



HAL
open science

Exploring digital conversation corpora with process mining

Dario Compagno, Elena V. Epure, Rebecca Deneckere, Camille Salinesi

► **To cite this version:**

Dario Compagno, Elena V. Epure, Rebecca Deneckere, Camille Salinesi. Exploring digital conversation corpora with process mining. *Corpus Pragmatics*, 2018, 2, pp.193-215. 10.1007/s41701-018-0030-6 . hal-01722928

HAL Id: hal-01722928

<https://hal.univ-lorraine.fr/hal-01722928v1>

Submitted on 9 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Exploring Digital Conversation Corpora with Process Mining

Dario Compagno¹ · Elena V. Epure² · Rébecca Deneckere-Lebas² · Camille Salinesi²

Abstract Understanding how sentences constitute conversations is still a matter of disagreement. The automatic analysis of corpora can bring new perspectives on this issue. We show that process mining, a computer science approach for analysing human behaviour, enables the inductive identification of within-turn regularities in digital conversations. We use speech acts as a way to discretise conversations and enable the analysis. To this end, we develop a fine-grained, corpus-independent classification of speech acts that we describe and validate. We perform a manual sentence-by-sentence multi-label contextual annotation of a corpus of digital conversations and we explore it with process mining. Our results suggest that speech act theory could be extended to include typicality conditions for sequences of utterances.

Keywords Corpus pragmatics · Speech acts · Conversation analysis · Process mining · Multi-label annotation · Online conversations

✉ Dario Compagno
dario.compagno@univ-lorraine.fr

Elena V. Epure
elena.epure@etu.univ-paris1.fr

Rébecca Deneckere-Lebas
rebecca.deneckere@univ-paris1.fr

Camille Salinesi
camille.salinesi@univ-paris1.fr

¹ CREM – Centre de Recherche sur les Médiations, UFR SHS-Metz, Université de Lorraine (France), BP 60228, 57045 Cedex 01, Île Du Saulcy, Metz, France

² CRI – Centre de Recherche en Informatique, Université Panthéon-Sorbonne (France), 90 rue de Tolbiac, 75013 Paris, France

Introduction

Understanding how sentences constitute conversations is still a matter of disagreement. While some scholars claim that conversations are only driven by non-linguistic factors, such as social habits and practices, others sustain the existence of an autonomous layer of regularities. Famously, on the one hand, John Searle (1992) argued against the possibility to define such a conceptual layer to explain conversations, on the other, Emanuel Schegloff (1992) defended the empirical existence of heuristic conversational rules. The development of models for conversation corpora bears therefore important consequences for theoretical and empirical pragmatics.

Advances in computer science have brought about automatic techniques for text processing and behavior inference that can allow for innovations in this field. Process mining (van der Aalst 2011) proposes a suite of techniques for discovering and modeling human behavior from digital traces, generated during the interaction with information systems. Conversations are a form of human behavior achieved through words (together with nonverbal means of communication) and conversation corpora can be seen as traces of human behaviour. We use speech acts to discretise these traces and allow for their analysis with process mining.

Specifically, we undertake the inductive identification of recurring patterns in conversations. For this end we create a fine-grained, corpus-independent classification of speech acts and validate it in an empirical annotation of a *Reddit*¹ corpus by 10 different groups of human annotators. Results confirm reproducibility and interpretation consistency for most classes. Process mining on the annotated *Reddit* corpus permits to formulate hypotheses on the typical development of conversations. In particular, we observe differences in how families of speech acts appear in turns, and identify a conversational preference in the occurring order of two specific speech acts.

Our results show that in between felicity conditions for isolated speech acts, at a smaller level, and discourse practices, at the larger one, we observe typicality conditions for the appearance of speech act sequences. Also, the method we propose to apply to linguistic corpora (process mining), could be integrated within the tools of conversation analysis. Finally, the classification of speech acts we elaborate for the aims of process mining can be used to annotate larger corpora.

Using speech acts for representing conversational units is motivated and follows a discussion of related works that approach corpora classification (“[Introduction](#)” and “[Theoretical Background and Related Work](#)” sections). Then a structured classification of speech acts suited for the process mining of conversation corpora is described, highlighting its groundings on speech act theory and conversation analysis (“[A Structural Classification of Speech Acts](#)” section). The validation of this classification is discussed, including the empirical results of a digital corpus annotation experiment (“[Assessing the Consistency of the Classification](#)” section).

¹ *Reddit* is a website organizing threads of discussion in sections called ‘subreddits’ according to thematic or other features. The subreddit we extracted our corpus from is: <https://www.reddit.com/r/autoimmunity/> (see below, “[Assessing the Consistency of the Classification](#)” section).

Finally, we present exploratory results obtained by applying process mining for the identification of regularities in our corpus (“[Exploring Digital Conversations with PM](#)” section).

Theoretical Background and Related Work

Do Conversations Have a General Structure?

Is it possible to find recurring patterns in conversations, allowing for a formal understanding of verbal interaction? Do conversations have a general “grammar”, guiding the speakers’ choices and structuring their behaviour into a meaningful rule-based exchange? Scholars disagree about the nature of conversations, whether they should be thought of as an epiphenomenal dimension of verbal discourse, actually determined by rules at other levels (sentences and speech acts at the smaller scale, genres and social practices at the larger one), or if, on the contrary, there is an autonomous layer of conversation rules with effective constraints that cannot be reduced to those governing smaller or larger units (Searle et al. 1992). Searle (1992) famously claimed against the existence of conversation rules, and successive debate, in particular with sociologist and Conversation analysis’ (CA) co-founder Schegloff defending rule-based conversational exchanges (1992), made the point of controversy emerge.

For Searle “conversation does not have an intrinsic structure about which a relevant theory can be formulated; and conversations are not subject to (constitutive) rules” (Parret and Verschueren 1992: 1). Searle’s point is that the kinds of regularities that can be found in conversations—for example those studied by CA (Hutchby and Wooffitt 2008; Sidnell and Stivers 2012)—cannot be considered rules, in the way we say that an English sentence is structured according to some syntactic rules, or a promise is defined by a precise set of codified engagements. For example, there are some rules allowing the sentence *Please, take a seat* to realize a felicitous speech act, by referring to a potential seat near the speaker and the hearer. Instead, there is no rule necessarily linking such sentence to the adjacent answer *Thank you*.

Searle states that politeness, used as explanatory principle by CA to account for real speech behaviour (such as the pair of sentences in our example), is not sufficiently predictive, and an adjacency pair is not a well-formed conversational unit. Moreover, for Searle, the eventuality that the interlocutor fails to answer *Thank you* does not affect the meaning of the first speech act (and the meaning of *Thank you*, too, is independent from whatever preceded it). The fact that CA can derive consequences from the compliance or non-compliance to politeness is relevant only as a way to study speech practices—politeness and not conversation, we could say. Searle gives a general explanation to this: if language production at a smaller scale (e.g. speech acts) is driven by specific and well-defined goals, conversations themselves do not have a single global purpose. Threads of discourse can be identified within one verbal exchange, but it would be a mistake to think that the entire exchange has one main aim and thus follows some regularities determined by the mutual foundational choice of this aim.

On the other hand, as found by CA, it is possible to empirically identify the exact moments in which turn-shifting is perceived as polite, impolite, aggressive, submissive and so on. In this way, rules of verbal exchange—or at least *norms*, intermediate between the linguistic system and individual execution (see Rastier 2011)—can be identified. How can such norms have no causal effect on behaviour, as stated by Searle, even if speakers may not be consciously aware of their existence? Jucker (1992) argues that if conversations are a particular type of unstructured discourse, in the sense that turns cannot be predicted from previous ones, they can still be modelled as *processes* with local organisation principles. This is to say that some conversational norms do exist, guiding how we interpret (at least short) exchanges, but they are not easy to identify with conceptual analysis. Moreover, speech acts scholars as Marina Sbisà (1992) argue that the production and interpretation of speech acts may not be completely independent from larger verbal exchange.² For example, *Thank you* may be used ironically or aggressively, in some specific contexts, such as when someone spilled coffee on someone else.

