



Università  
Ca' Foscari  
Venezia

Master's Degree  
in Language  
Sciences

Final Thesis

# Understanding Speech Acts: Towards the Automated Detection of Speech Acts

**Supervisor**

Ch. Prof. Guglielmo Cinque

**Assistant supervisor**

Ch. Dr. Rocco Tripodi

**Graduand**

Federico Vescovi

Matriculation number

842655

**Academic Year**

2018 / 2019

## CONTENTS

INTRODUCTION	p. 4
I - FUNDAMENTALS AND THEORY OF SPEECH ACTS	p. 5
1. Introduction: Semantics and Pragmatics	p. 5
2. Grice, Austin, and the Speech Act Theory	p. 11
2.1 Grice	p. 11
2.2 Austin and the Speech Act Theory	p. 15
3. An Introduction to Indirect Speech Acts	p. 22
4. Illocutionary Logic: F and P	p. 25
5. Performative Utterances and Illocutionary Force Indicating Devices	p. 27
6. Conclusion	p. 31
II - INDIRECT SPEECH ACTS	p. 33
1. Felicity Conditions	p. 33
2. A Parallel Analysis of Direct and Indirect Speech Acts	p. 41
3. Conventional, Semi-conventional, and Non Conventional Indirect Speech Acts	p. 60
III - ON CLASSIFICATION	p. 68
1. Introduction	p. 69
2. Ambiguity	p. 72
3. More Primitive vs. Less Primitive Devices	p. 74
4. Austin's Classification	p. 77
5. Searle's Classification	p. 82
6. Deep Structure Representations of Searle's Classes	p. 90
7. Computational Linguistics: Introduction and Motivation	p. 93
8. Overview of the Classifications (Tag-sets) in Computational Linguistics	p. 95
8.1 Synchronous Conversation Tag-sets	p. 96
8.2 Asynchronous Conversation Tag-sets	p. 98
9. DAMSL Standard	p. 99
10. SWBD-DAMSL	p. 111
11. MRDA	p. 134
12. MRDA: Adjacency Pairs	p. 169
13. Comparison Between SWBD-DAMSL and MRDA	p. 169
14. Email Speech Acts	p. 174
15. BC3, TA, and QC3	p. 175
16. Conclusion	p. 177
IV - PROBLEMS CONNECTED WITH SPEECH ACT IDENTIFICATION	p. 185
1. Statements	p. 187
2. Issues regarding other classes	p. 191
3. Structure of the Tags	p. 194
4. Conclusion	p. 195



## **Introduction**

The present work constitutes an attempt to analyze language in terms of the actions that we perform through speaking. Our work revolves around the speech act theory (Austin, 1962; Searle, 1969), a theory of language use that investigates the actions, or acts, that we perform when we utter linguistic expressions in conversation; a few examples of what such actions could be are: requesting, questioning, promising, threatening, and apologizing. Assuming that every utterance involves the performance of (at least) one speech act (Searle & Vanderveken, 1985), our goal is to determine what (and how many) types of speech acts we can efficiently classify, where each type or class of speech acts includes all the speech acts that share the same point or purpose in conversation (Searle, 1976). To do so, we will first need to define what a speech act is, and then determine which features of an utterance discriminate one speech act type from the other, or, in other words, which features can be used as indicators that one utterance is used with one purpose instead of another. We will perform an analysis of both the linguistic form of utterances and the context in which they are used. Our analysis results in the following two key observations: 1) elements of natural language can be used as indicators of speech act types; and 2) the use of such elements for utterance classification is as tempting as it is misleading since there are many ways to perform a speech act without using a corresponding natural language indicator. While from the point of view of pragmatics classifying speech acts might be of little use since speech act classification is to a large extent arbitrary and not always a necessary step for communicating successfully (Jaszczolt, 2002), there are many domains and research areas that benefit from having at hand an accurate classification of speech acts, as well as an effective way to systematically map utterances to speech act types or classes; for example: dialog systems, speech recognition (see Stolcke et al., 2000 and Paul et al., 1998), machine translation (see Levin et al., 2003), summarization (see McKeown et al., 2007), and question answering (see Hong and Davison, 2009). Moreover, if applied to emails (but also to other types of asynchronous communication), a classification of the so-called email acts (acts performed by sending an email) proves useful not only to speed up email communication overall, but also to predict leadership roles within email-centered work groups (Carvalho, 2008).

## CHAPTER 1 - FUNDAMENTALS AND THE THEORY OF SPEECH ACTS

The purpose of this chapter is to provide the reader with a concise yet informative introduction to the study of meaning and to the theory of speech acts. We will briefly introduce the fields of semantics and pragmatics, and familiarize ourselves with the relevant terminology. In doing so, we will elaborate on the reasons why semantics alone, without the intervention of pragmatics, falls short of accounting for what speakers actually mean when they communicate. We will then present the works of British philosophers of language H. P. Grice and J. L. Austin, which constitute the blueprint for contemporary research frameworks in pragmatics. Finally, we will focus on the speech act theory, a theory of language use that investigates the actions, or acts<sup>1</sup>, that we perform through speaking. At the end of this chapter, we will have at hand a full-fledged, pragmatics-aware theory of meaning, which will form the theoretical background of our proposal in the next chapters.

### 1. Introduction: Semantics and Pragmatics

What is meaning? The history of science and philosophy has witnessed numerous attempts to address this question, thus providing fertile ground for the birth and development of a number of theories of meaning. In contemporary language sciences, semantics and pragmatics are the branches of linguistics and philosophy that deal with the study of meaning. Semantic theories are typically concerned with the study of meaning as a component of the faculty of language, that is to say: the study of the literal<sup>2</sup> meaning of linguistic expressions, irrespective of the context in which they are used. Pragmatic theories, on the other hand, investigate the interaction between the context and the literal meaning of what is uttered, drawing particular attention to the role of interlocutors, i.e. the speaker and the addressee (Jaszczolt, 2002). In this respect, context is a general term encompassing numerous features of a circumstance of use and can be provisionally defined as the combination of physical and cultural setting, speaker intention, and discourse<sup>3</sup>. Semantic theories and pragmatic

---

<sup>1</sup> The Oxford Online Dictionary (2019) defines "act" and "action" very similarly: an "act" is "[a] thing done; a deed", and an "action" is "[a] thing done; an act". In the present work, we use "act" and "action" as synonyms.

<sup>2</sup> "Literal" can be defined as "derived from the core conventional meanings of words" (Jurafsky & Martin, 2018, p. 296) or as "taking words in their usual or most basic sense without metaphor or exaggeration" (Oxford English Dictionary, 2019).

<sup>3</sup> In our temporary definition of "context", we merge what are generally considered two distinct types of contexts. The term "context" is in fact usually intended as either 1) "a subjective, cognitive representation of the world"(Penco, 1999), made up of the subjective beliefs, intentions, psychological states, attitudes, and expectations of the interlocutors, or as 2) "an objective, metaphysical state of affairs" (Penco, 1999), made up of objective and external states of affairs or events, such as present or past social behavior (and the cultural-specific societal conventions that determine it), facts about material objects, etc., i.e. all that exists in the world.

theories are not necessarily in conflict with each other, but rather they have different purposes and fields of application. Semantics focuses on determining the *literal meaning* (for this reason also called semantic meaning) of linguistic expressions, whereas pragmatics involves a form of "higher order" reasoning on this literal meaning, as it tries to capture the information conveyed and the actions performed by uttering some expression in a particular context (Korta & Perry, 2015). Another way to clarify the distinctive roles of these two disciplines - semantics and pragmatics - is in terms of their fields of application: while the unit of analysis of pragmatics is the utterance<sup>4</sup>, a concrete product of speech and writing or a contextualized sentence, the unit of analysis of semantics is the sentence<sup>5</sup> understood as the abstract, grammatical unit that can be derived from an utterance by abstracting over contingent and contextual information. Utterances "come with information as to who the speaker is as well as information about the time, the place and other circumstances of the performed act of speaking" (Jaszczolt, 2002, p. 2); sentences, on the other hand, can be thought of as the "grammatical clothing" of utterances (Searle, 1969, p. 25). That being said, the present work is not concerned with semantics *per se*, nor it deals with that part of pragmatics, sometimes called "near-side pragmatics", that focuses on those pre-semantic (Levinson, 2000, p. 188; Recanati, 2004, p. 134) roles of context that concern the "facts that are relevant to determining what is said" (Korta & Perry, 2015) - such as disambiguation and reference resolution (cf. Grice 1989, p. 25) -, but instead focuses on the so-called "far-side pragmatics", that is to say: that part of pragmatics concerned with "what we do with language, beyond what we (literally) say" (Korta & Perry, 2015). Let's now clarify the notions of literal meaning, near-side pragmatics, and far-side pragmatics by considering the following utterance:

1. I am cold.

Roughly speaking, the literal or semantic meaning of 1 - what Grice calls "sentence meaning" (more in section 2.1) - is that "I", the subject of the sentence, predicates the attribute "cold" of him- or herself. Near-side pragmatics focuses on determining who "I" refers to (reference resolution) - let's say, for the sake of argument, that it refers to a person called Mary - and clarifies whether "cold" is meant as "cold hearted" or "low in temperature" (disambiguation) - let's say the latter. Therefore, the semantic or literal meaning of 1 enriched by contextual information provided by near-side pragmatics - what Grice calls "what is said" (more in section 2.1) - is that Mary, the subject of the sentence, predicates the attribute "low in temperature" of herself. Far-side pragmatic, on the other hand, is concerned with what the speaker communicates by uttering 1 in a specific context. Mary

---

<sup>4</sup> Here we use the term "utterance" to indicate the *result* of language production, whether spoken or written. We will sometimes use this term also to indicate the act of producing (spoken or written) language (as in "the utterance of a sentence").

<sup>5</sup> We will often refer to the sentence as the utterance's "linguistic form".

can in fact utter 1 and mean it literally, in which case she communicates what she says<sup>6</sup>, but she can also use 1 to do something else; for example, she can make an indirect request to John, her interlocutor, to switch off the air conditioning.

There is an ongoing debate about the extent to which semantics and pragmatics overlap, and about whether they overlap in the first place. In the present work, we will not partake in the debate. Rather, we will focus on demonstrating why a successful theory of meaning must be aware of the context in order to reliably account for what speakers actually mean when they communicate. That being said, we will definitely not disregard semantic theories altogether. On the contrary: recent works on speech acts - although arguably in contrast with Austin's (1962) original motivation behind the formulation of the speech act theory<sup>7</sup> - are built upon existing pragmatics-compatible semantic theories. We will focus in particular on the contributions of Searle and Vanderveken (Searle, 1969; Searle & Vanderveken, 1985), who incorporated the notion of proposition<sup>8</sup> into the speech act theory. Before delving into the study of speech acts, however, we first need to take a closer look at what semantic theories can and cannot achieve.

Generally speaking, semantic theories deal with sentences as decontextualized units of grammar and are particularly concerned with the propositions that they express. In Speaks' (2017) words, the current trend in semantics, can be described as follows:

"Most philosophers of language these days think that the (literal) meaning of an expression is a certain sort of entity, and that the job of semantics is to pair expressions with the entities which are their meanings. For these philosophers, the central question about the right form for a semantic theory concerns the nature of these entities. Because the entity corresponding to a sentence is called a *proposition*, I'll call these *propositional semantic theories*<sup>9</sup>" (Speaks, 2017).

Semantic theories can thus be broadly defined as those theories of meaning that are concerned with pairing sentences with propositions. At this point, the following question arises: "how do we

---

<sup>6</sup> In the present work, we use the term "say" in its narrow sense to mean "literally say".

<sup>7</sup> Austin (1962) formulated the speech act theory to bring about a revolution in the study of meaning. He fiercely opposed the study of meaning in terms of truth and was (arguably) also contrary to the use of propositions for describing meaning (for the full discussion see Sbisà, 2006).

