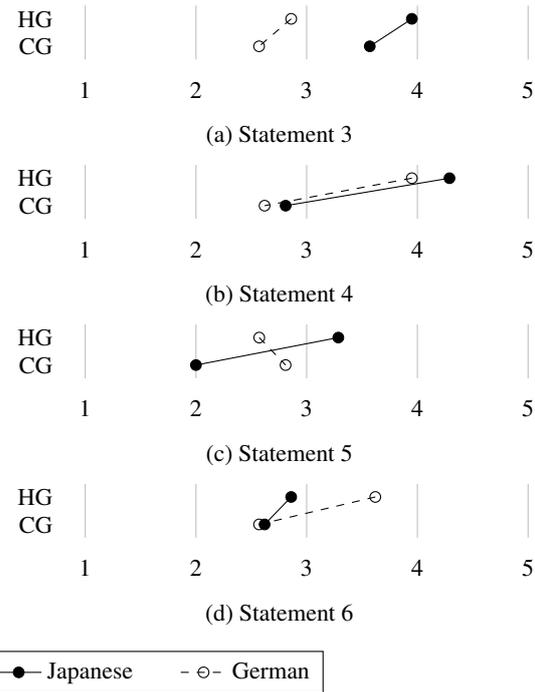(c) Statement 5

(d) Statement 6

Figure 5: Interactions between nationality and generation method for the question 'Which statement do you prefer?' (Elaborateness). Different patterns can be found: Nearly no difference between HG and CG, a sharp decline for CG, a sharp decline only for Japanese and a sharp decline only for Germans.

The generation method can have almost no impact or lead to a declining rating from either Japanese, Germans or both, as can be seen exemplary in Figure 5. For some statements, one of the cultures rates the CG statement in a similar manner as the HG one, while the rating of the other culture shows a sharp decline for CG statements. As a consequence of the different perceptions across cultures, it might be beneficial to consider the target culture during the generation process and thereby improve

its performance. Furthermore, a few original statements receive a worse rating for CG variants than for HG variants by both cultures. This suggests that while the generation of elaborateness often works well, potential for improvements still exists in those cases.

The ease with which the meaning of an indirect statement can be derived depends significantly on the interaction between generation method and nationality ($F(1, 37) = 10.469$, $p = .003$). The corresponding interaction pattern is depicted in Figure 6. If the indirect statement is HG, Germans seem to have an easier time to derive the original meaning than Japanese ($F(1, 37) = 5.547$, $p = .024$), who mostly rate this question neutrally. This difference between cultures disappears for CG statements. It is possible that Germans have an advantage with the HG statements, as those statements were extracted from German dialogues and the original statements were derived by a German. Hence, the implicit connections between original and indirect statement could be more obvious to Germans due to a similar cultural imprint. This advantage disappears when the implicit connection between original statement and indirect one is made automatically by an algorithm and therefore foreign to both cultures. This would suggest that an effort should be made to better capture the human approach to generating indirectness and thereby reduce the difficulty of interpreting it. In this endeavour, the target culture needs to be taken into account.

The preference for either the original or the indirect statement is influenced significantly by the interaction between generation method, nationality and original statement ($F(9, 333) = 3.124$, $p = .001$). Here, patterns similar to the ones found for the preference regarding elaborated statements can be observed: The generation method can have almost no impact or lead to a declining rating from either Japanese, Germans or both, depending on the original statement. This affirms the potential for improvements regarding the adjustment to the target culture as well as the overall performance.

### 5.3.2 Potential for Adaptation

To ascertain the ability of elaborateness and indirectness to contribute to the adaptability of a DM, we assess the impact of nationality and original statement on user preferences.