Our research aims to empirically identify some meaningful patterns in conversation corpora, starting from speech acts as minimal units. To find regularities, real world conversations have to be studied. Searle's scepticism over conversational structures does not make speech act theory (Austin 1962; Searle 1969) less useful for their search. In the end, empirical research may turn out to be useful for the development of speech act theory itself (Sbisà 2002a, b), by defining some *typicality conditions* (Violi 2001) for speech act sequences, existing next to formal felicity conditions for individual utterances, and possibly interacting with them. Such results may permit to better understand how speech acts are recognized within interactional contexts and build up complex conversational aims (Geis 1995; Kissine 2013).

The Process Mining of Conversation Corpora

Process mining (PM) is a bottom-up approach to analyse structured data, in order to extract knowledge as temporal models (van der Aalst 2011). Compared to other data mining techniques using statistical methods, PM wants to be a bridge to the production of explicit, easy to interpret, high-level models for human practices.

Data ('event logs') are traces of human practices, usually captured from information systems. These traces are seen as successions of *events* generated during the enactment of some *activities*. *Process models* are representations of activities and their relations. The idea grounding our research is that PM can be used to model conversations. Conversation is human behaviour, and corpora of written or recorded conversations can be seen as traces of this behaviour, from which regular patterns can be extracted inductively with PM.

² "It seems therefore that, when we want to assign a definite illocutionary force to a certain speech act, we should take the hearer's response into account. But if this is so, we have to admit that our consideration of isolated speech acts leads only to provisional results, until we know not only in which context a speech act has been issued, but also which response it has received. Thus the consideration of sequential phenomena seems to be relevant to the consideration of the very illocutionary force of an utterance." (Sbisà 1992: 101–102).

The most important choice for applying PM to verbal behaviour is to determine what is taken as unit of conversation (as event). PM has to start with a corpus of typed units (event log), in order to study their occurrence in regular patterns. As events have to be categorized before analysis, this pre-interpretation is the ground for any further result. For this reason, *speech acts* are a good choice as events for applying PM to the ends of linguistic theory. Speech acts are generic (corpus-independent) and not related to any particular topic of conversation, discourse genre or other specific situation of verbal production. They account for the “overt aim” of the speaker’s act (Strawson 1964; Kissine 2013). The overt speaker’s aim is the pragmatic level of description at which meaning is attributed to linguistic production in real-world contexts, by referring to the mutual understanding of communicative intentions. To use PM for conversation analysis, we map *instances of speech acts* onto events and *successions of instances of speech acts* onto traces (that is, conversation turns segmented into sentences, keeping the original order). Activities in process models are *types of speech acts*; process models therefore account for regular patterns of speech acts in conversations.

Process models are rich representations capturing several kinds of *relations* among activities: sequences (activity *a* follows activity *b*), parallelisms (activity *a* or part of it is realized concurrently with activity *b*), choices (either activity *a* or activity *b* follows activity *c*) and loops (activity *a* appears several consecutive times). These kinds of relations can be interpreted as recurring patterns of speech acts. PM aims to find relations *inductively and automatically* starting from their empirical occurrences in traces. For example, with PM, it could be possible to identify adjacency pairs (Schegloff and Sacks 1973), such as questions and answers, for several kinds of conversational units.

Winograd and Flores (1986) approached conversations as processes and speech acts as activities. Figure 1 shows a conversational workflow, analogous to a finite-state automaton, with speech acts as transitions and hidden states of conversation between them. The workflow moves from a start state with a request formulated by the speaker A, and through conversation with the speaker B (actually a negotiation of the request’s conditions), it can reach different end states, representing satisfactory or aborted conclusions of the negotiation. This workflow, however, is just a hypothetical model and not the result of empirical research.

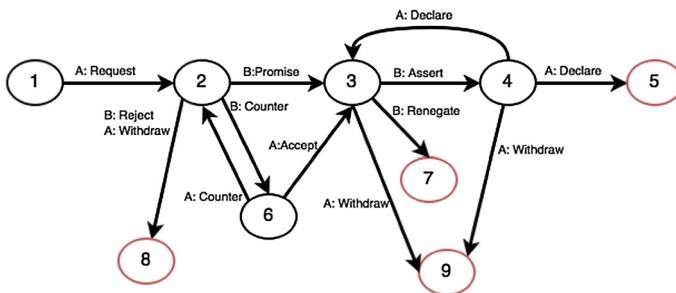


Fig. 1 A conversation process. Reproduced with permission from (Winograd and Flores 1986)

Wang et al. (2014) applied PM to threads of questions and answers in an online web forum. They managed to differentiate useful threads, leading to accepted answers, from inconclusive ones and also found correlations with the number of comments and users involved in their development. However, while Wang et al. discussed the idea of using speech acts to characterize events, they found them to be too complex for their ends, and preferred dialogue act tags instead (Kim et al. 2010). Such tags permit to differentiate answers and questions, but do not account for all illocutionary forces and are not suited for more detailed linguistic research on the structure of conversations.

Speech Acts Classifications for the Analysis of Conversation

Several works in natural language processing make use of speech acts to model various specialized verbal exchanges, including interactions via email (Carvalho and Cohen 2005; De Felice et al. 2013), *Twitter* messages (Zhang et al. 2011; Epure et al. 2017), web forums (Bhatia et al. 2012; Wang et al. 2014; Arguello and Shaffer 2015). These studies adopt one of two general strategies in approaching speech act theory: either they start from Searle’s five types of speech acts (Searle 1976) and extend them with custom classes as needed to fit their corpus (for example, in Carvalho and Cohen 2005; Zhang et al. 2011), or they constitute ad hoc classes of conversational units, dependent on some specific corpora, that can only to a lesser extent be considered speech acts (Wang et al. 2014; Bhatia et al. 2012).

Both strategies have potential weaknesses. First, they do not necessarily perform linguistic research to create the customized or ad hoc classes; therefore they cannot guarantee their consistency. Second, they are (sometimes heavily) corpus dependent, hence not generalizable.

Instead, we aim at the constitution of a corpus-independent, fine-grained classification, oriented towards the PM of conversation corpora. Such classification should be grounded on the development of speech act theory in corpus pragmatics (see Garcia McAllister 2015), validated empirically for consistency and reproducibility, and its heuristic power for the PM of a first corpus should be tested. This classification will be the basis for training algorithms for the automatic annotation of larger corpora, allowing for extensive application of PM.

The starting point of our classification is Searle’s original typology of speech acts (1976). Searle differentiated 5 classes of speech acts (assertives, directives, commissives, expressives and declarations) based on the action performed by the speaker with the utterance. This typology is a milestone in pragmatics, with several advantages: it is based on recognized philosophical research, while being simple to understand and interpret; it is well-known and widespread in linguistics as well as in computer science; it permits to differentiate discrete classes of speech acts and thus it can be used to annotate corpora. Searle’s typology presents one main inconvenience making it unsuitable for direct application to the PM of conversations: only five types of speech acts are too few for a detailed distinction of conversational events. We believe that, with more classes, more detailed relations can be discovered (see below, “[Exploring Digital Conversations with PM](#)” section).

