

dassBNC2014

Retrieving and mapping directive speech acts in an open-domain corpus of everyday conversations

Guying Zhou and Jiajin Xu
Beijing Foreign Studies University

This study introduces dassBNC2014, a manually annotated subset of the Spoken British National Corpus 2014 designed to capture directive speech acts in everyday conversation. In natural language processing (NLP), speech act annotations in open-domain dialog remain scarce, while manually annotated data are fragmented and inconsistently accessible. With dassBNC2014, we aim to address the persistent gaps in existing resources. Our annotation scheme combines the ISO 24617-2 framework with insights from speech act theory to achieve an intermediate level of granularity, which is both theoretically grounded and operationally feasible. Analysis of the annotated directives shows that they constitute a heterogeneous set of practices varying systematically in interactional locus, temporal orientation, deontic strength and action type. These findings refine the theoretical description of directive force, establishing dassBNC2014 as a resource for advancing corpus-based speech act research and supporting the development of NLP models capable of recognizing directive speech acts in authentic conversational data.

Keywords: dassBNC2014, directive speech act, pragmatic annotation, taxonomy for directives, large language models

1. Introduction

Speech acts are considered a fundamental unit of human communication (Searle, 1969:16; Searle & Vanderveken, 1985:1). Understanding how speech acts are realized in naturally occurring conversations is central to uncovering the interplay between linguistic forms, contextual factors, and communicative functions in the real world. However, the systematic identification and annotation of speech

acts remain a major challenge, largely due to their inherent form-function mismatch (see Kohnen, 2015: 54; Milà-Garcia, 2018: 265; Rühlemann, 2018: 19). Consequently, the development of large, pragmatically annotated corpora has lagged behind. However, such resources are essential not only for advancing theoretical research on speech acts but also for supporting practical applications in natural language processing (NLP), such as dialog systems, semantic interpretation, and automatic recognition of functional language in human-computer interaction.

Despite ongoing efforts, research-ready, pragmatically annotated corpora remain rare across both linguistics and NLP. Within linguistics, SPICE-Ireland (Kirk, 2016) represents one of the few large-scale efforts to cover speech acts across discourse types based on Searle's taxonomy. Other corpora, such as HKCSE-SAC (Seto, 2016) and SPAAC (Leech & Weisser, 2003), are restricted to specific registers such as business discourse and telephone interactions and employ tag sets created by individual researchers. In NLP, the most widely used dataset remains the SWDA from the 1990s (Jurafsky et al., 1997), which has become the *de facto* benchmark for dialog act classification. Other corpora, such as ICSI-MRDA (Janin et al., 2003), the AMI Meeting Corpus (Carletta, 2007), the HCRC Map Task Corpus (Anderson et al., 1991), Verbmobil (Jekat et al., 1995), and DialogBank (Bunt et al., 2016) under the ISO 24617-2 framework, are limited in size and scope and vary in taxonomy. These resources are geared toward computational tasks and fall short of supporting a detailed theory-informed linguistic analysis.

Taken together, existing corpora across both linguistics and NLP have yet to provide a sufficiently solid foundation to support the empirical and theoretical demands of speech act research. In linguistics, expert-annotated data are often fragmented and seldom integrated into accessible standardized resources. Manual annotations are frequently only described in publications and are not made available to the research community (e.g. Bilbow, 2002; Gagné, 2018; García, 2004; Xu & Hao, 2019, among others). With the exception of Weisser's series of studies (Leech & Weisser, 2013; Weisser, 2020), the field has largely neglected the development of a unified annotation framework. Most studies have adopted *ad hoc* schemes tailored to specific research goals, resulting in a highly inconsistent landscape that limits comparability and cumulative progress. Although more standardized frameworks exist in NLP, high-quality pragmatic annotation in everyday open-domain conversations is still lacking. These limitations highlight the need for a more integrated and theory-informed approach to corpus development.

This study introduces *dassBNC2014*: a directive-annotated subset of the Spoken BNC2014, developed to facilitate the investigation of directive speech acts in open-domain everyday conversation. We adopt ISO 24617-2 (2020), the current international standard for speech act annotation proposed by the

NLP community, as a guiding framework and integrated it with a linguistically grounded subclassification of directives suited to pragmatic research. This initial focus on directives reflects their centrality in pragmatic theory, and the distinctive challenges they pose to both linguistic analysis and automatic retrieval in NLP. Inherently functional and often indirect, directives are highly context-dependent, contributing to persistently low accuracy rates in automated identification, which is a limitation that is further exacerbated by the lack of high-quality annotated corpora. By foregrounding this category, this study seeks to establish a principled and extensible methodology for future annotations. In the long term, this work forms part of a broader initiative to annotate all five of Searle's (1979) major speech act categories, building comprehensive speech act-annotated corpora that can be fully exploited in both linguistics and NLP studies.

Central to the current research is a high-quality, manually annotated corpus of directive speech acts supported by a systematically developed classification scheme. This corpus captures a wide range of naturally occurring directives and enables fine-grained empirical analysis beyond traditional introspection-based accounts. After illustrating the background and goal of the research (Section 1), we review the existing classification schemes and introduce our own operational taxonomy, which is grounded in prior empirical observations and tailored to the needs of linguistic analysis (Section 2). We then describe the process of annotating directives in the Spoken BNC2014 (Section 3) and present an overview of how directives are realized in everyday conversations, highlighting key contrasts with traditional, theory-driven classifications (Section 4). Finally, we summarize the main contributions of this study and suggest directions for future research (Section 5).

2. Speech act taxonomies in previous research and the present study

In this section, we begin by reviewing the classic classifications of speech acts proposed by Austin and Searle, whose seminal work also inspired the development of a variety of bottom-up taxonomies in the field of NLP. Drawing on insights from both traditions, we then outline the operational taxonomy for directive speech acts employed in the present study.

2.1 Austin's and Searle's classifications

Although he did not use the term "taxonomy," Austin (1962) was the first to propose a structured classification of speech acts, identifying five types of speech acts based on performative verbs. Searle (1979) criticized this framework for

conflating performative verbs with illocutionary acts and for lacking clear classification principles. He introduced a revised taxonomy grounded in felicity conditions, distinguishing five categories: *assertives*, *directives*, *commissives*, *expressives*, and *declaratives*. Owing to its systematicity, clarity, and theoretical coherence, Searle's classification has become the most influential and widely adopted framework in speech act research.

Despite being developed through introspection, Searle's taxonomy remains prevalent in corpus-based studies that examine more complex naturally occurring discourse. For example, De Felice et al. (2013) adapted Searle's categories to annotate speech acts in business emails, whereas Bell (2020) applied them directly to a corpus of tweets. However, the framework has been criticized for its rigidity and inability to account for the fuzzy multifunctional nature of real speech acts. Flöck (2016) and Weisser (2020) argue that both Austin's and Searle's models are insufficiently grounded in empirical discourse data. Researchers have often modified Searle's taxonomy to accommodate the specificities of their corpora and alternative frameworks have been proposed (see Section 2.2). Recognizing the challenges involved in applying Searle's original classification to naturally occurring discourse, many studies have narrowed their focus to specific speech act types, such as apologies (e.g. Deutschmann, 2003; Su & Wei, 2018), thanks (e.g. Funke, 2022; Jautz, 2013), and requests (e.g. Flöck, 2016; Merrison et al., 2012), and investigated their contextual realization using corpus data. These studies have yielded important empirical insights into how particular speech acts are performed in context. At the same time, they remain rooted in Searle's "armchair" taxonomy (Levinson, 2017: 206), within which certain patterns of variation observable in authentic discourse may remain underexplored.

2.2 Bottom-up taxonomies

As previously discussed, Austin's and Searle's classifications are primarily theoretical. However, annotation of speech acts in the real world must address the complexities and nuances of authentic language usage. Over the past decades, two distinct lines of research on bottom-up taxonomies have emerged: speech act annotation taxonomies dedicated to corpus linguistic research, and those primarily applied in NLP tasks.

In the field of linguistics, apart from Weisser's contributions, few scholars have developed systematic speech-act annotation frameworks suited to authentic discourse. Most existing studies adopt *ad hoc* schemes based on Searle's taxonomy, which are designed to meet the specific needs of individual projects. These frameworks are typically limited in scope, not standardized, and lack reusability. García (2004) proposed 10 subtypes of directives for her own study, while Seto (2016)

developed 69 speech act categories for the business sub-corpus of the Hong Kong Corpus of Spoken English (HKCSE).

Weisser’s work, originally intended to support computational dialog modeling, has progressively shifted toward facilitating linguistic analyses. Under the EAGLES initiative, SPAAC (Leech & Weisser, 2003) introduced a semiautomatic annotation framework using C units and 41 functional tags for task-oriented dialog. This was later extended to the development of DART (Weisser, 2019), which expanded the tagset to 132 categories across four iterations, thereby enabling highly granular and nuanced annotations. However, despite sustained efforts by a single author to develop and refine DART, its uptake has remained limited, and it has not yet become a widely adopted annotation standard for speech acts in the field. Moreover, although supported by accompanying tools (SPAACy and DART), its logic-based approach lags behind the deep-learning methods that now dominate computational annotation, further restricting its relevance and scalability.