A significant interaction effect of nationality and original statement on the preference for ei-
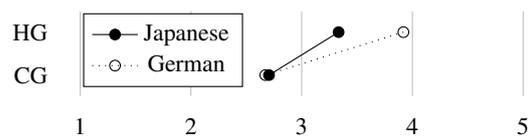


Figure 6: Interactions between nationality and generation method for the question 'How easy is it is to derive the original meaning?'.
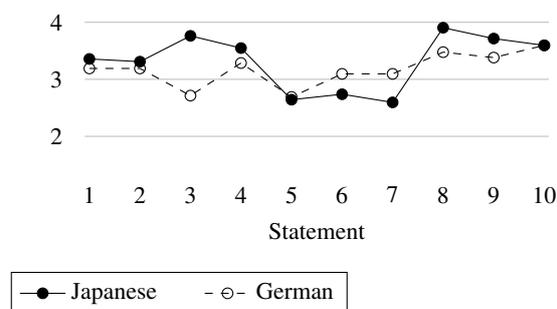


Figure 7: Interactions between nationality and original statement for the question 'Which statement do you prefer?' (Elaborateness).

ther the original or the elaborated statement can be found ($F(9, 333) = 2.578$, $p = .007$). As can be seen in Figure 7, Germans seem to be rather indifferent to the level of elaborateness. They mostly rate neutrally. In contrast, a clear distinction between topics can be found for Japanese. They tend to prefer concise statements if the topic of conversation is uncritical, such as the weather or day trips. This can be seen for Statements 5, 6 and 7. When talking about family members, more elaborated statements are preferred. Considering those findings, a culture and situation adaptive DM could utilise elaborateness as means to implement suitable adaptation.

As discussed in Section 5.3.1, there exist significant interactions of generation method, nationality and original statement on the preference of elaborateness/indirectness. If the generation methods are examined separately, the interaction of nationality and original statement still impacts the rating for both elaborateness (HG: $F(9, 333) = 2.197$, $p = .022$, CG: $F(9, 333) = 2.838$, $p = .003$) and indirectness (HG: $F(9, 333) = 2.143$, $p = .026$, CG: $F(9, 333) = 1.922$, $p = .048$). This implies that, while nationality and original statement always influence the user preference, the way they impact it is not the same for HG and CG statements. The different interac-
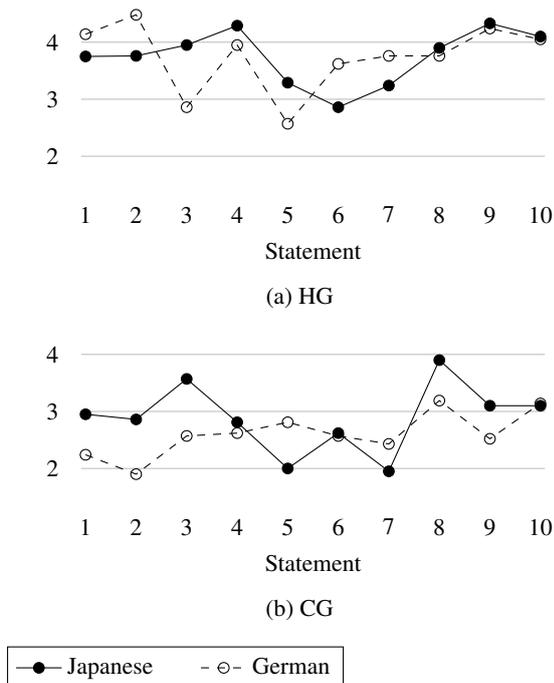
(a) HG



(b) CG

—●— Japanese    - ○- German

Figure 8: Comparison of the interaction pattern between nationality and original statement for the question 'Which statement do you prefer?' (Elaborateness) for HG and CG statements.

tion patterns are depicted exemplary in Figure 8. Those results suggests that an adaptive DM should not learn its dialogue strategy based on human-human dialogue containing examples of elaborateness/indirectness, as this can lead to wrong assumption about the user preferences. Instead, the dialogue strategy should be based on CG variants that reflect more accurately the elaborateness/indirectness variants that will be used.

### 5.4 Discussion

Overall, CG statements perform well compared to HG ones. However, there are cases when the rating of CG generated statements decreases for one or both of the considered cultures when compared to the HG statement. Therefore, it might be beneficial to consider both nationality as well as the topic of the original statement during the generation of system action variants. Another possibility would be to generate multiple elaborated/indirect variants and let the DM choose the most suitable with regard to the context and culture. Both approaches might improve the results of CG statements in cases where participants rated them worse than HG ones with the current approach.

