Dialogue & Discourse 16(2) (2025) 1–34 doi: 10.5210/dad.2025.201

Speech acts that support other speech acts

Katja Jasinskaja KATJA.JASINSKAJA@UNI-KOELN.DE
University of Cologne

Frank Zickenheiner FRANK.ZICKENHEINER@UNI-BONN.DE
University of Bonn

Editor: Jonathan Ginzburg

Submitted 06/2024; Accepted 11/2025; Published online 11/2025

Abstract

Theories of discourse structure and existing discourse structure annotation schemes (e.g. Mann and Thompson, 1988; Sanders, 1997; Asher and Lascarides, 2003; Webber et al., 2019) often make a distinction between *propositional*, *subject-matter*, or *semantic* coherence relations on the one hand, and *speech-act-level*, *presentational*, or *pragmatic* relations, on the other. While there have been several convincing attempts to circumscribe the space of all possible propositional relations and to subdivide it into theoretically motivated subcategories (Sanders et al., 1992; Kehler, 2002), to date there is no comparable comprehensive taxonomy for speech-act-level relations. This paper develops a fragment of such a taxonomy, which describes what we call *support relations*—relations that connect two speech acts of the same speaker iff one of them fails to achieve its goal and the other helps achieve that same goal, as for instance, in *Evidence* relations, where one speech act makes the proposition asserted in the other more believable. We provide conceptual motivation for the proposed categories grounded in insights from sociolinguistic, psychological, and philosophical studies of human communication and illustrate the categories with examples from naturally occurring discourse, some of which do not fit easily into any existing classifications.

Keywords: coherence relations, speech acts, communicative goals, dialogue, monologue, belief

1. Introduction

The central assumption of a large group of approaches to discourse structure (Mann and Thompson, 1988; Sanders et al., 1992; Kehler, 2002; Asher and Lascarides, 2003) is that a discourse is coherent to the extent that hearers or readers are able to connect the sentences and larger discourse units it consists of with meaningful links—coherence relations (*alias* rhetorical relations or discourse relations, see Jasinskaja and Karagjosova, 2020, for an overview). While details might differ between the approaches, many of them make a distinction between *propositional*, *subject-matter*, *event-level*, or *semantic* relations on the one hand (1-a), and *speech-act-level*, *presentational*, or *pragmatic* relations, on the other (1-b). So in (1-a), a causal coherence relation holds between the described events of pushing and falling. In (1-b), the relation can be characterised in two ways: either as a causal relation that holds between the speech act of asking the question and the fact that there is a good movie on (Sweetser, 1990), or as a relation between two speech acts, a question and an assertion, where the purpose of the assertion is to explain why the question is justified or relevant to the speaker.

(1) a. Max fell. John pushed him.
b. What are you doing tonight? Because there's a good movie on.

The focus of this paper is on cases like (1-b) and on pragmatic relations of the latter kind—those that connect two speech acts in discourse. This paper pursues three goals:

A comprehensive taxonomy: While a number of speech-act-level relation types have been identified in previous studies and existing discourse structure annotation schemes (Asher and Lascarides, 2003; Sanders et al., 2009; Webber et al., 2019), to date there is no comprehensive taxonomy of such relations. The research reported in this paper seeks to shed light on the question of what kinds of coherence relations between speech acts exist, *should* exist, and are *theoretically possible*. The emphasis is here on 'should' and on 'theoretically possible', as our approach is not empirical or exploratory in the sense of looking for speech-act-level relations in naturally occurring discourse and trying to define fitting categories for the cases we find. Our goal is to find out how speech acts should relate to each other based on a theoretical understanding of what a speech act is, what definitional properties it has, and how those properties impose constraints on speech act combinatorics.

The question of the existence of an exhaustive list of coherence relations is fundamental in relational theories of discourse coherence. Without an exhaustive list, there is no substance in the claim that a discourse is coherent only if we can connect all its parts by *some* coherence relation. If we fail to connect the discourse units with a relation from a given list, knowing that the list is not exhaustive, we still cannot say whether the discourse is therefore incoherent or we just have not yet encountered a suitable relation. This has been a major point of criticism directed at the empirical, bottom-up approach of Rhetorical Structure Theory (RST, Mann and Thompson, 1988), which has never claimed to provide a comprehensive taxonomy (see discussion in Knott and Dale, 1994; Kehler, 2002; Jasinskaja and Karagjosova, 2020).

On the other hand, the existing conceptually-driven top-down approaches have not provided a comprehensive taxonomy of pragmatic relations so far. Kehler (2002) offers a prime example of the theoretical methodology we adopt, but only covers semantic relations. While Sanders et al. (1992) propose a top-down taxonomy of pragmatic relations, in the next section we argue against their specific way of dealing with the problem. Some of the same criticism also applies to Segmented Discourse Representation Theory (SDRT, Asher and Lascarides, 2003).

In this paper we argue that coherence relations between speech acts are best understood in terms of relations between their communicative goals. At this point, we only develop a fragment of the taxonomy, focusing on what we call *support relations*, i.e. relations where one speech act is produced as a means to successfully achieve the communicative goal of another speech act of the same speaker. The class of support relations overlaps several prominent relation classes familiar from previous studies, in particular, presentational relations in RST and some kinds of self-repairs (Levelt, 1983; Clark, 1994). However, it also includes cases that do not seem to fit comfortably into any existing classifications or annotations schemes. The relation between (2-b) and (2-c) is a case in point:

(2) a. "Pa told Peter he wanted *him* to be chairman."
b. "Sure he did.
c. If you don't believe me, ask Peter."[1]

1. From *Night over Water* by Ken Follett. Note that (2-a) is uttered by one speaker, and (2-b) and (2-c) by another, as indicated by the quotation marks. A support relation holds between the utterances of the second speaker.

The speech act in (2-c) is a directive, which asks the addressee to seek evidence for the proposition asserted in (2-b). Its purpose is to make (2-b) more believable, but it does so without directly providing evidence, and therefore does not fall under the standard definitions of *Evidence* relations. The purpose of a comprehensive taxonomy of speech act relations is not only theoretical, but also empirical—to provide a better coverage of previously uncategorised cases, such as (2).

Coherence relations in dialogue: The focus of relational theories of discourse structure has traditionally been on written monologues that consist predominantly of assertions. Even though extensions of the approach to dialogue have been proposed (Taboada, 2004; Asher and Lascarides, 2003; Lascarides and Asher, 2009), the coverage of non-assertive speech acts remains limited. The framework developed in this paper naturally accommodates speech acts of all types. Besides, since the need to start a second attempt to achieve the same communicative goal usually arises when the first attempt was in some way unsuccessful, support relations help describe sequences where things do not go according to the speaker's plan and cover a range of cases that fall under the notions of disfluency and self-repair—phenomena typical in spontaneous spoken communication.

Grounding the structure of monologue in interaction: Support relations in the definition we propose are relations between speech acts of *the same speaker*. Therefore, they can occur both in conversation, for instance, where one speaker produces a turn at talk that consists of more than one utterance, and in *bona fide* monologue, like an email, a lecture, or a software user manual. However, by viewing support relations as a communicative practice that is called for when the speaker encounters or anticipates resistance from the addressee in adopting the speaker's communicative intentions, we look at this coherence relation type as something that exists in interaction, be that with an actual physically present addressee, or a projected hypothetical audience. This is based on the idea that even extended written monologue is embedded in interaction.

Furthermore, historically, writing and most monological text types are more recent inventions than spoken face-to-face dialogue. They are also acquired later in the course of individual development than face-to-face dialogue communication skills. So it stands to reason that at least some discourse-structural patterns that we find in monologue will have developed from communicative practices used in dialogue.

In this paper, we propose a way to look at coherence relations used in monologue through the lens of dialogue. While we are not going to provide proof or empirical evidence for this claim, we would like to put it out there as a hypothesis for future research to engage with.

This paper is structured as follows. In section 2, we argue that coherence between speech acts is governed by a different set of principles and should not be treated by analogy with propositional-level or semantic relations. We discuss previous work on speech-act-level coherence in dialogue and show how those findings can be applied to monologue. Finally, the notion of support relation is introduced in that section. In section 3, we define goals of speech acts and identify necessary subgoals that a speech act must achieve in order to be successful. Section 4 presents a fine-grained classification of support relations along two main dimensions: (a) the subgoal at which the initial speech act fails, which the supporting speech act tries to repair; (b) the means available for doing so, depending on the type of goal state to be achieved (belief, desire, emotion, action, etc.). Finally, in section 5 we outline the general idea for how our approach extends to other kinds of relations between speech acts, going beyond support.

2. Relations between speech acts

This section reviews some previous approaches to coherence at the level of speech acts, focusing on the question of how relations between speech acts make a discourse coherent, and, conversely, what it means for a discourse to be incoherent at the speech act level. Section 2.1 looks into the matter from the perspective of the by far better understood coherence relations at propositional level and argues against the common assumption that the same coherence principles apply at both levels. Section 2.2 summarises some relevant insights concerning coherence relations between speech acts across speakers in dialogue. Section 2.3 discusses some previous ideas on how these insights can be made fruitful to understand coherence between utterances of the same speaker, within a single conversational turn or in monologue more generally. Finally, building on these ideas, section 2.4 introduces the notion of *support relations*, a subclass of relations between speech acts of the same speaker whose contribution to discourse coherence is the main subject of this paper.

2.1 Semantic relations at speech act level

Many existing approaches to discourse coherence (Mann and Thompson, 1988; Redeker, 1990; Sweetser, 1990; Sanders et al., 1992; Sanders, 1997; Asher and Lascarides, 2003) make a distinction between coherence relations at propositional and at speech act level. The exact boundary might be drawn slightly differently depending on the approach, but the general idea is roughly the same. Here we will refer to Sanders et al.'s (1992, pp. 7–8) definitions of what they call *semantic* vs. *pragmatic* relations:

> '*A relation is semantic* if the discourse segments are related because of their propositional content. In this case the writer refers to the locutionary meaning of the segments. The coherence exists because the world that is described is perceived as coherent.'
>
> '*A relation is pragmatic* if the discourse segments are related because of the illocutionary meaning of one or both of the segments. In pragmatic relations the coherence relation concerns the speech act status of the segments. The coherence exists because of the writer's goal-oriented communicative acts.'

There are two relevant things to note about these definitions. First, note the asymmetry between them. The definition of semantic relations clearly specifies what it is about the locutionary meaning of the discourse segments that makes the sequence coherent—namely the coherence of the world described. Following David Hume's classification of relations between ideas, Kehler (2002) proposes that there are exactly three ways in which elements of the described reality can cohere or belong together—by causal relations, by spatio-temporal contiguity, or by resemblance (similarities and differences). If we can recognise such relations between objects, states, and events in the world, we perceive that world as coherent, otherwise we don't. If we can infer these relations from a description of a complex state of affairs, we perceive that description as a coherent discourse at the semantic level.

In contrast, Sanders et al.'s definition of pragmatic relations does not specify what it is about the goal-oriented communicative acts that creates coherence. However, in practice, the authors simply transfer the same set of relations that are motivated from the point of view of coherence in the world (e.g. causal relations) to the domain of speech acts. Although the issue is not explicitly addressed in that work, in our understanding, this basically amounts to saying that coherence between speech

acts is governed by the same principles as coherence between propositions and ultimately depends on coherence between states and events in some relevant world. (While this is a sensible null hypothesis, below we argue that it is wrong.)

Second, note that according to Sanders et al.'s definition, in pragmatic relations the discourse segments are related because of the illocutionary meaning of *one or both* of the segments. That is, we have two kinds of cases. The first one is where the relation holds between the performance of a speech act and the content, or the proposition expressed by another discourse segment. We will refer to this case as a *proposition–speech-act relation*. The second case is where the relation holds between two performances of two speech acts—a *speech-act–speech-act relation*.

Proposition–speech-act relations are the better studied of the two. This is the standard way of looking at examples like (1-b). The connective *because* encodes a causal relation between the content of its syntactic argument, expressed by the subordinate clause, and another argument recovered from the context. In (1-b), that argument happens to be the performance of the question speech act, so (1-b) can be paraphrased as (3-a). But since the second argument is a proposition rather than a speech act, (1-b) cannot be paraphrased as (3-b).