Taxonomies developed within NLP, from the initial generic scheme DAMSL to the international standard ISO 24617-2, share the common objective of serving as effective frameworks that facilitate the automatic recognition of speech acts. A key feature of these efforts is their collaborative development, resulting in systematic annotation schemes that are continuously tested and refined using authentic language data. The DAMSL (Allen & Core, 1997) introduced a multidimensional scheme that was later simplified into the widely used SWBD-DAMSL (Jurafsky et al., 1997). To better handle multiparty discourse, MRDA (Dhillon et al., 2004) incorporates syntactic and prosodic cues and allows multilabeling. MALTUS (Popescu-Belis, 2004) further streamlines annotations using constrained tag combinations. A more theoretically grounded model, DIT++ (Bunt, 2009), introduced ten dialog dimensions and distinguished between general- and dimension-specific functions. It laid the foundation for ISO 24617-2, the current international standard (2012, revised 2020), which retains DIT++’s multidimensional structure, defines 63 communicative functions, and uses the “functional segment” as its annotation unit. Supported by the creation of DialogBank, a multilingual corpus annotated according to ISO standards offers a robust, semantically grounded, and operationally viable framework. It currently represents the most comprehensive approach to speech act annotation and is widely adopted in both computational and linguistic research.

This study considers ISO 24617-2 to be the most suitable annotation scheme because of its rigorous design and widespread acceptance within the research community. In Verdonik’s (2023: 156) comparative analysis of several schemes, the ISO 24617-2 tagset turns out to be well-documented and strictly developed. It

follows the criteria of empirical and theoretical validity, requiring every communicative function to be clearly distinguished from other functions. The tagset is exclusive, and Verdonik (2023) has no problem using it in terms of the potential ambiguity of the tags. Weisser (2015:90) criticizes ISO 24617-2 for conceptually deviating from a pragmatic perspective, favoring a semantic view of speech acts as information-state updates rather than pragmatically relevant features. However, the successful application of the ISO 24617-2 scheme to various language data, along with Verdonik's observations (2023:154) indicates that ISO 24617-2 still retains the essential elements of the original pragmatic speech act theory. These elements ensure the continued relevance and utility of the scheme for linguistic analysis.

2.3 Our operational taxonomy for directives

The annotation guidelines for directives in our study build on the ISO 24617-2 scheme but are also adapted to refine certain linguistic definitions, particularly the subtypes of directives, as will be discussed in this section. By balancing the standardization afforded by ISO 24617-2 with the flexibility of linguistically motivated refinements, this approach supports consistent annotation across future studies, while preserving the adaptability necessary for detailed pragmatic analysis. It also facilitates integration with computational retrieval methods developed for NLP.

Directives are defined here as utterances in which the speaker intends the hearer to perform an action, regardless of whether the intention is sincere or whether the action is actually carried out. This definition includes indirect, humorous, and non-literal uses, thereby capturing a wide functional range. The annotation follows the two guiding principles proposed by Bunt (2009): (1) adopting the addressee's perspective and (2) prioritizing communicative function over form. To determine this function, annotators rely on linguistic and nonlinguistic indicators, and next-turn responses (Flöck, 2016).

As demonstrated in Figure 1, at the highest level of classification ISO 24617-2 distinguishes between two main general-purpose communicative functions: information transfer (e.g. providing or requesting information) and action discussion (e.g. initiating or negotiating actions). Directives belong to the latter category as they aim to elicit concrete actions from hearers. While linguistically oriented frameworks, such as Searle's taxonomy and SWBD-DAMSL, tend to classify all questions as directives, this study adopts a more fine-grained approach. Only action-initiating questions (e.g. *can you pass the salt?*) were treated as directives, whereas pure information requests (e.g. *how's the weather?*) are excluded.

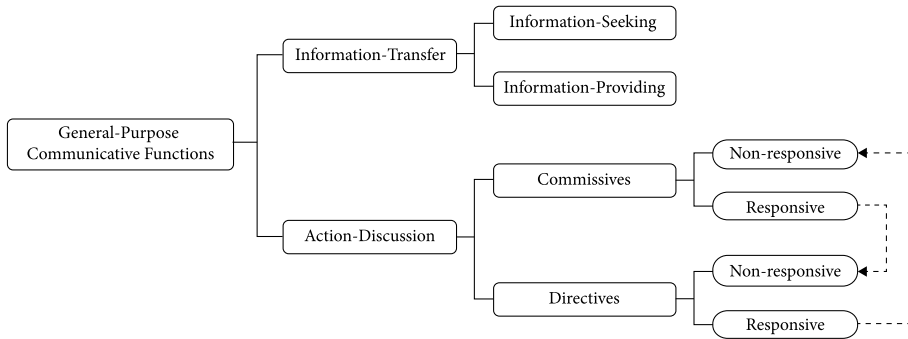


Figure 1. ISO 24617-2 classification of general-purpose communicative functions in conversation

Directives are further distinguished from commissives, where the speaker commits to a future action, in contrast to directives that target the hearer's actions. Although this distinction has long been recognized in classical speech act theory (Austin, 1962; Searle, 1969), such introspection-based classifications do not account for *responsive* acts that are prevalent in natural conversations. ISO 24617-2 addresses this by distinguishing between *non-responsive* and *responsive* variants of both directives and commissives. The speaker initiates non-responsive acts independently, as seen in Examples (1a) and (2a), whereas responsive acts arise in a direct reaction to a preceding speech act, as illustrated in Examples (1b) and (2b).¹

- (1) a. Would you like some coffee? (commissive — offer)
b. Yes, please. (response to offer — responsive directive)
- (2) a. Can I have a bit of sugar and a splash of milk, please? (directive — request)
b. Sure. Here you go. (response to request — responsive commissive)

This study focused exclusively on non-responsive directives that initiate new dialogic moves and influence the forward trajectory of interactions. In the DAMSL framework (Core & Allen, 1997), such acts fall under Forward Communicative Functions, a classification that is maintained in many subsequent annotation schemes. In contrast, responsive directives typically take formulaic forms such as *yes*, *sure*, or *okay*, which exhibit limited structural or functional variability. These formulaic expressions offer limited variability and communicative complexity, making them less relevant in linguistic studies.

1. For expository purposes, unattributed examples are constructed by the author. Attested corpus examples are marked with citations, indicating the corpus file, utterance number and sub-utterance number.

While ISO 24617-2 provides a high-level classification, further subcategorization of directives is essential for a detailed linguistic analysis. However, there is no consensus on the number of directive subtypes that should be distinguished. For instance, Searle and Vanderveken (1985) listed directive verbs without proposing a systematic categorization. Bach and Harnish (1979: 47–49), adopting a reason-based view of directives, identified six subtypes: requestives, questions, requirements, prohibitives, advisories, and permissives. Alston (2000) differentiated between institutional and non-institutional directives based on the nature of obligation, while Hindelang (1978) distinguished binding from non-binding directives, further subdividing the latter according to speaker or hearer preferences.

Although these theories provide valuable insights, corpus-based analyses require a more operational and empirically grounded taxonomy. In this study, we adopted a data-driven iterative categorization process that emerged during annotation. Our goal was to capture the functional diversity of directives, as realized in spontaneous conversations.

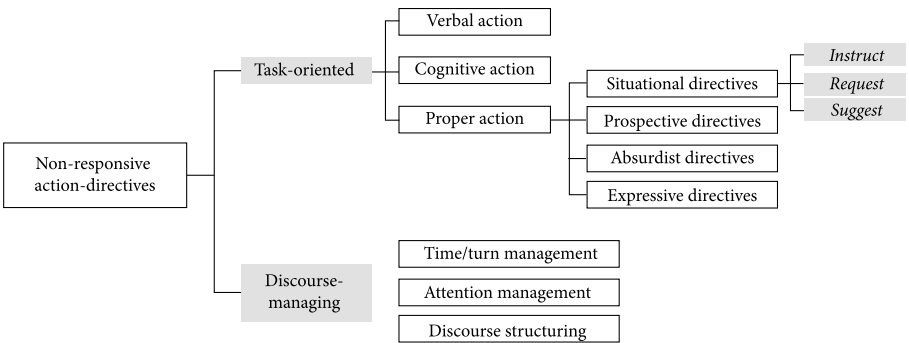


Figure 2. A data-driven subcategorization of directive speech acts

First, we draw a major distinction between task-oriented and discourse-managing directives (see Figure 2). Task-oriented directives are most commonly examined in speech act theory and are aimed at eliciting physical, verbal, or cognitive actions. In contrast, discourse-managing directives regulate interactions rather than task performance. They tend to be more conventionalized and exhibit less variation.

In task-oriented directives, further distinctions are made based on the nature of the actions involved. These distinctions include directives of verbal actions (e.g. *pardon, tell me more*), cognitive actions (e.g. *guess, imagine if*), and other proper actions with a broader spectrum of verbs. Within the category of proper actions, additional subtypes are recognized, including situational, prospective, absurdist,

and expressive directives. Notably, situational directives are central to traditional discussions on speech acts, whereas the other three subtypes occupy more peripheral positions in the directive family. Situational directives focus on immediate, context-dependent actions that often require a prompt response. For instance, *can you pass me the salt?* exemplifies this type as it necessitates immediate action in the current situation. Situational directives can be further delineated according to the intended beneficiaries of the action (ISO 24617-2, 2020: 49; see also Couper-Kuhlen, 2014: 628). A suggestion such as *you can stay there if you want* assumes that the proposed action is in the addressee's interest and presents it as an option rather than a mandate. A request, by contrast, typically benefits the speaker and requires the hearer's consent to act (e.g. *could you pass me a plate, please?*). Finally, an instruction represents a prototypical directive where the speaker assumes the hearer's capacity and willingness to act, as illustrated by *go right round until you get to just above that*. Although these subtypes are analytically distinct, their boundaries are often contextually fluid and nuanced. Prospective directives are less immediate and frequently adopt a suggestive or evaluative tone (e.g. *you could rent the room out to cover the mortgage*). Absurdist directives are characterized by humor, exaggeration, or deliberate absurdity, and are typically not intended to be interpreted as serious (e.g. *we should throw a party at the bottom of the ocean*). Lastly, expressive directives usually take the form of an imperative to communicate politeness, well-wishes, or emotional support. Examples include expressions like *have fun*, *take care*, or *enjoy yourself*.