In our study, the potential for adaptivity that is provided by system actions with different levels elaborateness and indirectness was investigated. For elaborateness, we could show that different preferences exist depending on the culture and original statement. Therefore, it is feasible for an adaptive DM to choose either the elaborated or the concise variant depending on both the culture of the user as well as the current context of the conversation to improve the user satisfaction. Furthermore, our results show that the interaction between generation method, nationality and original statement has an impact on the user's preference for elaborated/indirect statements. This implies that, while both elaborateness and indirectness can be used for adaptation, the DM should base its dialogue policy on experiences with CG statements instead of HG ones.

## 6 Conclusion

In this work, an approach to the automatic generation of more elaborated or indirect variants of a system action on the semantic level has been discussed. We proposed an algorithms for the acquisition of semantic data and the assessment of this data with regard to the dialogue contribution under consideration. Furthermore, a user study was performed to investigate the performance of our approach compared to humans and the applicability of elaborateness and indirectness for adaptiveness. The results show that, while the variants produced by the proposed algorithm are often perceived in a similar manner as human generated variants, complex interactions exist with both nationality and topic of the statement. Taking those into account can further improve the performance. Additionally, the study shows that differing preferences across cultures and statements exist and hence can be considered in adaptive DM.

In future work, the presented approach will be integrated into a fully functional DS, including the knowledge integration and language generation components that it relies on. Furthermore, the proposed algorithm will be further improved to better adjust to the user culture and the topic of the conversation.

### Acknowledgements

## References

Elisabeth André, Matthias Rehm, Wolfgang Minker, and Dirk Bühler. 2004. Endowing spoken language dialogue systems with emotional intelligence. In *Affective Dialogue Systems*, pages 178–187. Springer.

Ruth Aylett and Ana Paiva. 2012. Computational modelling of culture and affect. *Emotion Review*, 4(3):253–263.

Regina Barzilay and Mirella Lapata. 2005. Collective content selection for concept-to-text generation. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 331–338. Association for Computational Linguistics.

Gregor Bertrand, Florian Nothdurft, Wolfgang Minker, Harald Traue, and Steffen Walter. 2011. Adapting dialogue to user emotion-a wizard-of-oz study for adaptation strategies. *Proc. of IWSDS*, pages 285–294.

Martin Beveridge and John Fox. 2006. Automatic generation of spoken dialogue from medical plans and ontologies. *Journal of biomedical informatics*, 39(5):482–499.

Nadjet Bouayad-Agha, Gerard Casamayor, Simon Mille, Marco Rospocher, Horacio Saggion, Luciano Serafini, and Leo Wanner. 2012. From ontology to NL: generation of multilingual user-oriented environmental reports. In *Natural Language Processing and Information Systems - 17th International Conference on Applications of Natural Language to Information Systems, NLDB 2012, Groningen, The Netherlands, June 26-28, 2012. Proceedings*, pages 216–221.

Pablo A Duboue and Kathleen R McKeown. 2003. Statistical acquisition of content selection rules for natural language generation. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pages 121–128. Association for Computational Linguistics.

Milan Gnjatović and Dietmar Rösner. 2008. Adaptive dialogue management in the NIMITEK prototype system. In *Perception in Multimodal Dialogue Systems*, pages 14–25. Springer.

Nada Jaksic, Pedro Branco, Peter Stephenson, and L Miguel Encarnação. 2006. The effectiveness of social agents in reducing user frustration. In *CHI'06 extended abstracts on Human factors in computing systems*, pages 917–922. ACM.

Rudolf Kadlec, Martin Schmid, and Jan Kleindienst. 2015. Improved deep learning baselines for ubuntu corpus dialogs. *arXiv preprint arXiv:1510.03753*.

Kazunori Komatani, Shinichi Ueno, Tatsuya Kawahara, and Hiroshi G Okuno. 2005. User modeling in spoken dialogue systems to generate flexible guidance. *User Modeling and User-Adapted Interaction*, 15(1):169–183.

Raymond Kozlowski, Kathleen F McCoy, and K Vijay-Shanker. 2003. Generation of single-sentence paraphrases from predicate/argument structure using lexico-grammatical resources. In *Proceedings of the second international workshop on Paraphrasing-Volume 16*, pages 1–8. Association for Computational Linguistics.