Daniel Vanderveken (1990) studied how illocutionary forces (constituting speech acts) are lexicalized in contemporary English vocabulary. This resulted in a detailed description of a list of 300 verbs. Compared to Searle's, Vanderveken's work has the advantage of detail; but a such a list seems too detailed to be applicable, and 300 illocutionary forces too many for modelling conversations. Still, Vanderveken's verbs are organized in five trees where the roots are the 'primitive speech acts' (corresponding to Searle's typology) and the other verbs appear as nodes and leaves. This organisation is the reason why we selected Vanderveken's work as a basis to develop our own classification. Vanderveken chose some verbs (the non-leaves nodes) as representatives for more ample classes of verbs, including the nodes themselves and all of their children. For example, commissives are all instances of their main 'primitive', *commit*, then a path can be followed to more specific commissives such as *accept* and *renounce*. Vanderveken identifies 21 classes as immediate children of the roots (of the primitive speech acts). Our solution takes these 21 classes as a corpus-independent classification that is more detailed than Searle's without being too detailed, a middle level suited for the PM of conversations.

However, Vanderveken seems not to follow any explicit structuring principle in choosing the 21 direct children of the primitive speech acts. This makes the classification somehow challenging to interpret: for each type of speech act, the 21 classes have to be memorized individually, each one a special case in which the illocutionary force happens to have been lexicalized by the English language. This is why our solution adds a further layer of organisation to the 21 most general classes of verbs identified by Vanderveken. Conceptual analysis of these classes revealed that it is possible to identify a few very general *oppositional traits* to create a more understandable and applicable classification of illocutionary forces. Our intervention restructured Vanderveken's classes and ended up identifying 18 classes of speech acts.

Our classification is similar to that proposed by Garcia McAllister in her Ph.D. thesis (2004: 73–74). Like us, Garcia's classes propose a further organisation of Searle's families. She relied on empirical corpus exploration to identify some conceptual thresholds permitting a finer-grained consistent annotation. While we fully share the need for a structured finer-grained classification, we deviate from Garcia's taxonomy in some specific aspects. Mainly, we aimed at a corpus-independent classification, and Garcia's one is instead oriented towards a specialized domain (academic contexts). For this reason we recurred to Vanderveken's classes, as an—at least conceptually—corpus-free taxonomy. Also, we took into account the application of PM from the start, by identifying structuring criteria oriented towards its specific needs and aims.

The final classification, presented below, passed through several revisions and was evaluated in terms of inter-rater agreement. This resulted in criteria giving a main structure to the classification and leading to the merging or splitting of some of Vanderveken's classes.³

³ Merged classes: hypothesize and guess; congratulate and rejoice; condole and complain; subscribe, undertake, pledge, threaten and engage; warn and advise; propose and direct. Splitting classes: request and require; assert and sustain.

A Structural Classification of Speech Acts

To begin with, Searle's five types of speech acts can already be divided into two groups (see Fig. 2). The first includes assertives and expressives, the second commissives and directives, while declaratives are considered as a special case. The main difference is that speech acts in the first group are claims about the world, while those in the second group realize social engagements with words. An example of the distinction is given by the two sentences *The salt is on the table* and *Could you pass me the salt?*

In the first group, assertives and expressives differ because of the degree of subjectivity of their enunciation (Benveniste 1966; Ducrot 1984). This conception of assertives and expressives is different than the one proposed by Searle (1969, 1976, 1979). For Searle, expressives refer to intentions, conceived as mental entities, and assertives refer instead to other states of affairs existing "out there" in the world. Searle's sharp distinction reflects an ontology that is not necessary—and may even be an impair—for linguistic research.⁴ Both assertives and expressives are seen by us as claims realized by the speaker about the world (even when this does not seem to be the case, such as in the expressive *I feel sad*), and they differ because of the presence or absence of cues outlining the speaker's emotional stance on the described situation explicitly. The application of our speech acts classes to real-world corpora showed that there is a *continuum* between assertives and expressives, based on how we read words and punctuation as signs of enunciation, instead of a sharp divide based on the nature of their reference. Examples showing this distinction: *Twelve people were killed in the car crash* is an assertive; *What a horrible incident!* is an expressive; *Sadly twelve people were killed in the car crash* is in between the two.

In the second group, commissives and directives are complementary: directives ask for engagements and commissives bring them about. There is no continuity between the two classes (compared to the case for assertives and expressives), but conversational complementarity: ideally, for each directive a corresponding commissive can be imagined, and vice versa. Examples are: *Would you send me the manuscript?* and *Yes, I'm sending it now*. In a way, this shows that Searle's typology has already some conversational logics "hardwired" into it.⁵

For the ends of the analysis of conversation corpora, declaratives are treated as a special case, heterogeneous and rare in conversations, and do not require further detail.⁶ A classic example of declarative is *I declare you husband and wife*.

⁴ Searle's view has not to be considered the only correct interpretation of Austin's theory. Sbisà (2007) argues that in Austin (1962) there is a less substantial view of the speaking subject than in Searle.

⁵ "When asking ourselves how the study of conversation is to be related to the study of speech acts, we should not try simply to extend speech act theory to conversation, since the study of speech acts already presupposes some reference to conversational sequences." Sbisà (2002a).

⁶ It could be argued that also declaratives and assertives are poles of a continuum, having, on the one hand, facts that we believe exist before speech (assertives) and, on the other, facts that we believe come to existence only after speech (declaratives). However, Searle's typology aims to simplify the understanding of illocutionary forces, and our classification also aims to simplicity. Declaratives are, then, easy to define as a residual family of relatively rare utterances, with strict felicity conditions and conventional performative effects.

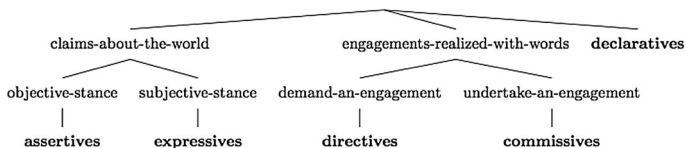


Fig. 2 Structure of the 5 types of speech acts

Let us now continue with the description of the 18 classes in which our proposed classification splits the five main speech acts (see Fig. 3). Assertives can be independent from previous statements or necessarily dependent from previous statements. In the latter case, there are two main illocutionary forces: one can *agree* or *disagree* with what was stated beforehand. In the former case, the degree of strength in something stated can be differentiated. The speaker can plainly *assert* something (with a medium strength), *sustain* it with arguments, examples and confidence (strong strength) or weaken its strength by stating the limitations of his or her belief and implying probability or possibility (class of illocutionary forces we have called *guess*). Examples of assertives are *You are right (agree)*, *I don't think so (disagree)*, *It costs 200\$ (assert)*, *I'm going to buy it because it's very convenient (sustain)*, *I'm unsure he's right for her (guess)*.

Regarding expressives, the general cases have been separated from some special ones. For the general expressives, the most relevant distinction to be traced is between positive and negative feelings, and so between the speech acts we called *rejoice* and *complain*.⁷ The special cases selected for the ends of conversation analysis are *greet* (subsuming all forms of salutation), *thank* (declare someone's merit for something and express a positive feeling about it) and *apologize* (declare one's fault for something and express a negative feeling about it). A last expressive is *wish*, any positive or negative feeling about the future. That a wish is necessarily about the future differentiates it well from *rejoice* and *complain* speech acts. Examples of expressives are: *Great! (rejoice)*, *This shouldn't have happened (complain)*, *Hi there (greet)*, *I am grateful for what you did (thank)*, *I'm sorry for yesterday (apologize)*, *I hope you recover soon (wish)*.

Directives can be differentiated by means of their strength of illocutionary force: strong in *require*, medium in *request* and weak in *suggest*. Requirements include orders and injunctions; requests include questions as well as invitations, while suggestions are said for the good of the addressed person.⁸ Examples of directives are: *Go to your room! (require)*, *Can you spare a dime? (request)*, *You should not smoke (suggest)*.

⁷ Some scholars treat the positive/negative polarity as a general category describing primitive emotions (Johnson-Laird and Oatley 1989: 91). Independent of its interpretation, our practical way of differentiating expressives implies that all expressions of subjectivity fall into either one of these sides, treating neutral (neither positive nor negative) claims about the world as assertives. We believe that sarcasm, and other forms of ambiguity, do not invalidate the feasibility of the distinction, as long as we take into account the conversational context, in which doubts about polarity are limited.