<sup>8</sup> By reason of the broad use of the term "proposition" in contemporary philosophy, it is challenging to devise a reliable definition of it (McGrath, 2018). Propositions are "commonly treated as the meanings or, to use the more standard terminology, the *semantic contents* of (declarative) sentences" (McGrath, 2018). For simplicity, we will adopt this very definition of "proposition", aware of the fact that it is an oversimplification of a rather technical term. We will use the terms "proposition", "propositional content", and "semantic content" interchangeably. For a complete discussion on the different uses of the term "proposition", see McGrath (2018) and Lewis (1980).

<sup>9</sup> We must acknowledge the fact that *non*-propositional semantic theories have also been formulated. Generally speaking, these theories challenge the idea that propositions are the right sort of entities for representing meaning (McGrath, 2018) and disagree with the view that the job of a semantic theory is that of systematically pairing expressions with entities representing their meanings (Speaks, 2017).

represent the meaning of a sentence?”, i.e. “what does a proposition look like?”. The issue of pairing linguistic expressions with entities corresponding to their meanings is in fact intertwined with the issue of giving form to these entities. Propositions are captured in formal structures called meaning representations, and their creation and assignment to linguistic inputs is called semantic analysis (Jurafsky & Martin, 2018, pp. 295-296). Propositions can be successfully represented thanks to a number of meaning representation metalanguages, such as first-order logic, that are designed to describe *literal* meaning in an unambiguous way (Jaszczolt, 2002). Let's consider the following sentence:

2a. Every man loves a woman.

This sentence has a semantic ambiguity caused by the unspecified scope of the verb "love". This ambiguity results in the sentence expressing two possible propositions, each represented unambiguously in first-order logic as follows:

2b.  $(\forall x)\text{man}(x) \rightarrow ((\exists y)\text{woman}(y) \wedge (\text{love}(x, y)))$

2c.  $(\exists y)\text{woman}(y) \wedge ((\forall x)\text{man}(x) \rightarrow \text{love}(x, y))$

According to 2b, for every man, there is a woman, and it's possible that each man loves a different woman, whereas according to 2c, there is one particular woman who is loved by every man. We can use logical representations to describe the logical structures of sentences. This enables us to see clearly their logical inferential properties, and precisely and unambiguously determine their truth conditions (more on truth below). While logical representations indeed prove useful in disambiguating sentences from a semantic perspective, that is in terms of lexicon, structure, and scope, they are not sufficient for determining with certainty what speakers communicate (or mean) by uttering those sentences in conversation. Propositions, being abstract entities, are in fact communicatively (or pragmatically) inert. While we will remain neutral on the appropriate conceptualization of propositions, we will examine the reasons why the use of propositions and semantic theories overall are in some sense deficient.

Truth-conditional semantics (see in particular Davidson, 1967), which is the current predominant approach in semantics (Jaszczolt, 2002), claims that knowing the meaning of a sentence means knowing what the world would have to be like for the sentence to be true (Jaszczolt, 2002). We can test whether sentences express different propositions by invoking the notion of truth. The proposition is evaluated to a truth value: the evaluation will return *true* if the sentence corresponds to the world, otherwise it will return *false*. According to truth-conditional semantics, the meaning of an expression is its contribution to the truth conditions of the sentence, that is the conditions the world has to fulfill for the sentence to be true (Jaszczolt, 2002). For example, the following utterance

3. I am in Cambridge.

expresses a proposition that is true if the speaker is in Cambridge. If the speaker substitutes "Cambridge" with "Oxford" and the speaker is in Cambridge, the proposition will instead be false, thus indicating a different meaning. This is, generally speaking, how meaning is understood in terms of truth.

We have opened this parenthesis on truth, and on semantic theories more in general, to demonstrate that a semantic, truth-conditional approach to meaning, despite working fairly well in representing the meaning of syntactically and semantically complete *declarative* sentences (sentences typically used to make statements), reveals itself fairly limited, not only because it does not say much about the meaning of each single word composing the sentence, but also because of its incapability of dealing effectively with *non-declarative* sentences, such as questions (e.g. "Are you coming to my birthday party?"), commands (e.g. "Shut the door!"), and modalities (e.g. "He may / must be in London"), as well as propositional attitude reports (e.g. "I believe that he will be late"), sentences without a clear propositional content (e.g. "Wow!"), sentences with explicit indicators of illocutionary force<sup>10</sup> (e.g. "I promise that I will come"), and sentences performing indirect speech acts (e.g. "Can you pass me the salt?") (more on all of these below and in the next chapters). These types of sentences are in fact not merely describing or reporting facts of the real world that can be evaluated as true or false (Austin, 1962), which makes them non susceptible to a satisfactory truth-conditional analysis (Jaszczolt, 2002). For this reason, it would be short-sighted to analyze utterances only in terms of their propositional contents as the bearers of truth values. Just to make a few examples: in which cases can we consider the propositional content of a question to be true? And in which cases false? And what about the propositional content of a command?

Analyzing utterances in terms of the truth of their propositions reveals itself problematic also in the case of declarative sentences. As Austin (1962) points out: "many utterances which look like statements are either not intended at all, or only intended in part, to record or impart straightforward information about the facts" (p. 2). Austin further argues that "specially perplexing words embedded in apparently descriptive statements do not serve to indicate some specially odd additional feature in the reality reported, but to indicate (not to report) the circumstances in which the statement is made or reservations to which it is subject or the way in which it is to be taken and the like" (p. 3). Simply put: not all declarative sentences are statements describing states of affairs (Austin, 1962). Let's consider the following examples (4 is from Austin, 1962, p. 5):

4. I bet you six pence it will rain tomorrow.

---

<sup>10</sup> "Illocutionary force" can provisionally be defined as "speaker's intended use". We will examine illocutionary force more in detail in section 2.2 and in chapter 3.

5. I state that I am in Oxford.

By uttering 4, the speaker is not describing or reporting what he or she is doing while uttering that sentence, but rather is *doing something* by uttering that sentence: the speaker is performing the action, or act, of making a bet (Austin, 1962). 4 cannot then be evaluated as a true or false proposition, but instead it should be subject to other conditions which make it successful or unsuccessful as an *action*, that is as being either a sincere or insincere bet, and so forth (more on sincerity conditions and other conditions of success in chapter 2). The proposition expressed in 5 does not have truth values either, or better it is true just in case the speaker stated it, irrespective of whether the speaker is indeed in Oxford: the speaker can replace "Oxford" with the name of any other location and the proposition will still be true. Ambiguities such as those arisen in 4 and 5 can be solved by identifying the verbs "state" and "bet" as playing a special role in the utterance. "State" and "bet" are in fact examples of so-called explicit indicators of illocutionary force (more precisely, performative verbs), and the propositions that they precede - assuming that we adopt the proposition-centric view of the speech act theory - are subject to that force in a way that impacts the overall meaning of the utterance (more in sections 2.2 and 5). As Austin (1962) points out, "once we realize that what we have to study is not the sentence but the issuing of an utterance in a speech situation, there can hardly be any longer a possibility of not seeing that stating is performing an act" (p. 138). Austin goes on to say that statements, just like the other types of action, take effect: "if I have stated something, then that commits me to other statements: other statements made by me will be in order or out of order" (Austin, 1962, p. 138). The fact that utterances, including statements, exert a certain influence on the future developments of the conversation suggests that each utterance can be understood even better if it is analyzed inside of the conversation in which it occurs.

In conclusion, we can say that truth-conditional semantics is incapable of accounting for what speakers mean when they communicate. Statements, just like bets, questions, and commands, are not sentences that express a proposition which is either true or false, but rather sentences that speakers utter to *do something* in conversation. Language use is in effect part and parcel of every utterance, including statements, and thus needs to be accounted for in some way. In order to actualize an efficient pragmatic analysis of utterances, however, we need a new set of theoretical tools. Grice (1957; 1975) will guide us along the journey from the structural, semantic analysis of the sentence to the communicative, pragmatic analysis of the utterance. We will in fact be concerned with understanding what the speaker means by uttering a given sentence in conversation, rather than what that sentence means out of context. Austin (1962), who first formulated the speech act theory, will take us a step further, towards the understanding of pragmatic meaning in terms of actions. Finally, the works of Searle and Vanderveken (Searle, 1969; Searle & Vanderveken, 1985)

will provide us with a new perspective on the study of speech acts, which integrates the concept of the proposition into the speech act theory: they elaborate on how the propositional content of a speech act can be thought of being under the scope of its illocutionary force.

## 2. Grice, Austin, and the Speech Act Theory

Contemporary research in pragmatics can be traced back to the works of Grice (1957) and Austin (1962), who are the two central figures of the "beyond saying" turn in philosophy of language in the second half of the Twentieth Century.

### 2.1 Grice

Grice (1957; 1975) distinguishes three levels of meaning: sentence meaning and what is said, jointly the object of study of semantics, and what is implicated, studied by pragmatics. In turn, what is said and what is implicated jointly constitute what Grice calls speaker meaning, as opposed to the abstract and decontextualized sentence meaning. Grice thus splits literal or semantic meaning into two: sentence meaning and what is said. Sentence meaning refers to what words, combined together to form sentences (according to the rules of syntactic and semantic composition), mean out of context. For example, the sentence meaning of a context-sensitive term such as "here" is simply the formal instruction to look into the context for the current location. Speaker meaning, on the other hand, indicates what people mean and refer to when using those words in conversation. Speaker meaning can correspond either to what the speaker says, i.e. to what is said, or to what the speaker implicates, i.e. to what is implicated, depending on the context. If what the speaker means corresponds to what the speaker says, we can retrieve the speaker meaning of "here" (or what the speaker means by "here") simply by solving for its referent, i.e. by finding what location "here" refers to in that particular context. In other words, if speaker meaning and what it said coincide, what the speaker means by "here" in a given context *c* will be the particular location referred to in *c*. What is said stands somewhere in-between semantics and pragmatics as it is determined by sentence meaning plus disambiguation and reference resolution (near-side pragmatics). That being said, there are also cases in which what the speaker means differs from what the speaker says (i.e. cases in which speaker meaning differs from what is said). In these cases, according to Grice, the speaker generates an implicature<sup>11</sup>.

---

<sup>11</sup> In the present work, by "implicature" we always mean "conversational implicature", as opposed to "conventional implicature". "Conventional implicatures are as much inferences as conversational implicatures" (Wayne, 2014), where "inference" can be defined as "conclusion reached on the basis of evidence and reasoning" (Oxford English Dictionary, 2019). However, there is a fundamental difference between conversational and conventional implicatures:

As we have mentioned above, Grice introduces the notion of implicature. As Horn puts it, "implicature is a component of speaker meaning that constitutes an aspect of what is meant in a speaker's utterance without being part of what is said" (Horn, 2004, p. 3). In other words, what is implicated is part of the global message intended by the speaker that remains unsaid and is left to the rational elaboration of the addressee. The central idea in Gricean pragmatics is that humans understand each others' communicative acts in terms of their underlying intentions. Meaning thus comes from the speaker's intention to convey information, to produce a belief in the addressee. In turn, speaker's intentions may be made explicit in the linguistic form of the utterance. Alternatively, recovery of communicative intentions may be left to the inferential elaboration of the addressee, based on the assumption that rational conversationalists share and abide by a number of "principles" and so-called "maxims" of conversation, which are generally aimed at enhancing rational co-operation and the maximization of communicated information with the least effort. Let's clarify the notion of implicature by considering the following exchange (from Grice, 1975, p. 32):

7a. A: Smith doesn't seem to have a girlfriend these days.

7b. B: He has been paying a lot of visits to New York lately.