Discourse-managing directives are geared toward shaping the conversational process. They operate at the turn-taking, attention management, and discourse organization levels. Utterances such as *hang on*, *wait*, or *hold up* regulate temporal sequencing and turn allocation; expressions like *listen*, *look*, or *mind you* guide the hearer's attention; while formulations such as *let's move on to...*, *we'll come to this later*, or *let's wrap this up* mark transitions, initiate topic shifts, or signal closure. Although often realized as brief and conventionalized imperatives, these directives play a crucial role in maintaining interactional coherence.

3. Corpus design and data annotation

This section details the corpus design and data annotation procedures. Specifically, it introduces the source corpus, the dataset sampling strategy, and the further segmentation of basic units, before outlining the annotation process and the assessment of inter-annotator agreement.

3.1 The source corpus: Spoken BNC2014

This study draws on the Spoken British National Corpus 2014 (Spoken BNC2014) to systematically extract and analyze instances of directive speech acts. The Spoken BNC2014 is an 11.5-million-word corpus of orthographically transcribed spontaneous conversations involving L1 speakers of British English from across the UK, recorded between 2012 and 2016 (Love et al., 2017). Designed to reflect informal, naturally occurring dialog, the Spoken BNC2014 primarily comprises interactions that are neither text-governed nor task-oriented, making it a valuable resource for analyzing speech acts in everyday conversation.

The present study relies on the transcribed textual component of Spoken BNC2014 rather than the audio recordings. This decision is motivated by two considerations. First, the transcribed texts are directly accessible for large-scale corpus processing. Second, existing studies on the automatic retrieval of speech acts predominantly operate on textual representations, which allows the present methodology to remain compatible with, and reusable for other comparable corpora. It is nevertheless acknowledged that reliance on transcription alone inevitably entails the loss of certain phonetic and prosodic cues that may contribute to the interpretation of directive force.

In the corpus, each transcribed text is encoded in XML format. The basic unit of transcription is the utterance, defined as a single speaker's uninterrupted stretch of speech, enclosed within <u> tags. These tags include metadata such as turn number (n="?"), speaker ID (who="?"), annotation confidence (whoConfidence="?"), and overlapping speech (trans="overlap"), where applicable. In addition to verbal content, the transcriptions incorporate basic paralinguistic features, including pauses (<pause dur="?"/>), background sounds (<vocal desc="?"/>), truncated speech (<trunc>), and unclear segments (<unclear>). This detailed structure facilitates reliable identification and contextual analysis of directive speech acts in authentic conversational settings. The corpus is enriched with detailed speaker metadata including gender, age, education level, and social class. This sociolinguistic information makes it particularly well-suited for linguistic analyses.

3.2 Data sampling and sentence segmentation

The Spoken BNC2014 comprises 1,251 transcribed recordings, totaling 987 hours of spontaneous British English and over 11 million tokens (Love et al., 2017). Although the corpus offers a rich and diverse dataset for spoken language research, analyzing it in its entirety is neither feasible nor methodologically appropriate for the current study. This is especially true in the case of directive

speech acts, which are characterized by a significant form-function mismatch. As directives often lack consistent surface markers and are heavily dependent on contextual interpretation, they cannot be reliably identified through simple keyword searches. Instead, accurate retrieval requires close manual reading, which is a labor-intensive but necessary step to ensure the construction of a high-quality, reusable dataset for future linguistic research.

The present study adopted a stratified sampling strategy, selecting a subset of 61 recordings with a total duration of 40 hours, 14 minutes, and 44 seconds (see Table 1). Sampling was based on the speaker relationship types categorized in the corpus metadata. Of the seven predefined categories, five were retained for the analysis: (1) close family, partners, and very close friends; (2) friends and extended family; (3) acquaintances; (4) colleagues; and (5) strangers. Categories with sparse or ambiguous data, such as “friends and one stranger” and “uncategorized” were excluded. In selecting the recordings, priority was given to maximizing demographic diversity across speakers. This stratified subset ensures broad coverage of interactional contexts, speaker roles, and directive realizations, while remaining manageable for in-depth, qualitative analysis.

Table 1. An overview of the sampled data from the Spoken BNC2014

Relationship between speakers	Number of recordings	Number of samplings	Length of samplings (hh:mm:ss)
Close family, partners, very close friends	911	21	14:29:08
Friends, wider family circle	263	14	11:10:40
Acquaintances	29	9	05:22:57
Colleagues	35	10	05:19:57
Friends and one stranger	1	–	–
Strangers	7	7	03:52:02
Uncategorized (NA)	5	–	–
SUM	1,251	61	40:14:44

Before annotating the directive speech acts, it is necessary to establish an appropriate unit of analysis. The default unit in the Spoken BNC2014, namely, the speaker utterance, is often too coarse, especially when the turns are long and contain multiple unrelated clauses. Actually, the analytical unit should not be so large that multiple speech acts are conflated within a single unit. This could introduce noise and reduces the reliability of speech act annotations. To improve precision, this study adopts sub-utterances as the basic annotation unit, targeting syntactically and pragmatically independent segments within speaker turns. We

implemented automated segmentation using the GPT-4o-mini prompted by the few-shot strategy. This ensured consistency, efficiency, and scalability across the datasets. A full example of the prompted and segmented outputs is provided in Appendices A and B.

To evaluate whether this segmentation strategy yields analytically stable units, we compared the distribution of token counts across different units within each file. Specifically, we compared three units: entire utterances, units segmented by the <pause dur="?"> tag and sub-utterances segmented using GPT-4o-mini. Importantly, this comparison is not intended to privilege uniform unit length as a linguistic ideal, nor to deny the natural variability of utterances in spontaneous conversation. Rather, token-count variability is used here as a diagnostic indicator to identify potential under-segmentation within the same discourse context. Excessive variability may signal that a segmentation method inconsistently groups multiple acts into a single unit.

Table 2. An overview of three units based on different segmentation approaches

Relationship between speakers	Token count	Number of utterances	Number of segments by pause	Number of segments by GPT-4o-mini
Close family, partners, very close friends	161,951	18,173	20,516	26,197
Friends, wider family circle	120,238	12,409	15,123	18,796
Acquaintances	69,268	6,972	8,484	10,002
Colleagues	76,080	7,026	8,347	10,739
Strangers	42,487	4,249	4,547	6,372
SUM	470,024	48,829	57,017	72,106

The boxplot in Figure 3 reveals notable variations in token counts across these methods. The coefficient of variation (CV), defined as the ratio of the standard deviation to the mean, was calculated for the three variables, T_U, T_S1, and T_S2, to assess their relative variability.

Across all five speaker relationship categories, GPT-4o-mini produced the most segments with the lowest token count variability ($CV=0.158$), outperforming the pause-based ($CV=0.284$) and utterance-level units ($CV=0.358$). This pattern suggests that the LLM-based segmentation yields more stable analytical units within comparable discourse conditions. Therefore, it was selected as the unit of analysis for identifying directive speech acts in spontaneous conversation.²

2. During subsequent manual annotation, each annotator examined and annotated sub-utterances individually. Inappropriate or ill-formed segmentation units would be expected to make annotation difficult or infeasible; however, no such issues were reported.

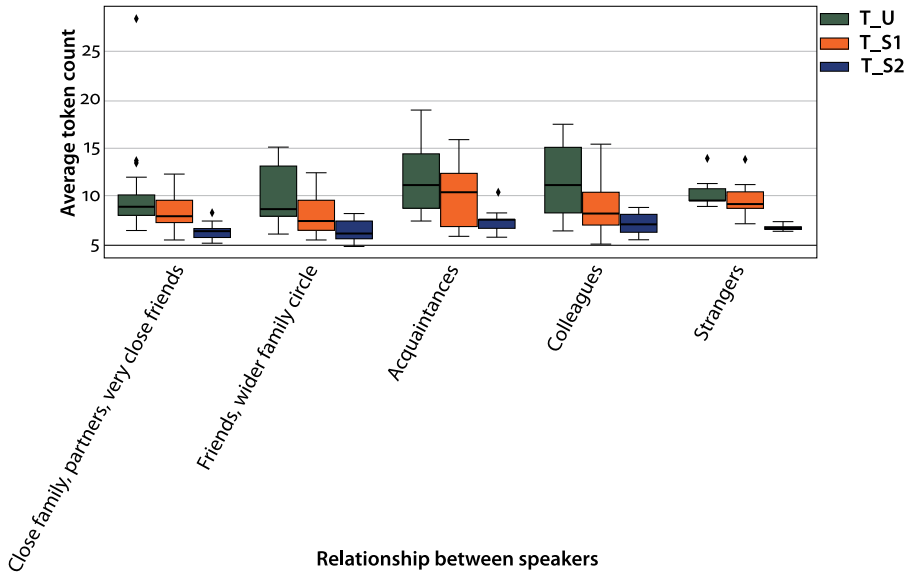


Figure 3. A boxplot for average token per file across three different units

3.3 Annotation process

As outlined in Section 3.2, we adopted sub-utterances as the annotation unit. A central concern in the annotation process is to define what constitutes a complete speech act. Following Blum-Kulka and Olshtain (1984: 200), a directive sequence may comprise three components: an address term, a head act, and an adjunct. The head act represents the nucleus of the directive and is independently sufficient to realize the act. Address terms index the speaker-hearer relationship, whereas adjuncts serve to strengthen or support the act, often by providing reasons (such as the cause of the request), consequences (such as potential outcomes if the act is not performed), or positive reinforcement. For instance, consider Example (3):

- (3) Babe, can you shut the door? There's a draft in here.