(3) a. *I ask you* what you are doing tonight because there's a good movie on.
b. #*I ask you* what you are doing tonight because *I inform* you that there's a good movie on.

The idea goes back to Sweetser (1990), but is also found in metatalk relations in Segmented Discourse Representation Theory (SDRT Asher and Lascarides, 2003), speech act relations in the annotation scheme of the Penn Discourse Tree Bank (Webber et al., 2019), and in Sanders et al.'s own work. As pointed out above, the tacit assumption here is that the same relations that exist at the propositional level will also be possible between a speech act and a proposition.

In the more than thirty years since Sweetser's and Sanders et al.'s seminal work one would expect to have encountered pragmatic versions of all sorts of semantic relations. However, this only appears to be the case for relations that have a causal or conditional component, primarily *Explanation*, as in (1-b), *Result* (4-a) and *Conditional* (4-b) (*Result** and *Consequence**, respectively, in Asher and Lascarides, 2003, p. 334), *Negative Conditional* (4-c) (Webber et al., 2019, p. 23) and *Concession* (4-d).

(4) a. I'm cold. Please close the window.
b. If you failed the test, then why should I listen to you?
c. Unless you're on a diet, there are some cookies in the cupboard.
d. Although I hate to say it, please don't panic purchase or hoard Decaf.[2]

On the other hand, we have not seen any discussion of proposition–speech-act versions of *Parallel*, *Contrast* (other than the denial-of-expectation type of contrast $\approx$ concession), or *Narration* so far. In fact, we even find it difficult to construct a coherent sequence where the whole point of one sentence is to describe an event that happens before or after the speech act expressed by the other, which would fit the definition of a proposition–speech-act *Narration*. We are also not familiar with any attempts to explain these gaps in the paradigm. SDRT might explain away the absence of metatalk *Parallel* and *Contrast* by pointing out that metatalk counterparts only exist for content-level relations in SDRT's narrow understanding of the term, whereas *Parallel* and *Contrast* are not

2. From https://www.mycuppa.com.au/blogs/news/aug-2022. Last checked on November 17, 2025.

content-level, but text-structuring relations (Asher and Lascarides, 2003, pp. 465–466). However, why isn't there a metatalk version of *Narration*, which is a content-level relation by all definitions?

Speech-act–speech-act relations have received less attention, but for instance PDTB treats them in the same way as proposition–speech-act relations. Webber et al. (2019, p. 25) characterise (5-a) as a *Concession* that holds between two speech acts and can be paraphrased as (5-b).

(5) a. He lived in Peking, or should I say Beijing, for 20 years.
b. While *I say* he lived in Peking, *it might be more accurate to say* he lived in Beijing.

The problem is that if we continue to follow the same logic, transferring the same set of relations from the domain of propositions to the domain of speech acts, for speech-act–speech-act relations this makes even less sense. For instance, one could argue that *Narration* holds between the two speech acts in (6-a), because one immediately follows the other in time and the sequence can be paraphrased as (6-b), but the sequence is nevertheless perceived as incoherent.

(6) a. # John broke his leg. I like plums. (Knott and Dale, 1994)
b. *I say* John broke his leg, and then *I say* I like plums.

Moreover, any pair of adjacent speech acts in discourse are spatio-temporally contiguous by definition. If spatio-temporal contiguity were enough to make a discourse coherent at speech act level, then all sequences of speech acts would be coherent. Clearly, such a notion of discourse coherence would not be very useful.

A possible objection to this argument could be that it is not the *actual* spatio-temporal contiguity or adjacency that is relevant to establish coherence between speech acts, but the *expected* adjacency (see Fetzer, 2013, on a related distinction between adjacency position and adjacency expectation). The sequence in (6-a) is incoherent, because the second speech act presents an unlikely, unexpected follow-up to the first. In contrast, (7) is coherent because the second speech act is a felicitous answer to the question and in that sense is expected in that context. One could even argue that A's question causes B to answer, so a causal coherence relation holds between these speech acts.

(7) A: What's your favourite fruit?
B: I like plums.

The problem with this view is that it does not really give an answer to the question what makes (7) coherent at speech act level, but simply restates the question. Intuitively, (7) is coherent for a different reason than (1-b), (2), or (5-a) is. We want to understand that difference rather than just saying that all four are coherent because some speech act presents a likely follow-up to another. Therefore, we will not attempt to save the idea that coherence at the semantic and the pragmatic level is governed by the same principles and described by the same set of coherence relations. Instead, we will approach the issue from a completely different, genuinely speech-act-oriented perspective, focusing on speech-act–speech-act relations. While our approach might ultimately also help explain the gaps in the paradigm of proposition–speech-act relations, this is not our goal in this paper.

2.2 Relations between speech acts in dialogue

Interestingly, the foundation for an answer to the question of what makes a sequence of speech acts coherent was already laid by Austin (1962, p. 160), who pointed out that many speech acts

have illocutionary forces that "invite by convention a response or sequel". For example, if one asks someone to do something, the hearer is invited to perform a certain action, if someone asks a question the hearer is invited to give an answer, and so on. Searle (1969, 1983) labels the relevant property of a speech act as the *condition of satisfaction*. For example, the condition of satisfaction of an order is that the hearer obeys that order, the condition of satisfaction of a question is that the hearer gives an answer to that question, and so on. Obviously, not all speech act types will be "satisfied" by another speech act, but some will (question–answer pairs being the paradigmatic case), and in those cases one could say that the sequence is coherent because the second speech act fulfils the satisfaction condition of the first.

In the philosophical tradition of speech act theory, this idea was further developed by Sbisà (2002), who argued that conventional effects of illocutionary acts make sequences of speech acts possible, and sequences, in turn, make it possible for individual speech acts to achieve their conventional effects. However, there have been altogether few attempts to apply the notions of traditional speech act theory to explain coherence of speech act sequences (notable exceptions being Sbisà (2002); Franke (1990), Fritz and Hundsnurscher (2009) who analysed possible reactions to accusations and Hindelang (2010), who classified speech acts in sequences based on the work of Franke, Fritz and Hundschnurscher).

The issue received more interest in conversation analysis. Sacks and Schegloff coined the term *adjacency pair*, referring to pairs of utterances where the second one constitutes a proper reaction to the first one. Typical examples are *question – answer*, *greeting – counter greeting* or *offer – acceptance/refusal*. Sacks and Schegloff (1973) did not attempt to explain these pairings in terms of the notions of speech act theory, but what is clear from the known listings is that different illocutionary forces are paired with different types of reactions. Within this approach one could say that a pair of speech acts is coherent if it constitutes an adjacency pair, and even if we do not fully understand which properties of the speech acts are responsible for the legitimate pairings, there clearly is a connection.

Finally, an utterance may constitute a coherent follow-up to another one, not only if it satisfies the expectations set up by the first utterance, but also when it frustrates those expectations, but only in a way that ultimately serves the purpose of the conversation. Franke (1990), for example, introduced *decision-preparing reaction moves*, which are performed to help decide between a positive and a negative response to a previous speech act. Franke's decision-preparing moves include clarification requests (which were also studied at length in computationally oriented approaches to dialogue, e.g. Ginzburg, 2012) and speech acts that raise other kinds of problems related to a previous speech act. With respect to speech-act-level coherence that means in particular that a clarification request can constitute a coherent follow up to a speech act of any type, as long as it helps solving communicative problems related to it.

Adjacency pairs and relations between a speech act and a clarification request it triggers are, in our view, genuine speech-act–speech-act coherence relations, because the reason why those speech acts constitute a coherent sequence can only be understood if we take into account their illocutionary forces and/or perlocutionary goals. Incidentally, all these relations connect speech acts produced by two different speakers where the second one is a reaction to the first. Franke's (1990) classification of 'speech acts of the second move' can be seen as a proposal towards a comprehensive taxonomy of such relations, and Ginzburg et al.'s (2022) response spaces is a fragment that describes reactions to questions. These relations will normally occur in interactive discourses between two or more speakers, so one might get the impression that coherence at speech act level is a phenomenon

restricted to dialogue. The next section shows what we can learn from dialogue about coherence relations between speech acts of the same speaker in dialogue or monologue.

2.3 From dialogue to monologue

The notions of dialogue and monologue can be understood holistically, as categories of discourse types; for instance, an interview is a dialogue, and a novel is a monologue. In this paper we adopt a more technical definition: a sequence of utterances is a dialogue if it contains an exchange of speaker and hearer roles between at least two participants. A sequence of utterances is a monologue if all those utterances are produced by the same speaker without interruption by another speaker. In this sense, an extended turn at talk is a monologue, so monologues can be part of dialogues.

Coherence relations between utterances of the same speaker will generally occur in monological sequences, but they can also hold between two non-adjacent utterances of the same speaker separated by an utterance of a different speaker in a dialogical sequence. In a dyadic exchange between speakers S1 and S2, after S2's reaction to S1's initial utterance, the turn usually comes back to S1. One might wonder whether the relationship between S1's initial speech act and her reaction to S2's reaction can be fruitfully analysed in speech-act-theoretic terms. For instance, S1 could use an ambiguous pronoun in the first move (8-a), S2 could react with a clarification request in the second move (8-b), to which S1 could provide the desired clarification in the third move (8-c). The question is then: How do the first and the third utterance, (8-a) and (8-c), cohere at speech act level? Which properties of these utterances as goal-oriented communicative acts are essential for the overall coherence of the sequence?

(8) a. $S1$: At that point, it was over across the road.
b. $S2$: What do you mean 'it'?
c. $S1$: The warehouse.

A systematic classification of 'speech acts of the third move' has been developed by Franke (1990). After a clarification request or some other kind of reactive move that frustrates the goals of the initial utterance, the speech act in the third move may belong to one of three classes according to the relationship between its goals and the goals of the initial speech act: *Retractive speech acts* signal that S1 completely gives up on the goal of her initial speech act; *revising speech acts* modify the communicative goal of the initial speech act in such a way that it is more likely to be achieved; finally, *re-initiative speech acts* present a second attempt to achieve the goals of the initial speech act without modification. The third move in (8) is an instance of a re-initiative speech act. In (8-a) S1 is trying to inform S2 that the warehouse was across the road; in (8-c) she is still trying to do so. In this respect, (8-c) presents a coherent follow-up to (8-a) and (8-b) because it continues to pursue the goals of (8-a) after (8-b) indicates that they were not reached in the first attempt.

Notice that re-initiation is a relation between utterances of the same speaker. But speakers often do not wait for an explicit clarification request from their interlocutor, but detect potential communicative problems by monitoring their own speech, and produce re-initiative speech acts without conversational turn transition, as is the case in most self-repairs (Levelt, 1983). Therefore re-initiation can function as a speech-act-level coherence relation in a monologue. This idea is developed by Ginzburg et al. (2014) as a general approach to self-repair moves and other kinds of speech disfluencies. In this approach, (9) has the same underlying structure as (8), with the

difference that in (8) the clarification request is made explicit, whereas in (9) it constitutes an implicit question under discussion (QUD).[3]

(9) At that point, it, the warehouse was over across the road.

As far as the source of coherence is concerned, we could say that *the warehouse* constitutes a coherent interjection within the utterance *At that point, it was over across the road* because it ultimately helps achieve the initial goal of informing the hearer that the warehouse was across the road. Coherence is established at the level of the communicative goals.

2.4 Support relations

Adjacency pairs, clarification requests, retractions, revisions, re-initiations, and (self-) repairs are relational categories developed in previous research that, in our view, genuinely pertain to the speech act level of coherence because the source of coherence (the reason why a sequence is coherent) in all these cases has something to do with the communicative goals of the speech acts involved and the relative success or failure in achieving those goals. Of course, causality also plays an important role here since one speaker communicating a certain goal can *cause* the other speaker to want to satisfy that goal, or the failure to achieve a certain goal can *cause* the speaker to start another attempt, but as we argued in section 2.1, in order to achieve a deeper understanding of coherence at speech act level, we should shift our focus from causality in general to more specific relationships between communicative goals. Therefore, the focus of this paper will be on a subclass of relations between speech acts that we will call *support relations*:[4]

(10) SUPPORT RELATIONS:
Speech act S of speaker A *supports* speech act N of the same speaker iff
a. A believes that N has failed or will fail to achieve its goal
b. A believes that S will help to achieve that goal.