⁸ As pointed out by Vanderveken (1990: 189), a test for discriminating requirements from requests is that requirements demand as complementary response to obey or to disobey them, while requests demand to grant or to refuse them (suggestions demand an even weaker response, like a thank).

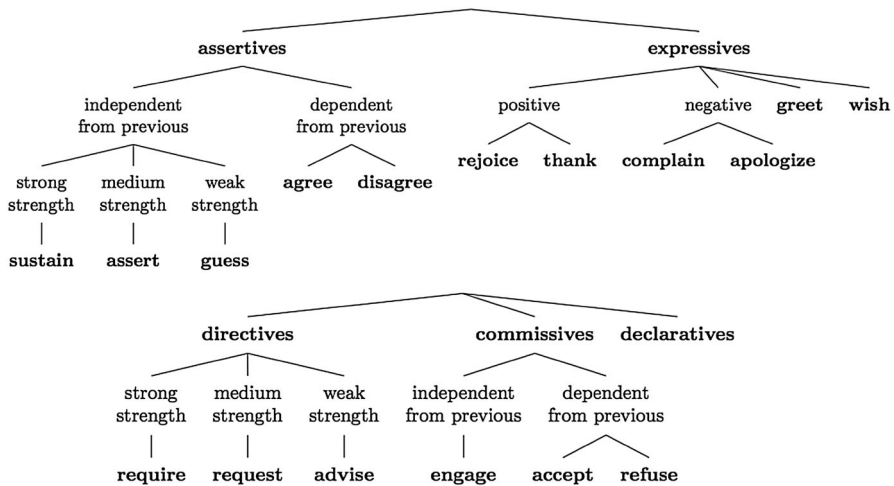


Fig. 3 Proposed classes of speech acts

Commissives can be independent from previous directives or necessarily dependent from previous directives. If dependent, we talk of *accept* or *refuse* speech acts (independently of the degree of force of the complementary directive). If independent, we just talk of *engage* speech acts. Examples of commissives are: *Ok let's do it (accept)*, *Sorry not this time (refuse)*, *Starting tomorrow no more alcohol for me (engage)*.

As previously mentioned, declaratives are too rare in conversations to need internal differentiation for the ends of PM.

These intentionally broad differential principles permit to define the classes of speech acts in a somehow fuzzy way: we rely on the speakers' spontaneous linguistic judgements and on their ability to intuitively grasp the thresholds between classes. In the end, our classes are defined by linguistically significant oppositional borders and by examples considered to be typical for each class. This approach is suitable for the practical application of speech act theory to corpora. On the contrary, it would be impractical to define a complete and sufficient number of precise linguistic features for each class of speech acts: these features may not even exist (Violi 2001), given that the classes subsume otherwise heterogeneous sentences.

Assessing the Consistency of the Classification

Procedure and Limitations of the Experimental Annotation

The consistency of the classification was assessed by applying it to a corpus of digital conversations (see Tudini 2014), realized on the website *Reddit*. Each conversation starts with a first post and continues with comments to the initial post or to other comments. The corpus was composed of anonymized threads about

autoimmune diseases. This particular ‘subreddit’ was chosen for evaluation because, in a comparative exploratory analysis, we noticed that most of our classes of speech acts could be found in it: when discussing their experiences with diseases, people may give advice, thank for it, ask for specific information, share positive or negative feelings about personal experiences and so on. For this first annotation, the choice of focussing on just one specific corpus (instead of on several subreddits) has been driven by the will to have sufficient data to adequately evaluate at least one application. In order to empirically prove the corpus-independency of the classification (and eventually its better fit to some particular cases compared to others) more annotations will have to be performed.

The classification used for this annotation had three differences with the general one presented above. First, a class *other* was added, because a generic class for speech acts not fitting into the classification was needed. Second, the classes *require* and *request* were merged, because of empirical evidence about the absence of *require* speech acts (such as orders) in *Reddit* threads, collected in the initial phase of our research. Third, the class *declare* was removed, for similar reasons, collecting occasional declarative speech acts in the *other* class.

21 conversations of various lengths were extracted from the subreddit,⁹ parsed into grammatical sentences considering punctuation and randomly assigned to annotators. There were two kinds of annotators: 2 ‘experts’ (the authors who have elaborated the classification) and 10 ‘raters’ (participants who had no previous knowledge of the classification). Raters were mainly academics, ranging from graduate students to professors, with an interest in language but specialized in various fields including computer science, sociology of communication and anthropology. None of them was a native English speaker, but all were proficient.

Raters were given a set of instructions explaining the task and the classification to be used, with examples. Each rater had to annotate 3 conversations, with a total of 120–150 conversational units; experts were paired with raters and annotated the same units independently. Raters and experts were asked to annotate individual sentences one by one with reference to their linguistic form but also by taking into account their context (the entire conversation).¹⁰ Exploratory pre-tests showed that giving access to the utterances’ uptakes made the classification task faster, perceived as easier and more accurate (better inter-rater agreement).

It is important to remark that the subjects were given the possibility to choose up to three classes for each sentence (*multi-label annotation*). Exploratory annotations showed that single-label annotation bore some negative consequences: the process of choosing only one class among multiple often equally legitimate possibilities could be long and frustrating, resulting sometimes in arbitrary choices by different annotators. In fact, sometimes a single grammatical sentence realizes more than one

⁹ Conversations contained in average 8 turns and 45 sentences. Most posts were made of between 1 and 13 sentences and most comments of between 1 and 7 sentences.

¹⁰ Raters were given a sheet of instructions, asking to recognize the aims of some sentences presented in threads. We specified that the context (the entire thread) had to be taken into account to understand the aim of individual sentences, and that our classification had to be used. We also stated that more than one class (up to three) could be chosen, if a sentence had more than one aim or if in doubt. A second sheet presented the classes with short definitions and examples.

consecutive speech act (for example *Hi, what time is it?* realizes a *greet* and a *request*). More generally, the fuzzy definition of our classes opened the possibility of a sentence belonging to more than one class.

Additionally, multi-label annotation enhances the actual application of speech act theory to real corpora. Univocal categorizations of sentences are bound to limited results because of the copresence of locutionary and illocutionary acts (Austin 1962) and of the occurrence of indirect speech acts (Searle 1979). For Austin, the intended overt aim of an utterance (illocutionary act) is realized by means of a sentence having an ostensible meaning (locutionary act); in some cases, the identification of only one intended force for a given utterance can be problematic. Searle proposed to differentiate two kinds of illocutionary acts (direct and indirect), on the basis of the effort of cooperation needed by the hearer or reader to infer an intended meaning from the ostensible one.

Practically, if it is easy to identify a very small number of classes describing the nature of a speech act, it becomes much harder to choose one force considered to be the most important.¹¹ On the contrary, multi-label annotation reduces cognitive distress and gives also a more comprehensive description of the individual sentences, useful to better grasp the complexity of human behaviour. In the instructions, we asked to identify the main aim(s) of the sentences, without making any reference to the locutionary/illocutionary and direct/indirect distinctions.

This is one first limitation of our study: more complex annotation schemes, such as DSAT (Geis 1995) or others (see Weisser 2014; Garcia McAllister 2015) would allow for the mining of more precise activities and interaction structures, connecting more than one sentence into complex communicative acts. Our implementation has a lower level of detail, but this makes it easier to apply and interpret. Furthermore, PM is aimed to the inductive identification of regularities that could finally result into interaction structures such as those discussed by Geis 1995.

