

Investigating Proactivity in Task-Oriented Dialogues

Sofia Brenna

*Fondazione Bruno Kessler - Trento, Italy
Free University of Bozen-Bolzano - Bolzano, Italy*

SBRENN@FBK.EU

Elisabetta Jezek

University of Pavia - Pavia, Italy

ELISABETTA.JEZEK@UNIPV.IT

Bernardo Magnini

Fondazione Bruno Kessler - Trento, Italy

MAGNINI@FBK.EU

Editor: Manfred Stede

Submitted 03/2024; Accepted 03/2025; Published online 03/2025

Abstract

This paper investigates *proactivity*, a characteristic phenomenon of collaborative human-human interaction, where a participant in the dialogue offers the addressee some useful and not explicitly requested information. More precisely, a proactive behaviour is: (i) self-prompted and not simply reactive, that is, the speaker does not act merely in response to the requests the other participant has made; (ii) somehow effective for the achievement of the dialogue goal, since the speaker has a long-term, goal-directed behaviour that predicts future states and needs. Proactivity has been poorly investigated from a theoretical point of view, and there is a general need of empirical data for both quantitative and qualitative research. The paper provides an extensive analysis of proactivity in several human-human task-oriented dialogic corpora, selected with different characteristics, including chat exchanges and telephone calls, collection modalities such as natural setting and Wizard of Oz, and two languages, Italian and English. The main result is the *D-Pro Corpus*, a new resource manually annotated at the utterance level with proactivity and dialogue acts, which allows to investigate proactivity in the context of task-oriented dialogues. There are several findings from our empirical investigation of proactivity: (i) we find that about 20% of turns in our corpus are proactive turns, showing that this is a very diffused and relevant phenomenon; (ii) we confirm the non-reactive nature of proactivity, highlighting the presence of a pattern where a turn in the dialogue triggers a reaction in a following turn and a proactive utterance is then added to the turn; (iii) we show that only a limited number of dialogue acts are actually involved in expressing proactivity, and we discuss the theoretical implications of this finding; (iv) we empirically confirm that proactivity has a crucial role in recovering from goal-failure situations, contributing to the effectiveness of the whole dialogue; (v) we support the intuition of a non-uniform distribution of proactive utterances throughout the dialogue. Our empirical findings and the D-Pro Corpus provide relevant insights for deeper theoretical investigations, as well as crucial resources for improving proactivity in current task-oriented dialogue systems.

Keywords: proactivity, task-oriented dialogue, annotated resources

1. Introduction

Human dialogue is a complex interaction characterised by systematic, coordinated behaviours and a collaborative effort on the part of each participant to communicate (Ellefson, 2021). Collaborative behaviours in dialogue refer to the various actions and strategies employed by participants to adapt appropriately to each other and work together towards effective communication, shared understanding, and the achievement of conversational goals. There are several collaborative behaviours that have been individuated, which include *grounding* (Clark and Schaefer, 1987, 1989; Clark and Brennan, 1991; Clark, 1996), *clarification requests* (Purver et al., 2003a,b), *backchanneling* (Shelley and Gonzalez, 2013), *proactivity* (Strauß and Minker, 2010; Balaraman and Magnini, 2020a,b), *reformulation* (Fetzer, 2006), *giving examples*, and *convergence / divergence / maintenance* (Giles and Ogay, 2006). However, although some of such collaborative behaviours have received attention, particularly from the perspective of the development of computational dialogue systems, most of them are still under-investigated in recent data-driven approaches to dialogue models, resulting in a substantial under-representation of collaborative behaviours in human-machine dialogues. This situation is quite evident in the area of task-oriented dialogues (Mctear, 2020; Louvan and Magnini, 2020; Balaraman et al., 2021) and conversational search (Radlinski and Craswell, 2017), where, despite the huge application-oriented interest, there is a gap of empirical studies on collaborative phenomena.

The purpose of this study is to investigate *proactivity*, a collaborative phenomenon representing a fundamental property of human interaction. Proactivity can be regarded as the ability to provide the addressee with some useful, yet not explicitly requested information. In Example (1), an excerpt from a human-human dialogue between a Client (C) and an Agent (A) is reported; the Agent, in utterance U19, reacts to a question about a point of interest (*Does it have an entrance fee?*) answering the question (*That information is not available to me*). In the same turn, in utterance U20, the Agent takes a non-requested, non-reactive initiative, providing a phone number, which was not explicitly required. We regard utterance U20, marked with PRO, as a proactive utterance.

EXAMPLE (1)

c: **U18** Does it have an entrance fee?
a: **U19** That information is not available to me.
 U20 [PRO] The phone number is 00872208000.¹

Although we have an intuition that proactive behaviours are widespread in human dialogues, to our knowledge there is a lack of quantitative and qualitative analysis supporting this intuition. In our investigation we focus on task-oriented dialogues, because we believe that their inherent collaborative nature (participants jointly aim at and contribute to the achievement of one or more communicative goals) should encourage proactivity. We address the following research questions: (i) what is the amount of proactive utterances in task-oriented dialogues? (ii) is there a relation between proactivity and the dialogue acts employed by the dialogue participants? (iii) are there typical linguistic markers of proactivity? (iv) how is proactivity distributed along the flow of a task-oriented dialogue?

To address our research questions, we start by formulating an operative definition of proactivity, that we then use to annotate a selected sample of human-human task-oriented dialogues. In the

¹Example taken from the MultiWOZ 2.2 corpus, cfr. Zang et al. (2020). A = Agent, C = Client.

annotation effort, we focus on a few relevant aspects of proactivity, including the relation between proactivity and dialogue acts, goal-failure situations, and the relation with utterances that typically precede proactivity. As for the data, we exploit five already existing task-oriented dialogue collections, with different characteristics in terms of language, conversational domain, media used for exchanging turns, and collection modalities. The resulting annotated corpus (called D-Pro Corpus²) is freely distributed for further research, thereby compensating the absence of quantitative studies on the presence of proactivity in task-oriented dialogue corpora, especially with regards to the Italian language.

There are several findings from our empirical investigation of proactivity: (i) we find that about 20% of turns in our corpus are proactive turns, showing that this is a very diffused and relevant phenomenon. In addition, we find that proactivity is more frequent in spontaneous corpora (e.g., social media chat) than in corpora collected through a guided process (i.e., Wizard of Oz); (ii) we confirm the non-reactive nature of proactivity, highlighting a pattern in which a turn in the dialogue triggers a reactive utterance in the subsequent turn, followed by the addition of a proactive utterance to the triggered response; (iii) we show that only a limited number of dialogue acts are involved in expressing proactivity, with the vast majority of proactive utterances serving the communicative intent of providing information (60%), suggestions (13.9%), or offers (12.5%); (iv) we empirically confirm that proactivity has a crucial role in recovering from goal-failure situations, contributing to the effectiveness of the whole dialogue; furthermore, that dialogues characterized by higher levels of proactivity experience the fewest instances of failure; (v) we provide evidence supporting the intuition that proactive utterances are non-uniformly distributed throughout the dialogue, with a higher concentration observed in the central segments. Our empirical findings and the D-Pro Corpus provide relevant insights for deeper theoretical investigations, as well as crucial resources for improving proactivity in current task-oriented dialogue systems.

The paper is structured as follows. Section 2 situates proactivity in the context of linguistics and natural language processing. Section 3 introduces the definition of proactivity we adopt in our study and presents our annotation scheme of proactivity, while Section 4 provides information about our source corpora and the resulting D-Pro Corpus, including statistics about the number of dialogues, turns and utterances it contains, and its lexical richness. In Sections 5 and 6, we discuss the results of the annotation of proactivity at the utterance-, turn-, and dialogue-act levels. Finally, in Section 7 we investigate how proactive utterances are positioned within the flow of a task-oriented dialogue. In 8 we report our concluding observations and ongoing work.

2. Background and Related Work

Collaborative behaviour in dialogue refers to participants' various actions and strategies to work together towards effective communication, shared understanding, and attainment of conversation goals. These behaviours help maintain the flow, coherence, and relevance of the dialogue while ensuring that all participants have the opportunity to contribute and be heard.

As referenced in the Introduction, among the most prominent linguistic techniques that participants can use for collaborative purposes in a dialogue, we find *proactivity*. Derived from the definition of proactivity in organisational behaviours (Grant and Ashford, 2008), the term proactivity has been used in Natural Language Processing since at least Li et al. (2016) to refer to conversational agents' capability to create or control the conversation by taking the initiative and anticipating the

²https://github.com/sofiabrenna/d-pro_corpus

impacts on themselves or human users, rather than passively responding to the user’s request (see Deng et al. (2023) for an overview).

From a theoretical point of view, the concept of collaborative behaviour - within which proactivity is couched - is not tied to a single theory but instead emerges under different terminologies in different traditions of studies, including philosophy and pragmatics. Among the first systematic attempts to specify the rules governing participants’ collaborative behaviour in human interactions, we recall H. Paul Grice’s *cooperative principle* and *maxims of conversation* (Grice, 1975, 1989), the latter often interpreted as a way of spelling out what the principle itself articulates. Grice’s main focus, however, was not to provide a fully-fledged theory of cooperation in human interactions but rather to account for how participants in a communicative exchange derive the implicated meaning of their interlocutors’ utterances, particularly in cases where there is no apparent relation between the utterances.³ The work of J.L. Austin (Austin, 1962) and J. Searle (Searle, 1969, 1975) integrates Grice’s contribution by identifying a typology of speech acts and illocutionary forces and by examining their application condition in detail. Their proposal has been taken up in computational linguistics under the label of *dialogue act* (Stolcke et al., 2000), *conversation act* and *intent* (Bunt et al., 2010; Bunt and Girard, 2005; Bunt, 2006; Traum and Hinkelman, 1992).

In other works originating from the social psychology of language, the concept of *accommodation* has been put forth, which offers a theoretical framework for analysing *proactivity* in NLP. *Accommodation* is the process of modifying one’s communication style, vocabulary, code, and tone (including politeness, Brown and Levinson (1987); Bargiela-Chiappini (2003)) to better align with a conversation partner (cf. *speech accommodation theory* (Giles et al., 1973; Giles, 1979; Giles et al., 1991; Giles and Powesland, 1997; Burt, 1994; Scotton, 1988) and *communication accommodation theory* (Giles and Ogay, 2006)). This adaptation facilitates understanding, promotes effective collaboration, and fosters a positive interactional atmosphere. Accommodation has already been investigated in the design of spoken dialogue systems (cf. vocal accommodation in Raveh (2021) and prosodic accommodation in De Looze et al. (2014)).

A significant body of work has also been dedicated to the concept of participant *initiative* in a dialogue. According to Traum (1997), *a speaker would have the initiative if the speaker had the choice as to the content of the utterance, while the other speaker would have the initiative if the speaker had to frame the utterance in response to speech by the other speaker*. This concept is closely connected to proactivity, since being proactive inherently requires taking the initiative to anticipate future needs, rather than responding reactively. Initiative has been studied in dialogue and discourse analysis in several contexts, for example in task-oriented and advisory dialogues (Whittaker and Stenton, 1988; Walker and Whittaker, 1990), in learning environments (Core et al., 2003; Kersey et al., 2009), in overlaps of speech (Yang and Heeman, 2010), in multi-party dialogues (Strauß and Minker, 2010), and in negotiation dialogues (Nouri and Traum, 2014). Research on mixed-initiative dialogues initially focused mainly on monitoring the *control* flow in dialogue (Whittaker and Stenton, 1988) showing that control shifts are predictable based on utterance type.

³See the following example taken from Grice (1975), 51:

A: Smith doesn’t seem to have a girlfriend these days.
B: He has been paying a lot of visits to New York lately.

In such cases, cooperation is needed as a requirement on the behaviour of speakers to reconstruct the *unstated connection* between the utterances, to go beyond what is said and to understand what is meant (B implicating that Smith has, or may have, a girlfriend in New York). Note that what Grice actually meant by cooperative is still controversial (see Ellefson 2021 for a thorough discussion).

Walker and Whittaker (1990) equate initiative to control, associate four utterance types with the allocation of control to either participant, and identify types of control shift⁴. They suggest that control transfer in mixed-initiative dialogues is often collaborative, even during interruptions when the non-controlling participant takes initiative. In such cases, interruptions help align mutual beliefs needed for the collaborative plan, supporting rather than obstructing the dialogue goal achievement. Guinn (1998) explores effective collaboration when participants rely on each other to achieve a common goal, viewing initiative as the decision-making power to manage sub-tasks. He argues that having initiative in task management corresponds to having initiative in dialogue management. Challenging this monolithic view, Smith (1992) introduces *variable initiative*, while Cohen et al. (1998) propose a non-binary perspective with varying degrees of initiative. Chu-Carroll and Brown (1999), followed by Kersey et al. (2009) and others, distinguish between *dialogue initiative* and *task initiative*: dialogue initiative is held by the participant guiding the conversation, while task initiative belongs to the one leading goal planning. This distinction separates the two types of initiative, aligning with Jordan and Di Eugenio (1997), who refute Walker and Whittaker (1990) arguing that control pertains to the dialogue level, whereas initiative relates to the problem-solving level.

We propose that proactivity operates at an additional level—the turn or utterance level. This view partly aligns with Nouri and Traum (2014), who develop an annotation scheme for initiative and *response* behaviour within dialogue turns. They distinguish between two aspects of initiative: establishing new discourse obligations and providing unsolicited material. On the response side, they examine fulfilling obligations and maintaining relevance to prior turns, reflecting Sperber and Wilson (1986)’s notion of *relevance*. We define proactivity as closely related to the second aspect of both initiative and response, involving unsolicited yet relevant contributions to the dialogue’s topics and goals. In Section 3.1, we will further elaborate on this by seeking to define proactivity more precisely.