Irene Langkilde and Kevin Knight. 1998. Generation that exploits corpus-based statistical knowledge. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 1*, pages 704–710. Association for Computational Linguistics.

Samuel Mascarenhas, Rui Prada, Ana Paiva, and Gert Jan Hofstede. 2013. Social importance dynamics: a model for culturally-adaptive agents. In *Intelligent virtual agents*, pages 325–338. Springer.

Georgios Meditskos, Stamatia Dasiopoulou, Louisa Pragst, Stefan Ultes, Stefanos Vrochidis, Ioannis Kompatsiaris, and Leo Wanner. 2016. Towards an ontology-driven adaptive dialogue framework. In *Proceedings of the 1st International Workshop on Multimedia Analysis and Retrieval for Multimodal Interaction*, pages 15–20. ACM.

Georgios Meditskos, Stefanos Vrochidis, and Ioannis Kompatsiaris. 2017. *Description Logics and Rules for Multimodal Situational Awareness in Healthcare*. Springer International Publishing, Cham.

Juliana Miehle, Koichiro Yoshino, Louisa Pragst, Stefan Ultes, Satoshi Nakamura, and Wolfgang Minker. 2016. Cultural communication idiosyncrasies in human-computer interaction. In *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, Los Angeles, USA. Association for Computational Linguistics.

Nobal B Niraula, Amanda Stent, Hyuckchul Jung, Giuseppe Di Fabbrizio, I Dan Melamed, and Vasile Rus. 2014. Forms2dialog: Automatic dialog generation for web tasks. In *Spoken Language Technology Workshop (SLT), 2014 IEEE*, pages 608–613. IEEE.

Timo Partala and Veikko Surakka. 2004. The effects of affective interventions in human–computer interaction. *Interacting with computers*, 16(2):295–309.

Johannes Pittermann and Angela Pittermann. 2007. A data-oriented approach to integrate emotions in adaptive dialogue management. In *Proceedings of the 12th international conference on Intelligent user interfaces*, pages 270–273. ACM.

Louisa Pragst, Juliana Miehle, Stefan Ultes, and Wolfgang Minker. 2016. Automatic modification of communication style in dialogue management. In *Proceedings of the INLG 2016 Workshop on Computational Creativity in Natural Language Generation*,

pages 36–40. Association for Computational Linguistics.

Louisa Pragst, Wolfgang Minker, and Stefan Ultes. 2017. Exploring the applicability of elaborateness and indirectness in dialogue management. In *Proceedings of the 8th International Workshop On Spoken Dialogue Systems (IWSDS)*.

Verena Rieser and Oliver Lemon. 2011. *Reinforcement learning for adaptive dialogue systems: a data-driven methodology for dialogue management and natural language generation*. Springer Science & Business Media.

Stefan Ultes, Matthias Kraus, Alexander Schmitt, and Wolfgang Minker. 2015. Quality-adaptive spoken dialogue initiative selection and implications on reward modelling. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, pages 374–383.

Stefan Ultes and Wolfgang Minker. 2014. Managing adaptive spoken dialogue for intelligent environments. *Journal of Ambient Intelligence and Smart Environments*, 6(5):523–539.

Leo Wanner, Josep Blat, Stamatia Dasiopoulou, Mónica Domínguez, Gerard Llorach, Simon Mille, Federico Sukno, Eleni Kamateri, Stefanos Vrochidis, Ioannis Kompatsiaris, et al. 2016. Towards a multimedia knowledge-based agent with social competence and human interaction capabilities. In *Proceedings of the 1st International Workshop on Multimedia Analysis and Retrieval for Multimodal Interaction*, pages 21–26. ACM.

Tsung-Hsien Wen, Milica Gasic, Nikola Mrksic, Pei-Hao Su, David Vandyke, and Steve Young. 2015. Semantically conditioned lstm-based natural language generation for spoken dialogue systems. *arXiv preprint arXiv:1508.01745*.

Ke Zhai and Jason D Williams. 2014. Discovering latent structure in task-oriented dialogues. In *ACL (1)*, pages 36–46.