A second limitation of our study regards the distinction between illocutionary forces and other kinds of conversational aims. Illocutionary forces, realized through speech acts, depend on the recognition of a communicative intention.¹² On the other hand, the speaker may want to obtain some effects depending on the fact that his or her intention is not recognized as such (Austin 1962; Strawson 1964). More generally speaking, we could cite Kissine (2013) and differentiate illocutionary (or ‘transactional’) conversational aims and other interactional ones. Now, our classification only takes into account illocutionary forces, but we asked the annotators to more generically identify the sentences’ conversational aims. This leaves space for some ambiguity in the interpretation of the annotated corpus, as it may have happened that a rater identified a non-illocutionary conversational aim and falsely categorized it as a speech act. We recognize the existence of such limitation; however, instructing raters to differentiate illocutionary from other

¹¹ De Felice et al. (2013: 77) write: “how to account for this distinction [between locutionary and illocutionary meanings] is one of the main challenges of speech act annotation.”.

¹² “For the illocutionary force of an utterance is essentially something that is intended to be understood. And the understanding of the force of an utterance in all cases involves recognizing what may be called broadly an audience-directed intention and recognizing it as wholly overt, as intended to be recognized.” Strawson (1964, p. 459).

conversational goals could have made the annotation harder. In any case, our classification and the exploratory results of PM presented below (“Exploring Digital Conversations with PM” section) regard only illocutionary forces and not all possible conversational aims.

Comparing Inter-Rater Reliability for Searle’s Types and for Our Classes

We aimed to validate the consistency of our classification (its “correctness” in an annotation) and not the exhaustive annotation of a corpus.¹³ This means that the units for which we calculated agreement were the *classes* and not the *sentences*. We wanted to know whether different annotators would use the classes in the same way (attributing the same meaning to them, we could argue). By measuring how many times annotators agreed, using the same class, we also identified the classes for which agreement was the lowest. It would have been a different task, beyond the scope of our research, to measure whether the annotators were able of finding all the illocutionary forces in each sentence.

Inter-rater reliability has been measured by comparing the annotations done by pairs of raters and experts. Given that multi-label annotation was allowed, some data transformations were necessary before evaluation. First, for each conversational unit, the labels used by the different annotators were aligned such that the matching classes were paired despite the order in which they were given. This preprocessing step resulted in generating between 1 and 3 pairs of classes per conversational unit. For example, if the expert reported “*assert, complain, empty*” and the rater “*complain, sustain, empty*”, the resulting pairs were “*assert–sustain*” and “*complain–complain*”. Sometimes, one annotator used more classes than the other. In this case, we excluded from the agreement calculation the pairs including *empty*. For example, if the expert reported “*assert, request, empty*”, and the rater “*assert, empty, empty*”, we considered that both annotators agreed on the use of the *assert* class, and do not take into account the “*request–empty*” pair.¹⁴ This allowed us to measure how many times the expert and the rater used the same class to describe a sentence.

For this data structure, it was possible to calculate the reliability with Fleiss’ Kappa statistical measure (Fleiss 1971), considering class-wise agreement and the probability of agreeing by chance.¹⁵ Table 1 shows the usual interpretation of Kappa results.

Fleiss’ Kappa was *moderate* to *substantial* for all annotators, as reported in Table 2. The null hypothesis $Kappa = 0$ (there is no agreement among annotators) is tested using the z-test. The values of the z-test are also provided in Table 2, with a *p* value of less than 0.01 for all cases. The same statistics were calculated for

¹³ The “completeness” of the classification, that is, whether or not all illocutionary forces are taken into account by it, cannot but rest on Searle’s and Vanderweken’s work, starting from which our classification was designed. If the annotators resorted often in using the residual class *other*, than this would have been an indiciary sign of incompleteness (but eventually this did not happen).

¹⁴ Annotators used more than one class for about 16.5% of the total sentences.

¹⁵ Fleiss’ Kappa (Fleiss 1971) is a statistical technique for calculating inter-rater agreement. We used the function *kappam.fleiss* from the R package *irr* (Gamer et al. 2012).

Table 1 Results interpretation for Fleiss' Kappa among annotators

| Judgement | Fleiss' Kappa |
|----------------|---------------|
| Poor | < 0.00 |
| Slight | 0.01–0.20 |
| Fair | 0.21–0.40 |
| Moderate | 0.41–0.60 |
| Substantial | 0.61–0.80 |
| Almost perfect | 0.81–0.99 |

Table 2 Fleiss' Kappa (F.K.) with z-test's values for all 10 raters and our 18 classes of speech acts

| Rater | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|--------|------|------|------|------|------|------|------|------|------|------|
| F.K. | 0.58 | 0.59 | 0.72 | 0.48 | 0.6 | 0.68 | 0.68 | 0.66 | 0.73 | 0.65 |
| z-test | 18.8 | 18.3 | 23 | 13.2 | 20.6 | 18.6 | 21.7 | 21.1 | 20 | 18.4 |

Searle's five original types of speech acts. This has been possible because our classes are specifications of Searle's types (so for example *sustain* and *assert* are both specifications of the type *assert*). The Fleiss' Kappa was *almost perfect* for 5 raters, *substantial* for 4 raters and *moderate* for 1, as reported in Table 3. The z-test's values are provided; the p-value (to reject the hypothesis of $Kappa = 0$) was less than 0.01 for all raters.

The number of speech acts identified in the corpus for every class and for every type varied substantially, as shown in Table 4. Assertives, and specifically the *assert* class, were by far the most common type of speech acts. Fleiss' Kappa was calculated for each class and type of speech acts and for each pair of raters, noticing that it varied substantially. Table 4 presents the median Fleiss' Kappa for all classes and types. The classes *thank* and *direct* show an *almost perfect* agreement. On the contrary, *accept* and *refuse* show no agreement. Apart from the classes *sustain*, *disagree*, *engage* and *other*, there was a *substantial* agreement in the identification of all other classes.

The low values of inter-rater agreement for some classes point to some potential issues in our classification. This may be due to unclear definitions; however, the very low number of occurrences found in the corpus for these classes has also to be considered. The classes *accept* and *refuse* were the subject of the lowest agreement and also among the least represented classes in the corpus. However, the class *sustain* was well represented in the corpus and still object of low agreement.

Assert appears to be the fallback class. *Rejoice* and *complain* were sometimes paired with assertives (respectively 29 and 27.5% of their occurrences, for example

Table 3 Fleiss' Kappa (F.K.) with z-test's values for all 10 raters and Searle's 5 types of speech acts

| Rater | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|--------|------|------|------|------|------|------|------|------|------|------|
| F.K. | 0.78 | 0.75 | 0.82 | 0.57 | 0.84 | 0.75 | 0.87 | 0.85 | 0.82 | 0.77 |
| z-test | 16.8 | 13.2 | 14.3 | 9.93 | 17.2 | 11.4 | 18.4 | 15.1 | 12.1 | 11.9 |

Table 4 Frequency information and Fleiss' Kappa score (F.K) for speech act classes

| Speech act | Frequency | F.K. | Group | Total |
|------------|-----------|---------|------------|-------|
| ASSERT | 1351 | 0.62 | ASSERTIVE | 2216 |
| SUSTAIN | 427 | 0.36 | | |
| GUESS | 290 | 0.66 | | |
| AGREE | 93 | 0.61 | | |
| DISAGREE | 55 | 0.25 | | |
| REJOICE | 253 | 0.71 | EXPRESSIVE | 936 |
| COMPLAIN | 432 | 0.77 | | |
| WISH | 90 | 0.70 | | |
| APOLOGIZE | 34 | 0.66 | | |
| THANK | 88 | 1 | | |
| GREET | 39 | 0.66 | | |
| DIRECT | 252 | 0.89 | DIRECTIVE | 548 |
| SUGGEST | 296 | 0.72 | | |
| ENGAGE | 72 | 0.51 | COMMISSIVE | 126 |
| ACCEPT | 40 | - 0.003 | | |
| REFUSE | 14 | - 0.004 | | |
| OTHER | 98 | 0.35 | OTHER | 98 |

in the sentence “*I was sick from it for 3 months!*”), but very rarely with other classes (4% of their occurrences). This can be seen as a sign of the continuity between assertives and expressives as we hypothesized in “[A Structural Classification of Speech Acts](#)” section. Within assertives, *sustain* was hard to differentiate from *assert* (mismatched in 41.6% of occurrences, for example in “*It was clearly just a conjecture, though, there’s no data out there to confirm this*”). We suspect that this could be due to unclear definition. The classes *agree* and *disagree* were often paired with assertives (17.5 and 30.7% of occurrences respectively, for instance in “*Same symptoms, too.*”). The raters might have not considered the preceding sentences in conversations, possibly because of the interface we provided for annotation. Finally, we remark that *suggest* was mistaken with assertives (24.5% of its occurrences, for example in “*For the sore joints, I try to workout twice a week to keep my body from feeling too stiff*”). We imagine that the reason can be the presence of indirect suggestions not always identified.