Several attempts have been made to design dialogue systems that enable the conversational agent to behave proactively, for example, by introducing new topics or useful suggestions during the conversation. Particularly, in task-oriented dialogues, proactivity has been addressed primarily in the so-called non-collaborative dialogues, where the system and the user may have divergent objectives or conflicting interests regarding the completion of the task (e.g., the price bargain negotiation), and in enriched task-oriented dialogues (Balaraman and Magnini, 2020a), where the Agent takes the initiative to provide useful supplementary information not explicitly requested by the user (e.g., additional knowledge or chitchats), which can improve the quality and effectiveness of conveying functional service in the conversation. Finally, Sun et al. (2021) constructed the AC-CENTOR dataset by adding topical chit-chats into the responses for task-oriented dialogues to make the interactions more engaging and interactive.

Despite the progress, the NLP community still lacks a comprehensive framework that brings together all the concepts related to proactivity under a unified theoretical perspective. We believe that a deeper understanding of proactivity is essential for improving dialogue system design aimed at simulating natural interactions, and our work is an effort to contribute in this direction.

⁴Walker and Whittaker (1990) note that *assertions*, *commands*, and *questions* are typically produced by the controlling participant, whereas prompts leave the control to the hearer. A parallel can be drawn between these controller utterance types and the dialogue acts that we identify as conveying proactivity: assertions with *inform*, *offer*, and *suggest*, commands with *request* and *instruct*, questions with *requests*. Yet, we believe that while proactivity always entails some sort of initiative, the same is not necessarily true for control. Criticism of the equivalence between initiative and control is addressed further ahead in this section.

Accordingly, there is a shortage of materials and resources specifically focused on proactivity that we can rely on. However, in recent years, possibly due to the impressive performance of Large Language Models, proactivity has gained significant attention and has been incorporated into data annotation efforts, also in neighbouring fields such as Human-Computer Interaction (HCI). A notable work in this area is ProDial (Kraus et al., 2022), a collection of mixed-initiative human-computer collaborative interactions including different levels of proactive dialogue actions, meant to create a proactive dialogue model. The dataset contains 3,696 system-user exchanges, collected in a serious game setting based on crowd-sourced interactions with an autonomous agent capable of modelling four actions of proactive behaviour (None, Notification, Suggestion, Intervention). The dialogue actions corpus has been annotated with user information and self-reported assessments of the user’s experience with the dialogue system’s behaviour. While the main ProDial focus is on the human-computer trust relationship, our goal is to investigate proactivity in human-human dialogues.

3. Annotating Proactivity

In this section we first present an operative definition of proactive behaviour and then we introduce the schema we developed to annotate proactivity in dialogues. The purpose is to extract useful quantitative and qualitative data about proactivity in human-human task-oriented dialogues.

3.1 Defining Proactivity

We have introduced proactivity (see Section 1) as *the ability to provide the addressee with some useful and not explicitly requested information*. A more operative definition is proposed in Balaraman and Magnini (2020a), where a proactive behaviour, in the context of a task-oriented dialogue system, is defined as any information that: (i) is introduced by the system; (ii) was not previously introduced in the dialogue by the user; and (iii) is assumed to be relevant to achieve the user needs. While this definition has the merit to relate proactivity with information content, which can be somehow located (i.e., *annotated*), it requires that *proactive units* (Balaraman and Magnini, 2020a) are exactly located within dialogue utterances, making the annotation effort excessively complex. In addition, the definition does not consider the proactive contribution of the user in the dialogue, which is instead a crucial one. In this study, we still base proactivity on information content, although adopting a more comprehensive and usable definition. We say that an utterance in the context of a task-oriented dialogue is considered as proactive when one of the participants, either Agent or Client:

1. does not act merely in response to the requests the other participant has made, so the behaviour is self-prompted and not simply reactive;
2. has a long-term, goal-directed behaviour that predicts future states and needs, so the behaviour is somehow effective for the achievement of the dialogue goal.

When these two conditions are satisfied, the corresponding utterance is marked with the tag PRO. In example (2), from the Italian corpus JILDA (Sucameli et al., 2021), utterance U8 (*Dopo aver fatto la triennale a Roma*, see translation in footnote) and utterance U11 (*però ci sono delle opportunità di lavoro su Roma.*) are marked as PRO, because they are not a direct answer to the interlocutor’s request, and, nonetheless, they provide a piece of useful information for the dialogue.

EXAMPLE (2)

- A:** **U7** Hai qualche preferenza riguardo al luogo di lavoro?
- c:** **U8 [PRO]** Dopo aver fatto la triennale a Roma,
 U9 mi piacerebbe tornare verso casa, a Firenze.
- A:** **U10** Al momento non abbiamo nessun annuncio che faccia al caso tuo nella
 zona di Firenze,
 U11 [PRO] però ci sono delle opportunità di lavoro su Roma.⁵

There are a few requirements that need to be satisfied for our annotation schema to be applied. We focus on written or transcribed, mixed-initiative task-oriented dialogues that, as mentioned earlier, provide the ideal context for investigating collaborative behaviours. In such dialogues we assume a turn-taking partition of the conversation, where each turn is assigned to a *participant*: in order to make our annotation homogeneous through different dialogues, participants are generally referred to as Client, the participant who provides the initial task to be addressed, and Agent, the participant who helps the Client to solve the task (see example (1)). Finally, we assume that each turn can be segmented into the utterances that compose the turn.

The annotation schema is based on four levels: utterance annotation, dialogue act annotation, goal failure annotation, and turn adjacency annotation. We describe them in the following sections.

3.2 Utterance Annotation

The basic units we consider for proactivity annotation in a dialogue are utterances, that is, according with (Traum and Heeman, 1996), continuous pieces of speech beginning and ending with a clear pause, possibly related to paralinguistic features, including facial expressions, laughter, eye contact, and gestures. More specifically, relating the notion of utterance to dialogue acts, we can state, referencing Traum (2004), that an *utterance* can be defined as a small unit of speech or text within a conversational turn corresponding to a single act that is bordered by the speaker’s silence and/or prosodic boundary tones. Thus, as far as written chats are concerned, an utterance is bordered by the writer’s sending of a single message, for instance by pressing ‘enter’ on the keyboard or tapping the ‘send’ button on a smartphone screen, and/or punctuation. This single-act-matching concept enables us to divide conversational turns into utterances within dialogue corpora that lack an initial segmentation into utterances and align all dialogues to the same splitting criterion (see 4.1 for details).

The annotation task includes both Agent and Client utterances. The task consists in marking as proactive an utterance in its totality: the utterance itself may as well contain some non proactive behaviour, but nonetheless it should be marked as proactive whenever it holds a piece of proactive content.

⁵Example taken from the JILDA corpus (Sucameli et al., 2020). It may be translated to English as follows:

- A:** **U7** Do you have any preference about where to work?
- c:** **U8 [PRO]** After completing my Bachelor’s degree in Rome,
 U9 I would like to move back towards home, to Florence.
- A:** **U10** Currently we do not have any offer in the Florence area that matches your
 requests,
 U11 [PRO] however, there are job opportunities in Rome.

3.3 Dialogue Act Annotation

Proactive utterances are then further classified according to the dialogue act they convey. Dialogue acts refer to the performative dimension of dialogue as originally investigated by Austin (1962), and subsequently adapted to the purposes of dialogue systems through the notion of conversation acts (Traum and Hinkelman, 1992), dialogue acts (Stolcke et al., 2000; Bunt, 2006; Bunt et al., 2010), and, more recently, through the notion of intent (Louvan and Magnini, 2020).

To simplify the manual annotation task, we use a limited number of high-level dialogue acts, selected from Bunt et al. (2010)’s ISO standard taxonomy developed for annotating dialogue with semantic information. Our purpose is to employ the same dialogue act annotation schema for each of the 5 sub-corpora, so we need high-level dialogue act tags. In particular, from the schema in Figure 1, we select the following dialogue acts (general-purpose communicative functions in the ISO taxonomy), which potentially can express proactive utterances.⁶ Each dialogue act is paired with an operative definition adapted to the proactivity annotation task.

- **INFORM** = a proactive utterance where the participant provides information;
- **OFFER** = a proactive utterance where the participant proposes to do something or to provide some further information;
- **SUGGEST** = a proactive utterance where the participant suggests that the addressee should do something;
- **REQUEST** = a proactive utterance where the participant demands that the addressee do something or where they demand that the addressee provide some information;
- **INSTRUCT** = a proactive utterance where the participant provides the addressee with instructions to follow.
- **OTHER** = a label made available to annotators for tagging utterances that could not be classified within the designated 5 dialogue act labels; however, it was never used in the final set of annotations.

Given this tagset, the annotator must observe what kind of dialogue act is meant and performed by the participant in utterances that have already been observed to present some proactive behaviour (both the Agent’s and the Client’s). As an example, consider the following dialogue, where the proactive utterance U20 has been annotated with the dialogue act **INFORM**, while utterance U21, in the same turn, has been annotated with **REQUEST**.

EXAMPLE (3)

P: U15 perfect!
 U16 can meet there at 8ish?
 R: U17 Sounds good ^.^
 P: U18 i’ll be there at 8.10 is that ok?
 R: U19 Yes perfect!
 U20 [**PRO**] [**INFORM**] I’m sitting inside with an Italian guy I met at a tandem last week ^.^
 U21 [**PRO**] [**REQUEST**] tell me when you arrive!

⁶The dialogue act selection process was based on a pilot annotation (see also 4.2).

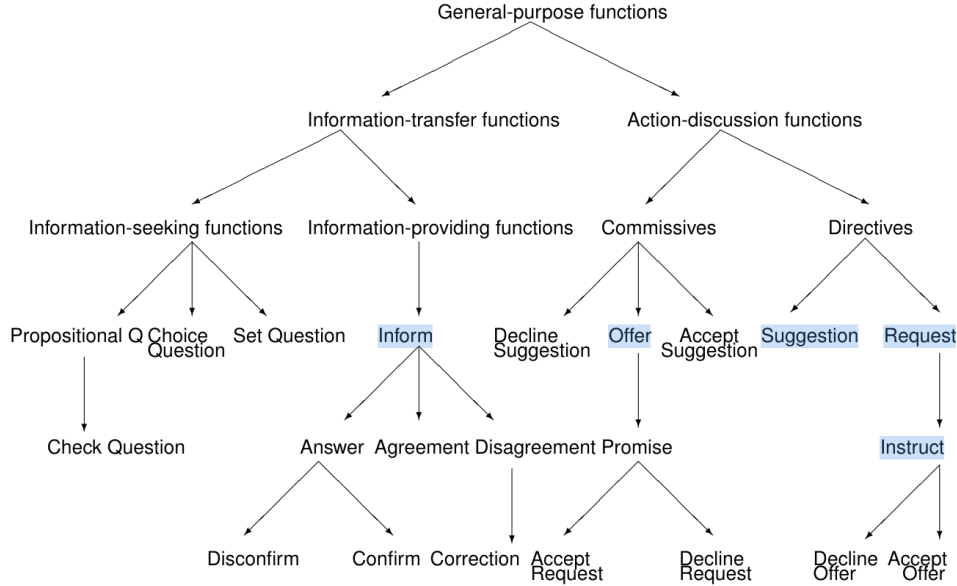


Figure 1: The ISO Standard taxonomy of general-purpose functions proposed by Bunt et al. (2010). Dialogue acts that are part of our tag-set are highlighted in blue.

P: U22 i'm here!⁷

3.4 Goal-Failure Annotation

We also annotate the presence of situations of *failure* in utterances where a participant fails in achieving a dialogue goal. Such goal failure situations are typically linguistically realised through negative answering to a question or through the impossibility to fulfil a request. Goal failure situations are interesting for proactivity because they often require some sort of repair that a proactive behaviour can conveniently bring (Balaraman and Magnini, 2020a). Thus, the annotation of failure situations can give us some insights on the correlation between these two phenomena. We annotate goal failure as follows:

- FAIL = a turn that contains a situation of failure, where the participant can not answer a question or fulfil a request.

For instance, in the following dialogue (already presented in example 2), utterance U10 at turn T5 (*"Currently we do not have any offer in the Florence area that matches your requests"*) is annotated as FAIL, and it is connected to the next utterance in the dialogue, U11, where the Agent is trying to recover from the goal failure with a practice utterance, informing the Client about a job offer of interest in the area of Rome.

EXAMPLE (4)

⁷Example taken from the Italian WhatsApp Corpus (Hewett, 2017). WhatsApp chat participants are identified by the first letter of their pseudonym; for example, "P" represents "Peter" and "R" represents "Raffaella".

A: T3 U7 Hai qualche preferenza riguardo al luogo di lavoro?
 c: T4 U8 [PRO] [INFORM] Dopo aver fatto la triennale a Roma,
 U9 mi piacerebbe tornare verso casa, a Firenze.
 A: T5 U10 [FAIL] Al momento non abbiamo nessun annuncio che faccia al caso
 tuo nella zona di Firenze,
 U11 [PRO] [INFORM] però ci sono delle opportunità di lavoro su Roma.⁸

3.5 Turn Adjacency Annotation

In the last annotation level, we consider the relation between a proactive utterance and the utterances of the previous (i.e., adjacent) turn. The intuition is that through turn adjacency annotation it will be possible to better investigate the elements in the dialogue that trigger proactivity. Practically, once a certain utterance is marked as PRO, the annotator has to look at the adjacent previous turn in the dialogue. If the utterances of the previous turn provide all required context to motivate the current PRO utterance, the ADJ tag (adjacent) is added to the current proactive utterance. On the other hand, when the adjacent turn does not suffice to provide all required context in order to motivate proactivity, the PRO utterance is labelled as NA (non adjacent).