This exchange demonstrates that a purely semantic analysis falls short of accounting for what speaker B globally means by uttering a sentence such as 7b (in response to 7a). Without taking into account Gricean implicatures, it is in fact impossible to conclude that, in the relevant context,

---

while conversational implicatures, as we will see in detail below, are inferences that "depend on features of the conversational context", conventional implicatures are inferences that are part of "the conventional meaning of the sentence used" (Wayne, 2014). Before we move on, a terminological clarification is in order: "inference" can also be used as a mass noun, in which case it can be defined as "[t]he process of inferring something" (Oxford English Dictionary, 2018), i.e. the process by which we reach a reasonable conclusion. In the present work, we use the term "inference" with its former definition, thus equating it with "reasonable conclusion". Instead, whenever we use "inference" with its latter definition (to refer to the process of inferring something), in order to avoid confusing it with the result of such process, we will call it explicitly "inferential process". That being said, since we are interested in the "beyond saying", we want to be able to distinguish conventional implicatures from conversational implicatures so as to put the former to one side and focus on the latter. Let's consider the following example (from Potts 2005; 2007, p. 668).

6a. Ravel, a Spaniard, wrote music reminiscent of Spain.

6b. Ravel was a Spaniard.

By uttering 6a and meaning it literally, the speaker conventionally implicates, but does not say, that 6b (Wayne, 2014). The conventional implicature 6b is generated syntactically by means of an appositive construction. In other words, the syntax of 6a together with the conventional (or literal) meaning of each of the words composing it generate the conventional implicature that Ravel was a Spaniard. In Wayne's (2014) words, "[t]he implicature is conventional because the sentence cannot be used with its English meaning without implicating that Ravel was a Spaniard". The addressee can infer the conventional implicature 6b on the basis of the literal meaning of 6a alone, without the intervention of the context. Since conventional implicatures are part of what is said, some - including Bach (1999; 2006) - have argued that conventional implicatures should have never been detached (or separated) from what is said in the first place (Wayne, 2014). We will not dive into this issue since it is out of the scope of the present work. We will limit ourselves to saying that conventional implicatures are conclusions that we reach reasoning on the literal meaning of the utterance alone (i.e. on what is said), whereas conversational implicatures are conclusions that we reach reasoning on the interaction between what is said and the context. From now on, we will focus only on conversational implicatures and we will always use the term "implicature" to mean "conversational implicature".

speaker B communicated his or her knowledge (or suspicion) that Smith has a girlfriend in New York (Jaszczolt, 2002). On Grice's view, this information - B's intended meaning - is available as an implicature that the addressee can rationally infer, reasoning on B's apparent violation of the maxim of relation. The maxim of relation (one of the four maxims of rational conversation proposed by Grice; more on Gricean maxims below) presupposes that the *rational* speaker is relevant, i.e. that his or her utterances are pertinent to the discussion; any intentional violations of this maxim - or of any other maxim for that matter - are to be interpreted by the addressee as a signal that an implicature has been generated, i.e. that some additional information, or some additional meaning, is available to be inferred. In the exchange reported above, speaker B, by *intentionally* not being relevant, makes available to speaker A some meaning which is additional to what he or she says. Speaker A can infer this additional meaning by reasoning on how the literal meaning of 7b interacts with that particular context of utterance.

The exchange above demonstrates that semantics alone is sometimes incapable of retrieving the actual meaning of an utterance and therefore a pragmatics-rich theory of meaning becomes necessary. In fact, any pragmatics-unaware theories of meaning would not be able to capture the meaning of semantically uninformative utterances like 7b. Entering the realm of pragmatics, however, comes with a number of problems: while there is always a direct correspondence between the sentence and its literal meaning, we must acknowledge the fact that there is no rigid correspondence between the utterance and what is implicated. This is because implicatures depend on the context and many aspects of the context are volatile. In Korta and Perry's (2015) words: "it is possible for different speakers in different circumstances to mean different things using (the same) words". We can prove this point by considering a different context for 7b; for example, if Smith works all the time and has no free time when he is in New York, speaker B, by uttering the same words, will communicate his or her knowledge (or suspicion) that Smith does *not* have a girlfriend in New York (because Smith would not have enough time for her as he is always working while he is in New York).

That being said, in order to determine what the speaker means, we first need to determine whether the speaker intends to generate an implicature or instead wants his or her utterance to be taken literally. If the speaker intends to generate an implicature, he or she can (attempt to) communicate this intention to the addressee by purposefully not being rational or cooperative. This is when Grice's Cooperative Principle comes into play. According to Grice, the governing dictum of rational interchange is the Cooperative Principle: "Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange" (Grice, 1975, p. 45). The Cooperative Principle can be instantiated by the following four maxims or

macroprinciples - one of which is the maxim of relation seen above - and their respective submaxims (Grice, 1975):

- 1) QUALITY: Try to make your contribution one that is true.
  - 1.1 Do not say what you believe to be false.
  - 1.2 Do not say that for which you lack evidence.
- 2) QUANTITY:
  - 2.1 Make your contribution as informative as is required (for the current purposes of the exchange).
  - 2.2 Do not make your contribution more informative than is required.
- 3) RELATION: Be relevant.
- 4) MANNER: Be perspicuous.
  - 4.1 Avoid obscurity of expression.
  - 4.2 Avoid ambiguity.
  - 4.3 Be brief. (Avoid unnecessary prolixity).
  - 4.4 Be orderly.

As we have said, any intentional violations of any these maxims or submaxims are to be interpreted by the addressee as a signal that the speaker intends to communicate an additional, non-literal meaning. Such additional meaning, according to Grice, takes the form of an implicature. To be more precise, there exist two kinds of implicatures: particularized implicatures and generalized implicatures:

1) in particularized implicatures, pragmatic inferences *enrich* the structure of the uttered sentence with additional constituents, so that the speaker's intended meaning is arrived at. Let's consider the following (unfortunate) scenario: John and Mary are painting a wall; John leaves temporarily; Mary falls from the ladder on which she was standing and begs for help; John runs towards her; once he has arrived, John says "I am here". By responding to Mary's request for help with "I am here", John will likely not intend to communicate (just) his geographical location (which is obvious to both interlocutors), but rather his willingness to help Mary (which is in turn intended to have the effect of comforting her). Therefore, John intends to communicate a global message akin to the following: "I am here (to help you)". By uttering "I am here" in that context, John is generating a particularized implicature ("to help you"), which Mary can infer from the context. "To help you" is part of what the speaker means without being part of what he says. "I am here", uttered in that context, has the additional meaning of "to help you" by virtue of the fact that it violates the second submaxim of the maxim of quantity ("Do not make your contribution more informative than

is required"). In fact, it would be over-informative for John to communicate his physical location when Mary is clearly aware of it;

2) in generalized implicatures, pragmatic inferences give rise to an entirely different proposition in the role of speaker's intended meaning, as in "He has been paying a lot of visits to New York lately", where "He might have a girlfriend in New York" is the entirely new proposition that the addressee can infer from the context.

To sum up, "utterances have a sentence-based meaning defined by semantics, and some additional meaning which is rendered by pragmatics" (Jaszczolt, pp. 207 and 208). According to Grice, this additional meaning takes the form of an implicature. Implicatures can be either particularized or generalized, and are generated when the speaker intentionally violates any of the maxims or submaxims of rational conversation. This brief overview of Grice is useful to our discussion on speech acts in that it provides us with two key notions: speaker meaning and implicature. Firstly, the idea of speaker meaning, which is at the foundation of the speech act theory, moves our attention from the structural, abstract analysis of the sentence being uttered to the speaker's communicative intentions behind the utterance of the sentence. Secondly, the idea of implicature clarifies that speakers sometimes mean something more with respect to what they say, and that meaning can be the result of a negotiation between the speaker and the hearer. We will see below, although not without some reservations, that an "accurate characterization of speech acts builds on Grice's notion of speaker meaning" (Green, 2017) since the performance of every speech act depends on the communicative intentions of the speaker. Moreover, as we will see more in detail in chapter 2, the notion of "indirect speech act" is similar in many respects to that of implicature: implicatures can easily be reanalyzed as indirect speech acts and vice versa.

## **2.2 Austin and the Speech Act Theory**

In the study of the "beyond saying", Austin (1962) concentrates on the *use* that the speakers make of utterances. His preliminary observation is that words can be used to do different things, such as asserting, suggesting, promising, persuading, arguing, and so forth. Moreover, the use of words does not only depend on their literal meaning, but also on what the speaker intends to perform with those words, as well as the social setting where the linguistic activity takes place (Korta & Perry, 2015). A speech act is an action, or act, that we perform through speaking: we perform the speech act of asserting when we utter a sentence with the intention of making an assertion, we perform the speech act of suggesting when we utter a sentence with the intention of making a suggestion, and so on and so forth. That being said, it is sometimes not sufficient for the speaker to intend to perform a certain speech act in order for that speech act to be performed

successfully. This is because some speech acts need to conform to a number of societal, group-specific conventions in order to take place (more below). These observations are the ideological foundation of the speech act theory, a theory of language use that focuses on the definition of general principles to capture the mapping between (types of) utterances and (types of) actions. The origin of the speech act theory can be made to coincide with the publication of Austin's monograph "How to Do Things with Words" in 1962. In this work, Austin elaborates ideas often associated with the later work of Ludwig Wittgenstein, whose main tenet is that "the meaning of a word is its use in the language" (Wittgenstein, 1953, §43). This Wittgensteinian research is embodied in the works of the so-called Ordinary Language Philosophy group, of which Austin was the most important representative. This research outlook investigates "meaning as use" (Wittgenstein, 1953), and is primarily interested in the role of speaker meaning for a theory of language and communication.

Speech acts, as we said, rely on the context in that their successful (or felicitous) performance depends on the satisfaction of a number of conditions that are contextual in nature. As we mentioned in page 1 (see footnote), we can distinguish two types of context: the subjective or cognitive context, made up of beliefs and intentions, internal to the speakers, and the objective context, made up of objective physical and metaphysical states of affairs, external to the speakers (Penco, 1999). The successful performance of a speech act depends on conditions that are both internal and external to the speakers, belonging respectively to the subjective and to the objective context. Internal contextual conditions are essentially a matter of belief and intention: if the condition that the speaker has a certain belief or intention is satisfied, then the performance of the speech act is successful. To make a couple of examples: by asserting, the speaker expresses his or her intentions to make the addressee believe that his or her sentence is true and/or his or her belief that the sentence is true; by giving orders, the speaker expresses his or her desire, intention, or wish that the addressee bring about the truth of the sentence; by promising, the speaker expresses his or her intention to bring about him- or herself the truth of the sentence and the belief that he or she is committed to do so by that utterance (Kissine, 2013, p. 4). The successful performance of every speech act is also dependent on a number of objective or external contextual factors. External contextual conditions are in a certain sense more heterogeneous than internal conditions as they include both physical states of affairs - roughly speaking, the reality perceptible through the senses, as well as present and past events - and metaphysical states of affairs, constituted by the conventions, peculiar to certain groups, that are in force or "invoked" for the performance of particular types of speech acts, what Strawson (1964) calls conventional or institutional speech acts. Such societal conventions arguably apply, at least to a certain extent, to other types of speech acts

which are usually not considered institutional speech acts per se, first and foremost the class of commissives (see Sperber and Wilson, 1995; more on speech acts types and classes in chapter 3). As we will see, institutional speech acts are culture-dependent and therefore cannot be analyzed in cognitive, intra-cultural terms, i.e. in terms of speaker's intention.