As should be clear from the discussion in previous sections, this definition primarily covers self-repairs and Franke's re-initiations. It is also closely related to the notion of *presentational relations* (11) in RST (Mann and Thompson, 1988):

(11) PRESENTATIONAL RELATIONS: Presentational relations are those whose intended effect is to *increase some inclination* in the reader, such as the desire to act or the degree of positive regard for, belief in, or acceptance of the nucleus. (Mann and Thompson, 1988)

All presentational relations in RST are relations between a nucleus (N), a discourse unit that is more central to the overall purpose of the discourse, and a satellite (S), a discourse unit whose function is defined in relation to the nucleus. The N and S notation in (10) is our adaptation of this distinction to support relations, where S stands for support, and N for the unit being supported.

There is a substantial overlap in the range of cases covered by the definitions (10) and (11). Both types of relations have in common that S has the function to increase the chances of the success of N. For instance, if the goal of N is that the hearer believe a proposition and S increases the hearer's

3. (9) is the original example cited by Ginzburg et al. (2014). The example in (8) is constructed from it.
4. Note that our notion of support is quite different from that of Grosz and Sidner (1986), who we otherwise have a lot in common with, see section 3 for more discussion.

belief in that proposition, as in the case of *Evidence* relations like that in (12), then S also helps achieve the goal of N.

(12) a. He must have been here recently.
b. There are his footprints.

However, in this paper, we pursue a methodological approach radically different from that of RST. The approach of RST is empirically driven and bottom-up, the proposed set of rhetorical relations is motivated by what can be found in naturally occurring texts, and is potentially open for new additions. This approach has been criticised for its failure to give a principled basis for saying what a theoretically (im)possible coherence relation is (Knott and Dale, 1994). As a consequence, the resulting list of relations does not provide a basis for making predictions about the coherence of texts. For an incoherent sequence like that in (6-a), we cannot claim that it is incoherent because there is no suitable relation in the list, as nothing prevents us from adding an *inform-accident-and-mention-fruit* relation, as Knott and Dale (1994) put it.

Our goal is to ensure that there are no inform-accident-and-mention-fruit relations at the speech act level. In this paper, we address this challenge by taking a strictly top-down approach. We first try to give answers to the following questions: What kinds of goals can speech acts have? What are the possible ways to fail those goals? What are the possible ways to achieve those goals? That provides the basis for our taxonomy of all theoretically possible support relations. As will become clear, this approach does not only reveal systematic similarities between self-repairs, Franke's re-initiations and RST's presentational relations, but also creates previously unknown categories for speech-act-level coherence relations.

One last remark before we embark on this endeavour: Presentational relations in RST do not impose any general constraints on the linear order of the nucleus and the satellite, both (N, S) and (S, N) sequences are possible. Repair moves can also generally follow the reparandum, or be linearly embedded in it, as in (9) (or even precede it, although in this case the reparandum is rarely fully articulated, cf. Forward Looking Disfluencies in Ginzburg et al., 2014). In principle, supporting speech acts can also follow, precede or be linearly embedded in their nuclei. However, this paper will mainly focus on the case of (N, S), where support follows the nucleus.[5]

3. The goals of a speech act

As defined in the previous section, supporting speech acts and support relations are produced in order to help achieve the goal(s) of a speech act that is (anticipated to be) unsuccessful. A goal of an action, and a speech act in particular, is its intended effect—a proposition that describes the way the agent wants the world to become as a result of the action (see Zickenheiner, 2020, for a formal implementation of the idea in Discourse Representation Theory).

There is a vast theoretical tradition that recognises the central role of speech act goals in the structuring of discourse, most notably Grosz and Sidner (1986), as well as Roberts (1996) and Ginzburg (2012) with questions under discussion as a specific conceptualisation of goals. Our approach follows the general architecture of Grosz and Sidner (1986) with respect to relationships between goals (e.g. dominance, or goal–subgoal relationships) and the way they are managed in a stack structure in the flow of communication (again, see Zickenheiner, 2020, for formal details).

5. An example of a support relation with a (S, N) order of segments is (25), discussed in section 4.1.

Unlike Grosz and Sidner, we do not commit to the existence of a single or main goal of a speech act. The way the world should become as a result of an action can be described in many different ways, emphasising different aspects and at different levels of specificity. In principle, all those descriptions can be seen as goal specifications for any given speech act. However, it is important to distinguish what constitutes the goal of a specific speech act, rather than a sequence of multiple speech acts. For instance, in (13) the speaker ultimately wants the hearer (H) to believe that Fred ate the beans (f) and Mary ate the eggplant (m), or $believe(H, f) \wedge believe(H, m)$. The first speech act in (13-a) is a step in that direction, but it is not meant to achieve that goal by itself. We would not consider (13-a) to be unsuccessful if the hearer does not also acquire the belief that m in addition to the belief that f. Therefore, $believe(H, f) \wedge believe(H, m)$ is not a goal of (13-a), only $believe(H, f)$ is.

(13) a. Fred ate the beans.
b. Mary ate the eggplant.

The difference is important for distinguishing between Grosz and Sidner's dominance and utterances dominated by the same goal on the one hand and our support relations on the other. Even though at a certain level of abstraction, (13-a) and (13-b) pursue the same joint goal of $believe(H, f) \wedge believe(H, m)$, they do not pursue that goal each individually. The goals they pursue individually are distinct, and therefore the relation between (13-a) and (13-b) is not support.

On the other hand, a whole cascade of goals have to be achieved for an individual speech act to be fully successful. The intermediate steps with their respective goals that we assume to lead to communicative success (inspired by Clark's 1994; 1996, levels of action in communication, going back to Allwood et al. 1992) are listed in figure 1. For illustration, consider the speech act *Off with their heads!* performed by the Queen of Hearts upon seeing that her gardeners had planted roses in the wrong colour in Lewis Carroll's *Alice's Adventures in Wonderland*:

(14) 'I see!' said the Queen, who had meanwhile been examining the roses. 'Off with their heads!' and the procession moved on, three of the soldiers remaining behind to execute the unfortunate gardeners, who ran to Alice for protection.

In spoken communication, the first step is the production of a certain acoustic signal. The goal is to get sound waves of appropriate shape to travel to the right place at the right time (signal transmission). Second, the acoustic signal must be processed by the hearer's auditory system, resulting in a certain pattern of activation in the hearer's brain and, presumably, a representation in the hearer's mind (signal processing).

The next steps correspond more or less closely to Clark's levels of action in communication.[6] After the sounds have been recognised as speech, the hearer must first of all pay attention to the signal (*attention*), which is a prerequisite for all further processing. While Clark treats attention as the first step in the sequence on a par with the subsequent steps, we see it as a state that must hold for the whole duration of communication and in that sense cannot be meaningfully ordered with respect to the other steps.

6. Clark emphasises the joint character of communicative action in which both the speaker and the hearer have a role to play. He identifies four levels of action: 1. vocalisation and attention, 2. presentation and identification 3. meaning and understanding and 4. proposal and uptake (the first term in each pair describes the speaker's part and the second term the hearer's part in the joint action). In this paper, we think of the hearer's part as the goal to be achieved by the speaker's action.

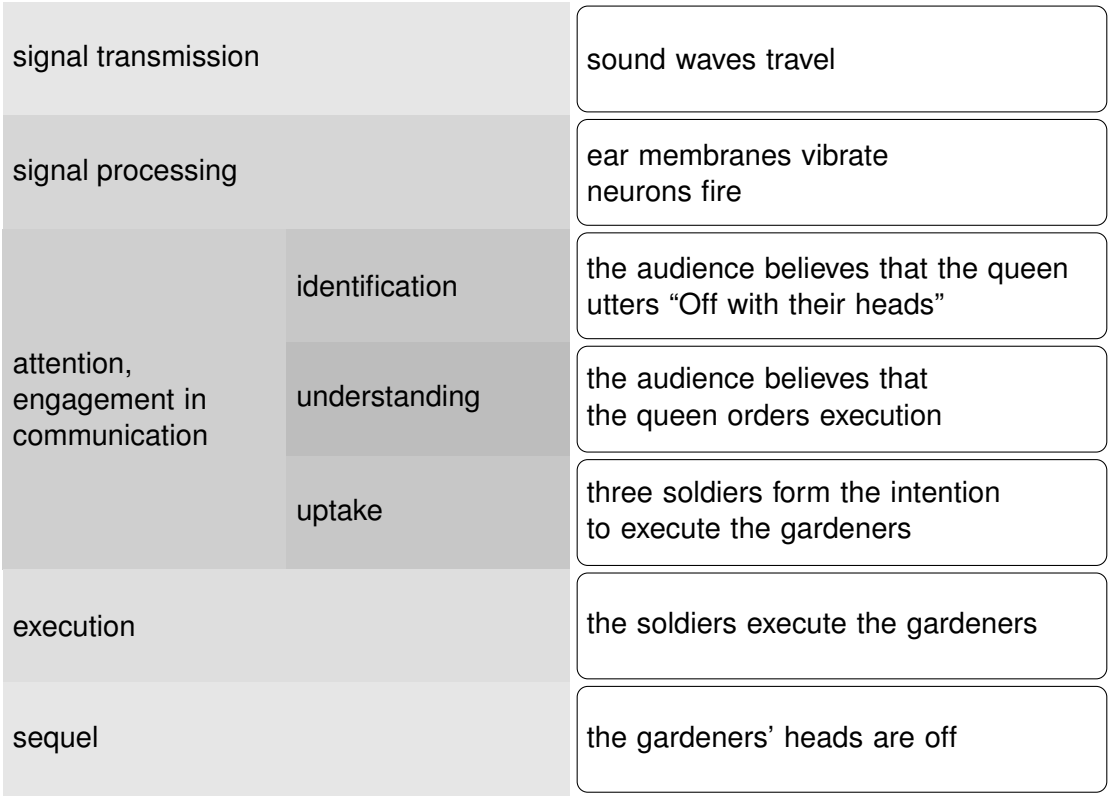

Figure 1: The workflow of the speech act *Off with their heads!* in (14).

Provided that the hearer is paying attention and engaging in the communication, they normally need to identify what the speaker *said* and what the speaker *meant*. In the next step, the queen's audience has to map the sounds they hear to the words *Off, with, their* and *heads*, i.e. create a representation of the linguistic form of the utterance (*identification*). Then, they must come to the belief that the queen produced these words because she wanted to order the execution of the gardeners. That includes recognition of both the semantic content of the utterance and the speaker's communicative intention (*understanding*). After the queen's message was understood, the ultimate success of the speech act will depend on whether or not the hearer reacts in the expected way. We divide the hearer's reaction into a 'mental' part (*uptake*) and a 'physical' or 'active' part (*execution*). The 'mental' part of the goal in the present example is that some of the soldiers (three out of ten taking part in the procession) form the intention to execute the gardeners.

A few clarifying remarks on our notion of uptake are in order. The term dates back to Austin (1962), who used it in a rather broad sense to refer to the whole range of effects a speech act can have upon the hearer, including Clark's identification, understanding, uptake in the narrow sense, and probably our execution as well. Clark (1996) draws a line between uptake and understanding: Understanding is the correct construal of the speaker's action, e.g. an order to behead the gardeners is recognised as an order to behead the gardeners. Uptake, in turn, is the hearer's action upon the proposal. However, Clark seems to include a number of rather different things in that.

On the one hand, Clark's notion of uptake includes appropriate reactions in dialogue, i.e. roughly, the second parts of adjacency pairs. For instance, an answer to a question constitutes the hearer's uptake of the question. To the extent that the notion of adjacency pair is taken to cover cases of non-speech action, the soldiers beheading the gardeners would constitute the uptake of the Queen's order (see also Hulstijn and Maudet, 2006).

