# Towards a Categorization of Natural Language Variability in Data for Spoken Dialog Systems

**Patricia Kittel, Wolfgang Maier, Maria Schmidt**
Daimler AG
Sindelfingen, Germany
patricia.kittel@daimler.com
wolfgang.wm.maier@daimler.com
maria.m.schmidt@daimler.com

**Wolfgang Minker**
Ulm University
Ulm, Germany
wolfgang.minker@uni-ulm.de

## Abstract

With the move towards more natural human-machine interfaces, recent spoken dialog systems are expected to understand all utterances that are associated with a specific semantic meaning. However, hard criteria which would define this set of utterances are lacking. In this paper, we address this point by contributing a set of language-independent criteria which can be used for quantifying the degree of natural language variability in a data set. We validate the suitability of the criteria with a real-world data set from the automotive domain.

## 1 Introduction

Recent R&D in spoken dialog systems (SDS) aims at achieving more intuitive and human-like user experience. New systems are now expected to understand a natural input style. This development can be observed both in academia and in the commerical sector (see McTear et al. (2016) for an overview).

A natural way of speaking presupposes that as system input, all natural language expressions which are associated with the specific semantic meaning of the user intent are allowed, such that users no longer have to speak predefined commands. With this approach, the amount of possible user input becomes possibly indefinite. This is a challenge for the spoken language understanding (SLU) module of SDS that has the task to map the user's spoken utterance to a representation of the meaning of that utterance, as described, e.g., by Henderson and Jurcicek (2012). In order to meet the expectation of a natural input, developers have turned to statistical data-driven models which allow systems to deal also with user input that is previously unseen (McTear et al., 2016; Bellegarda and Monz, 2016). The results of such models strongly depend on the quality of training and test data, i.e. the kind of utterances the data sets are composed of. As pointed out by Henderson and Jurcicek (2012), the data needs to cover both the range of variability in the semantics and the range of variability in the natural language expressions that convey that semantics.

However, there is a lack of works addressing the question what kind of utterances actually constitute the range of natural language expressions for a given semantic meaning. In fact, it currently seems to depend on the developers' opinions which kind of utterances SLU modules must be able to understand. The result is that evaluation scores of SLU performances are not meaningful and not comparable if they depend on subjective test data.

Motivated by this situation, we make two contributions in this paper. First, we define a set of language-independent criteria which allow for quantifying to what extent natural language variability is covered by a data set. The criteria can be used to evaluate SDS with regard to their capability of understanding the range of variability that natural language offers for task-oriented requests. This evaluation in turn can be used to model conversational human-machine interfaces accordingly. The criteria we propose are derived from existing cross-lingual works and from a study we conducted. Second, we contribute a method to check test sets with regard to their distribution of realistic utterances for a specific semantic meaning. In order to achieve this, we describe the linguistic elements that correspond with the criteria by means of a decision tree and we analyze the distribution of the criteria within our study data. This decision tree can also serve as a guideline for annotators.

The remainder of the paper is structured as follows. In Section 2 we review previous literature which aims to characterize naturally spoken user input. Next, in Section 3 we introduce the data our results are

*Proceedings of the 22nd Workshop on the Semantics and Pragmatics of Dialogue, November 8-10, 2018, Aix-en-Provence, France.*

based on. In Section 4 we first explain the theoretical considerations and second, the findings of existing works. In Section 5 we present the patterns that occur in our data set. Their distribution is presented in Section 6. Section 7 serves as the coda of the article.

## 2 Related Work

Within the research field of conversational human-machine interfaces, many works make use of dialog act categorization in order to describe phenomena of human conversation behavior, e.g., Pareti and Lando (2018). Some authors focus on specific domains or situations, e.g., Sinclair et al. (2017), and others take multi-party conversation into account, e.g., Marzuki et al. (2017). However, works that quantify the variability of how certain dialog acts can be realized, are lacking.