Strawson (1964) distinguishes between conventional (or institutional) and non-conventional (or non-institutional) speech acts. This distinction can be summarized as follows: "Understanding that an utterance amounts to a conventional speech act (...) requires knowing that certain conventions, peculiar to a certain group, are in force. By contrast, in order to recognise a non-conventional illocutionary act, it is sufficient (...) to grasp a certain multi-layered Gricean communicative intention" (Kissine, 2013, p. 2). In other words, while the successful performance of non-institutional speech acts depends solely on the subjective or cognitive context, i.e. on the intentions and beliefs of the speakers performing those acts, the successful performance of institutional speech acts also depend on a system of "rule- or convention-governed practices and procedures of which they essentially form parts" (Strawson, 1964, p. 457). One example is the utterance "I baptize you John", which counts as baptizing only if it is uttered conforming to certain group-specific conventions, that is to say: uttered by the priest as a "fixed and essential part to play within the frame of (the) ritual (of baptism)" (Kissine, 2013, p. 3). It must be noted that one can perform an institutional speech act also without making it explicit; for example, the speaker can appoint the addressee by saying "You are now Treasurer of the Corporation" instead of saying "I (hereby) appoint you Treasurer of the Corporation" (Green, 2017; more on explicitness in chapter 3). Austin (1962), who first formulated the speech act theory, focuses for the most part (but not exclusively) on institutional speech acts, reasoning on the conventional conditions that need to be met in order for the speaker to successfully perform speech acts such as naming a ship and indulging in marriage. He argues that not any speaker has the role or authority to name a ship or indulge in marriage as their successful performance depends on a number of cultural or group-specific norms, procedures, sanctions, habits, and practices, which must be in force and accepted not only by the interlocutors but also by society at large. As a consequence, one cannot name a ship simply by uttering "I name this ship the Queen Elizabeth" (Austin, 1962, p. 116), nor indulge in marriage by uttering "I do", despite being one's intention to do so. The condition that the speaker has the authority or is in the position within a certain ritual frame, recognized by society, to name a ship or indulge in marriage is a necessary condition for the successful performance of said institutional speech acts: if such condition is satisfied, then the institutional speech act can be performed successfully. If the speaker fails to perform a certain speech act because any of the necessary cultural, group-specific conditions is not met, the speech act is said to misfire: the speaker

has "performed an act of speech but no speech act" (Green, 2017). A speech act can misfire also in the absence of the appropriate uptake; for example, one cannot succeed in betting unless the interlocutor accepts the bet (Green, 2017). Institutional speech acts are equivalent to Searle's (1969) declarations or declaratives (more in chapter 3).

Searle, Vanderveken (Searle, 1969; Searle & Vanderveken, 1985), and Bach and Harnish (1979), as opposed to Austin (1962), focus instead on the subjective contextual conditions, internal to the speakers, that need to be satisfied for the successful performance of non-institutional or non-declarative speech acts. Their works revolve around the notion of speaker meaning as they are deeply influenced by Grice's intention-based and inferential view of communication (Sbisà, 2002). Their main tenet is that "the success of the speech act (qua communicative illocutionary act) is defined in terms of the recognition of the speaker's communicative intention by the hearer" (Sbisà, 2002, p. 422): a speech act is successful if the speaker intends to perform that speech act and the hearer recognizes that intention. To be even more specific, this intention-based view of speech acts, instead of focusing on speech acts as moves in the "language game", investigates the parallels between speech acts and states of mind. As we saw, by asserting a proposition, the speaker expresses his or her belief that that proposition is true, and by promising, the speaker expresses his or her intention to bring about a future state of affairs. We can find evidence of the relationship between what the speaker expresses and what the speaker thinks in the fact that the following utterances would be absurd: "It's raining, but I don't believe that it is", and "I promise to come to the party, but I have no intention of doing so" (Green, 2017). These utterances are nonsensical because, by asserting and promising, the speaker communicates his or her states of mind, respectively of belief and intention, but then proceeds to explicitly deny them. Asserting without believing and promising without intending are examples of so-called abuses. We call a speech act an abuse if it is performed but is still less than successful; for example, if the speaker promises to come to the party but has not the intention of doing so, he or she is not being sincere and his or her speech act is therefore an abuse (Green, 2017).

To conclude the discussion on institutional versus non-institutional speech acts, we must acknowledge the fact that the influence of Grice's intention-based view of communication on the speech act theory can also be seen in Austin (1962), especially in the first half of lecture IV (pp. 39 - 45), where he discusses about the intentions of the interlocutors to engage in certain procedures. Austin's overall prevailing emphasis is, however, on the objective metaphysical contextual requirements behind speech acts (Sbisà, 2002). That being said, because of the impracticality of detecting institutional speech acts due to their cross-cultural volatility, we will concentrate our efforts on analyzing speech acts that are, generally speaking, independent of group-specific

conventions and that can thus be explained to a satisfactory extent in intra-cultural terms, thanks to the Gricean notion of speaker meaning.

According to Searle and Vanderveken (1985), speech acts are the minimal units of human communication: whenever a speaker produces an utterance with the intention of communicating something, he or she performs a speech act (or more than one; more in chapter 2). On this premise, utterances can be redefined in a number of ways: as either "specific events, the intentional acts of speakers at times and places" (Korta & Perry, 2015), or "full-blown speech acts, performed on a specific occasion by a specific speaker with specific communicative intentions" (Leezenberg, 2001, p. 98), or again more broadly as "acts of doing something through speaking, or speech acts" (Jaszczolt, 2002, p. 294). Austin (1962) identifies three different types of acts that are connected with performing every single speech act: locutionary (the act of uttering a sentence with a certain sense and reference), illocutionary (the act of performing an action or a function<sup>12</sup>), and perlocutionary (the act of exerting an influence on the hearer). This trichotomy is not real, but merely theoretical (Jaszczolt, 2002). In fact, as Austin (1962) himself points out, every genuine speech act always subsumes all the three types of acts (Austin, 1962, p. 147). Therefore, every speech act is at the same time:

- locutionary in that it involves the speaker uttering something meaningful (it is not merely a physical or mental act);
- illocutionary in that it is intentionally performed by the speaker to serve a specific function or to perform a specific action; and
- perlocutionary in that it will inevitably trigger a reaction or influence on the hearer; human communication is inherently multidirectional, i.e. it is aimed at the sharing and modification of messages between two or more participants (Hymes, 1974).

Despite serving a theory-internal role, this distinction is however useful to demonstrate the dynamicity of speech acts and their dependence on conversational interaction: speech acts depend on the intentions of the speaker and on their interpretation by the hearer (Jaszczolt, 2002). By uttering a meaningful sentence (locution), the speaker performs an action - or more than one - through speaking (illocution - illocutionary force), which in turn has the effect of triggering a reaction or influence on the hearer (perlocution - perlocutionary effect). More specifically, a "locutionary act (...) is roughly equivalent to uttering a certain sentence with a certain sense and

---

<sup>12</sup> We use the terms "action" and "function" as synonyms to denote the things that people do with language.

reference, which again is roughly equivalent to 'meaning' in the traditional sense<sup>13</sup>" (Austin, 1962, p. 108). Searle and Vanderveken (Searle, 1969; Searle & Vanderveken, 1985), in their proposition-centric view of the speech act theory, call locutionary acts "propositional acts" - i.e. the acts of expressing a proposition - since, according to them, locutionary meaning can be equated with the proposition. In isolation, locutionary meaning is in fact as abstract and communicatively (or pragmatically) inert as the proposition, both being devoid of any intrinsic illocutionary force. Locutionary meaning and the proposition become communicatively significant when they are used in conversation, by virtue of the intentions of the speaker and of their interpretation by the hearer. By uttering a meaningful sentence, I may argue, warn, make a request, inform, etc., according to the use that I intend to make of that sentence, and in turn "by arguing I may persuade or convince someone, by warning him I may scare or alarm him, by making a request I may get him to do something", etc. (Searle, 1969, p. 25).

We have said that speech acts serve functions that reflect the intention of the speaker. For this reason, they can be classified in terms of the function they perform; a few examples of what such functions could be are the following (from Jaszczolt, 2002. p. 295):

- to convey information
- to ask for information
- to give orders
- to make requests
- to make threats
- to give warnings
- to make bets
- to give advice
- to make a promise
- to complain
- to thank

A terminological clarification is in order. Austin (1962) himself uses the terms "speech acts" and "illocutionary acts" (or "illocutions") as synonyms, thus equating the "speech act" with one of its three dimensions (Kissine, 2013). Following the same logic, "to illocute" is nowadays commonly used as a verb meaning "to perform a speech act" (Green, 2017). Austin (1962) also introduces the term "illocutionary force". This term comes from the colloquial question "What is the force of those words?" which we may ask to our interlocutor when we want know how the meaning of his or her

---

<sup>13</sup> We will see in chapters 3 and 4 that the speaker can successfully perform a speech act even without uttering a complete and meaningful sentence.

sentence is to be taken (Green, 2017); for example, by uttering a meaningful sentence such as (from Green, 2017):

8. You'll be more punctual in the future.

the speaker does not make clear whether he or she is making a prediction, issuing a command, or making a threat. In other words, even though we understand those words' literal meaning we still do not know how that meaning is to be taken (Green, 2017). Asking "What is the force of your words?" will indeed clarify whether that meaning is to be taken as a prediction, a command, or a threat. For this reason, besides being identifiable in terms of the function they perform, speech acts can be also seen as locutions having a certain force (Austin, 1962), such as the force of a question, the force of a request, and so on (Jaszczolt, 2002). We have not elucidated yet why we are concerned with illocutionary force in the first place and not, say, decibel level. As Green (2017) points out, semantic content underdetermines other components of the utterance, such as decibel level. However, illocutionary force, unlike decibel level, is a component of speaker meaning. Illocutionary force "is a feature not of what is said but of how what is said is meant; decibel level, by contrast, is a feature at most of the way in which something is said" (Green, 2017). We will see in chapter 3 that the illocutionary force of an utterance can be broken down into a number of components that determine it.

At this point, while we have explained what illocutionary force is and why it is of our interest, we still have to justify why perlocutionary effects are not held to the same standard. We say that a speech act has a perlocutionary effect and *not* a perlocutionary purpose in that perlocutionary effects do not necessarily involve a voluntaristic-intentional component; for example, a speech act can have the perlocutionary effect of being offensive even if it was not the intention of the speaker to offend anyone. Nonetheless, there could also be the case in which the speaker actually intends to offend the addressee. In this sense, perlocutionary acts are much more abstract than illocutionary acts since they can be the characteristic aim of an illocution but are not themselves illocutions. As Green (2017) points out: while I can both urge and persuade you to shut the door, I can urge just by saying "I hereby urge you to shut the door" but in no circumstances I can persuade just by saying "I hereby persuade you to shut the door". This is because urging is an illocutionary act, whereas persuading is a perlocutionary effect. We can say that perlocutions, as opposed to illocutions, are in some sense more volatile, which makes them more difficult to detect and classify (Jaszczolt, 2002). For these reasons, it seems more efficient to analyze communication from the perspective of illocutions, and to classify speech acts according to their illocutionary forces, or illocutionary points (more in chapter 3), rather than attempting the less tangible task of classifying and predicting their possible effects (perlocutions) on the addressee.