The first element *babe* functions as the address term. The head act, *can you shut the door?* directly refers to the intended action. The adjunct, *There's a draft in here* provides the reason for the request and justifies the desired action. In more implicit realizations, however, the adjunct itself may serve as the head act (e.g. *It's a bit cold here*).

To ensure analytical consistency, each directive must contain exactly one head act, which is labeled {head}. Address terms were not annotated, as person names and other identifying information are anonymized in the Spoken BNC2014.

Given the prevalence of interruptions and incomplete turns in spoken discourse, we introduced the label {incomplete} to capture fragmented head acts that cannot independently perform the intended function. The label {mixed} is assigned when a sub-utterance contains both a head act and an adjunct. Adjuncts are uniformly annotated as {adjunct} irrespective of whether they appear in complete or fragmented form. Consequently, each sub-utterance was assigned one of the four mutually exclusive tags: head act, incomplete, mixed, or adjunct.

When a directive spans multiple units, all contributing sub-utterances are annotated to ensure a comprehensive representation. For example, in file S6J2, the directive *find a body that you're comfortable in* unfolds across several units: *you just have to* {incomplete}, *it's about finding a body that you're comfortable in* {head}, and *that's you know/healthy and that you can be satisfied with* {adjunct}. Furthermore, in cases where multiple fragments representing different speech acts occur within a single unit, the entire sub-utterance is annotated as a directive if any portion performs the directive function. This approach is particularly relevant for discourse-managing directives such as *wait*, which frequently occur in utterance-initial positions and serve to regulate conversational flow while co-occurring with non-directive elements.

3.4 Annotators and inter-annotator agreement

In this study, each sub-utterance was evaluated by the annotators solely to determine whether it constituted a directive speech act, resulting in an exhaustive binary annotation of all sub-utterances. Despite its binary nature, annotating such pragmatic phenomena presents a significant challenge, particularly because of the inherently subjective nature of interpretations. This is particularly true for implicit directives, which heavily depend on contextual inferences and often lack overt linguistic forms. We employed a double-annotation protocol with cross-checking to ensure reliability and consistency. Each file was independently annotated by chief and assistant annotators. Following the initial annotation, the two annotators engaged in structured discussions to reconcile discrepancies and produce a final version. To assess inter-rater agreement, Gwet's AC1³ was used as the reliability metric. This measure is especially appropriate for highly imbalanced datasets, such as the current one, in which directive speech acts account for less than 5% of all utterances in natural conversations.

3. Gwet's AC1 was used to assess inter-annotator agreement because the dataset is highly imbalanced, with directive speech acts accounting for only about 3% of all sub-utterances.

Table 3. Inter-rater reliability between chief annotator and 12 assistant annotators

Assistant annotator	Annotated length (hh:mm:ss)	Inter-rater reliability (Gwet's AC1)
AN01	07:07:21	95.93%
AN02	05:05:02	91.85%
AN03	04:09:45	92.90%
AN04	03:07:41	94.74%
AN05	03:06:11	93.97%
AN06	03:02:47	90.46%
AN07	02:59:24	93.58%
AN08	02:51:13	91.65%
AN09	02:48:35	87.11%
AN10	02:41:47	90.33%
AN11	02:16:40	90.83%
AN12	00:58:18	90.73%

As shown in Table 3, the inter-rater reliability scores were consistently high across all 12 assistant annotators, ranging from 87.11% to 95.93%. These results indicate a high level of annotation consistency. The annotated corpus (dassBNC2014) is openly available on GitHub: (<https://github.com/buyingo/dassBNC2014>).

4. A landscape of directives in everyday conversation

Based on the annotated data from dassBNC2014, this section outlines the distribution of directive speech acts in everyday conversation. The analysis suggests that directives do not form a homogeneous class but rather a heterogeneous set of interactional practices.

4.1 Overview of dassBNC2014

Our annotated dataset, referred to as the Directive Annotated Subset of Spoken BNC2014 (dassBNC2014), comprises 40 h, 14 min, and 44 s of recorded conversations containing 72,106 sub-utterances. Of these, 2,333 were manually identified as directives, representing approximately 3.24% of all the sub-utterances in naturally occurring conversations. For brevity, the distribution of annotated directives in the dassBNC2014 is summarized in Table 4 by speaker relationship type.

Table 4. Distribution of annotated sub-utterances as directives in dassBNC2014

Relationships	Length (hh:mm:ss)	Sub-utterances	Directives
Close family, partners, very close friends	14:29:08	26,197	957 (3.65%)
Friends, wider family circle	11:10:40	18,796	699 (3.72%)
Acquaintances	05:22:57	10,002	239 (2.39%)
Colleagues	05:19:57	10,739	361 (3.36%)
Strangers	03:52:02	6,372	77 (1.21%)
SUM	40:14:44	72,106	2,333 (3.24%)

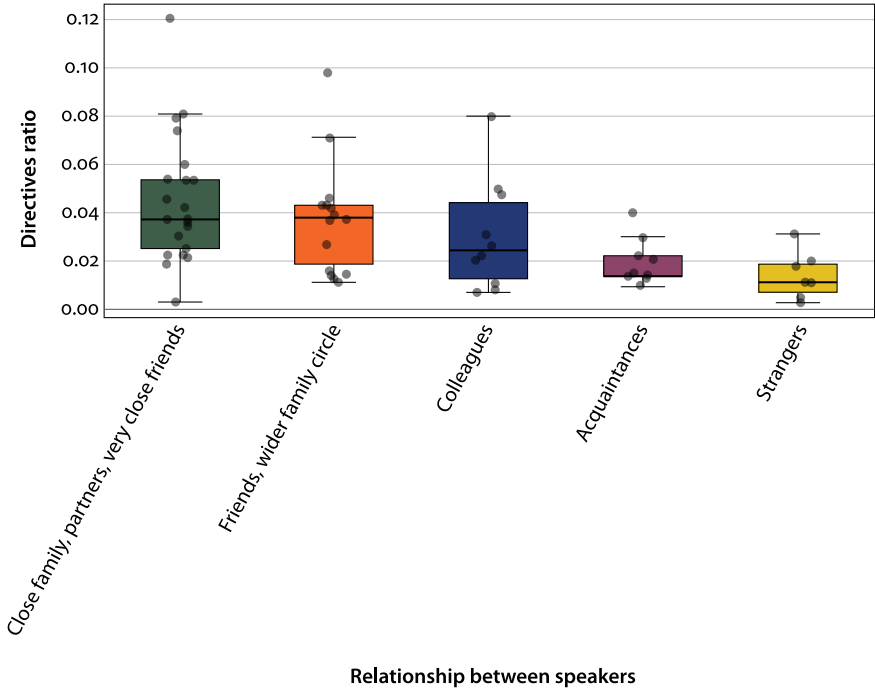


Figure 4. Directive ratios across different relationship groups

Analysis of these distributions revealed marked differences in the prevalence of directives across relationship types. The highest directive ratios occurred in interactions between friends and members of the wider family circle (3.72%), followed by those between close family members, partners, and very close friends (3.65%). In contrast, conversations between acquaintances (2.39%) and strangers (1.21%) showed substantially lower directive usage, whereas interactions between colleagues (3.36%) fell between these extremes. A one-way ANOVA confirmed

significant differences in directive ratios across relationship categories ($F=3.89$, $p=0.007$, $df_1=4$, $df_2=56$). A subsequent Tukey's HSD post hoc test identified significant pairwise differences between the most intimate group (close family, partners, and very close friends) and two other groups: (1) acquaintances ($p_{adj}=0.0428$) and (2) strangers ($p_{adj}=0.0174$).

Table 5 presents the distribution across the subtypes⁴ of annotated directive speech acts. It is presented as an exploratory attempt to map the range of directive practices observed in natural conversation, rather than as a definitive taxonomy. These subtypes were quantitatively analyzed based on both the number of sub-utterances and average tokens per speech act, offering valuable insights into their roles in natural conversation.

Table 5. Distribution of directive speech acts across subtypes

Categories				Number of sub-utterances	Average token per speech act
Task-oriented	Verbal action			75 (3.21%)	6.61
	Cognitive action			94 (4.03%)	9.43
	Proper action	Situational	Request	396 (16.97%)	8.71
			Suggest	482 (20.66%)	9.18
		Prospective	Instruct	217 (9.30%)	6.78
				822 (35.23%)	15.50
	Absurdist			69 (2.96%)	9.64
	Expressive			14 (0.60%)	5.77
Discourse-managing	Time/turn management			53 (2.27%)	6.00
	Attention management			78 (3.34%)	6.46
	Discourse structuring			33 (1.41%)	6.53

As demonstrated in the table above, task-oriented directives constitute the majority of the annotated directives in dassBNC2014. Of the 2,333 sub-utterances annotated as directive speech acts, 2,169 (92.97%) are task-oriented directives designed to facilitate external tasks or activities, thereby advancing goals beyond immediate conversational exchange. The remaining 164 sub-utterances (7.03%) are classified

4. The subtype classification was conducted by one of the authors based on the previously identified directive speech acts. It was developed through an iterative process and was intended as an analytical categorization rather than an independent annotation task. Accordingly, no inter-rater agreement was calculated for the subtype classification.

as discourse-managing directives, which regulate the flow, turn-taking, and structure of the interaction without eliciting actions external to the dialog.