On the other hand, Clark also counts mere consideration of the speaker's proposal by the hearer as uptake. In our present example it would mean that uptake already takes place when the soldiers contemplate whether or not they should follow the Queen's order. This weaker notion of uptake has been more widely adopted in subsequent research, to the point of complete replacement of the notion uptake by the notion of consideration (see e.g. Rodríguez and Schlangen, 2004).

In the very narrow sense of uptake that we adopt in this paper, neither counts as uptake. Consideration of the speaker's contribution, in our view, constitutes part of paying attention to the exchange and engagement in communication. As pointed out above, we assume that attention and engagement must be given at all levels of utterance processing, and consideration of the proposal should therefore be seen as engagement in communication at the level of uptake. Appropriate reactions in dialogue, such as answering a question or fulfilling an order, on the other hand, are *executions* of actions that may serve as evidence of the hearer's uptake, but they do not make up its essence.

In this paper, we will use the term 'uptake' to refer to the hearer's mental reaction to an utterance that is in accordance with the speaker's intentions for that utterance. This notion comes close to what Schlöder and Fernández (2015) call *intention adoption* (reaching mutual agreement), which they distinguish from *intention recognition* (understanding that goes beyond semantic interpretation). Along the same lines, we see intention recognition as part of pragmatic understanding, whereas the adoption of the speaker's intention is the hearer's mental reaction intended by the speaker.

What kind of mental reaction that is will generally depend on the speech act type. Uptake of a directive speech act, such as the Queen's order, is the adoption of the intention to fulfil that order. Uptake of an assertion is usually the belief of the asserted proposition. Uptake of an insult is feeling insulted, and so on. In any case, uptake goes beyond mere understanding of an utterance and includes mental compliance with the speaker's proposal, but does not go as far as the (physical) action that might result from that mental compliance, such as actually carrying out the order or signalling by means of a nod or a feedback utterance (e.g. *yes*, *mhm*) that the hearer accepts the asserted proposition. The latter belong to the level of execution.[7]

Finally, the goal of a speech act may go beyond the hearer's mental attitudes or physical actions and lie anywhere in the outside world. Any direct or indirect consequences of the speech act may constitute its ultimate intended effect. For instance, the state of affairs of the gardeners' heads being 'off' could be seen as the goal of the Queen's order at the level of the *sequel*. Note that the Queen does not specify how the gardeners' heads should get 'off', or who should do what to achieve that result. In fact, as we know from the further development of the story, Alice ensures that the gardeners' heads are 'off' without being parted from their bodies, and everyone appears happy with that solution. We could say that the *literal goal* of this speech act is that the gardeners' heads are

7. Admittedly, in some cases it might be difficult to draw a line between uptake and execution, especially in directives that concern mental acts. For instance (i) is literally a request to make the stated assumption, but it is hard to imagine forming an intention to make this assumption without making the assumption yet. It is not theoretically impossible, but something we probably very rarely do with simple mental actions of this kind.

(i) Suppose x is greater than 0.

'off'. This is also the *ultimate goal* of this speech act. That is, both the literal and the ultimate goal are located at the level of the sequel.[8]

Speech acts of different types expressed by different sentential moods will generally have literal/ultimate goals located at different levels. For instance, a typical directive expressed by an imperative sentence, such as (15), literally expresses the goal that the hearer perform the action of leaving the room. That is, the literal goal of this speech act (and taken at face value, also its ultimate goal) lies at the level of execution.

(15) Leave the room.

One might disagree on whether the goal that the hearer believe that it is raining is literally expressed by (16) or is the result of pragmatic inference. However, that goal clearly lies in the hearer's mind and therefore at the level of uptake. This is also its ultimate goal if the speech act is taken at face value, but if it is interpreted as an indirect speech act, e.g. as advice to put on appropriate clothing, then the ultimate goal lies at the level of execution.

(16) It is raining.

In other words, the ultimate goal of a speech act may pertain to the levels of uptake, execution or sequel, but reaching that goal will normally require reaching intermediate goals at the other levels leading up to it, which includes attention, identification and understanding. The linguistic form of the utterance, in particular its sentential mood, gives the hearer a clue of the intended ultimate goal, though it does not necessarily literally express it.

Things can go wrong at any of these levels. In the next section, we develop a taxonomy of support relations based on the location of the communicative problem as the main criterion for classification. However, first a few words are due on why we choose Allwood/Clark-style levels of action specifically as basis for our underlying classification of goals. Several related concepts that one could use instead or in addition come to mind.[9]

Since the fulfillment of a goal is a criterion for the success of a speech act, there is a certain overlap between our notion of goal and the speech-act-theoretic notion of *felicity condition* (Austin, 1962; Searle, 1969); however, the overlap is small. For instance, preparatory conditions of speech acts are not goals, because they need to be given before the speech act is performed, goals are achieved after. Besides, classical speech act theory focuses on illocution and has little to say about perlocutionary effects, such as the addressee's belief in the asserted proposition or emotional state triggered by an expressive speech act, while they are central in our approach. Clark's levels of action are a more useful framework for mapping the entire space of speech act goals because it takes the state of the addressee systematically into account.

Our notion of intended effect and the extension of Clark's hierarchy based on it is broader than the specific selection of intended effects (positive regard, belief, desire, acceptance of the right to present) that underlies RST's set of presentational relations, and therefore arguably is more compre-

8. To put these levels in relation to the standard speech-act-theoretic notions, reaching a goal at the level of identification corresponds roughly to performing the phonetic act in the sense of Austin (1962, p. 95). Reaching semantic understanding corresponds to the performance of a *locutionary act*, whereas reaching pragmatic understanding or what Schlöder and Fernández (2015) call intention recognition corresponds to the performance of an *illocutionary act*. Finally, achieving uptake, execution, or any more far-reaching goals that pertain to the sequel belongs in the domain of *perlocution*.

9. We thank the anonymous reviewers for bringing up these alternatives.

hensive. Clark's levels of action have been used as basis for functional classification of clarification requests and self-repairs by Rodríguez and Schlangen (2004) and by Clark (1994) himself. As we will see in the next section, our extended goals ladder provides necessary categories both for self-repairs and most of RST's presentational relations, bringing both sets of phenomena into a single conceptual framework.

In a taxonomy, there is no such thing as the correct level of granularity. For some applications, it might be enough to distinguish support from other kinds of speech act level relations. In that case, all the different categories presented in the next section are only interesting to the extent that they show what kinds of cases belong to the broad category of support. Accordingly, the distinctions between the types of goals underlying those categories would not be relevant for such an application. On the other hand, the categories can be subdivided further if necessary. For instance Traum (1994) additionally identifies a level for turn-taking acts (e.g. take turn, keep turn, etc.) and a level for argumentation acts (e.g. summarise, clarify, etc.). The former could be used in an approach like ours to further refine the taxonomy. The latter are relational by their nature. We hope that those or similar categories would follow from our definitions of speech-act-level relations rather than serving as input to them. The reason we believe that the level of granularity provided by Clark's levels of action is at least useful or interesting is because some of the resulting categories roughly match those identified in previous research, such as RST's *Evidence*, *Motivation*, and *Enablement* relations, as will become clear in the next section. If distinguishing those categories appeared relevant to analysts of naturally occurring discourse, then they might find the underlying categorisation of goals based on Clark relevant as well.

4. Towards a taxonomy

As we promised in section 2, the primary motivation for the taxonomy of support relations to be developed in this paper is theoretical rather than empirical. Our goal is to circumscribe all theoretically possible kinds of support relations based on the concept of support as a relation between speech acts. The classification criteria should therefore be based on the understanding of what support is and what a speech act is, on the constitutive parts of these notions and properties that are associated with them by definition.

The definition of support relations given in (10) is repeated below. Two main classification criteria follow from the two conditions in this definition. One can distinguish between different kinds of support depending on (a) how N failed, and (b) how S solves that problem.

(17) SUPPORT RELATIONS:
Speech act S of speaker A *supports* speech act N of the same speaker iff
   a. A believes that N has failed or will fail to achieve its goal, and
   b. A believes that S will help to achieve that goal.

In section 3, we offered our version of the levels of action that expands on and modifies that proposed by Clark (1996), cf. figure 1. However, the basic observation concerning these levels remains the same: a speech act can fail at any of the levels that it involves, and therefore a supporting speech act may be called for to repair a problem at any of these levels. So, the first dimension of our taxonomy is the location of the problem targeted by the supporting speech act. Section 4.1 gives an overview of the types of support relations according to this criterion. While we will have less to say about the second dimension—the types of support relations according to how they solve the problem at

hand—section 4.2 outlines our general strategy to approach this issue and showcases one group of support types aimed at affecting the hearer's beliefs.

4.1 Locating the problem

SUPPORT FOR ATTENTION / ENGAGEMENT IN COMMUNICATION

A classical example of a speech act that supports another speech act because the latter failed at the most basic level of securing the addressee's attention is (18) from Clark (1994). Here it seems that Bob did not even hear Ann's initial utterance (18-a). Ann solves the problem by repeating it more loudly (18-c). That is, (18-c) supports (18-a).

(18) a. Ann: Bob
b. Bob: [3 sec of no response]
c. Ann: Bob [louder]
d. Bob: What?

Clark's original ladder of levels of action creates the impression that attention, as the first step of that ladder, is on a par with the other steps in the sense that success at higher levels presupposes success at the attention level. This, in turn, seems to be based on the implicit assumption that once attention is achieved it cannot be taken away, similar to how understanding, once achieved, cannot be taken away.

However, as we pointed out in section 3, while identification, understanding and uptake are events where one is a precondition for the next one, attention is a state that needs to be given throughout the whole process of communication. Research on the psychology of attention has found that while attention can be induced as an automatic reaction to abrupt changes in the environment, it is generally given and maintained intentionally and is driven by the agent's domain-level goals and interests (Yantis and Jonides, 1984, 1990; Theeuwes, 1991; van der Lubbe and Postma, 2005).

Thus, the initial 'grabbing' of the addressee's attention that we see in (18) only marks the beginning of the attention state, after which attention needs to be maintained intentionally by the addressee. Example (19) shows that attention, or more generally the willingness to engage with the addressee, can be taken away after the utterance is identified and understood: (19-e) makes clear that the first person narrator has perfectly understood the lady's attempt at contact. Nevertheless, he denies her his attention and refuses to engage in the communication. It does not matter whether it is obvious to the speaker that the addressee has understood her utterances. By uttering (19-f) and (19-g) she is trying to (re)gain the addressee's attention. Therefore, (19-f) and (19-g) stand in a support relation to all or any of (19-a), (19-c) and (19-d).

(19) a. "Sorry,"
b. I hear somebody next to me say.
c. "Aren't you the man from the television?
d. The one who harassed those poor girls?"
e. Fuck... I don't answer [...]
f. "Hey,
g. I am talking to you."
h. The lady insists again.[10]

10. From *Almost* by Adriana LS Swift.

SUPPORT FOR IDENTIFICATION

This category of support relations mainly consists of self-repairs that target the hearer's failure to create the correct representation of the linguistic form of an utterance. In Clark's (1994) example, (20-c) supports (20-a).

(20) a. A: yes forty-nine Skipton Place
b. B: forty-one
c. A: n i n e . nine
d. B: forty-nine, Skipton Place,

SUPPORT FOR UNDERSTANDING

In section 3, we adopted a broad notion of understanding, which includes both semantic and pragmatic understanding. An utterance counts as fully understood only if the hearer is able to correctly identify the speaker's meaning behind it (Grice, 1957). That includes understanding the conventional meanings of the words and phrases, reference resolution, presupposition resolution, being able to draw the intended implicatures, understanding what kind of speech act the speaker intends to perform by means of the utterance, understanding whether the utterance is meant seriously or jokingly, etc.

Support relations that target semantic aspects of understanding are otherwise known as self-repair and reformulation. For instance, the self-repair in (21-c) clarifies the sense in which Ken uses the verb *evaluate* in (21-a).