As an example, the proactive utterances U20 and U21 in the following dialogue (Example (5)), are both marked as ADJ because the origin of the purpose of their proactivity can be found in utterance U18 in the previous turn (*i'll be there at 8.10 is that ok?*).

EXAMPLE (5)

P: T4 U15 perfect!
 U16 can meet there at 8ish?
 R: T5 U17 Sounds good ^.^
 P: T6 U18 i'll be there at 8.10 is that ok?
 R: T7 U19 Yes perfect!
 U20 [PRO] [INFORM] [ADJ] I'm sitting inside with an Italian guy I met at
 a tandem last week ^.^
 U21 [PRO] [REQUEST] [ADJ] tell me when you arrive!
 P: T8 U22 i'm here!⁹

By contrast, in Example (6), the proactive utterances U15 and U16 are annotated as NA because their proactivity is motivated by utterance U11 (*I also need to take a train on wednesday, leaving after 10:15.*), which is not the previous adjacent turn.

EXAMPLE (6)

c: T7 U11 I also need to take a train on wednesday, leaving after 10:15.
 A: T8 U12 Okay, we have a LOT of trains leaving after that time.
 U13 What is your starting point and destination?
 c: T9 U14 From Leicester to Cambridge, please.
 A: T10 U15 [PRO] [INFORM] [NA] OK, the TR9776 leaves at 11:09 and arrives at
 12:54, the cost is 37.80 pounds,
 U16 [PRO] [OFFER] [NA] do you want me to book you?
 c: T11 U17 Yes please book the train for 1 person
 U18 and make sure you give me the reference number.¹⁰

⁸Example taken from the JILDA corpus (Sucameli et al., 2020). See Example (2) for an English translation.

⁹Example taken from the Italian WhatsApp Corpus (Hewett, 2017).

¹⁰Example taken from the MultiWOZ 2.2 corpus (Zang et al., 2020).

4. Investigating Proactivity in a Task-Oriented Dialogic Corpus

As already mentioned, the focus of this study are task-oriented dialogues, under the assumption that such dialogues are likely to show collaborative phenomena among the interlocutors, including proactivity. In order to provide a representative sample of task-oriented dialogues we considered the following criteria:

- *Language.* The main language we use to investigate proactivity in task-oriented dialogues is Italian. However, in order to assess potential differences due to language, we include in our sample two English corpora: (i) one corpus (i.e., Ubuntu) with comparable (same communicative situation) Italian and English dialogues, and (ii) one English corpus (i.e., MultiWOZ). This choice allows for a comparison of proactivity in the two languages, reported in Section 5.3.
- *Medium.* We consider the diamesic dimension of dialogue, aiming at including in our sample a sufficient variety of communication media, including telephone call transcriptions (i.e., NESPOLE!), IRC chat (i.e., Ubuntu), social media chat (i.e., WhatsApp), and chat-based platforms (i.e., MultiWOZ and Jilda).
- *Data collection methodology.* We try to balance dialogues collected with several methods (e.g., Wizard of Oz, role-taking, ecological settings), in order to assess how proactivity might be influenced by different degrees of spontaneity and naturalness in the speakers—or writers, as well as dialogues showing different levels of lexical variety and syntactical complexity.
- *Participants.* All our dialogues are intended to represent human-human communication. We included both two-party dialogues and multi-party dialogues (for instance, Ubuntu and partially WhatsApp). Although MultiWOZ is collected through Wizard of Oz, this corpus is generally considered composed of human-human dialogues (see, for instance, Budzianowski et al. (2018); Paul et al. (2019); Wu et al. (2019)), as the Agent-Wizard dialogues were generated by humans, as opposed to human-machine dialogues.
- *Domain.* Our corpus selection covers a variety of interaction domains, including simulated professional support on different topics (e.g., NESPOLE!, MultiWOZ, and JILDA), technical support (e.g., Ubuntu), and informal social interactions (e.g., WhatsApp).

Table 1 summarizes the five source corpora we selected, as well as their main characteristics. More details for each dialogue corpus are reported in the next section.

4.1 Source Dialogic Corpora

In accordance with the outlined criteria, as sources for our study on task-oriented dialogues we have considered five existing corpora: the Italian NESPOLE! Corpus, the Italian WhatsApp Corpus, the Italian Ubuntu Chat Corpus, MultiWOZ 2.2, and the JILDA Corpus (Table 1).

NESPOLE! (Mana et al., 2003) is a VoIP human-human role-taking (Anderson et al., 1991) phone call dialogue collection, part of the multi-language and multi-modal NESPOLE! project. NESPOLE! dialogues span two domains: medicine and tourism. For our analysis, we focus on the 56 Italian dialogues from the tourism domain (total recording time: 7h 35'), where a tourist (Client) calls a travel operator, and the Agent's goal is to arrange a vacation for them in the Trentino region

Corpus	Year	Language	Medium	Methodology	Participants	Domain
NESPOLE!	2003	Ita	VoIP call	role-taking	human-human	tourism
Ubuntu	2013	Ita-Eng	IRC chat	natural	human-humans	tech support
WhatsApp	2017	Ita	social media chat	natural	human-human(s)	private chat
MultiWOZ 2.2	2020	Eng	chat	Wizard of Oz	human-wizard	multi domain
JILDA	2021	Ita	chat	role-taking	human-human	job offer

Table 1: A synoptic view on the dialogue corpora that have been analysed in our study, in chronological order; in those cases where only one specific part of a corpus has been used, such as with the Italian NESPOLE! corpus and the Italian Ubuntu Chat Corpus, the information is given about that specific part of the corpus.

of Italy.

Data preparation: dialogues were already segmented into both turns and utterances, and pauses and prosodic boundaries are transcribed. Minor edits have been performed in terms of utterance segmentation according to the single-act-matching criterion.

Ubuntu Chat Corpus (Uthus and Aha, 2013) is a multilingual Internet Relay Chat corpus of multi-party human-humans chats, composed of archived logs from Ubuntu’s IRC technical support channel for Ubuntu users, later assembled into dialogues by Lowe et al. (2015). The Italian channel chats is rather thin compared to the English chats: it collected 645,375 messages from over 10,300 users, while the English channel reached over 26,360,000 messages from almost 530,000 users. Users of the platform, identified by nicknames, ask other users to help them with technical issues related to the Linux-based operating system Ubuntu. The user initiating the help request is regarded as Client, users that help them are regarded as Agents.

Data preparation: the chats were already divided into messages, which roughly correspond to utterances, depending on the participant’s preference for shorter or longer messages. Minor edits to utterance division have been made.

Italian WhatsApp Corpus (Hewett, 2017) is composed of Italian two-party and multi-party chat dialogues, consisting in samples of WhatsApp private conversations from users based in Germany and Italy. We manually searched the 6,640 messages composing the corpus for well identifiable tasks and used excerpts from conversations that contained task-oriented dialogues, creating a Task-Oriented WhatsApp sub-corpus. Due to participants’ code-mixing and code-switching, minor parts of the corpus are in English and a few utterances are in German. Thus our WhatsApp sub-corpus is for the most part an Italian dialogue corpus, while containing 3 dialogues with one or more German utterances and 4 dialogues with one or more English utterances.

Data preparation: the same procedure was applied for utterance division as with Ubuntu.

MultiWOZ 2.2 (Zang et al., 2020) is an updated version of the widely used MultiWOZ corpus (Budzianowski et al., 2018), which gathers 8,438 English multi-domain written short dialogues (average turn per dialogue is 13.68), collected through the Wizard-of-Oz method (Kelley, 1984). Scripted conversations take place in various domains between a tourist Client (the User) and an information centre Agent (the System, namely the Wizard pretending to be a conversational machine). Notwithstanding the participants’ expectations caused by the Wizard of Oz framework, MultiWOZ is regarded as a human-human dialogue dataset in the literature and by its authors. A finer-grained

ontology than ours is used to annotate dialogue acts for the System turns only: Inform, Request, OfferBook, ReqMore, Bye, Offer, BookInform, Welcome, Recommend, NoOffer, Select, Greet. Data preparation: each dialogue turn corresponds to one single message, so messages were manually split into utterances.

JILDA (Sucameli et al., 2021) is an Italian corpus of 525 human-human dialogues in the job search and offer domain, collected through the role-taking method (Anderson et al., 1991), in a two-party online chat: a Client, who is looking for a job, is assisted by an Agent in his goal. The corpus is annotated for the presence of proactive information and with the following dialogue act ontology: greet, inform-basic, inform-proactive, request, select, deny. Data preparation: the same as for MultiWOZ applies.

4.2 The D-Pro Corpus

From the five source corpora presented in Table 1, we extract a smaller corpus called *Dialogue Proactivity Corpus (D-Pro)*, and manually annotate it in accordance with the schema presented in Section 3. As dialogues from different sources may have different length (e.g., dialogues in NESPOLE! are much longer than dialogues in MultiWOZ), from each source corpus, we sample a sub-corpus containing a total of about 600 turns. The resulting corpus consists in a total amount of 151 dialogues, divided into 2,855 turns and 6,028 utterances.¹¹

The annotation process of the D-Pro includes the following steps:

- Guidelines creation. For the purposes of annotation, a document with precise guidelines and examples for the annotators is realized, suited to clarify any doubts and to give a thread the annotator could follow in places where the parting line between what is considered proactive and what is considered not-proactive becomes blurred, and therefore a more subjective and annotator-dependent decision has to be made.
- Pilot annotations and guidelines revision. An expert annotator is selected and provided with the guidelines for a pilot annotation of 15 dialogues. After the pilot annotation, the Guidelines are slightly modified and the dialogue act annotation schema is consolidated, as a response to the feedback given by the annotator.
- Inter Annotator Agreement. After the pilot exercise, a second expert annotator is selected, and a portion of 15% of the D-Pro dialogues is annotated by the two annotators, in order to estimate their agreement. A detailed description of the Inter Annotator Agreement is presented in Section 4.3.
- Extensive D-Pro annotation. As the inter annotator agreement was high, in the last phase the two annotators are engaged in the extensive annotations of the whole D-Pro Corpus. This phase lasts for about two months, with the annotation of a single dialogue taking between half an hour to two hours of effort, depending on its length and complexity.

Table 2 presents the composition of the D-Pro Corpus.

¹¹The target of 600 turns per sub-corpus is not met for the WhatsApp sub-corpus due to the modest size of the Italian WhatsApp source corpus: 45 dialogues were extracted, yielding 401 turns overall. The WhatsApp sub-corpus was added at a later stage, and the annotation of 600 turns had already been completed for the other four sub-corpora by the time the WhatsApp sub-corpus was incorporated.

D-Pro Corpus	NESPOLE!	Ubuntu	WhatsApp	MultiWOZ	JILDA	D-Pro Tot.	D-Pro Micro Avg.
# dialogues	14	22	45	41	29	151	30.2
# turns	605	612	401	602	635	2855	571
# Agent turns	305	393	/	301	321	1320	330
# Client turns	300	219	/	301	318	1138	284.5
# utterances	1722	1181	959	963	1203	6028	1205.6
# tokens	11493	7437	4880	7907	9593	41310	8262
# types	1493	1983	1762	1048	1681	7966	1427.8
# lemmas	1092	1466	1249	688	1127	5622	1124.4
TTR	12.99	26.66	36.11	13.25	17.52	/	15.03
Avg. turns per dial.	43.21	27.82	8.91	14.68	21.9	/	18.90
St.Dev. turns per dial.	18.51	21.87	4.76	4.81	2.98	/	14.79
Avg. utt. per dial.	123	53.68	21.31	23.49	41.48	/	39.92
Avg. utt. per turn	2.85	1.93	2.39	1.60	1.89	/	2.11

Table 2: The composition of the D-Pro Corpus. The upper part of the table reports dialogue information and the middle one reports lexical information; the lower part reports statistical measures on dialogues. TTR = *type/token ratio*.

Statistics about dialogues. Dialogue numbers vary significantly across corpora. For instance, NESPOLE! contains approximately one-third of the dialogues found in WhatsApp. This variation is accompanied by notable differences in dialogue length, both between corpora—as reflected in the average turn counts (NESPOLE! 43.21 vs. WhatsApp 8.91)—and within individual corpora, as indicated by the standard deviation values. During annotation, it was also observed that each sub-corpus contains at least one dialogue that is twice as long as another, with the exception of JILDA, where dialogue lengths range more narrowly from 17 to 27 turns.

Statistics about turns and utterances. Shifting to a turn-level perspective, we can determine average turn length by looking at the average number of utterances contained in each turn. In accordance with the preceding discussion, NESPOLE! tops by far the other sub-corpora, followed by WhatsApp; both exceed the average threshold of 2.11 utterances per turn.

The division of dialogue turns per speaker, identified by their role of either Client or Agent, is included in Table 2. The two roles can not be easily assigned to users in most of WhatsApp dialogues, especially in group chats, where tasks are equally shared by the participants and where roles are flexible and can be taken on and left by any user during the course of the same dialogue. A similar situation occurs in multi-party dialogues of the Ubuntu sub-corpus, where, however, the chat room’s netiquette regulations demand that the user in need of aid directly announce their technical issue, hence facilitating the identification of the Client. The multi-party nature of Ubuntu dialogues justify the disproportion of turn allotted to the single Client versus the several Agents.

Statistics about lexicon. The counts of tokens, types, and lemmas provide an estimate of the size of each sub-corpus’s vocabulary, while the *type/token ratio* (TTR) offers insights into lexical richness and variety. TTR is considered an indicator of lexical diversity, with higher values indicating larger variability of the corpus vocabulary. TTR is affected by the length of the corpus, which makes comparisons between larger sub-corpora, like NESPOLE!, and smaller ones, like WhatsApp, less

meaningful. A higher TTR for Ubuntu compared to MultiWOZ and JILDA suggests greater lexical diversity, implying that Ubuntu has a more varied vocabulary per unit of text. This indicates that Ubuntu, despite having longer dialogues, contains a higher proportion of unique words compared to MultiWOZ and JILDA.