Nevertheless, there are several approaches that try to characterize the different kinds of linguistic elements that speakers might use to express a specific semantic meaning. These can be split into two groups which are briefly described in the following.

One group describes spoken user input by means of differentiating between a command-like and a natural way of speaking (Hofmann et al., 2012; Pang and Kumar, 2011; Berg, 2012; White et al., 2014). Commands are characterized by an incomplete sentence structure. A natural way of speaking is equated with human-directed speech which consists of full sentences, filler words and civility (Hofmann et al., 2012; Berg, 2012). The linguistic strategies that indicate civility are not explained further.

The second group of works linguistically investigates spoken user input. Large et al. (2017), e.g., describe different linguistic phenomena that occur when drivers interact with a natural language digital assistant. They identify back-channelling strategies, fillers and hesitation, vague language, ways of mitigating requests, politeness and praise. The range of variability within these phenomena is not explored further. Winter et al. (2010) examine their study data with regard to the degree of context information the utterances contain. The authors do not take other linguistic phenomena into account. Braunger et al. (2016) and Braunger et al. (2017) define sentence structures such as *imperative, declarative* or *infinitive* sentences. Braunger et al. (2017) additionally characterize freely spoken input with the help of measures commonly used in order to describe and compare corpora, e.g., *type-token ratio, content-function word ratio, syntactic complexity, POS tag frequencies*.

Most of the few characterization approaches applied by the literature so far are either too abstract, such as type-token ratio, or too specific, such as the sentence structures which are only related to a certain language. In addition, recent works do not combine different criteria to obtain a unified and quantifiable scheme which is suitable for system design and evaluation.

Therefore, we choose a language-independent approach which combines different linguistic phenomena to quantify the variability of expressing task-oriented requests. In advance, we explain the theoretical considerations that lead to the criteria we propose. Finally, we show the distribution of those criteria for a data set consisting of natural language requests directed at a human interlocutor. The data is introduced in the following section.

## 3 Data Set

Since interpersonal interaction is the most natural way of interaction, it is often taken as a baseline for the development of a natural and intuitive human-machine communication, cf. Bonin et al. (2015). Therefore, our work relies on 540 German requests directed at a human interlocutor. The utterances are acquired by a previous study, see Braunger et al. (2017). This study has aimed to examine how users would voice-control specific functions of an in-vehicle infotainment system in a natural and intuitive way. The experimental setup is briefly described in the following.

The functions the participants were to operate and the information they were to request were described graphically. This method was chosen in order to not bias the participants by putting words into their mouths. The pictures they were shown describe the following twelve tasks.

1. Listen to radio station SWR3
2. Play Michael Jackson Greatest Hits
3. Navigate to Stieglitzweg 23 in Berlin
4. Call Barack Obama on his mobile phone
5. Set temperature to 23 degrees
6. Send a text message to brother
7. Weather in Berlin today
8. Date of the European football championship final game
9. Population of Berlin
10. Score FC Bayern against VfB Stuttgart
11. Cinema program in Berlin today
12. Next Shell gas station

For our work, we divide the tasks into six action requests (1-6) and six information seeking requests (7-12). The study was split into two sessions. For every scenario described here, the participants' task was: How would you communicate this request to your passenger and how would you communicate this request to an in-car SDS? As for the passenger session, the participants were told that the passenger provided the information requested or activated the appropriate function with help of a tablet. Each participant took part in both sessions and solved all tasks. The tasks and the sessions were randomized. In this paper, we rely on the passenger session utterances.

In total, 45 subjects participated in the study. 46% of them were female and 54% were male. The average age was 39.5 years with a standard deviation of 13.5. 55.6% were aged between 20 and 39 years, 26.6% were 40 to 59 years old and 17.8% were older than 60 years.

The data was manually transcribed in such a way that the transcription exactly matched the spoken utterance. Afterwards, the data was annotated manually.