Within the task-oriented categories, three subtypes are distinguished according to the nature of the action being directed. Directives of verbal action account for 3.21% of the sub-utterances, with an average of 6.61 tokens per speech act. Directives of cognitive action represent 4.03% of the sub-utterances and have an average token length of 9.43. The largest proportion of task-oriented directives (85.73 %) fell under proper action directives, which can be subdivided into situational, prospective, absurdist, and expressive directives. Situational directives, including requests (16.97%), suggestions (20.66%), and instructions (9.30%) form a significant portion of the data. These directives are context dependent and typically call for immediate responses or actions in the short term. Requests and suggestions average 8.71 and 9.18 tokens, respectively, while instructions have an average length of 6.78 tokens. Prospective directives (35.23%), which focus on future-oriented actions or plans, exhibit the longest average token length (15.50), reflecting their evaluative and reflective nature. Other subtypes of task-oriented directives, such as absurdist directives (2.96%) and expressive directives (0.60%), while less frequent, convey humor or emotional support, with average token lengths of 9.64 and 5.77 tokens, respectively.

Discourse-managing directives, which regulate the flow of conversation, include attention management (3.34%), time/turn management (2.27%), and discourse-structuring directives (1.41%). These directives tend to be brief, averaging token lengths ranging from 6.00 to 6.53.

4.2 The heterogeneous nature of directives

Corpus evidence from *dassBNC2014* shows a clear “theory-corpus” misalignment. Classic accounts concentrate on task-oriented directives for proper actions, especially situational directives, which are context-bound moves requiring immediate uptake (e.g. requests, suggestions, and instructions), and prospective advice-like proposals, which are longer, evaluative, and future-oriented with attenuated obligation. In our data, these account for 46.94% and 35.23%, respectively, leaving 17.83% outside these widely discussed categories.

The remaining 17.83% comprises several less-frequently discussed types. Within task-oriented directives, directives for proper-action also include absurdist directives (2.96%), which maintain directive form but nullify uptake through hyperbole or implausibility, and expressive directives (0.60%), where imperatives convey goodwill (e.g. *have fun*, *take care*) with minimal deontic force. In addition, within the task-oriented domain, verbal action directives (3.21%) elicit explicit speech responses (e.g. *could you repeat that?*) and cognitive-action directives

(4.03%) prompt mental operations such as considering or evaluating (e.g. *consider this*). Finally, discourse-managing directives (7.03%) regulate participation, attention, and sequencing (e.g. *hang on, listen, let's move on*). Collectively, these categories demonstrate that directive practices extend well beyond the canonical focus of existing theories, underscoring the heterogeneous nature of directives. The following section begins with a discussion of the subtypes of task-oriented directives and then turns to discourse-managing directives.

4.2.1 Task-oriented directives

First, a macrolevel split of directive speech acts may be drawn according to their *interactional locus*, thus distinguishing between task-oriented and discourse-managing directives. The interactional locus concerns whether the illocutionary force of a directive is directed outward toward an action in the extra-linguistic world (task execution) or inward toward the organization and management of the conversation itself. Externally oriented directives typically involve actions such as moving an object, operating a device, or arranging a schedule and thus fall within the task-oriented category.

Among task-oriented directives, the subtype most extensively examined in both speech act theory and empirical research is *situational directives*, under the broader category of directives for proper action. Directives for proper action are those in which the target is an actual, concrete action as opposed to speaking (verbal action) or engaging in mental processes (cognitive action). Situational directives focus on immediate, context-dependent actions that often require a prompt response. For instance, *can you pass me the salt?* exemplifies this type as it necessitates immediate action in the current situation. In theoretical treatments, particularly those grounded in introspection, scholars have often and almost tacitly concentrated on this subtype as reflected in Searle's (1979) inventory of directive illocutionary verbs. Correspondingly, empirical studies, especially corpus-based ones, tend to focus on this subtype. In the literature on directive speech acts, requests have attracted the most attention, with substantial contributions from Flores-Salgado and Castineira-Benitez (2018), Flöck (2016), Gagné (2018), Merrison et al. (2012), and Urbanik (2017), among others.

Although Searle (1979: 11–12) criticized Austin's taxonomy for conflating performative verbs with illocutionary acts, he also relied on an unsystematic inventory of directive verbs, offering little internal differentiation among the five major speech act categories. Such a verb-based inventory conflates lexical semantics with illocutionary force, and when applied to naturally occurring conversational data, it may cause data sparsity. From a more practical and corpus-oriented perspective, we adopt a beneficiary-based distinction for situational directives (see Couper-Kuhlen, 2014: 628; ISO 24617-2, 2020: 49). A suggestion, such as that in

Example (4), assumes that the proposed action is in the addressee’s interest or promotes a common interest shared by both speaker and hearer, and presents it as an option rather than a mandate. By contrast, a request typically benefits the speaker and requires the hearer’s consent to act, as in Example (5). Finally, instruction represents a prototypical directive in which the speaker assumes the hearer’s capacity and willingness to act, as illustrated in Example (6).

- (4) (suggest) erm we can go this afternoon if you like (SUGJ, $u = 438, s = 1$) I’ve got nothing else to do once I’ve finished off the schedules (SUGJ, $u = 438, s = 2$)
- (5) (request) can you point that heater thing toward your window (SG3P, $u = 162, s = 7$) because I can’t see in my mirror? (SG3P, $u = 162, s = 8$)
- (6) (instruct) oh look there’s where I work (S2LC, $u = 4, s = 1$)

This three-fold subclassification offers analytic clarity while remaining sufficiently broad to ensure robust empirical coverage. This is unlike a finer-grained verb inventory that introduces semantically specific subtypes, such as *threaten* or *plea*, which may be intuitively appealing but are rarely attested in sufficient numbers.

Table 6. Structural characteristics of situational directive subtypes

	Average token per speech act	Average sub-utterance per speech act	Prototypical pattern
Instruct	6.78	1.11	{head}
Request	8.71	1.27	{adjunct} _{op} {head}{adjunct} _{op}
Suggest	9.18	1.25	

Even for situational directives, the three subtypes exhibit distinct characteristics. Regarding average tokens per speech act, instructs are the shortest, with an average length of 6.78 tokens, followed by requests at 8.71 tokens, and suggests at 9.18 tokens (see Table 6). These differences in length stem primarily from the structural variations. Instructs are typically realized by a single head act. Of the 217 sub-utterances annotated as instruct, 195 were identified as heads or mixed units, corresponding to 195 distinct directive acts. The remaining sub-utterances served only as adjuncts or incomplete fragments. This yields an average of 1.11 sub-utterances per instruct. By contrast, requests are somewhat longer and structurally complex. A total of 396 sub-utterances were annotated as requests, comprising 312 complete directive acts, equating to an average of 1.27 sub-utterances per request. Suggests are the longest subtype. Of the 482 sub-utterances annotated as suggest, 385 were identified as head or mixed units, yielding an average of 1.25 sub-utterances per act. Taken together, requests and suggests tend to be more

elaborated than instructs. While instructs are usually realized by a single-head act, requests and suggests often incorporate additional supporting fragments alongside the head. These adjuncts are optional rather than obligatory. They may precede the head, follow it, or occur in both positions. This optional expansion is reflected in their prototypical structural pattern $\{\text{adjunct}\}_{\text{op}}\{\text{head}\}\{\text{adjunct}\}_{\text{op}}$, which accounts for the comparatively greater average length of requests and suggests.

Unlike situational directives, *prospective directives* are less immediate and frequently adopt a suggestive or evaluative tone, encouraging the addressee to reflect on long-term possibilities or arrangements without necessarily implying a strong desire for action on the part of the hearer. In other words, they differ in temporal orientation and deontic strength. Prospective directives resemble advice, as described in speech act theory, and represent a borderline case between directives and assertives (see Ruytenbeek, 2017:26). As noted by Searle and Vanderveken (1985:202), advising can be either directive or assertive about the state of affairs represented by propositional content. Compared with situational suggestions, prospective advice generally exhibits a minimal degree of speaker intention for the addressee to act, focusing instead on evaluating the potential action; see Example (7).

- (7) exactly but if you're going to like if you're going to the UK (S2C9, $u=461$, $s=1$)
 you kind of need to go like June like July or August September (S2C9, $u=461$,
 $s=2$) cos you want it you want to take a stab at it being warm don't you?
 (S2C9, $u=461$, $s=3$)

This divergence is reflected in the linguistic forms. Prospective directives are also considerably longer than their situational counterparts, averaging 15.50 tokens. Their elaborated structure typically unfolds across multiple annotation units. Of the 822 sub-utterances identified as prospective directives, 523 were coded as heads or mixed units, corresponding to 523 distinct speech acts. On average, therefore, each prospective directive spans 1.57 sub-utterances, compared to just over one unit for situational directives. This extended span indicates their discursive function; rather than realizing a directive in a concise head act, prospective directives often combine evaluative commentary, explanatory justification, and elaboration. In many cases, it is difficult to determine whether the evaluative or directive component carries greater weight. This ambiguity illustrates how prospective directives mediate the relationship between stating one's opinions and inviting a course of action.