4.3 Inter Annotator Agreement

As referenced above, as an indication of the quality of the annotation, a portion of 15% of the D-Pro Corpus is selected to be manually annotated by two annotators, and their annotations are compared to calculate the agreement among them. Before entrusting the whole annotation work to the second annotator, some training is done, which implies a pilot annotation and confrontation on a selection of dialogues from each of the five sub-corpora. The annotation includes about 70 turns for each of the five source sub-corpora, with a total of 23 dialogues, 375 turns, and 896 utterances.

For the IAA calculation the metric used is Cohen’s Kappa coefficient, illustrated by Landis and Koch (1977) and Pustejovsky and Stubbs (2012).¹² The evaluation of the agreement through the resulting Kappa is made with reference to Landis and Koch (1977)’s graduated scale for the k value.

Table 3 reports the results of the Inter-Annotator Agreement over PRO and FAIL annotation computed at both utterance and turn level and of the dialogue act annotation. Each computation is made per single sub-corpus and with all sub-corpora combined together, namely on D-Pro Corpus as a whole.

Annotation level	NESPOLE!	Ubuntu	WhatsApp	MultiWOZ	JILDA	D-Pro
PRO Utterance	0.77	0.41	0.63	0.85	0.76	0.77
PRO Turn	0.81	0.45	0.66	0.84	0.81	0.72
FAIL Utterance	1.0	0.49	0.87	1.0	1.0	0.89
FAIL Turn	1.0	1.0	0.88	1.0	1.0	0.96
Dialogue Act Utterance	0.74	0.62	0.92	1.0	0.72	0.84

Table 3: IAA per single sub-corpus and on the whole D-Pro (all sub-corpora combined together) computed with Cohen’s kappa for both utterance-level and turn-level PRO and FAIL annotation, and for dialogue act annotation.

The outcomes for utterance-level PRO reveal an almost perfect agreement (0.85) on the annotation of the most structured sub-corpus, namely MultiWOZ, substantial agreement in NESPOLE!, WhatsApp, and JILDA, and moderate agreement in the least structured sub-corpus, Ubuntu (0.41). Combining the five sub-corpora together and computing the IAA over the whole D-Pro Corpus results in $k = 0.77$, while the simple and weighted mean of the k values for each sub-corpus is respectively 0.68 and 0.71 (*substantial agreement*). On the other hand, the overall combined turn-level PRO annotation IAA scored $k = 0.72$. Generally the PRO agreement computed at turn level

¹²Cohen’s Kappa brings more trustworthiness to the two-annotators agreement calculation since it takes into consideration the likelihood that a particular agreement situation accidentally occurred by chance. It is computed with the following formula:

$$k = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}$$

where $Pr(a)$ is the actual agreement observed between the two annotators and $Pr(e)$ is the expected agreement considering chance.

scores lower, as the number of turns is smaller than that of utterances in dialogues. Consequently, a disagreement on a single turn carries greater weight in the overall calculation, compared to a disagreement on an individual utterance.

Concerning the FAIL annotation, there is no substantial difference in the number of failure turns and the number of failure utterances, as no more than one failure situation occurs per turn. The almost perfect agreement on the annotation of failure situations corresponds to a k of 0.89 at utterance level and 0.96 at turn level: results suggest that identifying the specific utterance that conveys the failure is more challenging than determining where it occurs at a higher level, namely, in which turn.

The IAA for dialogue act annotation is calculated specifically on utterances where both annotators agreed on the proactive classification, yielding a kappa value of $k = 0.84$.

5. Turn-Level and Utterance-Level Proactivity Annotation Results

This section presents and analyses the results of the proactivity annotation (PRO label) described in Section 4, both at utterance-level and turn-level. We first discuss proactivity in the whole D-Pro Corpus, then we provide a detailed analysis of the individual sub-corpora in D-Pro, and, finally, we provide a cross-linguistic analysis related to the English and Italian portions of the Ubuntu corpus.

5.1 Proactivity in D-Pro

Sub-corpus-specific and overall results of the annotation of proactivity at both the utterance and the turn level are reported in Table 4.

D-Pro	NESPOLE!	Ubuntu	WhatsApp	MultiWOZ	JILDA	D-Pro Micro Avg.
PRO Turns %	19.01%	18.95%	35.91%	11.13%	18.27%	19.54%
Agent PRO Turns %	76.52%	64.66%	/	67.16%	49.14%	64.01%
Client PRO Turns %	23.48%	35.34%	/	32.84%	50.86%	35.99%
PRO Utterance %	14.59%	14.92%	25.34%	9.35%	13.55%	15.31%
Agent PRO Utt. %	83.67%	59.09%	/	76.67%	44.79%	67.06%
Client PRO Utt. %	16.33%	40.91%	/	23.33%	55.21%	33.94%
PRO Utt. per PRO Turn	2.18	1.52	1.69	1.34	1.41	1.65
Avg. Turn Length	2.84	1.93	2.39	1.60	1.89	2.11

Table 4: Percentages of sub-corpus-specific and total PRO annotation computed at turn and utterance level and divided by speaker. The number of proactive utterances per single proactive turn and average turn length reported for comparison.

Turn-level proactivity. In quantitative terms, the D-Pro Corpus annotation, made on a total of 2855 turns and 6028 utterances from 151 dialogues, resulted in the marking of 923 dialogue utterances with the PRO label¹³. Perhaps the most significant result of the D-Pro annotation is that proactivity is a relevant presence in the task-oriented dialogues that have been investigated, since

¹³Note that for our purposes, a dialogue turn is designated and, therefore, annotated as proactive if, and only if, it contains at least one proactive utterance. A proactive turn can, therefore, either consist of: (i) one single proactive utterance, (ii) only proactive utterances, or (iii) a mix of proactive and non-proactive utterances.

it can be found within a percentage of 19.54% over the total amount of dialogue turns of the D-Pro Corpus. While this is a significant finding, there are important individual differences (e.g., 11.13% in MultiWOZ and 35.91% in WhatsApp, St.Dev. = 9.15), which highlight how proactivity is influenced by different dialogue features. A Chi-Square test for independence was conducted to determine whether the distribution of PRO and non-PRO turns varied across the five sub-corpora. The test revealed a statistically significant association between sub-corpus and turn proactivity, $\chi^2(4) = 61.24$, $p < 0.001$. Particularly, examination of the residuals shows that the higher proactivity rate is present in the sub-corpus with higher natural setting (WhatsApp +5.39), while the lower proactivity is registered in MultiWOZ (-4.34), whose collection was highly guided through instructions to the Wizard. The other sub-corpora show smaller deviations that were not as significant. Overall, the high number of proactive turns confirm our initial intuition that proactivity plays a crucial role in human-human task-oriented dialogues.

PRO turn % in D-Pro generally reflects PRO utterance % outcomes. As reported in Table 4, there are subtle rises in the proportions of proactive turns across all sub-corpora, compared to the proportion of proactive utterances. Moreover, the data on average turn length suggests that WhatsApp and NESPOLE! in particular exhibit longer turns with more utterances. Consequently, within these extended turns filled also with non proactive utterances, proactivity is prone to dispersion: it loses part of its significance when analysed at the utterance-level. On the other hand, the 2% increase for MultiWOZ 2.2 correlates with the short turn length for this last sub-corpus. These observations are consistent with the average turn length data reported in the lower block of Table 4, where the metric is the average number of utterance in each turn.

Utterance-level proactivity. Overall, 15.31% of the D-Pro utterances have been annotated as proactive, with a lower standard deviation (St.Dev. = 5.91) among sub-corpora than for turns and with $\chi^2(4) = 73.51$, $p < 0.001$ (residuals: WhatsApp = +6.60, MultiWOZ = -4.21).

An additional datum reported in Table 4 is the *#pro-utterance/#pro-turn rate*, i.e., the proactive utterances count over proactive turns count, shows the average number of proactive utterances contained within one single proactive turn. This datum offers further details on the distribution of proactive utterances across turns. By comparison with the average turn length, computed by utterances composing each turn and reported in the last line of the table, we can draw two conclusions. First, proactive turns tend to be composed of both proactive and non-proactive utterances, since the average value of proactive utterance within proactive turn is below the average of the number of utterances typically composing one single turn. Second, results indicate a positive correlation between percentages of proactivity and turn length. Statistical significance tests were conducted to assess this correlation. Results indicated a positive, moderate correlation between the percentage of proactive turns and turn length, with Pearson's $\rho = 0.491$. Similarly, the percentage of proactive utterances was also positively correlated with turn length, showing a moderate association, Pearson's $\rho = 0.4997$. These findings suggest that both proactive turns and utterances are moderately correlated with turn length across sub-corpora, with correlation coefficients falling within the moderate range (0.3 to 0.7). In other words, a competent human speaker or writer will know how abundant or how relevant the information is that he or she can proactively provide in the unfolding of the conversation, or how many times he or she can exploit proactivity within the same dialogue, without violating collaborative rules of dialogue and thus without annoying his or her addressee and without deploying behaviours detrimental to achieving the conversational goal.

5.2 Proactivity in D-Pro sub-corpora

In this section we analyse proactivity in each individual sub-corpora composing D-Pro, according to the statistics presented in Table 4.

Analysis of NESPOLE! In NESPOLE!, 19.01% of turns are identified as proactive (PRO Turns %), which is lower only with respect to WhatsApp. Along with the PRO Utterance % (14.59%), this indicates that NESPOLE! is particularly rich in collaboration, a result that reflects the nature of telephonic interaction, where the speakers' proactive input is essential in managing and supporting the progress of the conversation effectively. The ratio of proactive utterances per proactive turn (2.18) is the highest among the sub-corpora, meaning that proactive turns in NESPOLE! often include multiple proactive utterances, emphasizing a more sustained proactive engagement. Additionally, the average turn length in NESPOLE! (2.84) is relatively long, possibly indicating more detailed guidance or suggestions by the Agent in this task-oriented setting. In fact, an interesting observation about the NESPOLE! sub-corpus comes from the comparison between the small percentage of proactive utterances offered by the Client and the (almost 5 times) larger proportion of proactive utterances provided by the Agent. Based on the analysed data, two primary reasons explain this pattern: (i) Clients' requests, made while explaining their desired accommodation, are informative and relevant, but are not considered proactive since they serve to set the dialogue goal, a necessary and introductory phase in user-initiated task-oriented dialogues like these, where the tourist (Client) calls the travel agency to arrange a vacation; (ii) The Agent proactively provides more information than the Client initially requests, particularly when describing all-inclusive vacation packages and accommodation options.

Analysis of Ubuntu. The Italian Ubuntu sub-corpus proactivity content is similar to both NESPOLE! and JILDA, showing a consistent presence of proactivity. In terms of proactive utterances annotation, 14.92% of all utterances in Ubuntu are proactive, and the average number of proactive utterances per proactive turn is 1.52. This pattern suggests that proactivity in Ubuntu is often situational and concise, focusing on specific points of support or troubleshooting advice rather than extensive instructional turns. Lastly, the average turn length in Ubuntu is 1.93 utterances, indicating short, to-the-point exchanges. The 172 proactive utterances in Ubuntu show a distribution of 104 proactive utterances offered by the Agents, and 72 provided by the Client. That seems fair if one considers that in each Ubuntu conversation, one Client potentially corresponds to several Agents, given that any user happening to be online inside the support chat at the moment when the Client poses his request, would be entitled to answer them and thereby become an Agent; besides, every dialogue actually unfolds with the intervention of one Client and at least 2 or 3 Agents. In addition to that, the nature of the Ubuntu dialogues itself contributes to these high values of proactivity offered by both the Agents and the Clients, since these most natural chat interactions regularly contain (i) overlapping conversations related to different tech support topics; (ii) new problems arising during tech assistance processes, resulting in either task-change or task-extension situation¹⁴; (iii) many attempts at solving the same problem in different ways or with different tools, or with the help of different Agents.

¹⁴By task-change we refer to those situations in dialogues where the task in force is essentially replaced by another, because the first one is either achieved, given up, or discarded for some reason; by task-extension we refer to places in dialogues where the main task is temporarily put aside to solve one or some more new side-task problems that have shown up during the interaction, and is eventually resumed.

Analysis of WhatsApp. WhatsApp dialogues hold the widest proportion of proactivity both at the utterance level (25.34%) and at the turn level (35.91%). Average turn length and the number of proactive utterances per proactive turn correlate, settling at 2.39 and 1.69, respectively. On the other hand, the average length of dialogues in the sub-corpus, about 9 turns, 21 utterances per dialogue, is the shortest found. This signals that proactivity may be the key to a fast and effective conclusion of the dialogue. The familiarity of the speakers with each other may also play a role in facilitating task-completion. Informants are always friends, sometimes close relatives: they perfectly know each other’s expectations and are capable of anticipating fast and effectively (especially in one-to-one conversations) the type of information that the other person needs. As a consequence, there is also little space for greetings, formalities, and courtesies. The exchange of information flows quite smoothly, with straightforward requests and without any particular misunderstanding or need for clarification. All of these factors combined facilitate highly collaborative, effective conversations.

Analysis of JILDA. In the JILDA dialogues, proactive behaviours are almost equally distributed among Clients and Agents turns. A dialogue where both parties can interject proactive guidance reflects a cooperative, balanced approach to the dialogue goal. Comparing the data to the PRO utterance % divided by speaker, we can argue however that Clients on average provide some more proactive utterances within proactive turns. This datum correlates with (i) the observed tendency shown by Agents to stick to repeating patterns of questions and answers: this behaviour was particularly emphasised in dialogues produced by specific individuals playing the role of the Agent, who created their own routinised conversation-management policy and tended to reproduce it even when their Client partner changed; furthermore it correlates with (ii) Clients being quite aware of which pieces of information were relevant for the Agent to select a suitable job offer so that the Clients proactively produced said relevant and concise information during the conversation.