(21) a. Ken: k who evaluates the property - - -
b. Ned: uh whoever you asked, . the surveyor for the building society
c. Ken: no, I meant who decides what price it'll go on the market -
d. Ned: (- snorts) . whatever people will pay

In (22), another example from Clark (1994), Sam's response *m* in (22-c) to Dar's clarification request supports (22-a) by confirming the reference of *this boy*:

(22) a. Sam: well wo uh what shall we do about uh this boy then
b. Dar: Duveen
c. Sam: m

These are instances of more or less spontaneous self-repair. But speech acts whose purpose is to improve the understanding of a previous utterance can also be planned. In (23), the speaker does not only replace the presumably unfamiliar term *anacrusis* with a more accessible definition, but the reformulation is used as a strategy to establish the equivalence of *anacrusis* and *an unaccented note which is not part of the first full bar* (Blakemore, 1993), i.e. to define the new term.

(23) a. This piece begins with an anacrusis,
b. an unaccented note which is not part of the first full bar.

Mann and Thompson's RST (Mann and Thompson, 1988, p. 273) includes a presentational relation *Background*, whose definition covers the essence of a support relation that targets a potential problem at the level of understanding (R: reader; N: nucleus; S: satellite):

(24) *Background*:
   a. R won't comprehend N sufficiently before reading text of S
   b. S increases the ability of R to comprehend an element in N

However, the cases that this definition is normally applied to are different from those mentioned above. In RST practice, the typical order of discourse units connected by a *Background* relation is (S, N), rather than (N, S). In (25) from Mann et al. (1989), the first sentence presents the event of media covering the results of ZPG's 1985 Urban Stress Test, which satisfies the existence presupposition of the definite NP *this remarkable media coverage* in the second sentence. Without (25-a) preceding it, (25-b) as it stands would be unacceptable in formally published written discourse, although with slight modifications the reverse order would also be possible, cf. (26).

(25) a. The results of ZPG's 1985 Urban Stress Test were reported as a top news story by hundreds of newspapers and TV and radio stations from coast to coast.
   b. I hope you'll help us monitor *this remarkable media coverage* by completing the enclosed reply form.

(26) a. I hope you'll help us monitor *the remarkable media coverage of the results of ZPG's 1985 Urban Stress Test* by completing the enclosed reply form.
   b. The results were reported as a top news story by hundreds of newspapers and TV and radio stations from coast to coast.

The second sentence in (26) reads like an afterthought that one would like to put in parentheses, and in that sense resembles the more spontaneous self-repairs in (21) and (22), which is probably why *Backgrounds* that follow their nuclei are less acceptable in written texts. However, what is common to (25) and (22) is that the purpose of the supporting speech act is to help establish the reference of a definite description, *this remarkable media coverage* and *this boy*, respectively.

Support relations that target pragmatic aspects of understanding have received less attention in previous research, or might have partly been handled under different unrelated categories. Understanding the speaker's intention behind the utterance involves many different layers, including the understanding of implicatures, illocutionary force, and perlocutionary object, among others. Example (27) is an instance of conversational implicature clarification. The first sentence (27-a) has two readings: the 'normal' one, without any notable quantity implicature, and the one where the quantity implicature *like* $\rightsquigarrow$ *not love* is drawn in the scope of negation. The second sentence (27-b) makes it clear that the more marked second reading is intended.

(27) a. Around here, we don't like coffee, (Horn, 1989, p. 382)
   b. we love it.

In (28), an excerpt from the novella *The Little Prince* by Antoine de Saint-Exupery, the Little Prince is in conversation with the King, who he meets on one of the planets he visits on his journey. In (28-d), the King orders the Little Prince to yawn. Just in case the exact illocutionary force of the imperative might be misunderstood, (28-e) supports (28-d) by clarifying that it is an order.

(28) a. It is years since I have seen anyone yawning.
b. Yawns, to me, are objects of curiosity.
c. Come, now!
d. Yawn again!
e. It is an order.

A supporting speech act may be called for if the perlocutionary goal of the nucleus needs to be clarified. For instance, (29-a) might be taken as an insult. In (29-b), the speaker makes clear that he does not intend to insult the hearer.

(29) a. Fuckin' I hate this guitar! I hate it so much!
b. No offence, Ed.[11]

Sweetser's (1990) classical examples of proposition–speech-act causality (30), as well as some instances of the RST relation *Justify* (31) might also belong to this category.

(30) a. What are you doing tonight?
b. Because there's a good movie on.

(31) *Justify*:
R's comprehending S increases R's readiness to accept W's right to present N

On the face of it, the perlocutionary goal of (30-a) is clear: The addressee should tell the speaker what they are doing tonight, but that might be seen as information the speaker is not entitled to. However, (30-b) explains why the question is relevant. That is, by uttering (30-b) the speaker shows that their underlying goal is to invite the addressee to go to see the film together, something that is meant for the addressee's benefit as much as their own. After understanding (30-b), the addressee does not even need to give a full answer to the original question. It is enough if they say whether or not they will go to the movies with the speaker as that is actually the ultimate goal.

Finally, the following example stems from a reference manual for Ableton Live 5, digital audio workstation software.[12] The purpose of user manuals is to give instructions about how to use a product. Instruction is a subtype of directive speech act that is appropriate to perform in a situation where the addressee has a certain goal in mind which he or she desires to achieve, but does not know how to do it. Therefore, by saying (32-a), the speaker insinuates that the addressee might want, among other things, to get rid of unwelcome house guests or terrifying small pets. In the supporting speech act (32-b), the speaker corrects for a potential misunderstanding, stating that (32-a) was meant as a joke.

(32) a. This can be very useful in creating new sounds and textures, as well as getting rid of unwelcome house guests, or terrifying small pets
b. (just kidding!).

It is important to note that sequences like (28-d)–(28-e), (29-a)–(29-b), (32-a)–(32-b) do not fit comfortably into the existing taxonomies of coherence relations. While they might literally fit the

11. From an interaction between Thom Yorke and his band during a studio recording: https://www.youtube.com/watch?v=6eRp97ZRwmk, time stamp 4:45–4:50. Last accessed on November 18, 2025.
12. URL: http://downloads.ableton.com/manuals/50/ableton_live_5_manual_en.pdf. Last accessed on November 18, 2025.

RST definition of *Background*, they do not fit the intuitive notion of background as something that is *behind* something in some sense (temporally, epistemically, or in the flow of communication) and they all show the non-canonical (N, S) order of segments. None of the PDTB speech act relations seems to apply in these cases. Perhaps, a case could be made that these are instances of proposition–speech-act, or metatalk *Elaboration*, but we are not familiar with any proposals that argue that.

SUPPORT FOR UPTAKE

The notion of uptake adopted in section 3 refers to the hearer's mental reaction that goes beyond mere understanding of the speaker's intention and encompasses cooperative adoption of that intention. It is the success of the utterance at the perlocutionary level, albeit limited to its mental component. The type of mental reaction in question will depend on the goal. For instance, the perlocutionary goal of an assertion is often to make the speaker believe a proposition. Therefore, the uptake of the assertion would consist in the hearer's adopting that belief. The perlocutionary goal of a directive is to bring the hearer to perform an action. The uptake of the directive would be the hearer's forming an intention to perform that action. Etc.

Support relations whose purpose is to secure uptake can be distinguished according to the type of mental act or state of the hearer that constitutes the goal of the utterance. If the goal is the addressee's belief in a certain proposition, then the supporting speech act should make that proposition more believable. The RST relation *Evidence* is geared towards this situation:

(33) *Evidence*
- a. R might not believe N to a degree satisfactory to W
- b. R's comprehending S increases R's belief of N

A typical *Evidence* is an assertion that states something more evident, i.e. something more directly observable, than the proposition expressed by the nucleus:

(34)
- a. He must have been here recently.
- b. There are his footprints.

But giving *Evidence* is not the only way to make a proposition more believable. In section 4.2 we will present a number of other support relations that also target the addressee's belief.

If the goal of the utterance is to bring the addressee to perform an action, the mental state that normally precedes that is the addressee's desire or intention to perform that action. This is typical for directive speech acts. If a directive is anticipated to fail, the speaker might perform a supporting speech act to make the addressee more willing to perform that action. The RST relation *Motivation* (35) covers this case neatly. For instance, (36-c) provides a reason why the google development team might want to fulfil the request expressed in (36-b).

(35) *Motivation*
- a. N presents an action in which R is the actor, unrealized with respect to the context of N
- b. comprehending S increases R's desire to perform action presented in N

(36) a. I have an idea,
b. why don't you guys add the pause and download on the playstore,
c. that way it would increase your downloads[13]

Belief and desire/intention are the first kinds of uptake that come to mind as they are directly associated with two kinds of speech acts—assertions and directives.[14] Speech acts can also pursue the goal of triggering emotions, which may or may not require the intermediate step of imparting a belief. For instance, the use of rude vocabulary can be insulting by itself, regardless of whether the addressee believes the proposition expressed by the utterance, or whether the utterance even expresses a proposition. Perlocutionary acts whose purpose is to elicit emotion are altogether less well studied, even less so are the speakers' discourse strategies when those acts fail. Support relations for emotional uptake will have to remain a question for future research. However, it is important to emphasise that they *must* exist because emotional perlocutionary acts exist, and therefore, a comprehensive taxonomy of support relations must include them.

SUPPORT FOR EXECUTION

Most speech acts do not require from the addressee the execution of any physical action, but those that do—directive speech acts, first and foremost—can also run into problems at the execution level. In (37), the uptake goes smoothly, speaker B readily agrees to perform the action requested in (37-a). However, in (37-c) speaker A provides information that will make it possible, or easier for B to perform that action. In RST, such support relations are called *Enablement*, (38).

(37) a. A: Please can you post these letters?
b. B: Sure.
c. A: The stamps are in the drawer.

(38) *Enablement*
a. N presents an action in which R is the actor, unrealized with respect to the context of N
b. R comprehending S increases R's potential ability to perform the action presented in N

CONCLUDING REMARKS

We have surveyed different types of support relations according to the location of the communicative problem that the supporting speech act is designed to solve. As it appears, some of them correspond closely to relations previously defined in the literature (*Evidence*, *Motivation*, *Enablement*), while others cross-cut known categories or are difficult to sort under any categories proposed in previous research.

13. A message posted on August 7, 2023 on the Google Play Help forum: https://support.google.com/googleplay/thread/229269275/i-have-an-idea-why-don-t-you-guys-add-the-pause-and-download-on-the-playstore-it-will-help-you-guys. Last accessed: November 18, 2025.
14. Belief can also constitute a perlocutionary goal of a commissive speech act. If commissives were only about regulating the speaker's own commitments, then what is the point of saying it? Just do what you are committed to do! (See Gärtner, 2021, for extensive discussion.) By making explicit promises we often try to influence the behaviour of others, for instance, to obtain something in return. That requires that the addressees trust our promises, i.e. believe the propositions they express.

We took a closer look at support relations at the levels of attention, identification, understanding, uptake, and execution, but did not cover signal transmission, signal processing, and sequel in the proposed ladder of levels of action in communication in figure 1. However, support relations should exist at those levels, too. For signal transmission, it is not even that difficult to find natural examples. We are all familiar with the situation in online meetings during the covid pandemic when a participant forgets to switch on their microphone before they start to speak. Once they notice their mistake, they usually switch on the microphone and repeat their utterance(s). The combined action of switching on the mic and the repetition stands in a support relation to the initial utterance. This is support for signal transmission.[15]

We are not going to try to find examples for the last remaining two levels. Once again, it follows from the definition of support relations and from the proposed levels of action ladder that such relations *must* exist. We will either encounter them sooner or later in naturally occurring discourse, or we will have to offer a principled explanation why these levels constitute an exception. Both tasks go beyond the scope of this paper.

4.2 Solving the problem

Now we turn to the second criterion for the classification of support relations that follows from the definition in (17)—the way in which the supporting speech act tries to solve the problem caused by the nucleus. It will not be possible to give nearly as broad a survey of possible types as the one we gave in the previous section, but we will present some basic considerations on which, we believe, the taxonomy of possible solutions should be based, and we will develop a fragment of the taxonomy for only one, albeit prominent, subclass of support relations, those whose purpose is to induce a belief.