Further annotations on different corpora may help with solving these issues. Also providing clearer rating instructions, a better annotation interface and formulating better class definitions are directions for future work.

Exploring Digital Conversations with PM

Finally, we tested our classification by applying PM on our annotated corpus. In order to work with the most reliable corpus possible, the two experts discussed the annotated data and produced a consensual classification of all utterances. The corpus was then processed in order to pass it to PM software. Mainly, multi-label

sentences had to be converted into single-label ones. We chose to adopt a two-step procedure: first, if *assert* was among the classes of a multi-label utterance, it was removed; second, each multi-label utterance was randomly assigned to one of its classes. This procedure aimed to obtain the least biased annotation, while at the same time reducing the impact of frequent multi-label asserts (many utterances perform a speech act while also asserting something).

We compared the results of using Searle's types and our finer-grained classification. The data were analysed with the software *Disco*.¹⁶ This software was chosen because it implements a fuzzy algorithm of inductive process discovery (Günther and van der Aalst 2007), while having a graphical interface that is easy to use for exploratory analysis. The corpus was also split into four overlapping subsets (see Table 5).

This subsetting depends on the fact that online conversations in *Reddit* allow for *concurrent threads*. The same question (post) can be answered by several people. Then each of these answers can develop its own thread, giving an arborescent structure to the whole. It is therefore not possible to consider complete conversation traces, starting with a question and moving forward in a linear exchange of sequential alternating replies. When analysing digital conversations, one has necessarily to take into account concurrency, and we chose to “cut” the conversation tree into levels lying at the same distance from their root (a post).

As a consequence, our analysis takes into account conversational processes happening *within-turns*, and *not across-turns*—otherwise, the count of turns common to multiple concurrent threads would have biased the study (a post at the root would have been repeated as many times as the number of its comments at the leaves). This limits the extent of our analysis' results; however the exact same analysis, conducted on non-concurrent (oral or written) conversations, would instead be able to also take into account across-turns sequences of speech acts, arguably leading to more interesting results for conversation analysis.

Comparing Searle's and Our Classes on Full Corpus Analysis

Let us consider, first, the relations that can be found among Searle's types of speech acts in the full corpus (Fig. 4). The diagram is a graph, in which each vertex is a type of speech act (an activity) and arcs represent adjacency relationships between speech acts; the number of turns in which they appear is shown and also represented graphically (the darker or thicker, the more frequent the activity or path). The green and red circles represent the beginning and the end of turns.

All activities are shown (excluding the residual class *other* for clarity); on the contrary, only the most representative arcs are visible (75% of paths in this case): this permits to have a synthetic (while non-exhaustive) view of the typical process in play. By looking at Fig. 4 we can already obtain some insights on intra-turn sequences of speech acts:

¹⁶ <https://fluxicon.com/disco/>.

Table 5 Subsets of the corpus

| Set | Number of turns |
|---|-----------------|
| Full corpus | 398 |
| Only posts | 50 |
| Only comments | 348 |
| “First level” comments (immediate replies to posts) | 189 |
| “Second level” comments (immediate replies to “first level” comments) | 84 |

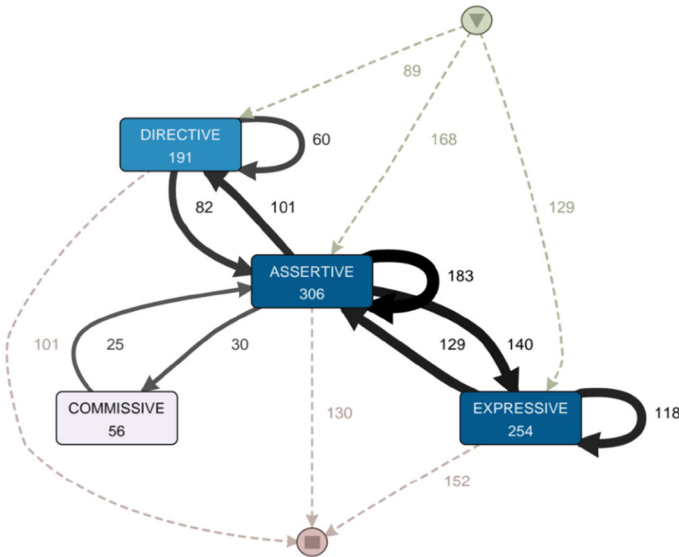


Fig. 4 Process mining of the full corpus using Searle's classification

- (a) Assertives represent the activity connecting all the others; on the contrary, directives, expressives and commissives do not usually follow each other directly;
- (b) One every two expressives is followed by another expressive, while this is true only of one directive out of three, and commissives tend not to appear in loops at all;
- (c) Assertive are the most common speech act at the beginning of turns (42.5%) while expressives at their end (38.5%); commissives are not found at beginnings nor at ends of turns.

While these considerations are very general, and should only be taken as guides to formulate hypotheses, they indicate the potential of PM for finding empirical regularities in conversation corpora. The “embedded” position of commissives, for example, suggests that in our corpus they are performed within an ongoing turn and

are surrounded by factual affirmations. A commissive could be perceived as improper, if used outside of its typical process (that is, not between assertives).

Another hypothesis that could be tested, regards differences in the length of speech act sequences. Commissives seem to be realized in one utterance, while the frequent loops of assertives and especially of expressives suggest that these activities need more time and detail to produce a complete illocutionary effect. Again, from Searle's formal perspective, each speech act can still be identified as an assertive or an expressive (as the annotators in fact did), but speech act theory does not give any account as for real world speech acts happening in "well formed", typical sequences. An isolated expressive, for example, may sound inappropriate, if expressives are typically not found alone.

Can our analysis permit to identify specific ways of *doing things with words*, *requiring more than one utterance*? Searle's classes are probably too ample for that. Let us now consider the processes that can be found in the same corpus by applying our finer-grained classification. Figure 5 shows only the 70% most common activities and the 30% most recurrent paths among them. It should be noted that arcs do not necessarily represent immediate followings as in Fig. 4 (where all activities were visible), and can hide longer chains of lesser common activities. Again, the aim is to find a way to typicality and not exhaustiveness. Compared to Fig. 4, we notice some important differences:

- (a) *Complain* and *rejoice*, subclasses of expressives, respectively about negative and positive feelings, occupy very different positions in the conversation process. While *complain* appears often at the beginning (11.6%) and at the end (14.7%) of turns (let us remember that our corpus is about medical advice), *rejoice* is less common and has a more peripheral position in the development of turns;
- (b) *Assert* and *complain* often loop with one another;
- (c) Among directives, *direct* (including questions) is more commonly found at the beginning of turns (12.4%), while *suggest* at the end (13.7%).