Analysis of MultiWOZ. Among the five analysed sub-corpora, MultiWOZ 2.2 stands out for having below-average proactivity proportions: 11.13% of turns and only 9.35% of utterances are proactive. This is possibly due to the Wizard of Oz methodology employed in the collection, which produces dialogues that follow given scripts and are poor in lexical and syntactical variety. The fact that the few cases of proactivity are mostly offered by the Agent (they are twice more frequent than the Client’s proactive utterances) agrees with the introduction of mid-talk changes of dialogue goal—a behaviour explicitly encouraged in the informants by the developers of the methodology. Overall, the MultiWOZ sub-corpus embodies an Agent-led form of proactivity appropriate for task completion, with minimal deviation from the dialogue main track, highlighting the participants’ role in maintaining the focus on the dialogue task.

5.3 A cross-linguistic analysis: the Ubuntu sub-corpus

Although the analysis of proactivity that we have conducted is mainly based on Italian, a relevant question is whether our findings can be extended to other languages. In addition, being the MultiWOZ corpus in English, it remains open the question whether the low proactivity rate in the sub-corpus (i.e., 11.13%) is due to the specific interaction modality, Wizard of Oz, or to the different language, English, with respect to the other D-Pro sub-corpora.

Although an extensive cross-language investigation is out of the scope of our work, we took advantage of the fact that the Ubuntu Chat Corpus is available in several languages, including Italian and English, which are comparable, as both the dialogue task and the collection methodology is the

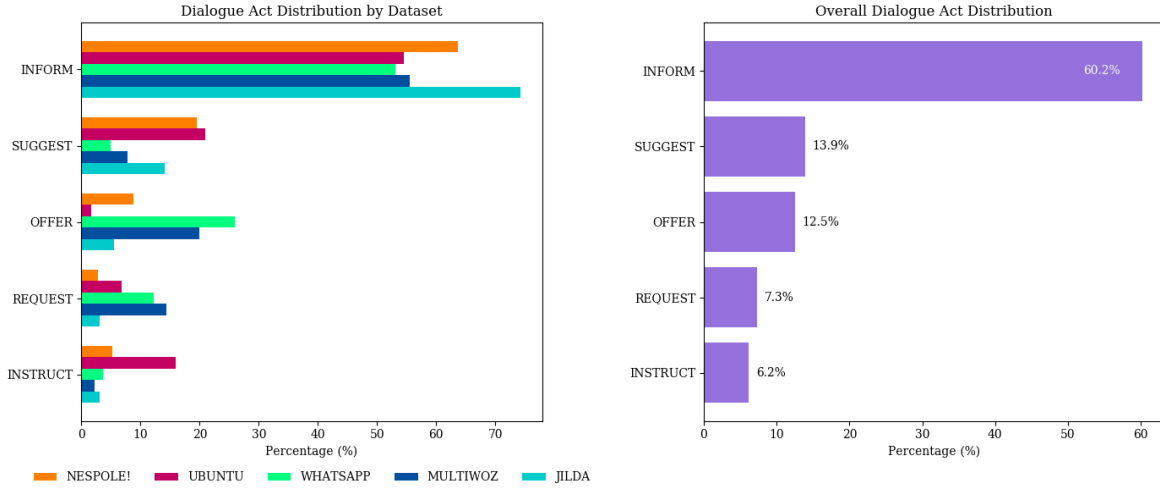


Figure 2: Percentage distribution of dialogue acts in proactive utterances divided by sub-corpus (left) and for the totality of D-Pro (right).

same. To investigate potential cross-linguistic differences in the use of proactivity among dialogue participants, a subsample of 200 turns from the English Ubuntu Chat Corpus was labelled. Out of 200 turns and 285 utterances, 20.5% and 17.54% respectively were identified as proactive, showing a statistically insignificant increase compared to the Italian Ubuntu percentages (18.95% and 14.92%): language does not significantly influence the results at the 0.05 significance level ($p\text{-value} = 0.1599$). These findings suggest that we can reject the hypothesis that the English language is to be accounted for major decreases in proactivity rates in the MultiWOZ corpus, which, on the contrary, can be explained on the basis of the Wizard of Oz interaction modality.

6. Dialogue Acts and Proactivity

In this section we discuss the results of the annotation of the dialogue act communicative functions performed on proactive utterances. First we provide statistics related to the dialogue acts involved in the annotation, and then we discuss the correlations between the linguistic structure of the utterance and the dialogue act annotation.

6.1 Dialogue Acts and Proactivity in D-Pro

In the following, we discuss the relation between proactivity and the dialogue acts used to annotate D-Pro. Statistics are reported in in Figure 2.

Inform. We notice a widespread prevalence of the INFORM tag: 60.2% of overall proactive utterances display a proactive information-giving attitude. This is somehow expected, considering that our definition of proactive behaviour (see Section 3.1) indeed focuses on the ability to add new, relevant, unsolicited *information*. JILDA has the most INFORM tags in proactive utterances (74.85%), which aligns with Section 5, where it is stated that JILDA’s proactivity is largely provided by Clients, who, as we presume, are aware of the type of information the Agent needs to

achieve the dialogue goal within the well-delimited domain of job search/offer (for instance, see example (4)).

Suggest. As for the SUGGEST label, which qualifies second for tagging frequency (13.9%), the results show that it appears overall thrice less than the INFORM tag, still keeping a noticeable advantage on the others. Proactive suggestions are more common in spontaneous talks than in structured dialogues, with the only exception of WhatsApp. In the first type of dialogues, interactions allow participants to negotiate Clients’ preferences, discuss benefits and drawbacks, and make suggestions among many solutions available. As far as Ubuntu is concerned, suggestions take the place that OFFERS hold in other sub-corpora, as in example (17) in Section 6.2: this makes sense because the ontology from which the Agents seek solutions to tech issues is made up of a one’s lifetime experience with the Ubuntu operating system, so there is not a set of pre-determined options to choose from.

Offer. In contrast to the two labels above, the OFFER label triggers the lowest visible value, that is, a percentage of 1.7%, provided by only three occurrences in Ubuntu, suggesting there could be a corpus-specific reason for that. In fact, that datum seems to be due to the peculiar nature of the Ubuntu interactions. Since the general goal of the dialogues is to resolve a technical problem that has arisen in the Client’s operating system, there is no need to select a final item amongst the available solutions as would instead be the case in MultiWOZ (e.g. selection of a restaurant), JILDA (selection of a job offer) or NESPOLE! (selection of an all-inclusive package). Consequently, there is no need for the Agent to offer such options for selection. On the contrary, the most OFFER dialogue acts are found in WhatsApp and MultiWOZ, where such acts can be performed, for instance, to meet the needs of the interlocutor or as an offer of action (example (6)).

Request. The MultiWOZ sub-corpus, though being the poorest in proactivity content among the five sub-corpora composing D-Pro, has the highest absolute and percentage value for the REQUEST tag. We argue that also this result is due to the corpus-specific nature of the conversations. As mentioned earlier, the dialogues were gathered through the Wizard of Oz simulation, which deceived the user participants into believing they were interacting with a dialogue system rather than a human being. The consequences of this collection method entail that users’ expectations about how the interaction will proceed differ from what they would be if the users believed to be interacting with an actual human being. As a consequence, there is an increased tendency to use requests that address the user’s needs straightforwardly, putting restrictions to the naturalness and linguistic richness of the dialogue and reducing collaborative phenomena and politeness mechanisms. The latter, typical of human dialogue, usually make people reluctant to make straightforward requests to mere acquaintances (rather than to close friends and relatives, as in WhatsApp: example (5)). Also, the generation of some sort of proactive behaviour is encouraged by MultiWOZ’s researchers thanks to the introduction of mid-conversation task changes: task change brings the need for even new requests made to the system to set the User’s preferences for the new dialogue goal. An exception is the Ubuntu sub-corpus, where requests are not so scarce as in NESPOLE! and JILDA. It is important to notice that in these latter cases, requests are mostly “requests of action”, that is, utterances where an Agent asks the Client to try some operation to narrow down the possible reasons for the issue or to attempt some solutions, with a procedure that advances by trials and errors.

Instruct. The arguments in the previous paragraph on REQUEST are helpful also to interpret the highest value for the INSTRUCT tag found in the Ubuntu sub-corpus. In proceeding by attempts, it

is not uncommon that Agents instruct Clients on doing a particular operation, often giving step-by-step instructions as those in U43 in example (7) below.

EXAMPLE (7)

c: U39 mi sono spiegato male:
 U40 l'icona nella dock unity non compare neanche quando Firefox è aperto...
 U41 *è aperto (correzione)
 A: U42 rocker: resetta unity allora:
 U43 [PRO] [INSTRUCT] unity --reset dato dopo aver premuto alt+f2.¹⁵

6.2 Dialogue Act-Related Linguistic Analysis

In this section we discuss the correlations that we observed between the utterance linguistic structure and the dialogue act annotation. We performed a qualitative analysis of the proactive utterances in the D-Pro Corpus to verify if they contain recurrent expressions that function as markers of proactivity. To this end, we uploaded our corpus in the Sketch Engine online platform and queried it through the functions *Wordlist* (that produces frequency lists of words) and *n-gram* (that produces frequency lists of sequences of tokens that tend to co-occur). Our analysis showed that proactive utterances contain several expressions that act as markers of proactivity and that some lexical-syntactical structures appear to be predominantly tied to one particular dialogue act annotation. In the following, we present some case studies: causal clauses, interrogative clauses, sentences introduced by modal verbs, and a selection of other recurrent patterns.

Causal Clauses. In the D-Pro Corpus causal connectors are often found, such as *perché*, *visto che*, *poiché* (ENG: *because*, *since*, *as*), that introduce proactive causal clauses usually labelled with the INFORM tag. This occurs, for example, when either (i) the Client makes a request and afterwards adds one proactive utterance in order to motivate his or her choice (as in example 8) or (ii) the Agent brings an offer or places a suggestion (as in examples 9 and 10) and motivates with the Client the choice of that particular offer/suggestion.

EXAMPLE (8)

c: U6 Il mio sogno sarebbe quello di fare l'insegnante
 U7 [PRO] [INFORM] perché mi piace lavorare con i bambini e ragazzi.¹⁶

EXAMPLE (9)

A: U65 quindi non so se lei ha le catene le consiglierei vivamente di portarle magari se non le vuole montare

¹⁵Example taken from the Ubuntu sub-corpus. ENG:

c: U39 I didn't make myself clear:
 U40 the icon in the unity dock does not appear even when Firefox is open...
 U41 *is open (correction)
 g: U42 rocker: reset unity then:
 U43 [PRO] [INSTRUCT] unity --reset run after pressing alt+f2.

¹⁶Example taken from the JILDA sub-corpus. ENG:

c: U6 My dream would be to be a teacher
 U7 [PRO] [INFORM] because I enjoy working with children and teenagers.

U66 {e} cioè non serve montarle
U67 comunque se le tenga nel se le porti
U68 [PRO] [INFORM] anche perché comunque poi salendo in montagna può sempre capitare una nevicata improvvisa.¹⁷

EXAMPLE (10)

R: **U20** comunque quando vuoi possiamo vederci anche io e te
U21 [PRO] [INFORM] visto che i tandem "ufficiali" sono solo il martedì e mercoledì :)¹⁸

D-Pro #?	Inform	Suggest	Offer	Request	Instruct	Tot.
#utterances	556	128	115	67	57	923
#?	1	4	31	3	0	39

Table 5: Interrogative clauses count within proactive utterances, grouped by dialogue act labelling.

Interrogative Clauses. Direct interrogative clauses have been collected through automatic search for question marks in the closing part of proactive utterances. Results on the distribution of interrogative clauses in co-occurrence with proactivity grouped by dialogue act labelling are presented in Table 5.

The vast majority (31 out of 39, about 80%) of interrogative utterances has been labelled with the OFFER tag: about 27% of total OFFERS are brought in an interrogative form. Example (11) is typical of proactive offers in the MultiWOZ corpus.

EXAMPLE (11)

c: **U8** I would like an expensive hotel if you can find one.
a: **U9** The express by holiday inn cambridge is located in the east and meet your criteria.
U10 [PRO] [OFFER] Shall I book you a room?¹⁹

As for the other dialogue acts, the interrogative clauses count is virtually negligible. The only interrogative sentence labelled with the INFORM tag is *Sai che forse non va fatta riposare??* (ENG: *Maybe you should'nt let the dough rest, you know??*), where the speaker is delivering information that she is not entirely confident about. With respect to SUGGEST and REQUEST dialogue acts,

¹⁷Example taken from the NESPOLE! sub-corpus. ENG:

a: **U65** so I don't know if you have snow chains I would strongly advise you to bring them maybe if you don't want to put them on
U66 {erm} I mean, you don't need to put them on
U67 anyway keep them in the bring them
U68 [PRO] [INFORM] also because then anyway going up into the mountains a sudden snowfall can always happen.

¹⁸Example taken from the WhatsApp sub-corpus. ENG:

R: **U20** anyway when you want we can also meet just you and me
U21 [PRO] [INFORM] since the "official" tandems are only on Tuesday and Wednesday :)

¹⁹Example taken from the MultiWOZ sub-corpus.

interrogative clauses include sentences starting with *Maybe...?*, *May I recommend...?*, *Sicuro che non...?* (ENG: *Are you sure you don't...?*), *Puoi...?* (ENG: *Could you...?*) and are used as a means of showing courtesy.