Another unique category of directives for proper action is *absurdist directives*, which are characterized by humor, exaggeration, or absurdity. They are not intended to prompt actual action but instead convey humor, playfulness, or

indirect emotional attitudes through exaggerated or non-sensical expressions, as in Example (8):

- (8) I'm throw can you throw that California book out the window? <vocal desc="laugh" /> (SQ9R, $u=287$, $s=1$)

Their temporal orientation remains indeterminate and their deontic strength is close to zero, as the speaker does not seek to impose any real-world obligation on the hearer. Rather, they exploit the directive format as a rhetorical resource to create shared amusement, mitigate tension, and indirectly express affective stances. Absurdist directives closely resemble situational directives in terms of form. They average 9.64 tokens per speech act. In total, 56 absurdist directives were identified and distributed across 69 sub-utterances, corresponding to an average of 1.23 sub-utterances per speech act.

A minor yet noteworthy category within the directive domain is what we term *expressive directives* (0.60%). They are typically imperative forms, delivered not to elicit action but to convey emotion, politeness, goodwill, or wishes (see Examples (9) and (10)). On average, expressive directives contain 5.77 tokens per speech act and span 1.08 sub-utterances, reflecting their concise and formulaic nature.

- (9) okey-doke my love have fun be good alright be good bye (S35K, $u=24$, $s=4$)
(10) please don't rain (SFNQ, $u=576$, $s=1$)

While these expressions retain their imperative structure, their pragmatic functions diverge sharply from canonical directive force. Previous research (e.g. Han, 2000; Jary & Kissine, 2014; Wilson & Sperber, 1988) has categorized such forms as good-wish imperatives, whose defining property lies in the addressee's lack of control over the desired state i.e. whether one enjoys oneself, remains safe, or rains. Therefore, the deontic obligation typically associated with imperatives is either attenuated or absent.

This can be understood in terms of pragmaticalization (Diewald, 2011), whereby pragmatic meanings become increasingly encoded and conventionalized. Through recurrent use in specific interactional contexts such as leave-taking or small talk, the original action-oriented meaning fades, while subjective and intersubjective meanings are foregrounded and conventionalized. This trajectory aligns with the processes of (inter)subjectification (Traugott & Trousdale, 2013), whereby speakers recast directive morphology as a vehicle for expressing a personal or relational stance. At the constructional level, such changes result in constructionalization, stabilizing a form-function pairing in which the imperative form is systematically associated with the wish expression.

Verbal- and *cognitive-action directives* are marked as peripheries. The former, directive speech act of verbal action, directs hearers to engage in speech-related tasks. Utterances such as *can you tell me...?*, *please explain this...*, or *could you repeat that?* explicitly directed the addressee to perform verbal acts, reflecting the speaker's intention to elicit a response. In our corpus, verbal-action directives average 6.61 tokens per speech act and span 1.04 sub-utterances; see Example (11).

(11) say it again? (STWC, $u=1721$, $s=1$)

The second subtype, directives of cognitive action, targets subjective cognitive engagement, encouraging the hearer to think, evaluate, or imagine, rather than undertake immediate physical actions. Common realizations include *consider this...*, *think about that...*, or *imagine if...* (see Example (12)). These directives average 9.43 tokens per speech act and 1.03 sub-utterances.

(12) imagine if the curtains were shut<pause dur="short" /> (SY8A, $u=138$, $s=1$)

These two subtypes are considered peripheral cases of task-oriented directives because, in everyday conversation, verbal and cognitive actions are typically assumed to occur by default. Speakers generally expect listeners to process their contributions and respond naturally without requiring explicit instructions. This implicit nature is considered as one of the reasons for distinguishing between information directives (e.g. interrogatives) and genuine action directives. Only when a speaker explicitly emphasizes the need for the addressee to perform verbal or cognitive actions using overt expressions, such as those listed above, do these actions attain a more prominent communicative role. In such cases, they are elevated to the status of action directives, which are comparable to those that target tangible or physical actions.

4.2.2 Discourse-managing directives

Traditional accounts of speech acts (e.g. Austin, 1962; Bach & Harnish, 1979; Searle, 1979) have concentrated almost exclusively on task-oriented acts, with little consideration given to those explicitly serving discourse-managing functions. However, researchers have realized that such discourse-managing acts play a significant role in annotating speech acts during naturally occurring conversations. Bottom-up speech act taxonomies have increasingly incorporated systematic treatments of this category. For instance, the early DAMSL scheme (Core & Allen, 1997) recognized Communication Management as a distinct component within its information level; the ISO 24617-2 standard (2020) offers a fine-grained differentiation of dimension-specific communicative functions, including time management, turn management, and discourse structuring; and Weisser (2003, 2020) designates dialog control as a discrete speech act type.

Defined in terms of interactional locus, discourse-managing directives are those whose illocutionary effects are oriented toward the organization of talk itself. Their functions are to regulate participation, allocate attention, and shape an evolving discourse structure. In this sense, they constitute a class of meta-communicative acts that operate within the conversation, rather than projecting actions beyond it.

They can be further categorized into three subtypes: time and turn management, attention management, and discourse-structuring directives. Directives for time and turn management regulate time and turn allocations. Examples include expressions such as *hang on*, *wait*, *hold up*, and *shut up*, which momentarily pause or redirect the flow of interaction to achieve smoother conversational coordination. Directives for attention management guide hearers' focus toward specific discourse elements. These directives emphasize key points and align the listener's focus with the speaker's communication priority. Examples include *listen*, *look*, *mind you*, and *you see*, all of which serve to emphasize the following content the speaker intends to highlight. Discourse structuring directives shape the overall organization and coherence of the conversation. These directives guide conversations by introducing new topics, transitioning between points, or signaling closures. Expressions such as *let's move on to...*, *we'll come to this later*, and *let's wrap this up* help speakers manage the macro structure of the interaction, ensuring that it remains logical and cohesive. As evidenced in our dataset, discourse-managing directives are notably concise and are often expressed as direct imperatives consisting of only a few words.

Despite these functional differences, the three subtypes share notable similarities. All are expressed with brevity, typically in the form of a minimal imperative: their average lengths ranged between 6.00 and 6.53 tokens, and almost all are realized as single sub-utterances. This compactness reflects what conversation analysts describe as sequence-organizing practices (Heritage, 2013; Schegloff, 2007), where minimal tokens perform important work in sustaining interaction.

4.3 Summary

Traditional speech act research, grounded primarily in introspection (Austin, 1962; Searle, 1969, 1979; Searle & Vanderveken, 1985), tended to treat directives as a homogeneous class, typically represented by prototypical commands, requests, etc. However, our systematic corpus-based investigation of naturally occurring conversations reveals that directive speech acts display a far greater degree of heterogeneity than previously recognized.

Situational directives remain closest to the canonical type. They are immediate in temporal orientation, moderate to strong in deontic force, and encompassed *requests*, *suggestions*, and *instructions*. Prospective directives, often realized

as *advice*, extend this domain by displaying a long-term orientation and weaker deontic strength. Their illocutionary profiles blend directive and assertive functions, as they involve evaluation rather than overt imposition.

The other types depart even further from traditional accounts. Expressive directives (e.g. *take care, please don't rain*) retain imperative form but have undergone pragmaticalization, serving as formulaic expressions of goodwill with minimal directive force. Absurdist directives, marked by humor or exaggeration, similarly nullify deontic strength and fell outside the sincere conditions presupposed in Searle's classical theory.

In addition, the data indicate peripheral but functionally significant subtypes. Verbal and cognitive directives explicitly solicit speech or thought rather than concrete actions. Their boundaries with information-seeking questions are often blurred, underscoring the need to treat them as distinct categories in their own right. Discourse-managing directives regulate turn-taking, attention, and the overall structure of the conversation. Although these forms exhibit a strong deontic orientation, they are often taken for granted and thus remain under-theorized despite their central role in sustaining conversational organization.

Taken together, these findings underscore that directives are not a monolithic category, but a heterogeneous family of practices that vary in interactional locus (*inward* or *outward*), temporal orientation (*immediate* or *long-term*), deontic strength (ranging from *zero* to *strong*), and action type (*proper action, interaction-supporting action* or *discourse-managing action*). Thus, a corpus-based perspective enables us to capture the full range of directive forms in everyday conversation and reconceptualize directivity as a gradient phenomenon encompassing prototypical, expressive, humorous, interaction-supporting, and discourse-organizational uses.

5. Conclusion

This study introduced dassBNC2014, a manually annotated corpus of directive speech acts in everyday British English conversations. The corpus addresses the major shortcomings of existing resources. In both linguistics and NLP, current corpora remain insufficient to meet the empirical and theoretical demands of speech act research. By integrating insights from ISO 24617-2 with categories from speech-act theory, the corpus provides a standardized and publicly available dataset that combines analytical precision with practical applicability.

Unlike earlier research based on introspection or constructed examples, this corpus grants direct access to directives as they unfold in authentic interactions, enabling us to move beyond static verb inventories to capture the heterogeneity

of directive practices in discourse. Our data demonstrate that directives are not a monolithic category, but a family of practices varying in temporal orientation, deontic strength, and interactional locus, with subtypes ranging from situational and prospective to expressive, absurdist, verbal, cognitive, and discourse management. These findings refine the typology of directives and highlight the overlooked dimensions of directive force from attenuated and expressive uses to metacommunicative control. Beyond these descriptive insights, *dassBNC2014* opens promising avenues for future research. Pragmatics enables a systematic investigation of directive subtypes neglected in theory. Computational linguistics provides data for modeling one of the most challenging cases of form-function mismatches in dialog and for advancing automatic recognition of pragmatically rich phenomena. As training data, it supports the development of models that can be transferred to other conversational corpora to retrieve directive speech acts beyond surface form. As evaluation data, it provides a benchmark for assessing the extent to which different computational approaches, ranging from traditional supervised classifiers to contemporary large language models, are able to recognize directive speech acts in natural conversation. In particular, it allows systematic testing of model performance on pragmatically challenging cases, such as implicit or weakly marked directives, where automatic recognition remains difficult. By creating a high-quality, pragmatically annotated corpus accessible to the community, *dassBNC2014* establishes a shared empirical foundation that can foster cumulative progress across disciplinary boundaries.