## 4 Towards Categorizing Natural Language Requests

In this paper, we aim to propose criteria which are applicable in different languages. The theoretical considerations of such a speech act driven approach are described in Section 4.1. Afterwards, in Section 4.2, we present the findings of existing works in that field.

### 4.1 Theoretical Considerations

Natural language offers various possibilities in order to verbalize user intents. As an example, the following utterances show the possibilities of expressing that someone has to turn the music down (borrowed from Meibauer et al. (2007)).

1. *Turn the music down!*
2. *Could you please turn the music down?*
3. *How about turning the music down a bit?*
4. *The music is too loud!*

According to the politeness theory of Brown et al. (1987) it depends on politeness strategies which option one decides for. Their theory is based on the assumption that everybody has a *face*. The face can be considered as the positive public image one seeks to establish in social interactions (Goffman, 1955). This consists of two components: on the one hand there is the desire that the self-image be appreciated and approved of (so called *positive face*) (cf. Brown et al. (1987)); on the other hand there is the need for freedom of action (so called *negative face*). Since the user intents that we are interested in aim to get the hearer (the system) to do something, those requests challenge the face the interlocutor wants to have. Those requests are by definition so called *face-threatening acts*. Hence, politeness is defined as the strategy to save faces. According to Brown et al. (1987), speakers either decide for a strategy that saves the positive or the negative face.

### 4.2 Pragmatic Scheme of Request Realizations

Based on these considerations Blum-Kulka et al. (1989) define cross-lingual[1] coding schemes for the realization of requests. Their proposed schemes are adjusted by Siebold (2010) for a contrastive analysis of Spanish and German requests. The findings of both works are considered in the following.

[1]The coding scheme is based on data of eight languages or varieties: Australian English, American English, British English, Canadian French, Danish, German, Hebrew, Russian.

| Strategy | Example |
|---|---|
| 1. Hedged performative | *I would like you to give your lecture a week earlier.* |
| 2. Explicit performative | *I ask you to clean up this mess.* |
| 3. Scope stating | *I really wish you'd stop bothering me.* |
| 4. Strong hint | *You've left this kitchen in a right mess.* |

Table 1: Strategy dimension, cf. Blum-Kulka et al. (1989).

| Perspective | Example |
|---|---|
| 1. Hearer oriented | Could *you* tidy up the kitchen soon? |
| 2. Speaker oriented | Do you think *I* could borrow your notes? |
| 3. Speaker and hearer oriented | So, could *we* please clean up? |
| 4. Impersonal | So it might be not be a bad idea to *get it cleaned up*. |

Table 2: Perspective dimension, cf. Blum-Kulka et al. (1989).

The coding scheme of Blum-Kulka et al. (1989) is mainly divided into *address term(s)*, *head act* and *adjunct(s) to head act*. As an example, the following utterance can be divided into three segments, cf. Blum-Kulka et al. (1989).

*Danny, could you lend me $100 for a week. I've run into problems with the rent for my apartment.*

a) *Danny*: Address term
b) *could you lend me $100 for a week*: Head act
c) *I've run into problems with the rent for my apartment*: Adjunct

Address terms are optional elements previous to the head act. Another example for an address term is the *attention getter "Pardon me"*. The head act is the nucleus of the speech act, i.e. that part which serves to realize the intent. Adjuncts are optional supplementary elements such as *grounders* which indicate the reasons for the request.

Within the head act part Blum-Kulka et al. (1989) identify nine strategies of such request realizations, ranging from a direct, explicit level over a conventionally indirect level to a non-conventional indirect level. Table 1 exemplarily shows some of the strategies that are mentioned by Blum-Kulka et al. (1989).

In order to mitigate a face-threatening act Blum-Kulka et al. (1989) and Siebold (2010) additionally identify syntactic as well as lexical modifications such as *understaters* that minimize parts of the proposition (e.g., *a bit*), or *intensifiers* (e.g., *Clean up this mess, it's disgusting*). These modifications are internal since they operate within the head act.