Modal Verbs. One set of very frequent expressions is the one containing the modal verbs *volere* (ENG: *want*), *potere* (ENG: *can, may*), *shall*, and *should* in the following constructions: *se vuoi / se vuole* (ENG: *if you want*), *ti posso / le posso* (ENG: *I can*) + infinitive clause, *shall / should / may I* + infinitive clause. Such phrases can be found in the context of use where the speaker, usually the Agent, offers either to provide some further information or to do something in order to help the addressee:

EXAMPLE (12)

c: U81 da kpakager ho soltanto adobe flash plugin
 a: U82 vai nel sito di adobe flash
 U83 e vedi se te lo installa da firefox
 U84 [PRO] [OFFER] oppure ti posso dire come installarlo in un altro modo.²⁰

EXAMPLE (13)

a: U13 I've found several restaurants that are located in the Centre with a moderate price range.
 U14 [PRO] [OFFER] May I recommend a British restaurant called the Oak Bistro?²¹

As seen in example (11), this construction often occurs in interrogative clauses reporting proactive utterances marked as OFFER. In fact, even in the case of modals, the linguistic pattern is used as a strategy to soften requests or offers and convey politeness (indirect speech act, Davison (1975)), whether or not they take the form of questions, as shown in (11).

Connectives. Another frequent structure is represented by sentences introduced by the connectives *ma / però, but*, or *anzi, invece, instead* as in (14) and (15). The connective introduces a logical relation of concession between discourse segments (Ferrari, 2010). In the case of example (14), the relation is between U5 and U6, while in example (15) is between U42 and U43. Proactivity introduced by this kind of connectives is not particularly tied to specific dialogue acts, but rather follows the general distribution of dialogue acts.

EXAMPLE (14)

J: U1 Ma per la pasta fresca si usa tutto l'uovo?
 U2 Va messa a riposare in frigor?
 U3 Un uovo ogni cento grammi?
 G: U4 Tutto l'uovo
 U5 Sarebbe un uovo ogni cento grammi
 U6 [PRO] [INSTRUCT] Ma io ne metto di meno
 U7 Tipo 3/4 uova per mezzo chilo.²²

²⁰Example taken from the Ubuntu sub-corpus. ENG:

c: U81 from kpakager I only have adobe flash plugin
 a: U82 go to the adobe flash website
 U83 And see if it lets you install it from firefox
 U84 [PRO] [OFFER] Or I can tell you how to install it in another way.

²¹Example taken from the MultiWOZ sub-corpus.

²²Example taken from the WhatsApp sub-corpus. ENG:

EXAMPLE (15)

a: U41 trus, ovvero: hai installato ubuntu "dentro" windows?
 c: U42 mi sa che hai ragione...anzi si...
 U43 [PRO] [INFORM] però mi sembra di ricordare che il disco in qualche modo me lo ha fatto partizionare lo stesso...²³

Other Recurrent Patterns. Additional recurrent expressions include verbs such as *consigliare*, *suggest*, or *provare*, *try*, in the constructions *(ti) consiglio (di)...*, *I suggest (that) you...*, *I recommend (that) you...*, and *prova (a)...*, *try (doing)...* These expressions are commonly employed to perform SUGGEST dialogue acts (as in 16 and 17) or REQUESTs, as for instance in utterances like "Prova e facci sapere", ENG: "Try (doing this) and let us know".

EXAMPLE (16)

c: U30 Mi potresti fornire informazioni sull'altra proposta di lavoro?
 a: U31 certo, attendi solo un momento per favore
 U32 [PRO] [SUGGEST] ti consiglio di informarti comunque presso la Munus s.r.l.²⁴

EXAMPLE (17)

c: U13 marcotux, puoi spiegarmi come si fa?
 a: U14 UnderFlea, provo a vedere se esiste in pacchetto
 U15 [PRO] [SUGGEST] prova a vedere nel gestore pacchetti se c'è lastfm.²⁵

7. Dialogue Structure and Proactivity

This section analyses how proactive utterances are positioned within the flow of a task-oriented dialogue. We discuss three aspects: (i) the relation between proactive utterances and goal failures; (ii) the relation between proactivity and the dialogue turn that originates a proactive utterance; and (iii) how proactive utterances are actually distributed throughout the whole dialogue.

j: U1 Do you use the whole egg to make fresh pasta?
 U2 Should it be put to rest in the fridge?
 U3 One egg for every hundred grams?
 g: U4 The whole egg.
 U5 That would be one egg for every hundred grams
 U6 [PRO] [INSTRUCT] But I put less than that
 U7 Like 3/4 eggs per pound.

²³Example taken from the Ubuntu sub-corpus. ENG:

a: U41 trus, that is: did you install ubuntu "inside" windows?
 c: U42 I guess you're right...actually yes....
 U43 [PRO] [INFORM] I think I remember, though, that it somehow let me partition the disk anyway...

²⁴Example taken from the JILDA sub-corpus. ENG:

c: U30 Could you provide me with information about the other job offer?
 a: U31 sure, just hold on a moment please
 U32 [PRO] [SUGGEST] I recommend that you still inquire with Munus s.r.l.

²⁵Example taken from the Ubuntu sub-corpus. ENG:

c: U13 marcotux, can you explain how to do that?
 a: U14 UnderFlea, I'll try to see if a package exists
 U15 [PRO] [SUGGEST] try seeing in the package manager if there is lastfm.

7.1 Goal-Failure Situations and Proactivity

As argued in Section 3.4, proactivity may help a participant recover from a failure situation, that is, when a communicative goal can not be satisfied. In task-oriented dialogues, such situations occur quite frequently when some expectations of the Client do not match with the knowledge of the Agent, most of the time simply because the Agent has a partial knowledge of the conversational domain (see example (4) in section 3.4). Here we analyse the FAIL tagging on the D-Pro Corpus, reported in Table 6, and investigate how proactivity is related to goal-failure situations.

D-Pro #FAIL	NESPOLE!	Ubuntu	WhatsApp	MultiWOZ	JILDA	D-Pro Tot.
#FAIL	13	19	14	27	32	105
#FAIL+PRO	11	12	10	6	21	60
#FAIL+PRO/#FAIL %	84.62%	63.16%	71.43%	22.22%	65.63%	57.14%
#FAIL/#utterances %	0.75%	1.61%	1.46%	2.80%	2.67%	1.74%
#FAIL+PRO/#utterances %	0.64%	1.02%	1.04%	0.62%	1.75%	1.00%
#FAIL+PRO/#PRO-utterances %	4.38%	6.82%	4.12%	6.67%	12.88%	6.50%

Table 6: Occurrences of failure situations related to proactive behaviour and to total number of utterances within the D-Pro Corpus (FAIL = tag for failure situations; FAIL+PRO = FAIL tag co-occurs with PRO tag within one single turn).

First of all, we notice that out of a total number of 105 failure situations, 60 of them (57%) originated a proactive initiative, supporting our hypothesis that failure turns are highly productive in terms of proactivity. In such a situation, typically, the Agent proactively offers an alternative solution with respect to the Client’s goals. In terms of distribution in our corpus, there is high variability. The proportion of proactivity in goal-failure situations is very high in NESPOLE! (85%) and in WhatsApp (71%), while it is low in MultiWOZ (22%). Notice that the proportion of FAIL utterances in MultiWOZ is the highest among our sub-corpora (2.80%), but, in spite of that, the large majority of them are not recovered through proactivity. A possible explanation may lay in the MultiWOZ design, which drives the Agent’s behaviour towards asking the Client to provide an alternative goal instead of offering a proactive solution.

Generally, our intuition is that proactivity, in addition to promoting recovery from failures, would also help to reduce the occurrence of goal-failure situations, namely, a proactive dialogue is less probable to manifest failures. This intuition is also supported by our data: the percentage of PRO utterance tags (see Table 4) inversely correlates to the number of FAIL tags (Pearson’s $\rho = -0.6$). Again, the most proactive sub-corpora are WhatsApp and NESPOLE!, with 25% and 15% of the utterances being proactive (refer to Table 4), while the two sub-corpora also had fewer failure utterances, 1.46% and 0.75%. On the other hand, MultiWOZ is both the least proactive (9% of proactive utterances) and the most failure-prone (2.80%, or almost four times more than NESPOLE!).

The last row in Table 6 shows the proportion of failures with proactivity out of the total proactive utterances in a dialogue. These numbers support the intuition that proactivity helps reducing failures in task-oriented dialogues. For instance, in Ubuntu only 6.82% of proactive utterances are employed to recover failure situations, indicating that the largely prevalent use of proactivity is outside failure situations: in other words, proactivity is indirectly used to prevent the insurgence of failures.

7.2 Adjacent and Non-Adjacent Turn Proactivity

In this section we analyse the relation between a proactive utterance and reactive (i.e., non proactive) utterances in adjacent turns and in the same turn. The aim is to validate the hypothesis that proactivity is not a direct response to a conversational stimulus (such as a question), but instead arises from a more autonomous initiative of a dialogue participant.

Adjacency	NESPOLE!	Ubuntu	WhatsApp	MultiWOZ	JILDA	D-Pro Micro Avg.
#adj %	74.10%	55.68%	79.01%	95.56%	95.09%	77.68%
#non adj %	25.90%	44.32%	20.99%	4.44%	4.91%	22.32%

Table 7: Proportion of PRO utterances occurring at a turn t_i following a reactive utterance triggered by the previous turn t_{i-1} .

Table 7 reports the outcomes of turn adjacency annotation with the ADJ and NA tags on the D-Pro Corpus. For two of our sub-corpora, namely MultiWOZ and JILDA, almost all proactive utterances are added to a reaction utterance triggered by the previous turn (95.56% and 95.09% of the cases, respectively). For instance, in Example (1) in Section 1, U20 is the proactive utterance (*The phone number is ...*), U19 in the same turn is the reactive utterance (*That information is not available to me.*), and U18 in the previous turn (*Does it have an entrance fee?*) is the utterance triggering U19. This pattern, [trigger utterance + reaction utterance + proactive utterance]²⁶, is by far the most frequent in our D-Pro Corpus, indicating that adjacent turns alone usually provide sufficient context for the following proactive turn to be understood.

By contrast, dialogues from NESPOLE! and, especially, from Ubuntu, contain a higher number of NA tags, implying longer dependencies between a proactive utterance, the reaction utterance and the triggering turn. A possible explanation is that longer turns in NESPOLE! usually correspond to higher numbers of asynchronous messages: these can be due to the medium (spoken phone conversation) in NESPOLE!, where speakers may miss cues for turn-taking dynamics, resulting in incorrect timing, interruptions, and overlap. In other cases, turn-adjacency is impeded by backchanneling turns, that is, expedients for participants to send feedback to the current speaker often emerging in the form of minimal verbal cues (such as *uh-uh*, *sì*, *yes*, *mm-hmm*, *bene*, *good*, *capisco*, *I see*) signalling active listening, engagement, and understanding. As for the Ubuntu dialogues, the participation of multiple Agents at once interrupts the flow of one-to-one conversations: since chat dialogues are necessarily asynchronous, this may happen especially in an uncontrolled environment, where users can easily wait minutes or hours before answering a message; incidentally, a behaviour that was discouraged by design in the creation of the MultiWOZ and the JILDA Corpus.

7.3 Proactivity Distribution in Dialogue

In this section we analyse proactivity from the point of view of the dialogue structure. In Table 4 we reported that, overall, proactivity accounts for almost 20% of the turns in our task-oriented

²⁶According to Nouri and Traum (2014)’s annotation scheme (R = turn that directly relates to previous turn; F = turn fulfils a pending discourse obligation; I = turn imposes an obligation; N = turn provides new optional material) our turn [reactive utterance + proactive utterance] would be labelled as [R, possibly F and/or I + N, possibly I].

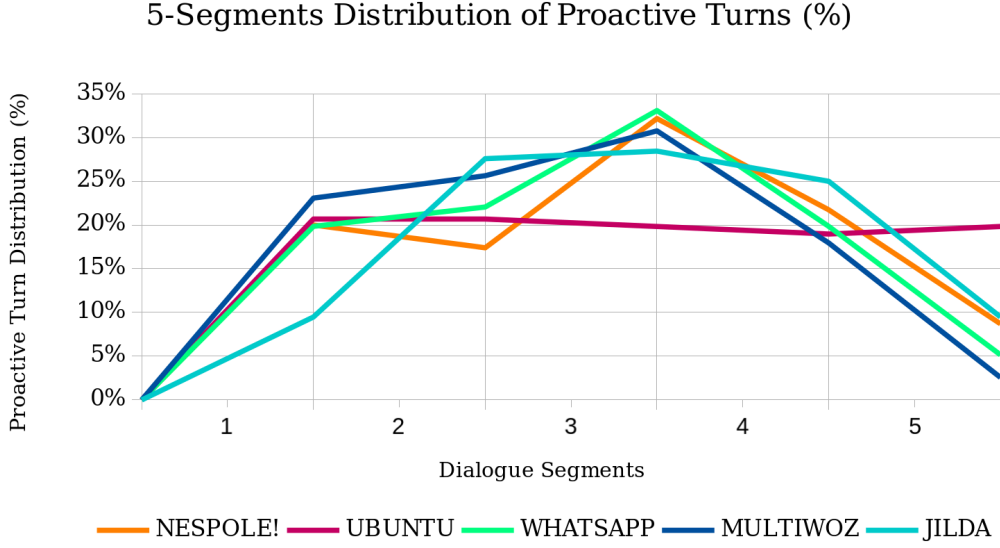


Figure 3: 5-segments distribution of proactive turns: each dialogue is divided into five equal parts (turn number is the standard measure for the division), and the percentage of proactive turns is computed within each of the five parts, so that dialogues of different lengths are comparable in a coarse grained analysis.

dialogues. However, intuitively, proactivity does not distribute uniformly across all portions of a task-oriented dialogue. For instance, the initial turns in a dialogue are typically introductory and used to reveal the communicative goals of the Client, while the last turns serve to finalize the dialogue (e.g., making a reservation) and for final greetings (Zhai and Williams, 2014). In these portions of the dialogue we expect to have less proactive utterances than in the central part of the dialogue. This intuition is confirmed by our findings on proactivity distribution in the D-Pro Corpus, as depicted in Figure 3.