Funding

The work was supported by the China MOE Major Project Fund of Key Research Center for Humanities and Social Sciences (22JJD740012).

Acknowledgements

We are grateful to the anonymous reviewers, editors and copy-editor for their constructive comments and careful reading of the manuscript that significantly improved the quality of the paper.

Declaration of generative AI

During the preparation of this manuscript, the authors used ChatGPT (GPT-5.5), for language polishing (grammar and clarity). AI tools were also used to assist with data annotation: GPT-4o-mini was used to support utterance segmentation, and GPT-4o was used to assist annotators in interpreting contextual information during the annotation process. All AI-

assisted outputs were reviewed and verified by the authors. After using the tools, the authors reviewed all language content and annotations, and retain full responsibility for the final decisions.

References

- doi Allen, J., & Core, M. (1997). *Draft of DAMSL: Dialog Act Markup in Several Layers*. University of Rochester. <https://www.cs.rochester.edu/research/cisd/resources/damsl/RevisedManual/>.
- doi Alston, W.P. (2000). *Illocutionary acts and sentence meaning*. Cornell University Press.
- doi Anderson, A.H., Bader, M., Bard, E.G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H.S., & Weinert, R. (1991). The HCRC map task corpus. *Language and Speech*, 34(4), 351–366.
- doi Austin, J.L. (1962). *How to do things with words*. Clarendon Press.
- doi Bach, K., & Harnish, R.M. (1979). *Linguistic communication and speech acts*. The MIT Press.
- doi Bell, L.B. (2020). Illocution on Twitter: The construction and analysis of a social media speech act corpus [Doctoral dissertation, Georgetown University]. Georgetown University. <https://hdl.handle.net/10822/1059502>.
- doi Bilbow, G.T. (2002). Commissive speech act use in intercultural business meetings. *International Review of Applied Linguistics in Language Teaching*, 40(4), 287–303.
- doi Blum-Kulka, S., & Olshtain, E. (1984). Requests and apologies: A Cross Cultural Study of Speech Act Realization Patterns (CCSARP). *Applied Linguistics*, 5(3), 196–213.
- doi Bunt, H. (2009). The DIT++ taxonomy for functional dialogue markup. In D. Heylen, C. Pelachaud, R. Catizone, & D. Traum (Eds.), *Proceedings of the AAMAS 2009 workshop "Towards a standard markup language for embodied dialogue acts" (EDAML 2009)* (pp. 13–24). The International Foundation for Autonomous Agents and Multiagent Systems. <https://dit.uvt.nl/general/people/bunt/docs/aamas3.pdf>.
- doi Bunt, H., Petukhova, V., Malchanau, A., Wijnhoven, K., & Fang, A. (2016). The DialogBank. In N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Groblenik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the tenth international conference on Language Resources and Evaluation (LREC '16)* (pp. 3151–3158). European Language Resources Association. <https://aclanthology.org/L16-1503/>.
- doi Carletta, J. (2007). Unleashing the killer corpus: Experiences in creating the multi-everything AMI Meeting Corpus. *Language Resources and Evaluation*, 41(2), 181–190.
- doi Core, M.G., & Allen, J. (1997). Coding dialogs with the DAMSL annotation scheme. In Traum, D.R. (ed.), *AAAI Fall symposium in communicative action in humans and machines* (pp. 28–35). Association for the Advancement of Artificial Intelligence. <https://www.eecis.udel.edu/~carberry/CIS-885/Papers/Core-Allen-Coding-Damsl.pdf>.
- doi Couper-Kuhlen, E. (2014). What does grammar tell us about action? *Pragmatics*, 24(3), 623–647.
- doi 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. https://icame.info/icame_static/ij37/Pages_71-106.pdf.

-  Deutschmann, M. (2003). Apologising in British English [Doctoral dissertation, Umeå University]. Digitala Vetenskapliga Arkivet. <https://www.diva-portal.org/smash/get/diva2:143441/FULLTEXT01.pdf>.
-  Dhillon, R., Bhagat, S., Carvey, H., & Shriberg, E. (2004). *Meeting recorder project: Dialog act labeling guide* (ICSI Technical Report TR-04-002). International Computer Science Institute. <https://www.academia.edu/download/38477264/tr-04-002.pdf>.
-  Diewald, G. (2011). Pragmaticalization (defined) as grammaticalization of discourse functions. *Linguistics*, 49(2), 365–390.
-  Flöck, I. (2016). *Requests in American and British English. A contrastive multi-method analysis*. John Benjamins.
-  Flores-Salgado, E., & Castineira-Benitez, T.A. (2018). The use of politeness in WhatsApp discourse and move 'requests'. *Journal of Pragmatics*, 133, 79–92.
-  Funke, N. (2022). Pragmatic nativisation of thanking in South Asian Englishes. *World Englishes*, 41(2), 136–150.
-  Gagné, C. (2018). Indirectness and entitlement in product requests in British service encounters. *Journal of Pragmatics*, 133, 1–14.
-  García, P. (2004). Meaning in academic contexts: A corpus-based study of pragmatic utterances [Unpublished doctoral dissertation]. Northern Arizona University.
-  Han, C. H. (2000). *The Structure and interpretation of imperatives: Mood and force in Universal Grammar*. Garland.
-  Heritage, J. (2013). Epistemics in conversation. In J. Sidnell & T. Stivers (Eds.), *The Handbook of Conversation Analysis* (pp. 370–394). Wiley-Blackwell.
-  Hindelang, G. (1978). *Auffordern: Die Untertypen des Aufforderns und ihre Sprachlichen Realisierungsformen*. Kümmerle.
-  ISO 24617–2. (2020). *Language resource management – Semantic annotation framework (SemAF), Part 2: Dialogue acts* (2nd ed.). Geneva. <https://www.iso.org/standard/76443.html>.
-  Janin, A., Baron, D., Edwards, J., Ellis, D., Gelbart, D., Morgan, N., Peskin, B., Pfau, T., Shriberg, E., Stolcke, A. & Wooters, C. (2003). The ICSI meeting corpus. In W.-C. Siu, A. G. Constantinides, & Y.-T. Chan (Eds.), *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (ICASSP '03)* (pp. 364–367). Institute of Electrical and Electronic Engineers.
-  Jary, M., & Kissine, M. (2014). *Imperatives*. Cambridge University Press.
-  Jautz, S. (2013). *Thanking formulae in English: Explorations across varieties and genres*. John Benjamins.
-  Jekat, S., Klein, M., Maier, E., Maleck, I., Mast, M., & Quantz, J. (1995). *Dialogue acts in VERBMOBIL. VM-Report 65*. DFKI GmbH. https://publikationen.sulb.uni-saarland.de/bitstream/20.500.11880/25208/1/report_65_95.pdf.
-  Jurafsky, D., Shriberg, E., & Biasca, D. (1997). *Switchboard SWBD-DAMSL Shallow-discourse function annotation. Coders Manual, Draft 13* (ICS Technical Report). University of Colorado. <https://www.colorado.edu/ics/sites/default/files/attached-files/97-02-part1.pdf>.
-  Kirk, J. M. (2016). The pragmatic annotation scheme of the SPICE-Ireland corpus. *International Journal of Corpus Linguistics*, 21(3), 299–322.

- doi Kohnen, T. (2015). Speech acts: A diachronic perspective. In K. Aijmer & C. Rühlemann (Eds.), *Corpus pragmatics: A handbook* (pp. 52–83). Cambridge University Press.
- doi Leech, G., & Weisser, M. (2003). Generic speech act annotation for task-oriented dialogues. In T. McEnery, A. Wilson, P. Rayson, & P. Baker (Eds.), *Proceedings of the corpus linguistics 2003 conference* (pp. 441–446). https://www.lancaster.ac.uk/fass/doc_library/linguistics/leechg/leech_and_weisser_2003.pdf.
- doi Leech, G., & Weisser, M. (2013). The SPAADIA annotation scheme. *Martinweisser.org*. https://martinweisser.org/publications/SPAADIA_Annotation_Scheme.pdf.
- doi Levinson, S. C. (2017). Speech acts. In Y. Huang (ed.), *The Oxford handbook of pragmatics* (pp. 199–216). Oxford University Press.
- doi Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics*, 22(3), 319–344.
- doi Merrison, A. J., Wilson, J. J., Davies, B. L., & Haugh, M. (2012). Getting stuff done: Comparing e-mail requests from students in higher education in Britain and Australia. *Journal of Pragmatics*, 44(9), 1077–1098.
- doi Milà-Garcia, A. (2018). Pragmatic annotation for a multi-layered analysis of speech acts: A methodological proposal. *Corpus Pragmatics*, 2(3), 265–287.
- doi Popescu-Belis, A. (2004). Abstracting a dialogue act tagset for meeting processing. In M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa, & R. Silva (Eds.), *Proceedings of the fourth international conference on Language Resources and Evaluation (LREC '04)* (pp. 1415–1418). European Language Resources Association. <http://www.lrec-conf.org/proceedings/lrec2004/pdf/268.pdf>.
- doi Rühlemann, C. (2018). *Corpus linguistics for pragmatics: A guide for research*. Routledge.
- doi Ruytenbeek, N. (2017). The Mechanics of indirectness: A case study of directive speech acts. [Unpublished doctoral dissertation]. Université Libre de Bruxelles.
- doi Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in Conversation Analysis*. Cambridge University Press.
- doi Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press.
- doi Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge University Press.
- doi Searle, J. R., & Vanderveken, D. (1985). *Foundations of illocutionary logic*. Cambridge University Press.
- doi Seto, A. W. (2016). A Corpus-driven study of speech acts in the Hong Kong Corpus of Spoken English (HKCSE) [Doctoral dissertation, The Hong Kong Polytechnic University]. PolyU Electronic Theses. <https://theses.lib.polyu.edu.hk/handle/200/8722>.
- doi Su, H., & Wei, N. (2018). I'm really sorry about what I said: A local grammar of apology. *Pragmatics*, 28(3), 439–462.
- doi Traugott, E. C., & Trousdale, G. (2013). *Constructionalization and constructional changes*. Oxford University Press.
- doi Urbanik, P. K. (2017). Requests in Polish and Norwegian informal conversation: A comparative study of grammatical and pragmatic patterns [Doctoral dissertation, University of Oslo]. Norwegian Research Information Repository. <https://hdl.handle.net/11250/4202256>.