Perlocutions must not be confused with indirect speech acts either: an indirect speech act, as the name suggests, is a speech act that is performed indirectly by virtue of the performance of another direct or literal speech act. In this case, both the direct and the indirect speech act belong to the *necessarily voluntaristic-intentional* illocutionary dimension. For example, the speaker can ask the literal question "Can you pass me the salt?" to indirectly make a request to the addressee to pass him or her the salt. The speaker performs an indirect speech act, in addition to a given literal or direct speech act, only if he or she intends to do so, and not as a perlocutionary effect of his or her literal act. That being said, the intention of the speaker needs to be feasibly discernible by the addressee; for example, the speaker cannot perform the literal speech act "It's raining" with the intention of making an indirect request to pass the salt and expect his or her utterance to be interpreted as intended. This is because the intention of the speaker must be made manifest in some way (Green, 2017). It is thus clear that the speaker, in order to be understood, needs to provide what Green (2017) calls "evidence justifying an inference to the best explanation", in such a way that literally asking whether the addressee can pass the salt will result in that utterance to be interpreted as an indirect request to pass the salt. As Green (2017) points out, "[t]hese considerations suggest that indirect speech acts (...) can be explained within the framework of conversational implicatures - that process by which we mean more (and on some occasions less) than we say". What the speaker means is different from what the speaker says if the speaker intentionally generates an implicature - or intentionally performs an indirect speech act - by providing evidence to the addressee that is sufficient for him or her to justify the inference of a different meaning than the meaning conveyed literally. In this sense, Searle's account of indirect speech acts is couched in terms of conversational implicature (Green, 2017).

### 3. An Introduction to Indirect Speech Acts

Having introduced Grice's and Austin's works and the terminology they use (in particular Grice's notion of "implicature" and Austin's notion of "speech act"), we can now refine our preliminary definition of "far-side pragmatics" as that part of pragmatics concerned with "what speech acts are performed in or by saying what is said, or what implicatures are generated by saying what is said" (Korta & Perry, 2015) in a specific context. As a matter of fact, many speech acts (if performed "indirectly") can be easily re-analyzed as implicatures, and vice versa. Let's consider the following example (from Wayne, 2014):

9a. Alan: Are you going to Paul's party?

9b. Barb: I have to work.

Barb implicates, but does not say, that she is not going to the party; that she is not going is her implicature (Wayne, 2014). "Implicating is what Searle (...) called an indirect speech act. Barb performed one speech act (meaning that she is not going) by performing another (saying that she has to work)" (Wayne, 2014). As we have seen in section 2.1, according to Grice, uttering a sentence with the intention of violating one of the maxims of rational conversation generates an implicature, i.e. makes available to the addressee some additional, non-literal meaning that can be inferred from the context. The speech act theory can be thought of going one step further as it investigates if and how that additional meaning influences the use that the speaker makes of that utterance. In other words, by intentionally violating one of the maxims of rational conversation the speaker can modify the *use* of an utterance, and thus the speech act that he or she performs by uttering it. As we will see more in detail below and in chapter 2, the speaker always performs a speech act which is tied to the semantic content of the utterance and, under certain circumstances, an additional speech act which is contextually generated (like Gricean implicatures). Contextually generated speech acts are always meant to overshadow the semantically generated speech acts from which they arise<sup>14</sup>.

Let's now consider the following utterances (produced in the context in which the two interlocutors are seated at the same table; adapted from Searle, 1975):

10a. Please, pass me the salt.

10b. Can you pass me the salt?

10c. Can you reach the salt?

The speaker utters sentences 10b and 10c to violate the maxim of relation: the speaker implicates either a different action to be applied to the same propositional content, i.e. implicates 10a by uttering 10b, or a different action to be applied to a different propositional content, i.e. 10a by uttering 10c. The speaker makes a request by way of making a question, and the question may or may not have a different propositional content than the request. It is in fact clear that, in a certain context, the speaker does not want to receive a yes/no answer about the addressee's ability to pass or reach the salt, nor wants the addressee to reach the salt without passing it. Instead, the speaker expects the addressee to perform the action of passing the salt. We can easily reanalyze 10b and 10c as indirect speech acts: in that context, the speaker can utter 10a, 10b, or 10c, indifferently, to perform the same speech act of making a polite request for action: to pass the salt. However, while 10a is literally a request for action, 10b and 10c are *literally* questions - or requests for information

---

<sup>14</sup> We need to bear in mind that an utterance by itself does not perform a speech act, but rather the speaker does by using that utterance in conversation.

(they request a yes|no answer) - and *contextually* requests for action. Who utters 10b and 10c is said to perform an indirect speech act of a polite order.

Ideally, every utterance requires that the context is investigated in order to determine with precision what speech act it performs, or speech acts if one is performed indirectly. Nonetheless, as we mentioned above, we must acknowledge the fact that there are elements of natural language which can be used as indicators that the utterance of a sentence containing those elements corresponds to a certain (type of) action or speech act. In the literature, these indicators of natural language are referred to as "speech devices" (Austin, 1962) or "illocutionary force indicating devices" (Searle and Vanderveken, 1985). Illocutionary force indicating devices cannot be used reliably on their own to determine illocutionary forces or speech act types. We will talk more in detail about speech devices in section 5 of this chapter. In chapter 3, we will clarify what we mean by speech act type or class. In chapter 2, we will focus on indirect speech acts and attempt to analyze them as a gradable category, that is we will divide them into conventional, semi-conventional, and non conventional indirect speech acts (Benincà et al. 1977); we will see how and to which extent we can leverage the context to determine that one speech act is performed by means of another (like 10a by means of 10b or 10c above).

We conclude this section on indirect speech acts by opening a brief parenthesis on speech act classification. We need to point out the fact that the speech act performed contextually or indirectly is of our interest *only* if it is of a different type - or if it has a different illocutionary force, or belongs to a different class - with respect to the speech act performed literally. This varies from classification to classification<sup>15</sup>. While some classifications include a large number of classes, where each class is defined in detail, other classifications have few coarse-grained classes. Let's consider the exchange of above, which we report here as 11a and 11b (from Wayne, 2014):

11a. Alan: Are you going to Paul's party?

11b. Barb: I have to work.

11b is literally an assertion (semantically unrelated to the previous utterance and to the context in general) and contextually a negative answer (pragmatically related to the previous utterance and to the context in general). If the classification (or tag-set) does not include "negative answer" as a possible type of speech act, it will not be able to capture the distinction between the literal and the indirect speech acts performed by 11b. As we will see in chapter 3, neither Austin's nor Searle's classifications distinguish answers from assertions.

---

<sup>15</sup> A classification, or tag-set, as they are often called in computational linguistics, is an arbitrary list of all possible types of speech acts.

Indeed we could dive deeper into the differences and similarities between Gricean pragmatics and the speech act theory, as well as between implicatures and indirect speech acts. However, for the purposes of the present work, while we treasure the contributions of Grice to contemporary pragmatics, we will focus our attention on Austin's work and on the works of his successors (in particular Searle, 1969; Searle & Vanderveken, 1985). In fact, we deem the speech act theory to feature a very effective hands-on bag of notions that will enable us to bridge the gap between utterances and actions. We will continue to talk about indirect speech act in chapter 2. The next sections of this chapter will further clarify the main properties of speech acts and of their successful performance.

#### **4. Illocutionary Logic: F and P**

While Austin (1962) claims that every speech act consists in the simultaneous performance of a locutionary, an illocutionary, and a perlocutionary act, Searle (1969) claims that every speech act is composed of an illocutionary force and a propositional content to which it is applied. The work of Searle and Vanderveken (1985) draws upon, or is a more up-to-date version of, Searle's (1969) proposition-centric view of the speech act theory. Searle and Vanderveken (1985) attempt a formalization of the theory of speech acts by proposing what they called "illocutionary logic". According to them, illocutionary acts have a logical form that determines their conditions of success. On their definition, "an illocutionary act consists of an illocutionary force F and a propositional content P" (Searle & Vanderveken, 1985, p. 1) and have the following symbolism: F(P). According to Searle (1969), "whenever two illocutionary acts contain the same reference and predication, provided that the meaning of the referring expression is the same, (...) the same proposition is expressed" (p. 29). In this regard, we must bear in mind that some statements, for example existential statements, have no reference (Searle, 1969); for example the utterance "there is a cat" does not point to any specific cats in the context. Finally, we must notice that "not all illocutionary acts have a propositional content, for example, an utterance of "Hurrah" does not, nor does "Ouch"" (Searle, 1969, p. 30).

Limiting ourselves (for now) to those illocutionary acts that do have a propositional content and a reference, let's see how an utterance can be broken up into propositional content (the embedded description of a state of affairs) and illocutionary force (reflecting the action performed on the propositional content). To explain the difference between the role of the two variables P and F, Searle and Vanderveken (1985, p. 1) make the following examples:

12a. You will leave the room.

12b. Leave the room!

13a. Are you going to the movies?

13b. When will you see John?

Utterances 12a and 12b share the same propositional content P (you will leave the room) but differ in terms of their illocutionary force F: 12a has the force F of a prediction and 12b has the force F of an order. Conversely, utterances 13a and 13b have the same force F of questions but differ in terms of their propositional content P (you go to the movies vs. you see John), i.e. they ask two different questions. A similar case is the following (from Green, 2017):

14a. Is the door shut?

14b. Shut the door!

14c. The door is shut.

These utterances have in common the same proposition (the door is shut), which is queried in 14a, commanded (to be true) in 14b, and asserted in 14c (Green, 2017). It is thus clear that many possible propositional contents can have the same illocutionary force, and many possible illocutionary forces can be applied to the same propositional content. Let's now consider the following utterances (from Searle, 1969, p. 22):

15a. Sam smokes habitually.

15b. Does Sam smoke habitually?

"In uttering any of these the speaker refers to or mentions or designates a certain object Sam, and he predicates the expression 'smokes habitually' (or one of its inflections) of the object referred to" (Searle, 1969, p. 23). By referring to Sam and predicating "smokes habitually" of him, i.e. by expressing the proposition that Sam smokes habitually, the speaker performs two different speech acts: an assertion in 15a, and a question in 15b. Searle (1969) maintains that "[p]ropositional acts (the acts of referring and predicating) cannot occur alone; that is, one cannot just refer and predicate without making an assertion or asking a question or performing some other illocutionary act" (p. 25). In the case of assertions, for example, the proposition by itself is not the assertion: "a proposition is *what* is asserted in the act of asserting [emphasis added]" (Searle, 1969, p. 29). By asserting, the speaker is committing him- or herself to the truth of the proposition (Searle, 1969). As Green (2017) points out: "merely expressing the proposition (...) is not to make a move in a 'language game'. Rather, such a move is only made by putting forth a proposition with an illocutionary force such as assertion, conjecture, command, etc.". Along these lines, in the case of questions, the proposition is *what* is questioned; in the case of requests, the proposition is *what* is requested, and so on. To sum up, "[w]hen a proposition is expressed it is always expressed in the performance of an illocutionary act" (Searle, 1969, p. 29).

Let's now consider the roles of F and P in a complex sentence. Searle argues that "clauses beginning with "that..." (...) are a characteristic form for explicitly isolating propositions" (Searle, 1969, p. 29). The utterance:

16. I assert that Sam smokes habitually.

is in a certain pragmatic sense - but not in a truth-conditional sense (as we saw in section 1) - equivalent to 15a ("Sam smokes habitually"). In fact, by uttering either of these sentences, the speaker asserts the same proposition. However, in 16, the proposition is explicitly isolated from the complete speech act by the employ of a that-clause. Another thing that the speaker makes explicit in 16 is the illocutionary force of the utterance by employing a so-called illocutionary force indicating device, in particular what Austin (1962) calls a performative verb (more in section 5). In conclusion, we can say that, in order to capture the global message intended by the speaker, "it is not sufficient (...) simply to assign propositions (...) to sentences" (Searle and Vanderveken, 1985, p. 7) in that speakers can perform different actions by expressing the same proposition. Instead, assuming that "every complete sentence, even a one-word sentence, has some indicator of illocutionary force" (Searle & Vanderveken, 1985, p. 7), we need focus on identifying illocutionary force, by taking advantage of both linguistic and contextual evidence.