To investigate this aspect, each dialogue in the D-Pro Corpus was split in five segments containing an equal amount of turns (e.g., given a dialogue with 15 turns, each segment contains exactly three turns). Taking advantage of the PRO annotations of utterances, we calculated the proportion of proactive utterances for each segment of the dialogue, and for each sub-corpus in D-Pro. According to our hypothesis, results show that segment 1 and segment 5 are less proactive than the central segments. On average, about 15% of utterances in segment 1 are proactive, and about 10% in segment 5, while the average proactivity in segment 3 is about 30%, showing that the central turns in a task-oriented dialogue are the most proactive. This distribution holds for all our sub-corpora, NESPOLE!, WhatsApp, MultiWOZ and JILDA, with the exception of Ubuntu, whose proactivity distribution is almost uniform across the dialogue segments. A Chi-square test has been conducted to compare the distributions across the five sub-corpora at five observation points. For the full corpus, the Chi-square statistic is $\chi^2 = 33.30$ with 16 degrees of freedom, and the p-value is $p = 0.0067$, indicating a significant difference between the groups. After removing the Ubuntu sub-corpus, the Chi-square statistic is $\chi^2 = 14.92$ with 12 degrees of freedom, and the p-value is

$p = 0.2459$, indicating there is not any more significant difference in proactivity distribution among the groups. A possible explanation for this, is that Ubuntu participants are instructed to join the chat without introducing themselves, thus avoiding initial and final greetings that are instead present in the other four types of dialogues. In addition, Ubuntu interactions are multi-party dialogues, where different participants bring their contribution at different points in time (given the asynchronous nature of the chat).

8. Conclusions and ongoing work

In this research, we focus on proactivity in task-oriented dialogues. We take advantage of investigations of different traditions, including the cooperation principle in language pragmatics, accommodation in social psychology of language, and the notion of initiative in dialogue defined in computational linguistics. We provide an operational definition of proactivity at the utterance level, as a collaborative behaviour occurring when: (i) a participant does not act merely in response to a previous request; and (ii) the participant’s behaviour is somehow effective for the achievement of the dialogue goal.

We annotate proactive language behaviours in human-human task-oriented dialogues with the goal of quantifying the extent of the phenomenon, clarifying which dialogue acts are expressed by proactive utterances, and identifying under what conditions proactivity tends to occur. To reach these targets, we gather human-human task-oriented dialogues from five pre-existing corpora with different characteristics in terms of language, conversational domain, media used for exchanging turns, and collection modalities. We develop an annotation scheme and the guidelines to label the presence of proactivity in dialogue utterances and turns, and to tag the dialogue act displayed by proactive utterances. These activities result in the creation of a new corpus (called D-Pro) annotated for proactive behaviours.

Findings. Our investigation of proactivity in human-human dialogues enables us to have a much clearer definition of the phenomenon, from both a quantitative and qualitative point of view. First, we are able to quantify that about 20% of turns in the D-Pro Corpus are proactive turns, showing that this is a pervasive phenomenon. Second, we show that only a limited number of dialogue acts are actually involved in expressing proactivity, a result that opens interesting theoretical perspectives. In addition, we confirm the non-reactive nature of proactivity, highlighting the presence of a pattern where a turn t_i triggers a reaction in a following turn t_{i+1} and a proactive utterance is then added to t_{i+1} . Moreover, we empirically confirm the hypothesis that proactivity has a crucial role in recovering from goal-failure situations, contributing to the efficacy of the whole dialogue. Finally, we demonstrate the non-uniform distribution of proactivity throughout the dialogue.

Limitations and ongoing work. There are several aspects of proactivity that we could not address in this paper, and that we plan for future research. A first aspect is the impact of proactivity on the efficacy of task-oriented dialogues: this analysis would imply a clear methodology to measure dialogue efficacy (e.g., in terms of goal achievement), which, however, is still a challenging research topic. A second aspect involves analysing proactivity in different kinds of dialogues, including argumentative dialogues. Here a potential issue is the lack of clear communicative goals, which helped us to characterize proactivity in task-oriented dialogues.

Potential impact on computational models of dialogue. Finally, in the long term, we are interested in developing computational models of proactive conversational agents, based on Large

Language Models (LLMs). Current LLMs, such as GPT-4 and the open source Llama family, are specifically instructed to execute interactive tasks, such as question answering and chat-based exchanges. However, although LLMs achieve excellent performance in information seeking tasks, their conversational abilities when participants need to collaborate to jointly achieve a communicative goal (e.g., booking a restaurant, fixing an appointment) are still far from those exhibited by humans. In order to model collaborative behaviours, recent approaches investigate how LLMs can be fine tuned to address dialogue pragmatics. For instance, Shaikh et al. (2024) show that grounding acts can be identified and annotated by a Large Language Model and modelled through appropriate fine tuning of the model itself. As a first step in this direction we used the D-Pro Corpus to exploit a language model to predict whether the last utterance in a task-oriented dialogue is either proactive or not-proactive. In Brenna and Magnini (2024) we show that a few-shot approach with GPT-4o achieves encouraging performance on a test set composed of dialogue snippets collected from the five D-Pro sub-corpora, and that, in particular for the NESPOLE! corpus, the agreement between the model labels and the human-annotated gold labels is nearly equivalent to the agreement between humans. As a following step on proactivity, once we have sufficiently-accurate results, we plan to collect a large number of training dialogue snippets (e.g., 100K) both proactive and not-proactive, and use them to instruction-tune an open-source model like Llama. We expect that new LLMs will be able to manifest much better proactive behaviours, and, in general, better collaborative behaviours, than the current LLMs.

References

- Anne H. Anderson, Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, et al. The HCRC map task corpus. *Language and Speech*, 34(4):351–366, 1991. doi: 10.1177/002383099103400404. URL <https://doi.org/10.1177/002383099103400404>.
- John Langshaw Austin. *How to do things with words*. William James Lectures. Oxford University Press, 1962. URL http://scholar.google.de/scholar.bib?q=info:xI2JvixH8_QJ:scholar.google.com/&output=citation&hl=de&as_sdt=0,5&ct=citation&cd=1.
- Vevake Balaraman and Bernardo Magnini. Proactive systems and influenceable users: Simulating proactivity in task-oriented dialogues. In *Proceedings of the 24th Workshop on the Semantics and Pragmatics of Dialogue-Full Papers, Virtually at Brandeis, Waltham, New Jersey, July. SEMDIAL*, 2020a.
- Vevake Balaraman and Bernardo Magnini. Investigating proactivity in task-oriented dialogues. *Proceedings of the Seventh Italian Conference on Computational Linguistics CLiC-it 2020*, 2020b. URL <https://api.semanticscholar.org/CorpusID:229294115>.
- Vevake Balaraman, Seyedmostafa Sheikhalishahi, and Bernardo Magnini. Recent neural methods on dialogue state tracking for task-oriented dialogue systems: A survey. In Haizhou Li, Gina-Anne Levow, Zhou Yu, Chitralkha Gupta, Berrak Sisman, Siqu Cai, David Vandyke, Nina Dethlefs, Yan Wu, and Junyi Jessy Li, editors, *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue, SIGdial 2021, Singapore and On-*

- line, July 29-31, 2021, pages 239–251. Association for Computational Linguistics, 2021. URL <https://aclanthology.org/2021.sigdial-1.25>.
- Francesca Bargiela-Chiappini. Face and politeness: New (insights) for old (concepts). *Journal of Pragmatics*, 35(10-11):1453–1469, 2003.
- Sofia Brenna and Bernardo Magnini. Last utterance proactivity prediction in task-oriented dialogues. In *Proceedings of the Eighth Workshop on Natural Language for Artificial Intelligence (NL4AI 2024) co-located with the 23rd International Conference of the Italian Association for Artificial Intelligence (AI*IA 2024)*. CEUR-WS.org, 2024.
- Penelope Brown and Stephen C. Levinson. *Politeness: Some universals in language usage*, volume 4. Cambridge University Press, 1987.
- Paweł Budzianowski, Tsung-Hsien Wen, Bo-Hsiang Tseng, Inigo Casanueva, Stefan Ultes, Osman Ramadan, and Milica Gašić. Multiwoz—a large-scale multi-domain wizard-of-oz dataset for task-oriented dialogue modelling. *arXiv preprint arXiv:1810.00278*, pages 5016–5026, October–November 2018. doi: 10.18653/v1/D18-1547. URL <https://www.aclweb.org/anthology/D18-1547>.
- Harry Bunt. Dimensions in dialogue act annotation. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC’06)*, Genoa, Italy, May 2006. European Language Resources Association (ELRA). URL http://www.lrec-conf.org/proceedings/lrec2006/pdf/428_pdf.pdf.
- Harry Bunt and Yann Girard. Designing an open, multidimensional dialogue act taxonomy. In *DIALOR’05, Proceedings of the Ninth Workshop on the Semantics and Pragmatics of Dialogue*, Nancy, pages 37–44, 2005.
- Harry Bunt, Jan Alexandersson, Jean Carletta, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Kiyong Lee, Volha Petukhova, Andrei Popescu-Belis, Laurent Romary, et al. Towards an iso standard for dialogue act annotation. In *Seventh conference on International Language Resources and Evaluation (LREC’10)*, 2010.
- Susan Meredith Burt. Code choice in intercultural conversation: Speech accommodation theory and pragmatics. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)*, 4(4):535–559, 1994.
- Jennifer Chu-Carroll and Michael K. Brown. An evidential model for tracking initiative in collaborative dialogue interactions. In Susan Haller, Alfred Kobsa, and Susan McRoy, editors, *Computational Models of Mixed-Initiative Interaction*, pages 49–87. Springer Netherlands, Dordrecht, 1999. ISBN 978-94-017-1118-0. doi: 10.1007/978-94-017-1118-0_2. URL https://doi.org/10.1007/978-94-017-1118-0_2.
- Herbert H. Clark. *Using language*. Cambridge University Press, 1996.
- Herbert H. Clark and Susan E. Brennan. Grounding in communication. *Perspectives on socially shared cognition*, 1991.

- Herbert H. Clark and Edward F. Schaefer. Collaborating on contributions to conversations. *Language and cognitive processes*, 2(1):19–41, 1987.
- Herbert H. Clark and Edward F. Schaefer. Contributing to discourse. *Cognitive science*, 13(2): 259–294, 1989.
- Robin Cohen, Coralee Allaby, Christian Cumbaa, Mark Fitzgerald, Kinson Ho, Bowen Hui, Celine Latulipe, Fletcher Lu, Nancy Moussa, David Pooley, Alex Qian, and Saheem Siddiqi. What is initiative? *User Modeling and User-Adapted Interaction*, 8:171–214, 1998.
- Mark G. Core, Johanna Moore, and Claus Zinn. The role of initiative in tutorial dialogue. In *10th Conference of the European Chapter of the Association for Computational Linguistics*, pages 67–74. Association for Computational Linguistics, 2003.
- Alice Davison. Indirect speech acts and what to do with them. In Peter Cole and Jerry L. Morgan, editors, *Speech acts*, pages 143–185. Brill, 1975.
- Céline De Looze, Stefan Scherer, Brian Vaughan, and Nick Campbell. Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction. *Speech Communication*, 58:11–34, 2014. ISSN 0167-6393. doi: <https://doi.org/10.1016/j.specom.2013.10.002>. URL <https://www.sciencedirect.com/science/article/pii/S0167639313001386>.
- Yang Deng, Wenqiang Lei, Wai Lam, and Tat-Seng Chua. A survey on proactive dialogue systems: problems, methods, and prospects. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI '23*, 2023. ISBN 978-1-956792-03-4. doi: 10.24963/ijcai.2023/738. URL <https://doi.org/10.24963/ijcai.2023/738>.
- Gretchen Ellefson. Conversational cooperation revisited. *The Southern Journal of Philosophy*, 59 (4):545–571, 2021.
- Angela Ferrari. Connettivi. In *Enciclopedia dell'italiano*. Istituto della Enciclopedia Italiana, Roma, 2010. URL [https://www.treccani.it/enciclopedia/connettivi_\(Enciclopedia-dell%27Italiano\)/](https://www.treccani.it/enciclopedia/connettivi_(Enciclopedia-dell%27Italiano)/).
- Anita Fetzer. Reformulation and common grounds. In *Lexical markers of common grounds*, pages 159–181. Brill, 2006.
- Howard Giles. Accommodation theory: Optimal levels of convergence. In *Language and social psychology*, pages 45–65. Basil Blackwell, 1979.
- Howard Giles and Tania Ogay. Communication accommodation theory. In Bryan B. Whaley and Wendy Samter, editors, *Explaining communication: Contemporary theories and exemplars*, pages 293–310. Lawrence Erlbaum Associates Publishers, 2006.
- Howard Giles and Peter Powesland. Accommodation theory. In Nikolas Coupland and Adam Jaworski, editors, *Sociolinguistics*, pages 232–239. Macmillan Education UK, London, 1997. ISBN 978-1-349-25582-5. doi: 10.1007/978-1-349-25582-5_19. URL https://doi.org/10.1007/978-1-349-25582-5_19.