GENERAL CONSIDERATIONS

What means are available for solving a problem will strongly depend on the problem at hand and the nature of the state that the agent desires to achieve. So far we have identified five types of (intermediate or ultimate) goals that directly concern the state of the addressee: (a) attention/engagement

15. Note that we have systematically defined the ways in which a speech act can fail in terms of missing a certain effect on the addressee. Matej Drobňák (p.c.) pointed out that speakers might also perceive their utterances as failing and might feel the need to produce a supporting speech act when they do not comply with a socially defined standard of propriety or correctness, especially in institutional contexts. For instance, a judge can only sentence the accused in a valid and legally binding way in a courtroom. Suppose the judge pronounces the sentence while standing in the doorway of the courtroom. They might step back inside the courtroom to pronounce the sentence again to ensure that it has legal consequences. Importantly, it does not matter how the audience takes it. They might have completely understood and agreed with the sentence. Nevertheless, there is some goal the speech act has obviously failed to achieve if the judge feels the need to take another try at it.

    Another example might be self-corrections for spelling and grammar in instant messaging communication with WhatsApp or other applications that allow for later editing of sent messages. Again, the addressee might have understood and adopted the intention of the utterance regardless of the incorrect spelling and grammar. Nevertheless, the speaker feels the need to repair the utterance because it does not comply with some ideal standard of correctness.

    We are not sure yet whether these cases could be reduced to failing some less immediate addressee-related goals (e.g. goals in the sequel) or whether they call for a separate category. This is a point where our taxonomy of support relations might still be incomplete.

in communication; (b) belief; (c) desire/intention; (d) emotion; and (e) action.[16] Eliciting these reactions or behaviours in another person requires conscious or intuitive domain knowledge about attention, belief, desire and intention, different emotions, and different types of action.

For instance, we seem to intuitively know that abruptly increasing the volume or making an abrupt movement is likely to draw the addressee's attention (Theeuwes, 1991; Yantis and Jonides, 1984, 1990). Therefore, if our initial attempt to draw the addressee's attention did not work, we might want to increase the speech volume, as in example (18) in the previous section, or accompany our speech with a more expressed gesture. In order to keep the addressee engaged for a longer period of time, more sophisticated strategies might be necessary.

Making people want and intend to perform an action is what persuasion is about. For instance, O'Keefe (2006) distinguishes several major strategies of persuasion depending on the addressee's initial state of mind. One strategy targets the addressee's positive regard of the action itself and its outcomes for the addressee. As in (36), getting more downloads is presumably something the addressee wants, so the speaker chooses that as an argument to persuade them to fulfil his or her request. Other strategies target (a) the influence of the action on the addressee's perception by others; (b) the addressee's perceived ability to perform the action when the desire to perform it is a given and the only thing keeping them from turning the desire into a specific intention is the belief that they cannot successfully perform it; and (c) turning a general intention to perform the action into an intention to do it right now. Accordingly, one could distinguish further subtypes of the *Motivation* relation depending on which strategy is used by the speaker.

If contrary to the speaker's intention, a speech act fails to insult, frighten, console, or amuse the addressee, one needs specific knowledge in the domain of interpersonal emotional (dis)regulation to be able to produce a supporting speech act that will help induce the desired emotion. Some of that knowledge might be intuitively available to the majority of neurotypical population, some may require talent and/or professional training in therapy, advertisement, propaganda, or creative storytelling (Thoits, 1996; Ochsner and Gross, 2005; Niven et al., 2009; Reddy, 2012; Zaki and Williams, 2013)

To ensure successful execution of an action, one requires domain knowledge about that kind of action. For instance, posting letters requires domain knowledge about posting letters and, in particular, the fact that letters need stamps. That is what makes (37-c) a good supporting speech act for (37-a).

In other words, there is no single unified taxonomy of solutions for all kinds of communicative problems. The criteria for more fine-grained classification of supporting speech acts according to the solution strategy they employ will depend on the nature of the domain knowledge required. Therefore, it will not be possible to give a comprehensive survey of such criteria and types in this paper. However, to illustrate the basic idea of the approach, we will give a brief overview of strategies for inducing belief and the range of support relations that can be distinguished according to the chosen strategy.

16. This list excludes states that lie outside the addressee's mind or immediate control, those pertaining to the levels of signal transmission, signal processing and the sequel. For reasons of space, those levels will not be discussed any further in this paper.

SUPPORT FOR BELIEF

Belief plays a role at different levels of a speech act's workflow, cf. figure 1. The hearer must form the correct belief about the intended linguistic form of the utterance (i.e., *identification*), she must form the correct belief about the communicative intent of the speech act (i.e., *understanding*), and so on. In some cases, most typically in assertions, forming a certain belief is the intended ultimate goal of a speech act. Since assertions are the most common speech act type in monologues and most previous research on coherence relations has concentrated on assertions, it is useful to look more closely at their typical goal—belief. The most important means of inducing belief is evidence, and several existing frameworks include some version of an *Evidence* relation in their taxonomies (e.g. Mann and Thompson, 1988; Asher and Lascarides, 2003). However, since previous research did not consider the full range of possibilities for giving support for a belief, many pragmatic relations conceptually adjacent but not identical to *Evidence* were overlooked. In this section we give several examples of such relations.

There are two main approaches to the study of beliefs. The normative approach grounded in philosophy attempts to define the correct, or *proper* ways of managing our belief states (e.g. Feldman, 2000), driven by the goal of holding only true beliefs as well as by other ethical considerations. The empirical approaches in psychology and cognitive science try to answer the question how humans really acquire, store, and change their beliefs (see Porot and Mandelbaum, 2021, for a recent overview). Both trains of research can give us clues as to what speakers might consider as appropriate or effective ways to induce a certain belief in the addressee, and those can be used as basis for classification of discourse patterns of support for belief.

Evidence-based vs. pragmatic: The philosophical debate has identified two main reasons why an agent may hold a particular belief: evidence-based reasons and pragmatic reasons. According to *evidentialism*, it is only permissible to believe a proposition if there are a sufficient number of evidence-based reasons for the truth of that proposition. As Clifford (1877) put it in his seminal paper on the ethics of belief, "it is always, everywhere, and for everyone, wrong to believe anything on insufficient evidence".

The philosophical counterpoint to this view is that in some cases one might believe a proposition because of the benefits of believing it. For example, one might choose to believe that climate change is not anthropogenic because this belief does not force one to think about the consequences of one's choices, which is a more comfortable attitude than thinking about one's own responsibility and possible past mistakes. Jordan (1996) argues that sometimes it is morally and rationally permissible to form beliefs on the basis of *pragmatic* reasons.

From a psychological point of view, Porot and Mandelbaum (2021) point out that belief updating is partly governed by a 'psychological immune system' that makes it easier to believe propositions that are consistent with one's self-image, and to reject beliefs that contradict it. So the psychological immune system helps the development of pragmatic beliefs.

We expect this dichotomy to be reflected in the strategies to induce belief employed in supporting speech acts. We have already seen an example of evidence-based belief (34) in section 4.1, repeated in (39). Here the speaker wants to induce belief by providing evidence:

(39) a. He must have been here recently.
b. There are his footprints.

On the other hand, supporting moves can offer pragmatic reasons for adopting a particular belief. For example, in (40-b) Sue urges Ann to adopt an evidently false belief because that will help them finish the task they are at before the deadline.

(40) a. Ann: I'm hungry.
b. Sue: No, you're not.
c. Sue: We need to finish this before seven.

Note that (40-b)–(40-c) is not an *Evidence* relation, because the fact that Sue and Ann have urgent work to do has no bearing on the truth of Ann being hungry. One might wonder if it could be an instance of RST *Motivation*, as (40-c) attempts to make the belief that Ann is not hungry more desirable, cf. the definition in (35). However, (40-b) does not describe a future action of the addressee, so *Motivation* does not fit either. That means that we need a new category to be able to describe this relation, and pragmatic support for belief does the job.

Providing vs. asking to seek: Feldman (2000) points out that our talk about beliefs mirrors our talk about actions and moral judgement. We say that people should do certain things in certain situations, and we blame them when they commit certain actions. On the other hand, we say that, given a certain amount of evidence, people should believe something, and we blame them if they don't. Accordingly, Hall and Johnson (1998) argue that agents not only have a moral duty to do the right thing, but also an epistemic duty to seek evidence for uncertain propositions—the *epistemic duty thesis*.

Applied to a cooperative communicative situation, this thesis implies that communication partners have a duty to seek evidence for the truth of the proposition in question. In the case where one of the interlocutors wants the other to believe something, the joint duty can be fulfilled by *providing* the missing evidence or by *asking* the hearer *to seek* for evidence.

The first case is exemplified in (39). The second case can occur, for instance, where the speaker refers the hearer to another, perhaps more trustworthy epistemic authority (Zagzebski, 2012), as in (2), repeated in (41). Here the speaker urges the hearer to seek evidence from another source.

(41) a. "Pa told Peter he wanted *him* to be chairman."
b. "Sure he did.
c. If you don't believe me, ask Peter."

We have already pointed out that this example does not fall under the standard definitions of *Evidence* relations, since (41-c) does not directly provide evidence. This example has some similarity to the SDRT *Source* relation (Hunter et al., 2006; Hunter, 2016), which relates the content of a message (41-b) to a statement about who said that. However, (41-c) does not directly say that Peter said (41-b), maybe he never did before being asked. Therefore the *Source* relation does not fit here either. Moreover, (43)—another instance of the 'asking to seek' category that we will discuss presently—is not even similar to *Source*, but is likewise lacking a fitting category in existing classifications.

Perception vs. testimony vs. inference: Epistemologists have argued that we form our beliefs based on sources of evidence that can be broadly categorised as perception (what we directly perceive), testimony (what other people tell us), or inference (what we conclude based on other evidence), as summarised by Lesage et al. (2015, see also Millar 2011 and Davies and Matheson 2012). Competent speakers are intuitively aware of this difference, as shown by distinct evidential-

ity marking available in many languages of the world (see e.g. Faller, 2002). So it stands to reason that speech acts of support for belief would make recourse to these three sources of evidence and could be categorised along this dimension.

The standard *Evidence* relation in (39) involves the speaker *telling* the addressee the evidence. Thus for the addressee, the evidence comes from testimony. Similarly, in (41) the speaker asks the addressee to seek further testimony of the proposition expressed by the nucleus.

In contrast, the speaker of (42) *shows* the evidence. By singing the iconic tenor piece from Verdi's opera *Rigoletto* the speaker demonstrates the range of his voice, which the addressee can directly perceive. Contrary to what the RST definition of *Evidence* (33) would require, there is no need to comprehend the supporting speech act. In fact, the addressee need not have any knowledge of Italian to be convinced by the "argument" in (42-b).

(42) a. I'm a tenor.
b. [sings:] *La donna è mobile qual piuma al vento...*

In the majority of cases where the target of support is a belief at one of the lower levels in the speech act workflow, cf. figure 1, direct perception will be the most natural way of acquiring evidence by the addressee. For instance, identification of the utterance is essentially the result of processing the auditory and visual input. If the addressee fails to form the belief that the speaker uttered *nine* as part of (20-a), the most straightforward way to solve the problem is to simply re-enact the same utterance, whose direct perception will help the addressee form the correct belief.

Showing often requires actions that go beyond speech. For instance, performing a double toe loop is a way to show evidence for the statement *I can perform a double toe loop*. If the proposition to be shown concerns facts external to the speaker, showing will typically require drawing the addressees attention to events and states that serve as evidence to those facts. Pointing gestures or presenting pictures, while not strictly *speech* acts, are communicative acts that have this function in multimodal communication.