Furthermore, speakers have the chance to avoid naming the addressee in order to soften the impact of the imposition, cf. Blum-Kulka et al. (1989). They distinguish between four request perspectives. The patterns for the perspective dimension and examples are given in Table 2.

To sum up, Blum-Kulka et al. (1989) propose five dimensions: *Address term, strategy, modification, perspective, adjunct*. Each dimension consists of different patterns and the patterns are realized by language-specific elements.

Based on these considerations we investigate our data set in terms of the patterns the participants used for expressing task-oriented requests whereby we focus on the dimensions within the head act. The results are presented in the next sections.

## 5 Criteria for Natural Language Variability in Data for SDS

In this section, we propose a set of criteria which allow to quantify to what extent data sets cover the variability of natural language expressions. The criteria are derived from the scheme previously described and from the findings of our study.

Within our work, we mainly focus on the variation patterns of the head act. However, we want to mention that our data includes many uses of address terms such as the attention getters *Ähm, Ach, Mensch, Sag mal* (eng. *"Um, Oh, Gosh, Tell me"*) (28.7%). Adjuncts do not occur within our data set.

| Head act | | | |
|---|---|---|---|
| **Strategy** | **Modification** | | **Perspective** |
| | **Syntactic** | **Lexical** | |
| 1. Mood derivable<br>*Write a text message to my brother.* | 1. Subjunctive<br>*Could you...* | 1. Politeness marker<br>*please*<br>(*bitte*) | 1. Hearer oriented<br>*you* |
| 2. Direct question<br>*Where is the next Shell gas station?* | 2. Negation<br>*Wouldn't you...* | 2. Understater<br>*a bit*<br>(*einmal, mal, schnell, kurz*) | 2. Speaker oriented<br>*I* |
| 3. Wish<br>*I would like to have 23 degrees.* | 3. Past tense<br>*I wanted to ask...* | 3. Downtoner<br>*perhaps*<br>(*vielleicht*) | 3. Speaker and hearer oriented<br>*we* |
| 4. Reference to preparatory conditions (RPC)<br>*Could you call Barack Obama?* | | | 4. Impersonal |
| 5. Locution derivable<br>*We'll have to write a text message to my brother.* | | | |
| 6. Suggestion<br>*How about calling Barack Obama on his mobile phone?* | | | |
| 7. Keywords<br>*Radio SWR3.* | | | |
| 8. Hint<br>*I need some fuel.* | | | |

Table 3: Criteria for natural language variability in data for task-oriented SDS.

Out of the nine strategies defined by Blum-Kulka et al. (1989) five strategies appear in our data. The four strategies which do not occur are described in Table 1. In addition, we identify two strategies defined by Siebold (2010) (*wish* and *direct question*) and one strategy that has not been mentioned so far (*keywords*). Table 3 shows the strategies that appear in our data.

The first strategy, *mood derivable*, refers to German and English imperative constructions. The grammatical mood of the verb marks the utterance as a request. With the second strategy, *direct question*, speakers pose direct questions, as the name already says. The third strategy, *wish*, expresses the speaker's desire. Utterances of the fourth strategy, *reference to preparatory conditions (RPC)*, contain reference to preparatory conditions, i.e. the ability or willingness. With the fifth strategy, *locution derivable*, it is directly derivable from the semantic meaning of the locution what has to be done. Utterances of the sixth strategy, *suggestion*, contain a suggestion to do something. The seventh strategy, *keywords*, consists of keywords that represent the minimal information needed and does not contain a finite verb. Utterances of the eighth strategy, *hint*, contain reference to elements needed for the implementation of the act.

The decision tree in Figure 1 presents the distinguishing features of the strategies based on what our data analysis has revealed. An implementation of the decision tree helps to automatically categorize most of the utterances in a German data set. The suitability of this method has been validated with our data.