## **5. Performative Utterances and Illocutionary Force Indicating Devices**

Before delving into illocutionary force indicating devices, we dedicate a few lines to performative utterances and illocutionary denegation so as to demonstrate how an illocutionary force can be made explicit by a single element - a so-called performative verb - and how such illocutionary force can be explicitly negated. Performative verbs are illocutionary force indicating devices that only occur in a particular kind of sentences called performative sentences. A performative sentence underlies a performative utterance and always contains a main verb "in the first person, present tense, indicative mood, active voice, (and) describ(ing) its speaker as performing a speech act" (Green, 2017). A few examples of performative sentences are:

17. I assert that he is not to blame.

18. I apologize for the misunderstanding.

19. I promise to do it.

Jaszczolt (2002) explains how the logical form of illocutionary acts works by discussing the so-called "illocutionary denegation" on performative sentences. Illocutionary denegations are complex acts in which negation is used to deny the illocutionary force, rather than the propositional content, of a given utterance (Jaszczolt, 2002); 20a exemplifies a case of illocutionary denegation,

whereas 20c is an instance of ordinary sentential negation (from Jaszczolt, 2002. p. 299; in logic, the symbol " $\neg$ " indicates negation; note that in illocutionary logic F takes P as its argument):

20a. I do not promise to do it.

20b.  $\neg F(P)$

20c. I promise not to do it.

20d.  $F(\neg P)$

As Searle and Vanderveken (1985) assert, "an act of illocutionary denegation is one whose aim is to make it explicit that the speaker does not perform a certain illocutionary act" (p. 4). Illocutionary denegations can be achieved by negating a performative verb (as in 20a) or by using a performative verb of denegation; for example, "forbid" and "prohibit" correspond to the denegations of "permit", "refuse" is the denegation of "accept", and "disclaim" is the denegation of "claim" (Jaszczolt, 2002, p. 300).

The notions of illocutionary force and of illocutionary force indicating devices have been subject to a number of critiques. As an early critique of Austin's (1962) notion of illocutionary force, Cohen (1964) argues that illocutionary force is superfluous since we already have at hand the notion of a sentence's meaning, which, according to him, already determines illocutionary force. Cohen's (1964) conclusion can be summarized as follows: "meaning already guarantees force and so we do not require an extra-semantic notion to do so" (Green, 2017). Let's consider the following utterance:

21. I promise to come to your birthday party.

According to Cohen (1964), the literal meaning of this utterance already guarantees that it is a promise (Green, 2017). Cohen (1964) continues by saying that the same applies to utterances that are not performative, such as "I will come to your birthday party", in which case the promise is implicit in the sentence's meaning (Green, 2017). Similarly to Cohen (1964), Searle (1969) claims that, as Green sums up, "some locutionary acts are also illocutionary acts, and infers from this in turn that for some sentences, their locutionary meaning determines their illocutionary force" (Green, 2017). While it is true that a serious and literal utterance of "I hereby promise to climb the Eiffel Tower", made under the contextual conditions that guarantee its success, counts as a promise, it would be a non sequitur to infer from this that some locutionary acts are also illocutionary acts, i.e. that a sentence's locutionary meaning can determine the illocutionary force with which it was uttered (Green, 2017). The locutionary meaning or propositional content of an utterance cannot determine its illocutionary force as illocutionary force is determined by locutionary meaning *together with* contextual factors (Green, 2017), i.e. propositional content *plus* a number or contextual conditions being met. Bearing in mind that locutionary meaning by itself cannot

determine illocutionary force, we can still say that 21 "is designed to be used to make promises, just as common nouns are designed to be used to refer to things and predicates are designed to characterize things referred to" (Green, 2017). In addition to this, just like locutionary meaning underdetermines illocutionary force, conversely illocutionary force underdetermines locutionary meaning: "just from the fact that a speaker has made a promise, we cannot deduce what she has promised to do" (Green, 2017).

To sum up, the conclusions drawn by both Cohen (1964) and Searle (1969) ignore the fact that literal meaning or propositional content *by itself* cannot determine illocutionary force. As a consequence, a performative sentence is nothing more than a type of sentence, which can be uttered without actually performing a speech act (Green, 2017). Green (2017) makes the example of someone uttering in their sleep "I hereby promise to climb the Eiffel Tower", which clearly does not constitute a valid promise, nor would it constitute a valid promise if it was uttered without the speaker intending to be sincerely committed to that action (it would in fact be an abuse). We can thus say that, while a performative utterance must always have as its linguistic form a performative sentence, not every utterance of a performative sentence constitutes the performance of the speech act that is suggested by the performative verb; for example, the performative verb "promise" suggests, but does not guarantee, the performance of a promise. Green (2017) thus defines a performative utterance as "an utterance of a performative sentence that is also a speech act". That being said, we will not discard Cohen's (1964) and Searle's (1969) views completely: while on the one hand locutionary meaning underdetermines illocutionary force, on the other hand some locutionary acts are actually also illocutionary acts if they are backed by the speaker's intention to perform them literally (Green, 2017), plus the satisfaction of a number of other contextual conditions. As we said, it is not true that the speaker can perform any speech acts by uttering any sentences whatsoever so long as those sentences are backed by the speaker's intention. It is difficult to envisage a situation in which the speaker can utter "I do not promise to come" or "I apologize for the inconvenience" with the intention to perform the speech act of promising, and actually perform the promise successfully.

As we have mentioned above, the elements of natural language that can be used as the indicators (or, more appropriately, hints) that an utterance of a sentence containing those elements has a certain illocutionary force are called "illocutionary force indicating devices" (Searle & Vanderveken, 1985). We have seen the employment of one of such devices in 22, where the verb "promise" makes explicit the making of a promise. Searle (1969) writes the following on illocutionary force indicating devices: "the illocutionary force indicator shows how the proposition is to be taken, or to put it another way, what illocutionary force the utterance is to have; that is, what

illocutionary act the speaker is performing in the utterance of the sentence. Illocutionary force indicating devices in English include at least: word order, stress, intonation contour, punctuation, the mood of the verb, and the so-called performative<sup>1</sup> verbs" (p. 30). Searle (1969, p. 31) goes on to say that "in natural languages illocutionary force is indicated by a variety of devices, some of them fairly complicated syntactically". Austin's (1962) "pragmatic" view of illocutionary force opens us to consider the analysis of more complex cases where it is significantly more difficult to identify the force *F* of an utterance since *F* depends on the context. As Searle (1969) himself points out "[o]ften, in actual speech situations, the context will make it clear what the illocutionary force of the utterance is, without its being necessary to invoke the appropriate explicit illocutionary force indicator" (p. 30). In the next chapter, and in particular in chapter 2, we will examine more in depth how the context can be used to retrieve the illocutionary force of an utterance. For now, we limit ourselves to explaining *why* illocutionary force indicating devices are not sufficient and therefore the context has to be consulted.

Searle and Vanderveken (1985) point out that there are many possible illocutionary forces that do not have a corresponding performative verb, nor even a corresponding illocutionary force indicating device. Jaszczolt (2002) phrases it as follows: there are many ways to perform a speech act with a certain illocutionary force without using a corresponding verb or without using any other direct indicators available for its identification. At the same time, non-synonymous verbs may name the same force, which means that two non-synonymous illocutionary verbs do not necessarily name two different illocutionary forces (Searle & Vanderveken, 1985); for example the non-synonymous "mutter" and "shout" name the same illocutionary force in that they are both used to make assertions despite being different in terms of features connected to their utterance act. Moreover, even if ideally every element of natural language is a speech device, the distinction has to be made between performative verbs and the other elements of natural language (what Austin (1962) calls "more primitive devices"). In fact, performative verbs are to a larger extent bound up with specific illocutionary forces if compared to the other elements of language. In chapter 3, we will see that Austin (1962) identifies performative verbs as the most advanced devices for performing speech acts and as the most reliable indicators of illocutionary force. Other natural language indicators of illocutionary force, on the other hand, such as word order and modals, are more implicit and thus more difficult to associate systematically with particular illocutionary forces. That being said, a number of contextual conditions also apply in order for a speech act to be of a particular type. A promise, not only must be sincere, but it also must be beneficial to the addressee in order to be a promise. This means that a promise such as "I promise that I will hit you" is actually not a promise but it is in fact a threat. Its logical structure is thus: threat(I will hit you) despite containing the

performative verb "promise". This demonstrates that illocutionary force and propositional content are related and that the utterance needs to be analyzed in its entirety in order to accurately assessing its force: illocutionary force is, to a certain extent, dependent on the propositional content. We will see more in detail contextual conditions of these kinds in chapter 2

As Green (2017) points out, "[j]ust as content underdetermines force and force underdetermines content; so too even grammatical mood together with content underdetermine force". We have demonstrated it by pointing out that "You'll be more punctual in the future", despite being in the indicative mood, is not necessarily a prediction but it can also be a command or a threat, depending on the context. On the other hand, we need to acknowledge that mood and the other illocutionary force indicating devices play a role in influencing our final assessment on the type of speech act that has been performed. Green (2017) continues, "grammatical mood is of the devices we use, together with contextual clues, intonation and the like to indicate the force with which we are expressing a content". At the end of the day, an utterance in the indicative mood is a prediction rather than a command if it efficaciously manifests the intention of the speaker to be so taken (Green, 2017). In other words, there exist no infallible indicators of illocutionary force because there are no conventions that make the utterance of a particular expression unequivocally the performance of a certain illocutionary act (Green, 2017). That being said, we can summarize by saying that natural language contains devices that indicate illocutionary force conditional upon the speaker's intention to use them with that particular force (Green, 2017). As we will see in chapters 2 and 3, the fact that the context needs to be investigated for the determination of the illocutionary force of an utterance will raise a number of problems in the detection of speech acts performed by computers. The automatic detection of speech acts in fact detects the illocutionary force of an utterance solely on the basis of its linguistic form and a portion of the discourse (a few preceding and succeeding utterances). It would in fact be impossible to verify other elements of the context the way humans do (more on chapter 3).

## **6. Conclusion**

In the light of our observations, pragmatics can be redefined as that branch of linguistics and philosophy that deals with "regularities in language use that are guided by speaker's intentions" (Leezenberg, 2001, p. 98). We believe that linguistic expressions have the meanings that they have by virtue of their use in conversation. In this regards, we are also aware of the cultural differences that come into play in the performance of certain types of speech acts. For this reason, a distinction has been made between speech acts that, generally speaking, "do not depend on any group-specific

convention" (Kissine, 2013, p. 2), such as constatives, directives, and commissives, and speech acts that do depend on said cultural conventions, such as declaratives or institutional speech acts. Searle and Vanderveken's (1985) illocutionary logic - with its distinction between illocutionary force and propositional content - despite being particularly helpful for understanding the logical form of speech acts, it is arguably overly concerned with detail when it comes to automated speech act detection (more in chapters 3 and 4). Our ultimate goal is to be able to systematically and automatically map speech act types (or categories, classes) to utterances (or utterance types) in discourse. The theoretical foundations of the speech act theory, despite being useful for understanding what speech acts are, will slowly fade away in the next chapters to make room for the different implementations of the speech act theory in computational linguistics. As we will see, using the speech act theory as a theoretical background for studies in computational linguistics has led to a number of adaptations. We will argue that only the notion of speech act has survived, and, in particular, only the notion of illocutionary point.

## CHAPTER 2 - INDIRECT SPEECH ACTS

The purpose of this chapter is to provide the reader with an in-depth account of indirect speech acts. Firstly, we will focus on the conditions of success - or felicity conditions - that underlie the performance of speech acts, with a particular focus on the successful performance of promises. Since the same conditions of success are shared by all the speech acts - both direct and indirect - with the same force, felicity conditions will become crucial in our parallel analysis of direct and indirect speech acts. Secondly, we will clarify the notion of indirect speech act through the work of Searle (1975): we will examine the circumstances under which indirect speech acts are performed and discover how we can leverage the context to identify their type. Finally, we will focus on the different degrees of conventionality of indirect requests for action thanks to the contribution of Benincà et al. (1977). Conventionality of use, as we will see, is a spectrum: while there is strong linguistic evidence of the performance of conventional indirect speech acts, there is little to no linguistic evidence of the performance of non conventional indirect speech acts.