Rather than providing, the speaker can also ask the addressee to seek direct perceptual evidence. In (43-f), from *Harry Potter and the Goblet of Fire* by J. K. Rowling, Draco Malfoy asks Professor Snape to look, i.e. to seek perceptual evidence for his assertion that Harry Potter hit his, Malfoy's, friend Goyle in (43-e). And in fact, Snape does look, as is made clear in (43-g). So the support relation between (43-f) and (43-e) belongs to the 'perception' and 'asking to seek' category.[17]

(43) a. Snape pointed a long yellow finger at Malfoy
b. and said, "Explain."
c. "Potter attacked me, sir —"
d. "We attacked each other at the same time!" Harry shouted.
e. "— and he hit Goyle —
f. look —"
g. Snape examined Goyle, whose face now resembled something that would have been at home in a book on poisonous fungi.

17. Directive speech acts like (43-f) are often accompanied by pointing gestures, but they are distinct from pointing gestures in our taxonomy. Pointing gestures belong to the 'providing' category as they facilitate access to perceptual evidence rather than directly tasking the addressee with an action.

In other words, the distinction between different sources of evidence is orthogonal to the distinction between providing and asking to seek. We should also be able to find instances of support relations that provide and ask to seek inferential evidence, thus amounting to six possible types based on these two features.

Once again, we would like to stress the fact that (40), (41), (42), and (43-e)–(43-f) do not fit any existing definitions of *Evidence*, or of *any* coherence relations, to the best of our knowledge. Previous studies overlooked such cases probably because, whether by design or not, they focused on common phenomena in the types of discourse they happened to use as their main empirical basis. In contrast, our top-down approach forces us to look at the entire range of possible ways in which a belief can be induced in the addressee. That makes it possible to predict the existence of coherence relations based on strategies that have not been considered before, and once we know what to look for, it turns out that we can also find instances of such relations in naturally occurring discourse, resulting in better empirical coverage.

To summarise, we have distinguished between eight possible types of support relations based on the location of the communicative problem within the cascade of goals, or levels of action, of a speech act (cf. figure 1). We have described five of those types in more detail: (a) support for attention, (b) support for identification, (c) support for understanding, (d) support for uptake, and (e) support for execution. We have made a further distinction between five types of goals according to *what* needs to be achieved, what kind of state or event, giving rise to further types of support relations: (a) support for attention, (b) support for belief, (c) support for desire/intention, (d) support for emotion, and (e) support for action. We have argued that further subdivisions of support relations according to the means the speaker uses to achieve their goal will depend on the type of goal. That is, the means relevant for support for belief and those relevant for support for, let's say, attention will generally not be the same. Finally, we have identified three features that characterise support for belief relations according to the means used to induce belief: (a) evidence-based vs. pragmatic; (b) providing vs. asking to seek; (c) the source of evidence—perception vs. testimony vs. inference.

Crucially, these are all and only support relations that (should) exist. Of course, one can define more features to further subdivide this space, but for the features discussed so far one could already say that if a relation is a support relation, it must belong to one of the proposed categories at each level, otherwise it is not a support relation. For example, the classical *Evidence* relation (39) is support for uptake, support for belief, evidence-based, providing, and based on testimony. The relation in (43-f)–(43-e) is support for uptake, evidence-based, asking to seek, and based on perception. The relation in (20-a)–(20-c) is support for identification, support for belief, providing, and based on perception. Finally, to give one more example, the relation in (37-a)–(37-c) is support for execution, support for action, and we have not defined further subtypes based on the solution method for support for action relations.

As should have become evident in the meantime, the features this classification is based on are only partly independent. Obviously, attention as a stage in the workflow of a speech act and attention as a type of mental state to be achieved inherently go together. Desire and emotion as types of goals only seem to be relevant for support for uptake relations. However, belief constitutes the goal at several stages—identification, understanding, and uptake. These asymmetries result from our concept of the levels of action in communication. The fact that there are (probably) no support relations that belong both to the support for understanding and support for desire category is a consequence of the nature of understanding—that the result state of understanding is a belief,

rather than a desire. In other words, the asymmetries are exceptions that prove the rule as they are a logical consequence of our top-down, concept-driven approach.

Contrary to tradition, we have not given names (e.g. *Evidence*, *Motivation*, *Enablement*) to all possible combinations of the proposed features, and we would indeed need too many to label all the boxes we have opened. If a short label is needed for ease of reference or for the purposes of corpus annotation, one way to go could be to extend the existing labels to broader categories. For instance, one could agree to refer to all evidence-based support for belief relations as *Evidence*. However, one should keep in mind that this is neither RST nor SDRT *Evidence* any more, and that it covers cases like (43-f)–(43-e) and (20-a)–(20-c) that were not thought of as *Evidence* relations previously. Users of this taxonomy are welcome to come up with relation names as needed.

5. Conclusion and outlook

We started out by expressing our general dissatisfaction with previous definitions of speech-act-level coherence relations which did not specify clearly what it is about speech acts that make some combinations of them coherent while not others. We are now in a position to give a partial answer to that question. In general, coherence between speech acts can be characterised in terms of relations between the goals of those speech acts. There is a limited number of ways in which goals can be coherently related. In this paper, we have investigated one such way—support—where one speech act fails or is anticipated to fail to achieve one of its goals, while the other speech act of the same speaker helps achieve that goal.

Based on theoretically motivated considerations about levels of action in communication and the nature of different types of goals, we have offered some further subdivisions of the broad category of support relations and have shown how some speech-act-level relations known from previous studies, e.g. self-repairs or RST's presentational relations, fit into the proposed categories. However, unlike RST, our top-down approach also made it possible to predict what other types of support relations *should* exist, even if they might not be common in written texts—the type of data that predominantly informed early relational approaches to discourse structure. Moreover, for many of the predicted types we were able to find naturally occurring examples, some of which would be really hard to categorise within previously proposed classifications. This is especially true for sequences that include non-assertive speech acts, which have been barely taken into account by approaches to discourse structure based on coherence relations. This is why the proposed taxonomy is particularly relevant for the study of coherence relations in dialogue and will likely turn out useful for the annotation of dialogue corpora.

Another important advantage of the top-down approach is that it defines the limits of what is theoretically possible. Our ultimate goal is to have a comprehensive taxonomy of coherence relations at the speech act level, so that if a sequence of speech acts does not fit one of the categories in that taxonomy, then we should be able to say that the sequence should therefore be incoherent. The taxonomy proposed in this paper is not comprehensive at that level. If a sequence does not fit our definition of a support relation, it only means that it is not a support relation, but it could still be coherent because it features a speech-act-level relation of a different type. However, it is crucial that the range of other possibilities is also theoretically motivated and very limited. Apart from support relations, we mentioned adjacency pairs and Franke's (1990) retractions as other ways in which speech acts can be related. Yet another possibility, briefly mentioned in section 3, cf. (13), is when two speech acts (usually of the same type, e.g. two assertions) are designed to achieve distinct sub-

goals of a bigger communicative goal of the speaker. This case has been studied extensively within the approach to discourse structure based on Questions under Discussion (Roberts, 1996; Büring, 2003). As per Roberts' (1996) original intention, Questions under Discussion are one specific way to operationalise the more general notion of communicative goal. In (13), one could say that the speaker's goal is to make a 'big' assertion in which they inform the hearer about who ate what, but they split it up into two 'small' assertions, one to inform the hearer about what Fred ate, and the other to inform them about what Mary ate.

Many of the classical semantic or propositional-level coherence relations, e.g. *Parallel* in (13), *Contrast*, and *Narration*, would fall into this category. But then, we believe that that should be all. All coherent sequences of speech acts should be reducible to these few possibilities and combinations thereof, and if they cannot be characterised in these terms, they should be incoherent. It remains a task for future research to review other types of speech-act-level relations along the same methodological guidelines.

From the above it should be clear that our taxonomy of support relations, or speech-act-level relations more generally, is not intended to replace existing taxonomies of semantic relations at propositional level. Propositional-level taxonomies would still be relevant particularly for sequences of assertions. One could attempt to reformulate them in speech-act-theoretic terms, but it would not necessarily bring new insight. The notions of causality, contiguity, and resemblance have turned out useful to capture generalisations about relations between assertions just by looking at the relations between the communicated propositions. Our taxonomy becomes most relevant where propositional level taxonomies hit their limits, that is, especially for heterogeneous speech act sequences, interactive sequences of more than one speaker, and sequences that do not go according to plan.

Our approach still needs to be tested in an empirical setting. It is true that applying the proposed relation definitions to corpus annotation would require a lot of reasoning with the speakers' cognitive states, which are accessible to the analyst only to a limited extent. However, in cases where coherence at the speech act level is all we have, i.e. where the more familiar relations at propositional level do not seem to apply, hearers are bound to form at least some hypotheses about the speakers' communicative intentions to be able to perceive the sequence as coherent. If hearers can do it, then analysts should be able to do it too. How reliable the recognition of support relations is, both for hearers and for corpus annotators, is another open question.

Finally, it remains to be seen to what extent reasoning about the addressee's understanding and acceptance of a speech act plays a role in the production of support relations in real time or in the formation of rhetorical patterns in the course of the historical development of specific discourse genres. One of the fundamental ideas underlying our proposal is that even monologue—an uninterrupted sequence of utterances of the same speaker or writer—is in its essence an interactive practice. Our notion of support relations substantiates this idea and is a step towards a comprehensive theory of coherence in both dialogue and monologue.

Acknowledgments

The research reported in this paper was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 281511265 – SFB 1252 "Prominence in Language" in the project C06 at the University of Cologne, Department of German Language und Literature I. We would also like to thank the audiences of the workshop on *At-Issueness, Scope, and Coherence*

(ASC) 2019 in Cologne, the SPAGAD 2 workshop on *Speech Acts in Grammar and Discourse* 2020 at ZAS, Berlin, and the 24th *Szklarska Poręba Workshop on the Roots of Pragmasemantics*, especially Matej Drobňák, for inspiring discussions of previous versions of this paper.

References

Jens Allwood, Joakim Nivre, and Elisabeth Ahlsén. On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9(1):1–26, 1992. doi: 10.1093/jos/9.1.1. URL https://doi.org/10.1093/jos/9.1.1.

Nicholas Asher and Alex Lascarides. *Logics of Conversation*. Cambridge University Press, 2003.

John Langshaw Austin. *How to do things with words*. William James Lectures. Oxford University Press, 1962.

Diane Blakemore. The relevance of reformulations. *Language and Literature*, 2:101–120, 1993. doi: 10.1177/096394709300200202. URL https://doi.org/10.1177/096394709300200202.

Daniel Büring. On d-trees, beans, and b-accents. *Linguistics and Philosophy*, 26(5):511–545, 2003. doi: 10.1023/a:1025887707652. URL https://doi.org/10.1023/a:1025887707652.

Herbert H. Clark. Managing problems in speaking. *Speech communication*, 15(3-4):243–250, 1994. doi: https://doi.org/10.1016/0167-6393(94)90075-2. URL https://www.sciencedirect.com/science/article/pii/0167639394900752.

Herbert H. Clark. *Using Language*. Cambridge University Press, 1996.

William K. Clifford. The ethics of belief. *Contemporary Review*, 1877.

Jim Davies and David Matheson. The cognitive importance of testimony. *Principia: An international journal of epistemology*, 16(2):297–318, 2012. doi: 10.5007/1808-1711.2012v16n2p297. URL https://doi.org/10.5007/1808-1711.2012v16n2p297.

Martina Faller. Remarks on evidential hierarchies. In David I Beaver, Luis D Casillas Martínez, Brady Zack Clark, and Stefan Kaufmann, editors, *The construction of meaning*, pages 89–111. CSLI Publications, 2002.

Richard Feldman. The ethics of belief. *Philosophy and Phenomenological Research*, 60(3):667–695, 2000. doi: 10.2307/2653823. URL https://doi.org/10.2307/2653823.

Anita Fetzer. The structuring of discourse. In Marina Sbisà and Ken Turner, editors, *Pragmatics of speech actions*, pages 685–711. De Gruyter Mouton, Berlin, Boston, 2013. doi: 10.1515/9783110214383.685. URL https://doi.org/10.1515/9783110214383.685.