The modifications we detect in our data can be divided into lexical elements and syntactic variations, see Table 3. The modifications the participants used are also mentioned by Siebold (2010). The syntactic modifications that occur include *subjunctive, negation* and *past tense* and the lexical elements include

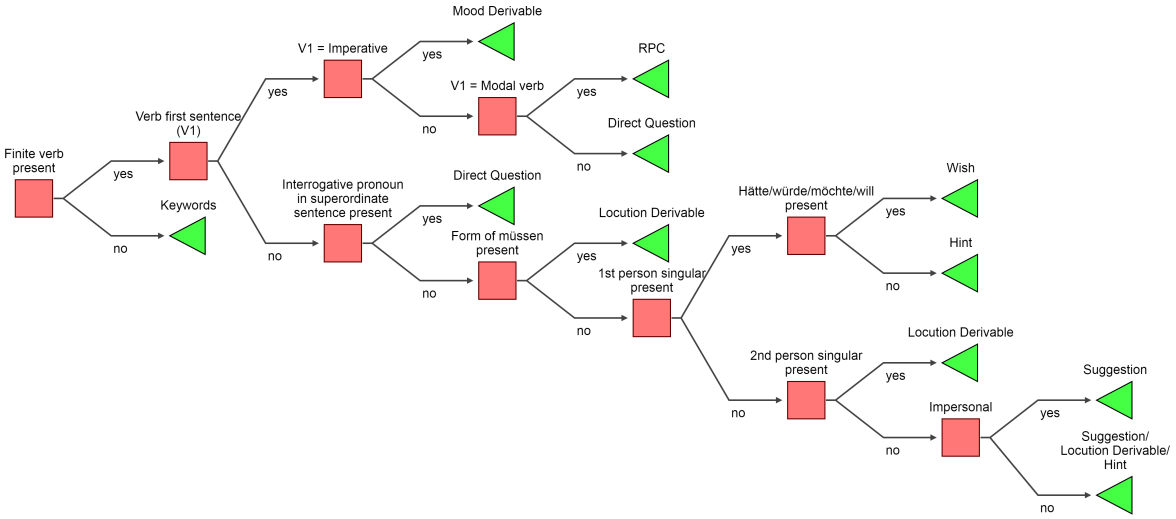


Figure 1: Decision tree for strategy categorization in German.

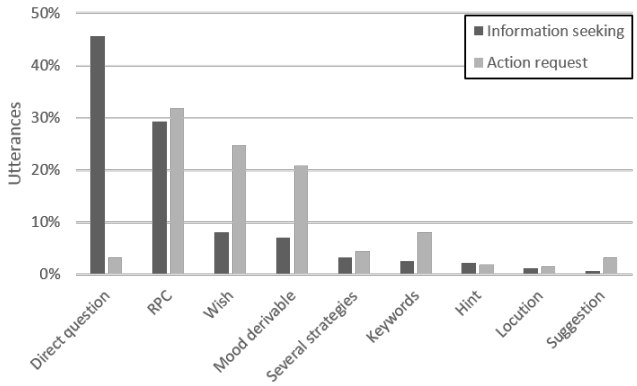


Figure 2: Strategies distribution.

*politeness markers, understaters* and *downtoners*. There is no use of so called *upgraders* that increase the compelling force of the speech act as reported by Blum-Kulka et al. (1989). The elements with which to detect the lexical modification criteria are given in Table 3 enclosed in brackets.

The analysis of the request perspectives reveals that all perspectives mentioned by Blum-Kulka et al. (1989) appear in our data.

We have shown that task-oriented requests are mainly realized in eight different ways. The study participants often mitigated such a face-threatening act by making use of six different modification patterns - lexical and syntactic ones. In addition, the study shows that people make use of all four perspectives when expressing a request. We conclude that the proposed criteria (cf. Table 3) are the most important for task-oriented requests towards SDS since they occurred within our actual utterances. Data sets that conversational task-oriented SDS have to deal with should at least cover these patterns.