We can observe that our finer-grained classification does give a different perspective on conversation turns, for example via the differentiation of positive and negative expressives. If negative feelings are not expressed in the same way as positive ones, because they appear in different typical sequences within turns, this suggests again that a comprehensive description of speech acts should take into account inter-utterance behaviour. *Complain* and *rejoice* are not the same speech act, also because they do not typically appear in the same conversational routines. It would not be enough to substitute *I feel happy* with *I feel sad* to switch between speech acts, because the two utterances cannot be situated at the same place in real-world conversations.

As Searle's efforts reveal, conceptual analysis of however numerous isolated utterances risks to be limited to grasp only a subset of felicity conditions. Typical behaviour of speech acts in conversations is in fact only visible through the inductive study of large real world corpora. If further studies show that indeed typicality conditions of appearance should be considered next to formal felicity

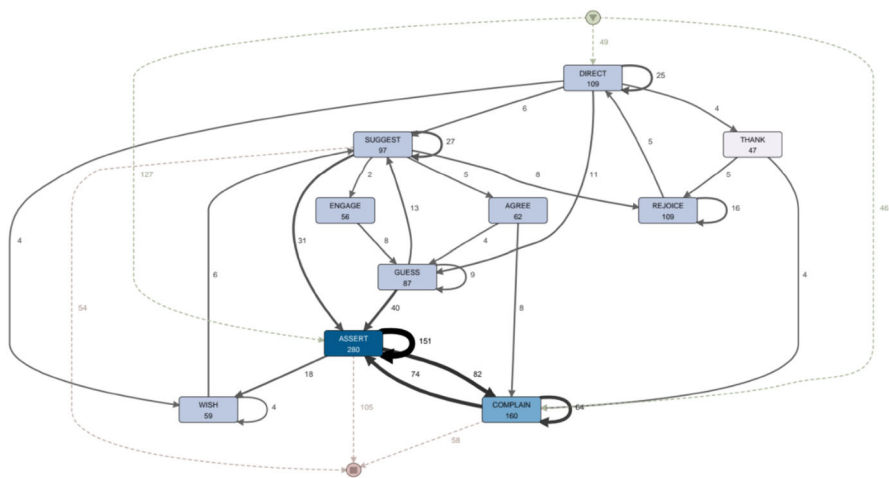


Fig. 5 Process mining of the full corpus using our classification

conditions when describing speech acts, this would mean that theoretical pragmatics cannot avoid corpus analysis to fully understand speech behaviour.

The advantage of finer-grained classifications of speech acts becomes more evident if we focus on the subsets of our corpus. Posts most often begin with a *direct* (46%), while “first level” comments with an *assert* (40.7%), a *complain* (15.3%) or a *suggest* (13.8%), and “second level” comments with a *thank* (23.8%), an *assert* (20.2%) or an *agree* (13.1%). This makes sense, as posts’ first utterances (their titles) are often formulated in interrogative terms, then they are followed by answers that can take the form of plain asserts, recommendations or expressions of empathy, and these are in turn often replied by the post’s author, expressing gratitude or agreement with what was suggested to him or her. This is not really an insight about typicality conditions, but more of ground truthing. Still, if we only considered Searle’s classification, it would be harder to spot such regularities.

“Direct Follows Complain” Preference in Posts

To conclude this exploratory application of PM to conversation corpora, let us focus on a last unexpected regularity that can be observed in one of our sub-corpora: in posts, *complain* precedes *direct* much more frequently than the reverse. *Direct* is eventually (that is, not necessarily immediately) followed by *complain* in the 36% of posts, and in the 8% *complain* is also never eventually followed by *direct* (that is, the two speech acts are found only in one order). On the other hand, *complain* is eventually followed by *direct* in the 48% of posts, and in the 20% *direct* is also never eventually followed by a *complain*. It seems therefore that there is a conversational preference for giving a specific order to the two speech acts. Figure 6 compares the processes in which *direct* is eventually followed by *complain* (on the left) and vice versa (on the right).

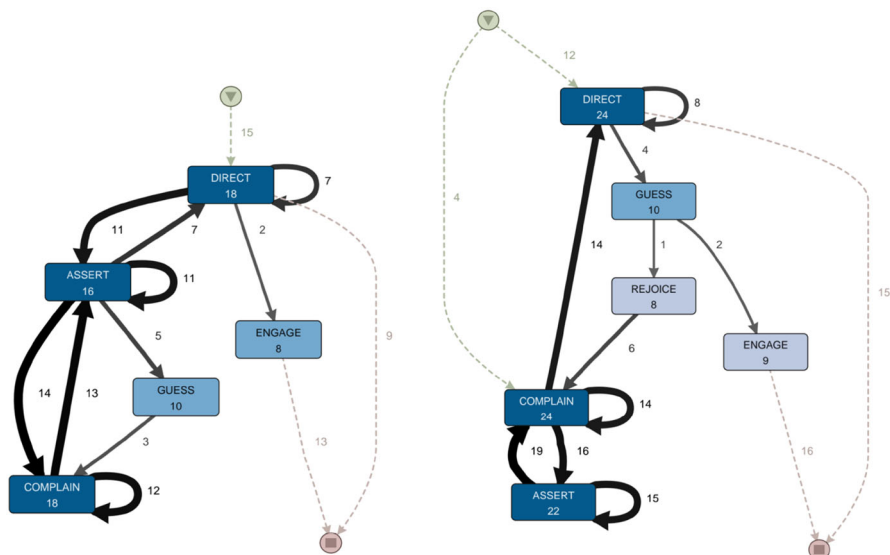


Fig. 6 Comparing posts in which *direct* is eventually followed by *complain* and vice versa

We can notice that *direct* follows *complain* immediately, while if *complain* follows *direct* then a third activity (*assert*) is found between them.¹⁷ This may be seen as an indicatory confirmation of such “*direct* follows *complain*” conversational preference: going against the norm demands a mediation.

As stated above, this is only a hint to further confirmatory research about the typical order in which these speech acts appear. However, if confirmed, it would show that we do complain and pose questions, but not in random order. A complain opens up the possibility of formulating questions; on the contrary, a question often demands other kinds of continuation, to be detailed and completed. The possibility of finding such conversational preferences leads to new scenarios for linking speech acts and conversation analysis.

Further work analysing linear conversations, and so capable of taking into account across-turns sequences, along with within-turns ones, could reveal much more. For a start, it would be important to see whether certain speech act sequences are more commonly followed by other speech acts in the same turn or instead tend to be followed by another turn. PM could help to differentiate “turn-ending sequences” from “turn-keeping sequences”. Generalising, this could permit to associate speech act sequences to turn-taking punctuation, and so to identify new norms of conversational grammar. Our application of PM to conversation corpora highlights these and other potential developments for conversation analysis.

¹⁷ In the posts in which the two speech acts are found in only one order, *complain* never follows *direct* immediately (0 out of 4 cases), and *direct* follows *complain* immediately in 7 cases out of 10.

Conclusion and Future Work

This paper put process mining to the test for the analysis of conversations. As this procedure is new, we faced several challenges and had to cope with some limitations. We aimed at a foundational step capable of sustaining further development, by creating, validating and testing a suitable speech acts classification. We needed a fine-grained, corpus-independent, easy to learn and interpret classification (“[A Structural Classification of Speech Acts](#)” section). We chose a multi-label, contextual setting for the annotation of a small corpus. To empirically prove the corpus-independence of our classification more annotations are needed—as for now, it can be presumed to be corpus-independent with reference to its design.

We worked on digital conversations about a very specific subject (medical advice). In the exploration of such a small corpus (“[Exploring Digital Conversations with PM](#)” section), we already found some evidence going against Searle’s assumptions: some typicality effects seem to exist in conversation processes, such as preferences in the ordering of speech acts. If these effects are confirmed, then corpus pragmatics may prove itself irreplaceable for the comprehensive understanding of speech acts: in fact, the definitions of speech acts may include a reference to norms about how to properly insert them in conversations. Eventually, a stronger idea may also need to be taken into consideration: that specific sequences of utterances realizing complex speech acts exists—as suggested by Sbisà (2002a, b). Further work may be aimed at understanding whether such sequences play precise roles in conversations, functioning as a sort of punctuation for turn-taking.