 Verdonik, D. (2023). Annotating dialogue acts in speech data: Problematic issues and basic dialogue act categories. *International Journal of Corpus Linguistics*, 28(2), 144–171.

 Weisser, M. (2003). SPAACy — A semi-automated tool for annotating dialogue acts. *International Journal of Corpus Linguistics*, 8(1), 63–74.

 Weisser, M. (2015). Speech act annotation. In K. Aijmer & C. Rühlemann (Eds.), *Corpus pragmatics. A handbook* (pp. 84–116). Cambridge University Press.

 Weisser, M. (2019). The DART annotation scheme: Form, applicability & application. *Studia Neophilologica*, 91(2), 131–153.

 Weisser, M. (2020). Speech acts in corpus pragmatics: Making the case for an extended taxonomy. *International Journal of Corpus Linguistics*, 25(4), 400–425.

 Wilson, D., & Sperber, D. (1988). Mood and the analysis of non-declarative sentences. In J. Dancy, J. Moravcsik, & C. Taylor (Eds.), *Human agency: Language, duty and value* (pp. 77–101). Stanford University Press.

 Xu, J. & Hao, X. (2019). Jiajin yanyu xingwei neibu tiaojie shouduan de yuyong tiaokong [Pragmatic regulation on the internal modification of suggestive speech acts]. *Shijie Hanyu Jiaoxue [Chinese Teaching in The World]*, 33(3), 353–368.

Appendix A. Prompt used to conduct batch segmentation via the OpenAI API

You are an expert in sentence segmentation.

I will give you an English conversation. Please read and understand the conversation. Your task is to analyze and segment each utterance based on the context meaning. You should break each utterance down into sentences that represent independent and complete units of meaning. Only perform segmentation WITHOUT making any changes to the original content. DO NOT delete any XML tags in the sentence.

One example:

Conversation:

'''

<u n="33" who="So282">it kind of fits in logically with <pause dur="short" /> the other services that are available there</u>

<u n="34" who="So651">yeah</u>

<u n="35" who="So282">so <pause dur="short" /> I want I need to understand from you what your thinking was about location cos you suggested a specific place</u>

<u n="36" who="So651">I did <pause dur="short" /> I think erm talking about the business <pause dur="short" /> plan then</u>

'''

Result you return:

'''

<u n="33" who="So282">

<s n="1">it kind of fits in logically with <pause dur="short" /> the other services that are available there</s>

</u>

```

<u n="34" who="So651">
<s n="1">yeah</s>
</u>
<u n="35" who="So282">
<s n="1">so <pause dur="short" /></s>
<s n="2">I want I need to understand from you what your thinking was about location</s>
<s n="3">cos you suggested a specific place</s>
</u>
<u n="36" who="So651">
<s n="1">I did <pause dur="short" /></s>
<s n="2">I think erm talking about the business <pause dur="short" /> plan then</s>
</u>
'''

```

Appendix B. Annotation statistics for each file in dassBNC2014

File	Length	Number of tokens	Number of utterances	Number of sub- utterances	Sub-utterances annotated as <i>Directives</i>	Assistant annotator
S2C9	02:12:00	25,137	1,836	3,305	47	ANo2
S2CY	00:14:25	2,647	299	401	14	ANo7
S2LD	01:07:58	11,856	1,511	2,095	148	AN10
S32W	01:00:29	14,563	1,297	1,938	27	ANo1
S35K	01:28:53	18,768	2,544	3,204	127	ANo9
S3A9	00:15:01	1,894	231	333	40	AN11
S3VM	00:10:04	2,066	148	283	3	ANo1
S56S	01:06:24	14,694	1,630	2,333	7	ANo5
S7QE	00:46:17	6,672	904	1,126	38	AN11
S7QU	00:22:58	5,579	196	676	25	ANo3
S8NX	00:41:29	7,707	816	1,168	3	AN11
SAVN	01:05:33	9,203	1,019	1,425	85	ANo1
SBQ6	01:11:25	15,885	1,570	2,395	44	ANo8
SEUM	00:12:00	2,720	339	438	3	AN11
SF7M	00:21:16	3,049	365	531	42	ANo9
SFXE	00:45:12	7,663	791	1,192	5	ANo1
SKGU	00:34:00	8,520	617	1,161	62	ANo8
SMBC	00:16:06	2,784	320	448	7	ANo6
SP84	00:49:01	8,316	735	1,187	13	ANo1
SQ63	00:15:41	3,354	325	449	6	ANo6

Appendix B. (continued)

File	Length	Number of tokens	Number of utterances	Number of sub- utterances	Sub-utterances annotated as <i>Directives</i>	Assistant annotator
SQRA	00:18:25	4,060	517	725	10	ANo9
SR7T	00:21:53	5,591	727	1,018	37	AN11
SRGA	00:47:38	6,480	774	1,107	43	ANo7
STQF	00:25:21	4,196	352	578	17	ANo9
STWC	01:59:47	18,069	2,924	3,787	163	ANo5
STXT	01:19:36	13,412	1,569	2,308	104	ANo7
STY2	00:37:45	6,429	537	975	52	ANo7
SUGJ	01:00:56	15,613	1,010	1,765	140	AN10
SXFR	00:12:38	2,387	327	452	24	ANo3
SY5K	00:55:01	11,529	816	1,430	29	ANo6
SYEZ	01:00:17	12,585	1,570	2,094	45	ANo6
SZKE	00:35:42	5,479	658	906	20	ANo6
S2A5	00:12:20	1,834	233	323	26	AN10
S2LC	00:57:57	9,448	883	1,344	30	ANo3
S2PS	00:20:52	5,644	876	1,042	32	AN12
S55B	00:20:58	4,928	694	819	21	ANo4
S66E	00:26:41	5,273	457	711	8	ANo1
S7JF	00:27:49	6,340	503	828	41	ANo1
S8GL	00:23:26	6,865	901	1,185	26	ANo3
S8Q6	00:14:06	2,519	257	363	11	ANo8
S9WP	00:22:19	3,649	209	445	21	ANo2
SC7H	00:14:26	3,043	192	381	3	ANo3
SG3P	00:14:40	2,610	192	378	37	ANo9
SHXF	00:30:27	6,113	325	588	12	ANo1
SN2R	00:30:38	4,771	470	718	53	ANo3
SPY2	00:49:42	8,914	940	1,331	23	ANo1
SRBZ	00:51:42	9,434	624	1,146	30	ANo8
SRVR	00:30:32	6,529	480	886	22	ANo1
SY8A	00:12:51	1,958	244	357	15	ANo3
SZT4	00:17:42	3,615	403	548	17	ANo4
S2EF	01:25:40	16,139	2,496	3,123	115	ANo2
S2TP	00:12:59	2,919	444	532	22	ANo1

Appendix B. (continued)

File	Length	Number of tokens	Number of utterances	Number of sub- utterances	Sub-utterances annotated as <i>Directives</i>	Assistant annotator
S38P	00:54:38	12,276	1,862	2,324	98	ANo4
S6J2	01:09:22	13,546	1,019	1,755	80	ANo4
S8WS	00:37:26	7,242	503	948	9	AN12
SBBP	00:18:52	4,206	416	663	13	ANo1
SFNQ	01:05:03	18,182	1,897	2,918	64	ANo2
SJDK	00:16:13	4,157	473	626	8	ANo3
SN7D	00:20:33	4,432	490	673	7	AN10
SQ9R	00:58:38	10,521	820	1,432	52	ANo3
SUU4	00:25:01	4,010	252	486	7	ANo4

Address for correspondence

Jiajin Xu
National Research Centre for Foreign Language Education
Beijing Foreign Studies University
19 Xisanhuan North Road
Haidian District
Beijing, 100089
China

xujiajin@bfsu.edu.cn

 <https://orcid.org/0000-0003-3454-9352>

Co-author information

Guying Zhou
School of German Studies
Beijing Foreign Studies University

guying.zhou@bfsu.edu.cn

 <https://orcid.org/0009-0007-4385-4259>

Publication history

Date received: 27 August 2025

Date accepted: 17 March 2026

Published online: 7 September 2026