## 6 Criteria Distribution

In this section, we analyze the frequency of occurrence of the criteria.

Figure 2 shows the distribution of strategies the participants used when speaking to the passenger. Since the tasks the participants had to fulfill consist of two main kinds of tasks we show the distribution broken down by information seeking tasks and action request tasks.

Figure 2 shows that most of the participants posed direct questions when they seek for information (45.6%). Many speakers also used the strategy with reference to preparatory conditions (29.3%). A few

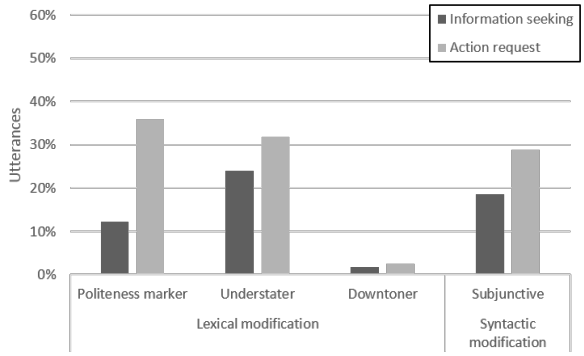


Figure 3: Modifications distribution.

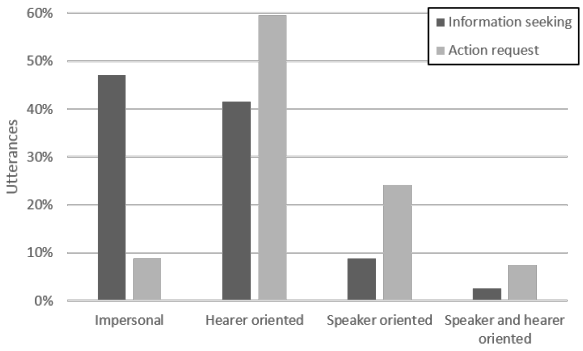


Figure 4: Perspectives distribution.

information seeking requests (2.6%) consist of a sequence of keywords. This might be an effect of being biased by the system interaction session that preceded the passenger session for half of the participants. Only 2.2% of the information seeking utterances consist of hints and 1.1% are of the locution strategy.

When requesting the passenger to perform an action, the participants mostly refer to preparatory conditions (31.9%). Expressing a desire is ranked second with over 15% and an imperative occurred in 20.7% of the action requests. 3.5% of the action requests are suggestions, 1.9% hints and 1.5% are classified as locution strategy.

We additionally found that 3.9% of all utterances consist of two strategies. The distribution between information seeking requests and action requests is nearly balanced. The first strategy of these utterances mostly expresses a desire or poses a direct question and the second strategy mostly refers to preparatory conditions or contains an imperative. An example is given in the follwing.

*Wie hat eigentlich der VfB gespielt? Kannst du das mal gucken?*
*"How was the game of VfB Stuttgart? Could you check that?"*

The first part in the example is a direct question and the second part contains reference to preparatory conditions.

The presented strategies can be internally modified by lexical and syntactic elements. The elements we identified are part of Table 3. The occurrence of these modifications is shown in Figure 3. Nearly one third of the utterances contain the politeness marker *please*. The politeness marker occured significantly ($p<0.05$) more often with action requests. Understaters occurred also very often - 27.9% of all utterances contain understaters. The most frequently used understater was the German *mal*. We found more understaters within the action requests (31.9%) than within the information seeking requests (24.1%). Downtoners, such as *vielleicht* (eng. *"perhaps"*), occurred only a few times. A subjunctive construction was also often used (23.7% of all utterances) but was less used with information seeking requests (only 18.5%) than with action requests (28.9%). Past tense was used two times. Negation was only used once, see the following utterance.