The exploration of real-world corpora may be used to formulate new questions and to produce evidence relevant to conversation analysis as well as to speech act theory. Our analysis was purely exploratory as we did not test specific hypotheses. Still, it led to the formulation of plausible hypotheses about typical sequences of specific speech acts, such as *direct* and *complain*. Further work may want to prove whether there is continuity between assertives and expressives, or instead they lie on different sides of an ontological divide (as in Searle’s view). Empirical evidence may also be gathered to understand whether the illocutionary force of some speech acts should be reinterpreted (wishes be considered directives instead of expressives, among others).

Digital conversations have some peculiarities that required ad hoc treatment: they are concurrent and develop an arborescent structure, that had to be “cut” into homogeneous levels. This implied that we could only look for within-turn regularities, leaving out across-turns ones. Also, compared to the kind of behaviour customarily studied by PM, conversations have a highly non-linear structure: there are no activities (speech acts, in our case) always appearing at the beginning or at the end of turns. However, PM proved capable of highlighting some potential empirical regularities.

The next step will be to train machine learning algorithms for automatic classification of larger and less specialized corpora of digital conversations, more suitable for supporting empirical research in these directions.

References

- Arguello, J., & Shaffer, K. (2015). Predicting speech acts in MOOC forum posts. In *Proceedings of the 9th international AAAI conference on web and social media*. Oxford: Oxford University Press.
- Austin, J. L. (1962). *How to do things with words*. Cambridge, MA: Harvard University Press.
- Benveniste, E. (1966). *Problèmes de linguistique générale*. Paris: Gallimard.
- Bhatia, S., Biyani, P., & Mitra, P. (2012). Classifying user messages for managing web forum data. In G. I. Zachary & Y. Velegarakis (Eds.), *Proceedings of the 15th international workshop on the web and databases* (pp. 13–18). ACL.
- Carvalho, V. R., & Cohen, W. W. (2005). On the collective classification of email 'speech acts'. In R. Baeza-Yates & N. Ziviani (Eds.), *Proceedings of the 28th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 345–352). New York, NY: ACM.
- De Felice, R., Darby, J., Fisher, A., & Peplow, D. (2013). A classification scheme for annotating speech acts in a business email corpus. *ICAME Journal*, 37, 71–105.
- Ducrot, O. (1984). *Le dire et le dit*. Paris: Minuit.
- Epure, E. V., Deneckere, R., & Salinesi, C. (2017). Analyzing perceived intentions of public health-related communication on twitter. In A. ten Teije, C. Popow, J. Holmes & L. Sacchi (Eds.), *Artificial intelligence in medicine*. Berlin: Springer.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378–382.
- Gamer, M., Lemon, J., Fellows, I., & Singh, S. (2012). *irr R Package*. CRAN. <https://cran.r-project.org/web/packages/irr/>.
- Garcia McAllister, P. (2004). *Meaning in academic contexts. A corpus-based study of pragmatic utterances*. Ph.D. dissertation, University of Northern Arizona.
- Garcia McAllister, P. (2015). Speech acts. A synchronic perspective. In K. Aijmer & C. Rühlemann (Eds.), *Corpus pragmatics. A handbook* (pp. 29–51). Cambridge: Cambridge UP.
- Geis, M. L. (1995). *Speech acts and conversational interaction*. Cambridge: Cambridge UP.
- Günther, C. W., & van der Aalst, W. M. P. (2007). Fuzzy mining: adaptive process simplification based on multi-perspective metrics. In *Proceedings of the 5th international conference on business process management* (pp. 328–343). Berlin: Springer.
- Hutchby, I., & Wooffitt, R. (2008). *Conversation analysis*. Cambridge: Polity.
- Johnson-Laird, P. N., & Oatley, K. (1989). The language of emotions: An analysis of a semantic field. *Cognition and Emotion*, 3(2), 81–123.
- Jucker, A. H. (2002). Conversation: Structure or process? In Searle, Parret & Verschueren (Eds.), 1992 (pp. 77–90).
- Kim, S. N., Wang, L., & Baldwin, T. (2010). Tagging and linking Web forum posts. In *Proceedings of the 14th conference on computational natural language learning* (pp. 192–202). Stroudsburg, PA: ACM.
- Kissine, M. V. (2013). *From utterances to speech acts*. Cambridge: Cambridge UP.
- Parret, H., & Verschueren, J. (1992). Introduction. In Searle, Parret & Verschueren (Eds.), 1992 (pp. 1–5).
- Rastier, F. (2011). *La mesure et le grain. Sémantique de corpus*. Paris: Champion.
- Sbisà, M. (1992). Speech acts, effects and responses. In Searle, Parret & Verschueren (Eds.), 1992 (pp. 101–111).
- Sbisà, M. (2002a). Speech acts in context. *Language & Communication*, 22(4), 421–436.
- Sbisà, M. (2002b). Cognition and narrativity in speech act sequences. In C. Meierkord & A. Fetzer (Eds.), *Rethinking sequentiality*. Amsterdam: John Benjamins.
- Sbisà, M. (2007). How to read Austin. *Pragmatics*, 17(3), 461–473.
- Schegloff, E. A. (1992). To Searle on conversation: A note in return. In Searle, Parret & Verschueren (Eds.), 1992 (pp. 113–128).
- Schegloff, E. A., & Sacks, H. (1973). Opening up closings. *Semiotica*, 8, 289–327.
- Searle, J. (1969). *Speech acts. An essay in philosophy of language*. Cambridge: Cambridge University Press.
- Searle, J. (1976). A classification of illocutionary acts. *Language in Society*, 5(1), 1–23.
- Searle, J. (1979). *Expression and meaning*. Cambridge: Cambridge UP.
- Searle, J. (1992). Conversation. In Searle, Parret & Verschueren (Eds.), 1992 (pp. 7–29).

- Searle, J., Parret, H., & Verschueren, J. (1992). *(On) Searle on conversation*. Amsterdam: John Benjamins.
- Sidnell, J., & Stivers, T. (Eds.). (2012). *The handbook of conversation analysis*. Oxford: Wiley.
- Strawson, P. F. (1964). Intention and convention in speech acts. *The Philosophical Review*, 73(4), 439–460.
- Tudini, V. (2014). Conversation analysis of computer-mediated interactions. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–7). Hoboken, NJ: Wiley-Blackwell.
- van der Aalst, W. (2011). *Process mining*. Berlin: Springer.
- Vanderveken, D. (1990). *Meaning and speech acts: Principles of language use* (Vol. 1). Cambridge: Cambridge UP.
- Violi, P. (2001). *Meaning and experience*. Bloomington, IN: Indiana UP.
- Wang, G. A., Wang, H. J., Li, J., Abrahams, A. S., & Fan, W. (2014). An analytical framework for understanding knowledge-sharing processes in online Q&A communities. *Transactions on Management Information Systems*, 5(4), 1–13. <https://doi.org/10.1145/2629445>.
- Weisser, M. (2014). Speech act annotation. In K. Aijmer & C. Rühlemann (Eds.), *Corpus pragmatics. A handbook* (pp. 29–51). Cambridge: Cambridge UP.
- Winograd, T., & Flores, F. (1986). *Understanding computers and cognition: A new foundation for design*. Norwood: Ablex.
- Zhang, R., Gao, D., & Li, W. (2011). What are Tweeters doing: Recognizing speech acts in Twitter. In *Papers from the 2011 AAAI workshop*, Menlo Park, CA: AAAI.