Wilhelm Franke. *Elementare Dialogstrukturen: Darstellung, Analyse, Diskussion*. Germanistische Linguistik. Niemeyer (De Gruyter), 1990. ISBN 978-3-484-31101-5. doi: 10.1515/9783111631905. URL https://doi.org/10.1515/9783111631905.

Gerd Fritz and Franz Hundsnurscher. Sprechaktsequenzen : Überlegungen zur vorwurf/rechtfertigungs-interaktion. *Der Deutschunterricht*, 27(2):81 – 103, 2009.

Hans-Martin Gärtner. On the utility of commissive signals and the "promissive gap". 2021. URL http://real.mtak.hu/id/eprint/137527.

Jonathan Ginzburg. *The Interactive Stance: Meaning for Conversation*. Oxford University Press, 2012. ISBN 9780198722991. doi: 10.1093/acprof:oso/9780199697922.001.0001. URL https://doi.org/10.1093/acprof:oso/9780199697922.001.0001.

Jonathan Ginzburg, Raquel Fernández, and David Schlangen. Disfluencies as intra-utterance dialogue moves. *Semantics and Pragmatics*, 7, 2014. doi: 10.3765/sp.7.9. URL https://doi.org/10.3765/sp.7.9.

Jonathan Ginzburg, Zulipiye Yusupujiang, Chuyuan Li, Kexin Ren, Aleksandra Kucharska, and Pawel Lupkowski. Characterizing the response space of questions: data and theory. *Dialogue & Discourse*, 13(2):79–132, 2022. doi: 10.5210/dad.2022.203. URL https://doi.org/10.5210/dad.2022.203.

H. Paul Grice. Meaning. *Philosophical Review*, 66(3):377–388, 1957. doi: 10.2307/2182440. URL https://doi.org/10.2307/2182440.

Barbara J. Grosz and Candace L. Sidner. Attention, intentions and the structure of discourse. *Computational Linguistics*, 12(3):175–204, 1986.

Richard J. Hall and Charles R. Johnson. The epistemic duty to seek more evidence. *American Philosophical Quarterly*, 35:129–139, 1998. URL https://www.jstor.org/stable/20009926.

Götz Hindelang. *Einführung in die Sprechakttheorie: Sprechakte, Äusserungsformen, Sprechaktsequenzen*. De Gruyter, 2010. ISBN 9783110231472.

Laurence R. Horn. *A Natural History of Negation*. University of Chicago Press, 1989.

Joris Hulstijn and Nicolas Maudet. Uptake and joint action. *Cognitive Systems Research*, 7(2):175–191, 2006. ISSN 1389-0417. URL https://doi.org/10.1016/j.cogsys.2005.11.002. Cognition, Joint Action and Collective Intentionality.

Julie Hunter. Reports in discourse. *Dialogue & Discourse*, 7(4):1–35, 2016. doi: 10.5087/dad.2016.401. URL https://doi.org/10.5087/dad.2016.401.

Julie Hunter, Nicholas Asher, Brian Reese, and Pascal Denis. Evidentiality and intensionality: Two uses of reportative constructions in discourse. In *Proceedings of the Workshop on Constraints in Discourse 2*, Maynooth, Ireland, 2006.

Katja Jasinskaja and Elena Karagjosova. Rhetorical relations. In Daniel Gutzmann, Lisa Matthewson, Cécile Meier, Hotze Rullmann, and Thomas Ede Zimmermann, editors, *The Companion to Semantics*. Wiley, Oxford, 2020. doi: 10.1002/9781118788516.sem061. URL https://doi.org/10.1002/9781118788516.sem061.

Jeff Jordan. Pragmatic arguments and belief. *American Philosophical Quarterly*, 33(4):409–420, 1996.

Andrew Kehler. *Coherence, Reference, and the Theory of Grammar*. CSLI Publications, 2002.

Alistair Knott and Robert Dale. Using linguistic phenomena to motivate a set of coherence relations. *Discourse Processes*, 18:35–62, 1994. doi: 10.1080/01638539409544883. URL https://doi.org/10.1080/01638539409544883.

Alex Lascarides and Nicholas Asher. Agreement, disputes and commitments in dialogue. *Journal of Semantics*, 26(2):109–158, 2009. URL https://doi.org/10.1093/jos/ffn013.

Claire Lesage, Nalini Ramlakhan, Ida Toivonen, and Chris Wildman. The reliability of testimony and perception: Connecting epistemology and linguistic evidentiality. In *CogSci*, volume 37, pages 1302–1307, 2015. URL https://www.researchgate.net/publication/275582507_The_reliability_of_testimony_and_perception_connecting_epistemology_and_linguistic_evidentiality.

Willem J.M. Levelt. Monitoring and self-repair in speech. *Cognition*, 14(1):41–104, 1983. ISSN 0010-0277. doi: 10.1016/0010-0277(83)90026-4. URL https://doi.org/10.1016/0010-0277(83)90026-4.

William C. Mann and Sandra Thompson. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8(3):243–281, 1988. doi: 10.1515/text.1.1988.8.3.243. URL https://doi.org/10.1515/text.1.1988.8.3.243.

William C. Mann, Christian M. I. M. Matthiessen, and Sandra A. Thompson. Rhetorical structure theory and text analysis. Technical report, University of Southern California, Information Sciences Institute, 1989. URL https://doi.org/10.1075/pbns.16.04man.

Alan Millar. How visual perception yields reasons for belief. *Philosophical Issues*, 21:332–351, 2011. doi: 10.1111/j.1533-6077.2011.00207.x. URL https://doi.org/10.1111/j.1533-6077.2011.00207.x.

Karen Niven, Peter Totterdell, and David Holman. A classification of controlled interpersonal affect regulation strategies. *Emotion*, 9(4):498–509, 2009. doi: 10.1037/a0015962. URL https://doi.org/10.1037/a0015962.

Kevin N. Ochsner and James J. Gross. The cognitive control of emotion. *Trends in Cognitive Sciences*, 9(5):242–249, 2005. ISSN 1364-6613. doi: 10.1016/j.tics.2005.03.010. URL https://doi.org/10.1016/j.tics.2005.03.010.

Daniel J. O'Keefe. Persuasion. In *The Handbook of Communication Skills. 3d edition*, pages 323–341. Routledge, 2006. ISBN 9780761925392.

Nicolas Porot and Eric Mandelbaum. The science of belief: A progress report. *Wiley Interdisciplinary Reviews: Cognitive Science*, 12, 2021. doi: 10.1002/wcs.1539.

Vasudevi Reddy. Moving others matters. In *Moving ourselves, Moving others*, Consciousness and emotion book series, pages 139–162. Philadelphia, 2012. ISBN 9789027241566.

Gisela Redeker. Ideational and pragmatic markers of discourse structure. *Journal of pragmatics*, 14(3):367–381, 1990. doi: 10.1016/0378-2166(90)90095-U. URL https://doi.org/10.1016/0378-2166(90)90095-U.

Craige Roberts. Information structure in discourse: Towards an integrated formal theory of pragmatics. *OSU Working Papers in Linguistics*, 49:91–136, 1996.

Kepa Joseba Rodríguez and David Schlangen. Form, intonation and function of clarification requests in German task oriented spoken dialogues. In *Proceedings of Catalog '04 (The 8th Workshop on the Semantics and Pragmatics of Dialogue, SemDial04)*, pages 101–108, 2004.

Harvey Sacks and Emanuel A. Schegloff. Opening up closings. *Semiotica*, 8(4):289–327, 1973. doi: 10.1515/semi.1973.8.4.289. URL https://doi.org/10.1515/semi.1973.8.4.289.

Ted Sanders. Semantic and pragmatic sources of coherence: On the categorization of coherence relations in context. *Discourse Processes*, 24:119–147, 1997. doi: 10.1080/01638539709545009. URL https://doi.org/10.1080/01638539709545009.

Ted Sanders, Wilbert Spooren, and Leo Noordman. Toward a taxonomy of coherence relations. *Discourse Processes*, 15(1):1–35, 1992. doi: 10.1080/01638539209544800. URL https://doi.org/10.1080/01638539209544800.

Ted Sanders, José Sanders, and Eve Sweetser. Causality, cognition and communication: A mental space analysis of subjectivity in causal connectives. In Ted Sanders and Eve Sweetser, editors, *Causal categories in discourse and cognition*, pages 19–59. De Gruyter Mouton, Berlin, New York, 2009. ISBN 9783110224429. doi: doi:10.1515/9783110224429.19. URL https://doi.org/10.1515/9783110224429.19.

Marina Sbisà. Cognition and narrativity in speech act sequences. In Anita Fetzer and C. Meierkord, editors, *Rethinking Sequentiality*, pages 71–97. John Benjamins Publishing Company, Amsterdam, 01 2002. doi: 10.1075/pbns.103.04sbi. URL https://doi.org/10.1075/pbns.103.04sbi.

Julian J. Schlöder and Raquel Fernández. Clarifying intentions in dialogue: A corpus study. In Matthew Purver, Mehrnoosh Sadrzadeh, and Matthew Stone, editors, *Proceedings of the 11th International Conference on Computational Semantics*, pages 46–51, London, UK, 2015. Association for Computational Linguistics. URL https://aclanthology.org/W15-0106.

John R. Searle. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, 1969. doi: 10.1017/CBO9781139173438. URL https://doi.org/10.1017/CBO9781139173438.

John R. Searle. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge paperback library. Cambridge University Press, 1983. doi: 10.1017/cbo9781139173452. URL https://doi.org/10.1017/cbo9781139173452.

Eve Sweetser. *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge University Press, 1990. doi: 10.1017/CBO9780511620904. URL https://doi.org/10.1017/CBO9780511620904.

Maite Taboada. Rhetorical relations in dialogue. In *Discourse across languages and cultures*, Studies in Language Companion Series, pages 75–97. John Benjamins Publishing Company, 2004. ISBN 9789027295262. URL https://www.sfu.ca/~mtaboada/docs/publications/Taboada_RST_2004.pdf.

Jan Theeuwes. Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception & Psychophysics*, 49(1):83–90, 1991. doi: 10.3758/BF03211619. URL https://doi.org/10.3758/BF03211619.

Peggy A. Thoits. Managing the emotions of others. *Symbolic Interaction*, 19(2):85–109, 1996. URL https://doi.org/10.1525/si.1996.19.2.85.

David R. Traum. *A Computational Theory of Grounding in Natural Language Conversation*. PhD thesis, University of Rochester, 1994.

Rob H. J. van der Lubbe and Albert Postma. Interruption from irrelevant auditory and visual onsets even when attention is in a focused state. *Experimental brain research*, 164(4):464–471, 2005. doi: 10.1007/s00221-005-2267-0. URL https://pubmed.ncbi.nlm.nih.gov/15785951/.

Bonnie Webber, Rashmi Prasad, Alan Lee, and Aravind Joshi. The Penn discourse treebank 3.0 annotation manual, 2019. URL https://catalog.ldc.upenn.edu/docs/LDC2019T05/PDTB3-Annotation-Manual.pdf.

Steven Yantis and John Jonides. Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human perception and performance*, 10(5):601–621, 1984. doi: 10.1037/0096-1523.10.5.601. URL https://doi.org/10.1037/0096-1523.10.5.601.

Steven Yantis and John Jonides. Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human perception and performance*, 16 (1):121–134, 1990. doi: 10.1037/0096-1523.16.1.121. URL https://doi.org/10.1037/0096-1523.16.1.121.

Linda Trinkaus Zagzebski. *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief*. Oxford University Press, 2012. doi: 10.1093/acprof:oso/9780199936472.001.0001. URL https://doi.org/10.1093/acprof:oso/9780199936472.001.0001.

Jamil Zaki and W. Craig Williams. Interpersonal emotion regulation. *Emotion*, 13(5):803–810, 2013. doi: 10.1037/a0033839. URL https://doi.org/10.1037/a0033839.

Frank Zickenheiner. *Goals in Discourse: From Actions to Rhetorical Relations*. PhD thesis, University of Cologne, Germany, 2020. URL http://kups.ub.uni-koeln.de/id/eprint/64204.