*Kannst du nicht mal das Album rein tun von Michael Jackson?*
*"Couldn't you play the album of Michael Jackson?"*

Sometimes, lexical elements are combined within an utterance. 8% of the human-directed utterances contain the politeness marker *please* in combination with an understater. In addition, we identified a combination of understater and downtoner, and downtoner and politeness marker each two times. 13% of the utterances contain a lexical element and a subjunctive. 44% of the utterances do not contain any modification elements.

Figure 4 displays the perspectives distribution. Most of the utterances are of a hearer oriented perspective (50.6%). 28% of the utterances are impersonal, i.e. there is no perspective explicitly marked. About 15% of the utterances are speaker oriented. Only 5% of the utterances tried to create a team feeling using a speaker and hearer oriented perspective. There are quite striking differences between information

| Criteria combination | Occurrence |
|---|---|
| Direct question | 44.4% |
| RPC - Understater | 12.2% |
| RPC | 7.0% |
| Wish | 4.8% |
| RPC - Subjunctive - Understater | 4.4% |

Table 4: Combinations - Information seeking tasks.

| Criteria combination | Occurrence |
|---|---|
| Wish | 10.7% |
| RPC - Subjunctive - Politeness | 7.8% |
| Wish - Subjunctive | 5.6% |
| Imperative | 5.2% |
| RPC - Politeness | 4.8% |

Table 5: Combinations - Action request tasks.

seeking requests and action requests. As for action requests, speakers tend to prefer a hearer oriented perspective (59.6%) whereas with information seeking requests they prefer an impersonal perspective (47%). A speaker oriented perspective occurred much more frequently with action requests (24.1%) than with information seeking requests (8.9%).

The perspective and the strategy a speaker chooses to realize a request are strongly interdepended. An imperative strategy, e.g., can not be realized without addressing the hearer. Also, a wish, e.g., is always formulated either in a speaker's or speaker and hearer's point of view. Therefore, the perspective dimension is disregarded in the analysis of the criteria combinations. Table 4 shows the five most frequent combinations for both, information seeking requests and action requests. The five most frequent combination patterns over all tasks represent 41.1% of the utterances. The most frequent combination pattern within the information seeking requests is the direct question strategy without any modification. This is followed by the RPC strategy combined with an understater. 7% of the information seeking requests are of an RPC strategy and 4.8% express wishes without any modification. RPC, subjunctive and understater is the fifth frequent combination pattern.

The most frequent combination pattern within the action request tasks was the wish strategy without any mitigating element (10.7%). 7.8% of the action requests are realized by an RPC strategy combined with a subjunctive and a politeness marker. This pattern is followed by the combination of the wish strategy and a subjunctive, then by an imperative and fifth, the combination of RPC and politeness marker.

## 7 Conclusion

In this paper, we have contributed criteria which can be used for quantifying the degree of natural language variability in a data set for SDS.

The criteria we have proposed are based on an existing speech act driven, language independent approach. The criteria we derived from the approach are composed of a strategy dimension, modification dimension and perspective dimension. We have presented the kind of strategies, modifications and perspectives our study participants used. The analysis of our study data has revealed that the most frequent strategies are direct questions, mood derivable, keywords, wish and reference to preparatory conditions. These strategies are modified by politeness marker, downtoners, understaters, subjunctive, past tense and negation. The perspectives we have identified include a hearer oriented, a speaker oriented, a hearer and speaker oriented, and an impersonal perspective.

We suggest these patterns to serve as criteria for the variability of natural language expressions which a task-oriented SDS must be able to understand. With the help of the criteria, their distinguishing features and the reference distribution within a real-world data set, developers are able to check the completeness as well as the representativeness of data sets for task-oriented SDS.

Our further goal is to take other languages into account. We have already collected a large amount of utterances for 150 task-oriented requests in twelve languages, european and non-european. We will examine the data with regard to the proposed criteria. This will be subject of future work.

## References


Jerome R. Bellegarda and Christof Monz. 2016. State of the art in statistical methods for language and speech processing. *Computer Speech and Language 35*, 35:163–184.