### 1. Felicity Conditions

Before diving into indirect speech acts, we deem it necessary to focus on the fact (already mentioned in chapter 1) that *every* utterance, even if it has explicit indicators of its illocutionary force, needs to satisfy a number of conditions that are pragmatic in nature in order to have a certain force. We saw that, as a consequence, illocutionary force indicating devices are not sufficient, on their own, to determine illocutionary force. In chapter 1, we focused for the most part on the intentions and beliefs of the speaker behind the performance of non-institutional speech acts. We saw that, by asserting, the speaker expresses his or her belief that that the proposition is true, and by promising, the speaker expresses his or her intention to bring about a future state of affairs. The beliefs and intentions of the speaker are a necessary condition for the successful performance of the speech acts, respectively, of asserting and promising. Searle (1969) calls the beliefs of the speaker the "sincerity condition" for asserting, and the intentions of the speaker the "sincerity condition" for promising. More specifically, according to Searle (1969), there is total of nine conditions - one of which is the sincerity condition - that are necessary (and as a set sufficient) for the successful performance of most<sup>16</sup> speech acts. He calls them "felicity conditions" (Searle, 1969). In this section, we will focus on the successful performance of promises and therefore our analysis will revolve around the felicity conditions for promises. Out of the nine conditions of success for

---

<sup>16</sup> We will see that some speech acts, e.g. greeting, have fewer conditions of success.

promises, three apply to all speech acts (and not just promises), and six are peculiar to promises. The six felicity conditions characteristic of promises in turn boil down to four conditions, namely: propositional content condition, preparatory condition, sincerity condition, and essential condition. Jaszczolt (2002) summarizes how these conditions are met (when the speech act performed is that of a promise) as follows (p. 296; below a more in-depth analysis of all conditions):

"in the case of a promise there has to be a sentence used with the content of a promise (this is the propositional content condition), the promise must be about an event beneficial to the addressee, otherwise it would be a warning or a threat, and about an event that is not going to happen anyway (preparatory condition). (...) The intentions of the promiser are also relevant (sincerity condition), as well as the awareness of putting oneself under an obligation to perform the action (essential condition)."

These four conditions are shared by all the speech acts - both direct and indirect - with the force of a promise. In other words, every promise, regardless of whether it is performed directly or indirectly, must satisfy all of the felicity conditions above in order to be successful. As we said, since the same conditions of success are shared by all the speech acts - both direct and indirect - with the same force, felicity conditions will become crucial in our parallel analysis of direct and indirect speech acts. We will see more in detail below how we can leverage felicity conditions to identify indirect speech acts. For now, our concern is that of giving an accurate description of each of the nine felicity conditions for promises.

In Searle's (1969) words: "Given that a speaker S utters a (grammatical well-formed) sentence T in the presence of a hearer H, then, in the literal utterance of T, S sincerely and non-defectively<sup>17</sup> promises that p to H if and only if the following conditions 1-9 obtain" (Searle, 1969, pp. 56-57). We summarize Searle's (1969, pp. 57-61) felicity conditions for the performance of a promise as follows:

- 1) S and H speak the same language, are conscious, have no physical impediments to communications, and are not acting or playing;
- 2) S expresses the proposition that p in the utterance of T, which isolates the proposition from the rest of the speech act;

---

<sup>17</sup> Speech acts that satisfy all the conditions of success except for any of the preparatory conditions are sometimes considered successful but defective (Searle & Vanderveken, 1985); for example, asserting without sufficient evidence for the truth of the proposition, or promising something that will happen regardless of the promise. In the present work, we will not give them special treatment and consider them simply as unsuccessful. Similarly, as we have seen in chapter 1, speech acts that satisfy all the conditions of success except for the sincerity condition are sometimes called abuses (as a particular type of unsuccessful speech acts); for example, asserting without believing the truth of the proposition, or promising without intending to fulfill the promise. We will not treat abuses differently from the other types of failures. More on unsuccessful speech acts below.

3) In expressing that p, S predicates a *future* act A of S, which means that the scope of the illocutionary force indicating device includes certain features of the proposition: the act must be predicated of the speaker and cannot be a past act;

---- Conditions 2 and 3 are what Searle calls *propositional content conditions* ----

4) H would prefer S's doing A to his not doing A, and S believes H would prefer his doing A to his not doing A, that is to say: a promise needs to be beneficial to the addressee and both S and H need to recognize it as such, or else it would be a threat (a promise is a pledge to do something for you, not to you); also, a promise needs some sort of occasion or situation whose crucial feature is that the promisee wishes (needs, desires, etc.) something to be done, or else it would be an invitation;

5) It is not obvious to both S and H that S will do A in the normal course of events (the act must have a point), that is to say: if S promises to do something that it is obvious to all concerned that he or she is going to do anyhow, or that is going to happen regardless of the act, then the act is pointless;

---- Conditions 4 and 5 are what Searle calls *preparatory conditions* ----

6) S intends to do A, which makes A a sincere promise.

---- Condition 6 is what Searle calls *sincerity condition* ----

7) S intends that the utterance of T will place him under an obligation to do A, that is to say: the essential feature of a promise is that it is the undertaking of an obligation to perform a certain act.

---- Condition 7 is what Searle calls *essential condition* ----

8) The speaker intends to produce a certain illocutionary force by means of getting the hearer to recognize his intention to produce that force, and he also intends this recognition to be achieved in virtue of the fact that the meaning of the item he utters conventionally associates it with producing that force. In the case of a promise, the speaker assumes that the semantic rules (which determine the meaning) of the expressions uttered are such that the utterance counts as the undertaking of an obligation. The rules, in short, as we shall see in the next condition, enable the intention in the essential condition 7 to be achieved by making the utterance. And the articulation of that achievement, the way the speaker gets the job done, is described in condition 8;

9) The semantical rules of the dialect spoken by S and H are such that T is correctly and sincerely uttered if and only if conditions 1-8 obtain. This condition is intended to make clear that the sentence uttered is one which, by the semantical rules of the language, is used to make a promise. The meaning of a sentence is entirely determined by the meaning of its

elements, both lexical and syntactical. And that is just another way of saying that the rules governing its utterance are determined by the rules governing its elements<sup>18</sup>.

At this point, Searle (1969) extracts from the conditions above five rules for the use of any illocutionary force indicating device for promising. Since conditions 1, 8, and 9 apply to most illocutionary acts and are not peculiar to promising, Searle focuses on conditions 2 to 7. For simplicity, we can equate Pr with the performative verb "promise", but Pr ideally stands for any indicator of illocutionary force for promising. The rules Searle (1969) defines are the following (p. 63):

Rule 1. Pr is to be uttered only in the context of a sentence (or larger stretch of discourse) T, the utterance of which predicates some future act A of the speaker S. I call this the propositional content rule. It is derived from the propositional content conditions 2 and 3.

Rule 2. Pr is to be uttered only if the hearer H would prefer S's doing A to his not doing A, and S believes H would prefer S's doing A to his not doing A.

Rule 3. Pr is to be uttered only if it is not obvious to both S and H that S will do A in the normal course of events. I call rules 2 and 3 preparatory rules, and they are derived from the preparatory conditions 4 and 5.

Rule 4. Pr is to be uttered only if S intends to do A. I call this the sincerity rule, and it is derived from the sincerity condition 6.

Rule 5. The utterance of Pr counts as the undertaking of an obligation to do A. I call this the essential rule.

Now that we have laid out the felicity conditions for promises and extracted from them a set of rules that account for the form of behavior of making promises, we can consider a few examples of unsuccessful promises and go through the reasons why they failed. All of the utterances below, except for 22a, 22c, and 22g (which are successful promises), do not meet (at least according their linguistic form) at least one of the conditions of success for promises.

22a. I promise I will come.

22b. I promise I will hit you.

22c. I promise I will come, and I really intend to.

22d. I promise I will come, but I have no intention to.

22e. I promise that the sun will rise tomorrow.

22f. I promise I came.

22g. I promise I will come, and I undertake the obligation to come.

---

<sup>18</sup> With regards to these last two conditions, we will see below that the speaker can get the hearer to recognize his or her intention to produce a certain illocutionary force not only in virtue of the conventional literal meaning of the sentence uttered, but also in virtue of the conventions of use in place for that sentence.

22h. I promise I will come, but I do not undertake the obligation to come.

Despite the fact that both 22a and 22b contain the illocutionary force indicating device "promise" (a performative verb), while 22a has the force of a promise, 22b has the force of a warning or a threat. This conclusion can be partially drawn from the linguistic form of the utterance: we can in fact assign to "come" the semantic property of being beneficial and to "hit" that of not being beneficial to the hearer. If this is the case, 22b is not a promise in that it does not satisfy one of the preparatory conditions for promises; the speaker in fact violates Rule 2. That being said, we still have no means to determine whether the hearer actually finds it beneficial that the speaker will come or (though less likely) non beneficial to be hit. One can for example say "I promise I will hit you" in the context of "if that's what's necessary to bring you back to consciousness" and actually make a promise (and not a threat) to hit somebody. It is easier, on the other hand, to imagine a context in which the hearer does not want the speaker to come (to an event, to a trip, to a birthday party, and so on) in such a way that 22a becomes a threat instead of a promise. 22a and 22b are further evidence of the fact that linguistic form underdetermines illocutionary force as they unravel the ineffectiveness of binding performative verbs to illocutionary forces. As Jaszczolt (2002, p. 302) points out: "the verb is not a reliable guide to the type of the speech act". In addition to this, even if we correctly assign to "come" the semantic property of being beneficial, we still do not know, from the utterance's linguistic form alone, whether the speaker utters 22a sincerely (and thus really intends to make a promise). In other words, we have no linguistic means to determine whether the speaker respects the sincerity rule (Rule 4). Even if the speaker made explicit his or her intentions, such as in 22c or (in an interesting nonsensical way) 22d, we still would not know whether the speaker is being sincere in externalizing his or her intentions. It is thus clear that factual background information (including information as to whether the speaker is trustworthy) becomes necessary to determine the sincerity behind 22a. Moving on, utterance 22e is not a promise in that, just like 22b, it does not meet one of the preparatory conditions: it is about an event that is going to happen anyway, whether or not the speaker commits to it. By uttering 22e, the speaker is in violation of Rule 3. 22f, on the other hand, cannot be a promise because it does not satisfy the propositional content condition: the proposition of a promise cannot be in the past tense. The speaker thereby violates Rule 1. We need to precise that 22f, despite not being a promise, is not nonsensical: it can in fact be interpreted as the expression of a strong belief of the truth of the propositional content on the part of the speaker, which makes it an assertion roughly equivalent to "I swear I came". The last two utterances are examples of the speaker making it explicit that the essential condition is (22g) and is not (22h) satisfied. By uttering 22h, the speaker violates Rule 5 (again, in an interesting nonsensical way). For both 22g and 22h, we have no means to determine whether the speaker is

being sincere. Finally, one can argue that, by uttering 22c or 22g, the speaker might intentionally be making his or her contribution more informative than it is required thus violating the maxim of quantity (Grice, 1975), in such a way as to communicate that he or she will not come. Of course, intonation plays an important role in the performance of 22c or 22g.