- Howard Giles, Donald M. Taylor, and Richard Bourhis. Towards a theory of interpersonal accommodation through language: some Canadian data. *Language in Society*, 2(2):177–192, 1973. doi: 10.1017/S0047404500000701.
- Howard Giles, Nikolas Coupland, and Justine Coupland. Accommodation theory: Communication, context, and consequence. *Contexts of accommodation: Developments in applied sociolinguistics*, 1:1–68, 1991.
- Adam M. Grant and Susan J. Ashford. The dynamics of proactivity at work. *Research in Organizational Behavior*, 28:3–34, 2008. ISSN 0191-3085. doi: <https://doi.org/10.1016/j.riob.2008.04.002>. URL <https://www.sciencedirect.com/science/article/pii/S0191308508000038>.
- Paul Grice. Logic and conversation. In Peter Cole and Jerry L. Morgan, editors, *Speech acts*, pages 41–58. Brill, 1975.
- Paul Grice. *Studies in the Way of Words*. Harvard University Press, 1989.
- Curry I. Guinn. An analysis of initiative selection in collaborative task-oriented discourse. In Judith Masthoff, editor, *User Modeling and User-Adapted Interaction*, volume 8, page 255–314. Springer Publishing Company, USA, 1998. doi: 10.1023/A:1008359330641. URL <https://doi.org/10.1023/A:1008359330641>.
- Freya Hewett. *Sequential Organisation in WhatsApp Conversations*. Unpublished Bachelor’s Thesis, Free University of Berlin, Summer Semester, 2017.
- Pamela W. Jordan and Barbara Di Eugenio. Control and initiative in collaborative problem solving dialogues. In *Working Notes of the AAAI Spring Symposium on Computational Models for Mixed-initiative Interaction*, pages 81–84, 1997.
- John F. Kelley. An iterative design methodology for user-friendly natural language office information applications. In John O. Limb, editor, *ACM Transactions on Information Systems (TOIS)*, volume 2(1), pages 26–41. ACM New York, NY, USA, January 1984. doi: 10.1145/357417.357420. URL <https://doi.org/10.1145/357417.357420>.
- Cynthia Kersey, Barbara Di Eugenio, Pamela Jordan, and Sandra Katz. Knowledge co-construction and initiative in peer learning interactions. In *Proceedings of the 2009 Conference on Artificial Intelligence in Education: Building Learning Systems That Care: From Knowledge Representation to Affective Modelling*, page 325–332, NLD, 2009. IOS Press. ISBN 9781607500285.
- Matthias Kraus, Nicolas Wagner, and Wolfgang Minker. ProDial – an annotated proactive dialogue act corpus for conversational assistants using crowdsourcing. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 3164–3173, Marseille, France, June 2022. European Language Resources Association. URL <https://aclanthology.org/2022.lrec-1.339/>.

- J. Richard Landis and Gary G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, 1977. ISSN 0006341X, 15410420. URL <http://www.jstor.org/stable/2529310>.
- Xiang Li, Lili Mou, Rui Yan, and Ming Zhang. Stalematebreaker: A proactive content-introducing approach to automatic human-computer conversation. In Subbarao Kambhampati, editor, *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016*, pages 2845–2851. IJCAI/AAAI Press, 2016. URL <http://www.ijcai.org/Abstract/16/404>.
- Samuel Louvan and Bernardo Magnini. Recent neural methods on slot filling and intent classification for task-oriented dialogue systems: A survey. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 480–496, Barcelona, Spain (Online), December 2020. International Committee on Computational Linguistics. doi: 10.18653/v1/2020.coling-main.42. URL <https://www.aclweb.org/anthology/2020.coling-main.42>.
- Ryan Lowe, Nissan Pow, Iulian Serban, and Joelle Pineau. The Ubuntu dialogue corpus: A large dataset for research in unstructured multi-turn dialogue systems. In Alexander Koller, Gabriel Skantze, Filip Jurcicek, Masahiro Araki, and Carolyn Penstein Rose, editors, *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 285–294, Prague, Czech Republic, September 2015. Association for Computational Linguistics. doi: 10.18653/v1/W15-4640. URL <https://aclanthology.org/W15-4640/>.
- Nadia Mana, Susanne Burger, Roldano Cattoni, Laurent Besacier, Victoria MacLaren, John McDonough, and Florian Metze. The NESPOLE! VoIP multilingual corpora in tourism and medical domains. In H. Bourlard, editor, *Proceedings of the 8th European Conference on Speech Communication and Technology, EUROSPEECH 2003, Geneva, Switzerland, 01-04 September 2003*, pages 1589–1592. International Speech Communication Association (ISCA), 2003. doi: 10.21437/Eurospeech.2003-464.
- Michael Mctear. *Conversational AI: Dialogue Systems, Conversational Agents, and Chatbots*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, United States, October 2020. ISBN 9781636390314. doi: 10.2200/S01060ED1V01Y202010HLT048.
- Elnaz Nouri and David Traum. Initiative taking in negotiation. In Kallirroi Georgila, Matthew Stone, Helen Hastie, and Ani Nenkova, editors, *Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, pages 186–193, Philadelphia, PA, U.S.A., June 2014. Association for Computational Linguistics. doi: 10.3115/v1/W14-4325. URL <https://aclanthology.org/W14-4325/>.
- Shachi Paul, Rahul Goel, and Dilek Hakkani-Tür. Towards universal dialogue act tagging for task-oriented dialogues. In *Proc. Interspeech 2019*, pages 1453–1457, 2019. doi: 10.21437/Interspeech.2019-1866.
- Matthew Purver, Jonathan Ginzburg, and Patrick Healey. On the means for clarification in dialogue. In Jan van Kuppevelt and Ronnie W. Smith, editors, *Current and New Directions in*

- Discourse and Dialogue*, pages 235–255. Springer Netherlands, Dordrecht, 2003a. ISBN 978-94-010-0019-2. doi: 10.1007/978-94-010-0019-2_11. URL https://doi.org/10.1007/978-94-010-0019-2_11.
- Matthew Purver, Patrick G.T. Healey, James King, Jonathan Ginzburg, and Greg J. Mills. Answering clarification questions. In *Proceedings of the Fourth SIGdial Workshop of Discourse and Dialogue*, pages 23–33, 2003b. URL <https://aclanthology.org/W03-2103/>.
- James Pustejovsky and Amber Stubbs. *Natural Language Annotation for Machine Learning: A guide to corpus-building for applications*. O’Reilly Media, Inc., 2012.
- Filip Radlinski and Nick Craswell. A theoretical framework for conversational search. In Ragnar Nordlie, Nils Pharo, Luanne Freund, Birger Larsen, and Dan Russel, editors, *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval, CHIIR 2017, Oslo, Norway, March 7-11, 2017*, pages 117–126. ACM, 2017. doi: 10.1145/3020165.3020183. URL <https://doi.org/10.1145/3020165.3020183>.
- Eran Raveh. *Vocal accommodation in human-computer interaction: modeling and integration into spoken dialogue systems*. PhD thesis, Saarländische Universitäts- und Landesbibliothek, 2021.
- Carol Myers Scotton. odeswitching as indexical of social negotiations. In Monica Heller, editor, *Codeswitching: Anthropological and Sociolinguistic Perspectives*, pages 151–186. De Gruyter Mouton, Berlin, New York, 1988. ISBN 9783110849615. doi: doi:10.1515/9783110849615.151. URL <https://doi.org/10.1515/9783110849615.151>.
- John R. Searle. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, 1969. doi: 10.1017/CBO9781139173438.
- John R. Searle. Indirect speech acts. In Peter Cole and Jerry L. Morgan, editors, *Speech acts*, pages 59–82. Brill, 1975.
- Omar Shaikh, Kristina Gligoric, Ashna Khetan, Matthias Gerstgrasser, Diyi Yang, and Dan Jurafsky. Grounding gaps in language model generations. In Kevin Duh, Helena Gomez, and Steven Bethard, editors, *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 6279–6296, Mexico City, Mexico, June 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.naacl-long.348. URL <https://aclanthology.org/2024.naacl-long.348/>.
- Leah Shelley and Fernando Gonzalez. Back channeling: Function of back channeling and L1 effects on back channeling in L2. *Linguistic Portfolios*, 2(1), 2013. URL https://repository.stcloudstate.edu/stcloud_ling/vol2/iss1/9.
- Ronnie W. Smith. *A computational model of expectation-driven mixed-initiative dialog processing*. PhD thesis, Duke University, USA, 1992.
- Dan Sperber and Deirdre Wilson. *Relevance: Communication and Cognition*, volume 142. Harvard University Press Cambridge, MA, 1986.

- Andreas Stolcke, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema, and Marie Meteer. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics*, 26(3):339–374, 2000. URL <https://aclanthology.org/J00-3003/>.
- Petra-Maria Strauß and Wolfgang Minker. *Proactive Spoken Dialogue Interaction in Multi-Party Environments*. Springer New York, NY, 1 edition, 2010. ISBN 978-1-4419-5991-1. doi: <https://doi.org/10.1007/978-1-4419-5992-8>. URL <https://doi.org/10.1007/978-1-4419-5992-8>.
- Ilaria Sucameli, Alessandro Lenci, Bernardo Magnini, Manuela Speranza, and Maria Simi. Toward data-driven collaborative dialogue systems: The jilda dataset. *Italia Journal of Computational Linguistics IJCoL (Torino)*, 7(1 — 2):67–90, 2021. doi: 10.4000/ijcol.842. URL <https://doi.org/10.4000/ijcol.842>.
- Irene Sucameli, Alessandro Lenci, Bernardo Magnini, Maria Simi, and Manuela Speranza. Becoming JILDA. In Johanna Monti, Felice Dell’Orletta, and Fabio Tamburini, editors, *Proceedings of the Seventh Italian Conference on Computational Linguistics CLIC-it 2020*, volume 2769 of *CEUR Workshop Proceedings*, Bologna, 2020. CEUR-WS. URL http://ceur-ws.org/Vol-2769/paper_69.pdf.
- Kai Sun, Seungwhan Moon, Paul Crook, Stephen Roller, Becka Silvert, Bing Liu, Zhiguang Wang, Honglei Liu, Eunjoon Cho, and Claire Cardie. Adding chit-chat to enhance task-oriented dialogues. In Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, Dilek Hakkani-Tur, Iz Beltagy, Steven Bethard, Ryan Cotterell, Tanmoy Chakraborty, and Yichao Zhou, editors, *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1570–1583, Online, June 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.naacl-main.124. URL <https://aclanthology.org/2021.naacl-main.124/>.
- David R. Traum. Views on mixed-initiative interaction. In *AAAI97 Spring Symposium On Mixed-Initiative Interaction*, pages 169–171, 1997.
- David R. Traum. Issues in multiparty dialogues. In Frank Dignum, editor, *Advances in Agent Communication*, pages 201–211, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg. ISBN 978-3-540-24608-4.
- David R. Traum and Peter A. Heeman. Utterance units in spoken dialogue. In Elisabeth Maier, Marion Mast, and Susann LuperFoy, editors, *Dialogue Processing in Spoken Language Systems, ECAI’96 Workshop, Budapest, Hungary, August 13, 1996, Revised Papers*, volume 1236 of *Lecture Notes in Computer Science*, pages 125–140. Springer, 1996. doi: 10.1007/3-540-63175-5_42. URL https://doi.org/10.1007/3-540-63175-5_42.
- David R. Traum and Elizabeth A. Hinkelman. Conversation acts in task-oriented spoken dialogue. Technical report, University of Rochester, USA, 1992.
- David C. Uthus and David W. Aha. The ubuntu chat corpus for multiparticipant chat analysis. In *Analyzing Microtext, Papers from the 2013 AAAI Spring Symposium, Palo Alto, California, USA*,

- March 25-27, 2013, volume SS-13-01 of *AAAI Technical Report*. AAAI, 2013. URL <http://www.aaai.org/ocs/index.php/SSS/SSS13/paper/view/5706>.
- Marilyn Walker and Steve Whittaker. Mixed initiative in dialogue: An investigation into discourse segmentation. In *28th Annual Meeting of the Association for Computational Linguistics*, pages 70–78, Pittsburgh, Pennsylvania, USA, June 1990. Association for Computational Linguistics. doi: 10.3115/981823.981833. URL <https://aclanthology.org/P90-1010/>.
- Steve Whittaker and Phil Stenton. Cues and control in expert-client dialogues. In *26th Annual Meeting of the Association for Computational Linguistics*, pages 123–130, Buffalo, New York, USA, June 1988. Association for Computational Linguistics. doi: 10.3115/982023.982038. URL <https://aclanthology.org/P88-1015/>.
- Chien-Sheng Wu, Andrea Madotto, Ehsan Hosseini-Asl, Caiming Xiong, Richard Socher, and Pascale Fung. Transferable multi-domain state generator for task-oriented dialogue systems. In Anna Korhonen, David Traum, and Lluís Màrquez, editors, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 808–819, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1078. URL <https://aclanthology.org/P19-1078/>.
- Fan Yang and Peter A. Heeman. Initiative conflicts in task-oriented dialogue. *Computer Speech & Language*, 24(2):175–189, 2010. URL <https://api.semanticscholar.org/CorpusID:14637469>.
- Xiaoxue Zang, Abhinav Rastogi, Srinivas Sunkara, Raghav Gupta, Jianguo Zhang, and Jindong Chen. MultiWOZ 2.2 : A dialogue dataset with additional annotation corrections and state tracking baselines. In *Proceedings of the 2nd Workshop on Natural Language Processing for Conversational AI*, pages 109–117, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.nlp4convai-1.13. URL <https://www.aclweb.org/anthology/2020.nlp4convai-1.13>.
- Ke Zhai and Jason D. Williams. Discovering latent structure in task-oriented dialogues. In Kristina Toutanova and Hua Wu, editors, *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 36–46, Baltimore, Maryland, June 2014. Association for Computational Linguistics. doi: 10.3115/v1/P14-1004. URL <https://aclanthology.org/P14-1004/>.