Markus Berg. 2012. Survey on spoken dialogue systems: User expectations regarding style and usability. In *XIV International PhD Workshop OWD*.

Shoshana Blum-Kulka, Juliane House, and Gabriele Kasper. 1989. *Cross-cultural pragmatics: Requests and apologies*. Ablex Publishing.

Francesca Bonin, Ronald Böck, Nick Campbell, and Ronald Poppe, editors. 2015. *Multimodal Analyses enabling Artificial Agents in Human-Machine Interaction*. Springer.

Patricia Braunger, Hansjörg Hofmann, Steffen Werner, and Maria Schmidt. 2016. A comparative analysis of crowdsourced natural language corpora for spoken dialog systems. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC)*, Portorož, Slovenia. European Language Resources Association (ELRA).

Patricia Braunger, Wolfgang Maier, Jan Wessling, and Steffen Werner. 2017. Natural language input for in-car spoken dialog systems: How natural is natural? In *Proceedings of the 18th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*.

Penelope Brown and Steven Levinson. 1987. *Politeness: Some universals in language usage*. Cambridge University Press.

Erving Goffman. 1955. On face-work: an analysis of ritual elements in social interaction. *Psychiatry: Journal for the Study of Interpersonal Processes*, 18:213–231.

James Henderson and Filip Jurcicek, 2012. *Data-Driven Methods for Adaptive Spoken Dialogue Systems*, chapter Data-Driven Methods for Spoken Language Understanding, pages 19–38. Springer.

Hansjörg Hofmann, Ute Ehrlich, Andre Berton, and Wolfgang Minker. 2012. Speech interaction with the internet - a user study. In *Proceedings of the Eighth International Conference on Intelligent Environments*.

David R. Large, Leigh Clark, Annie Quandt, Gary Burnett, and Lee Skrypchuk. 2017. Steering the conversation: A linguistic exploration of natural language interactions with a digital assistant during simulated driving. *Applied Ergonomics*, 63:53–61.

Ernisa Marzuki, Chris Cummins, Hannah Rohde, Holly Branigan, and Gareth Clegg. 2017. Resuscitation procedures as multi-party dialogue. In *Proceedings of the 21th Workshop on the Semantics and Pragmatics of Dialogue (Semdial)*.

Michael McTear, Zoraida Callejas, and David Griol. 2016. *The Conversational Interface. Talking to Smart Devices*. Springer.

Jörg Meibauer, Ulrike Demske, Jochen Geilfu-Wolfgang, Jürgen Pafel, Karl Heinz Ramers, Monika Rothweiler, and Markus Steinbach. 2007. *Einfhrung in die germanistische Linguistik*, volume 2. Auflage. J.B.Metzler.

Bo Pang and Ravi Kumar. 2011. Search in the lost sense of "query": Question formulation in web search queries and its temporal changes. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2*.

Silvia Pareti and Tatiana Lando. 2018. Dialog intent structure: A hierarchical schema of linked dialog acts. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC)*.

Kathrin Siebold, 2010. *Bestandsaufnahme der Germanistik in Spanien: Kulturtransfer und methodologische Erneuerung*, chapter Wie Spanier und Deutsche ihr Gesicht wahren - Eine kontrastive Untersuchung verbaler Hflichkeit, pages 123–136. Peter Lang.

Arabella Sinclair, Jon Oberlander, and Dragan Gasevic. 2017. Finding the zone of proximal development: Student-tutor second language dialogue interactions. In *Proceedings of the 21th Workshop on the Semantics and Pragmatics of Dialogue (SemDial)*.

Ryen White, Matthew Richardson, and Wen tau Yih. 2014. Questions vs. queries in informational search tasks. Technical report, July.

Ute Winter, Tim J. Grost, and Omer Tsimhoni. 2010. Language pattern analysis for automotive natural language speech application. In *Proceedings of the Second International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI)*.