At this point, we deem it useful to extend our analysis, although very briefly, beyond the speech act of promising, by considering how felicity conditions apply to other speech acts, such as ordering, asserting, and greeting. Doing so, will indeed help us see the big picture. With regards to the felicity conditions for giving orders, Searle (1969, p. 64) writes: "[t]he preparatory conditions include that the speaker should be in a position of authority over the hearer, the sincerity condition is that the speaker wants the ordered act done, and the essential condition has to do with the fact that the speaker intends the utterance as an attempt to get the hearer to do the act". With regards to assertions he writes: "the preparatory conditions include the fact that the hearer must have some basis for supposing the asserted proposition is true, the sincerity condition is that he must believe it to be true, and the essential condition has to do with the fact that the proposition is presented as representing an actual state of affairs" (Searle, 1969, p. 64). Finally, if we consider the "much simpler kind of speech act" (Searle, 1969, p. 64) of greeting, and in particular of the utterance of "Hello", Searle (1969) writes: "there is no propositional content and no sincerity condition. The preparatory condition is that the speaker must have just encountered the hearer, and the essential rule is that the utterance counts as a courteous indication of recognition of the hearer" (pp. 64-65). For the conditions of success of more speech acts, see Searle, 1969, pp. 66-67.

We conclude this section on felicity conditions with Searle's (1969) general hypotheses about speech acts. His hypotheses can be seen as a further development of the points he made thus far about the felicitous performance of speech acts. We summarize Searle's general hypotheses as follows (Searle, 1969, pp. 65-71):

1. Wherever there is a psychological state specified in the sincerity condition, the performance of the act counts as an *expression* of that psychological state. Thus to assert, affirm, state (that p) counts as an expression of belief (that p). To request, ask, order, entreat, enjoin, pray, or command (that A be done) counts as an expression of a wish or desire (that A be done). To promise, vow, threaten or pledge (that A) counts as an expression of intention (to do A). To thank, welcome or congratulate counts as an expression of gratitude, pleasure (at H's arrival), or pleasure (at H's good fortune).
2. The converse of the first law is that only where the act counts as the expression of a psychological state is insincerity possible. One cannot, for example, greet or christen insincerely, but one can state or promise insincerely.

3. Where the sincerity condition tells us what the speaker *expresses* in the performance of the act, the preparatory condition tells us (at least part of) what he *implies* in the performance of the act. To put it generally, in the performance of any illocutionary act, the speaker implies that the preparatory conditions of the act are satisfied. Thus, for example, when I make a statement I imply that I can back it up, when I make a promise, I imply that the thing promised is in the hearer's interest. When I thank someone, I imply that the thing I am thanking him for has benefited me (or was at least intended to benefit me), etc.

4. It is possible to perform the act without invoking an explicit illocutionary force-indicating device where the context and the utterance make it clear that the essential condition is satisfied. I may say only "I'll do it for you", but that utterance will count as and will be taken as a promise in any context where it is obvious that in saying it I am accepting (or undertaking, etc.) an obligation. Seldom, in fact, does one actually need to say the explicit "I promise". Similarly, I may say only "I wish you wouldn't do that", but this utterance in certain contexts will be more than merely an expression of a wish, for, say, autobiographical purposes. It will be a request. And it will be a request in those contexts where the point of saying it is to get you to stop doing something, i.e., where the essential condition for a request is satisfied. This feature of speech - that an utterance in a context can indicate the satisfaction of an essential condition without the use of the explicit illocutionary force-indicating device for that essential condition - is the origin of many polite turns of phrase. Thus, for example, the sentence, "Could you do this for me?" in spite of the meaning of the lexical items and the interrogative illocutionary force-indicating devices is *not* characteristically uttered as a subjunctive question concerning your abilities; it is characteristically uttered as a request [emphasis added].

5. Wherever the illocutionary force of an utterance is not explicit it can always be made explicit. Of course, a given language may not be rich enough to enable speakers to say everything they mean, but there are no barriers in principle to enriching it.

6. The overlap of conditions (among different speech acts) shows us that certain kinds of illocutionary acts are really special cases of other kinds; thus asking questions is really a special case of requesting, viz., requesting information (real question) or requesting that the hearer display knowledge (exam question). This explains our intuition that an utterance of the request form, "Tell me the name of the first President of the United States", is equivalent in force to an utterance of the question form, "What's the name of the first President of the United States?". It also partly explains why the verb "ask" covers both requests and questions, e.g., "He asked me to do it" (request), and "He asked me why" (question).

7. In general the essential condition determines the others. For example, since the essential rule for requesting is that the utterance counts as an attempt to get H to do something, then the propositional content rule has to involve future behavior of H.

8. The notions of illocutionary force and different illocutionary acts involve really several quite different principles of distinction. First and most important, there is the point or purpose of the act (the difference, for example, between a statement and a question); second, the relative positions of S and H (the difference between a request and an order); third, the degree of commitment undertaken (the difference between a mere expression of intention and a promise); fourth, the difference in propositional content (the difference between predictions and reports); fifth, the difference in the way the proposition relates to the interest of S and H (the difference between boasts and laments, between warnings and predictions); sixth, the different possible expressed psychological states (the difference between a promise, which is an expression of intention, and a statement, which is an expression of belief); seventh, the different ways in which an utterance relates to the rest of the conversation (the difference between simply replying to what someone has said and objecting to what he has said). Because the same utterance act may be performed with a variety of different intentions, it is important to realize that one and the same utterance may constitute the performance of several different illocutionary acts. There may be several different non-synonymous illocutionary verbs that correctly characterize the utterance. For example suppose at a party a wife says "It's really quite late". That utterance may be at one level a statement of fact; to her interlocutor, who has just remarked on how early it was, it may be (and be intended as) an objection; to her husband it may be (and be intended as) a suggestion or even a request ("Let's go home") as well as a warning ("You'll feel rotten in the morning if we don't").

9. Some illocutionary verbs are definable in terms of the intended perlocutionary effect, some not. Thus requesting is, as a matter of its essential condition, an attempt to get a hearer to do something, but promising is not essentially tied to such effects on or responses from the hearer.

While all of Searle's general hypotheses about speech acts are - though to different extents - relevant to our discussion on indirect speech acts, we are particularly interested in hypothesis 4. Here, Searle (1969) discusses indirect speech acts, and in particular indirect promises and indirect requests. He observes that the speaker can make a promise or a request without necessarily using explicit indicators of illocutionary force, as long as the context makes it clear that what is uttered counts as either the undertaking of an obligation (promise) or as an attempt to get the hearer to do

something (request), i.e. as long as the essential condition is satisfied (Searle, 1969). In the appropriate context, "I'll do it for you" can thus be taken as a promise, and "I wish you wouldn't do that" as a request (Searle, 1969, p. 68). The remainder of this chapter focuses almost exclusively on the indirect performance of requests for action because of the literature that already exists on the subject (will will extent our analysis to other types of speech acts in the next chapters). In the next section, we will examine the inferential steps that the hearer goes through to determine: 1) that the speaker has performed an indirect speech act, and 2) the type of indirect speech act that the speaker has performed.

## 2. A Parallel Analysis of Direct and Indirect Speech Acts

We have already come across indirect speech acts on different occasions in chapter 1. We have seen that the speaker can perform a speech act indirectly by virtue of another; for example, one can indirectly make the request "Please, pass me the salt" by virtue of directly, or literally, asking a question with the same propositional content "Can you pass me the salt?" or even a question with a different propositional content "Can you reach the salt?". We have also seen that, in such cases, the intervention of pragmatics is necessary to retrieve the actual force of the utterance as it is impossible to grasp what the speaker globally means from the literal meaning of the sentence in isolation. Searle (1975) introduces the notion of indirect speech act as follows (p. 59):

The simplest cases of meaning are those in which the speaker utters a sentence and means exactly and literally what he says. In such cases the speaker intends to produce a certain illocutionary effect in the hearer (...), and he intends to get the hearer to recognize this intention in virtue of the hearer's knowledge of the rules that govern the utterance of the sentence. But notoriously, not all cases of meaning are this simple: In hints, insinuations, irony, and metaphor - to mention a few examples - the speaker's utterance meaning and the sentence meaning come apart in various ways. One important class of such cases is that in which the speaker utters a sentence, means what he says, but also means something more. For example, a speaker may utter the sentence *I want you to do it* by way of requesting the hearer to do something. The utterance is incidentally meant as a statement, but it is also meant primarily as a request, a request made by way of making a statement. In such cases a sentence that contains the illocutionary force indicators for one kind of illocutionary act can be uttered to perform, IN ADDITION, another type of illocutionary act. There are also cases in which the speaker may utter a sentence and mean what he says and also mean another illocution with a different propositional content. For example, a speaker may utter the

sentence *Can you reach the salt?* and mean it not merely as a question but as a request to pass the salt.

We can reformulate Searle's point as follows. There are two types of utterances: 1) utterances by which the speaker means literally what he or she says, by which the speaker generates (more or less explicitly) one single illocutionary force that is recognizable thanks to the knowledge of the literal meaning of the words being used, and 2) utterances whose literal illocutionary force is overshadowed by an additional indirect force which can only be retrieved from the context. Utterances of the second type are said to be used to perform indirect speech acts: they have a literal use (what Searle (1975) calls secondary illocutionary act), which is tied to the linguistic form of the utterance, and a non-literal use (what Searle (1975) calls primary illocutionary act), which needs to be inferred from the context and ultimately takes effect. Searle continues by saying (1975, pp. 60-61): "In indirect speech acts the speaker communicates to the hearer more than he actually says by way of relying on their mutually shared background information, both linguistic and nonlinguistic, together with the general powers of rationality and inference on the part of the hearer". Searle (1975) specifies that the apparatus necessary for understanding indirect speech acts is composed of the speech act theory, Gricean maxims of cooperative or rational conversation, factual information about the world, and about the speaker and the hearer, and the inferential ability of the hearer.

Let's now consider the following exchange (from Searle, 1975, p. 61) - which, in some respects, is similar to examples 11a and 11b of chapter 1 - and reconstruct the inferential steps that the hearer goes through to derive the indirect illocution from the literal illocution:

23a. A: Let's go to the movies tonight.

23b. B: I have to study for an exam.

By uttering 23a, speaker A makes a proposal by virtue of the utterance's literal meaning, in particular the meaning of "Let's". By uttering 23b, speaker B rejects the proposal of A by virtue of the context, rather than the utterance's literal meaning. In fact, speaker B's literal utterance of sentence 23b would instead constitute a statement. In order to derive the indirect rejection of the proposal (indirect illocution) from the literal statement (direct locution), one unconsciously goes through the following steps (from Searle, 1975, p. 63, our comment will follow):

STEP 1: I have made a proposal to B, and in response he has made a statement to the effect that he has to study for an exam (facts about the conversation).

STEP 2: I assume that B is cooperating in the conversation and that therefore his remark is intended to be relevant (principles of conversational cooperation).

STEP 3: A relevant response must be one of acceptance, rejection, counterproposal, further discussion, etc. (theory of speech acts).

STEP 4: But his literal utterance was not one of these, and so was not a relevant response (inference from Steps 1 and 3).

STEP 5: Therefore, he probably means more than he says. Assuming that his remark is relevant, his primary illocutionary point<sup>19</sup> must differ from his literal one (inference from Steps 2 and 4).

STEP 6: I know that studying for an exam normally takes a large amount of time relative to a single evening, and I know that going to the movies normally takes a large amount of time relative to a single evening (factual background information).

STEP 7: Therefore, he probably cannot both go to the movies and study for an exam in one evening (inference from Step 6).

STEP 8: A preparatory condition on the acceptance of a proposal, or on any other commissive<sup>20</sup>, is the ability to perform the act predicated in the propositional content condition (theory of speech acts).

STEP 9: Therefore, I know that he has said something that has the consequence that he probably cannot consistently accept the proposal (inference from Steps 1, 7, and 8).

STEP 10: Therefore, his primary illocutionary point is probably to reject the proposal (inference from Steps 5 and 9).

Our first observation is that Grice's Cooperative Principle, and his intention-based and inferential view of communication play a strong role in the derivation of indirect speech acts. Speaker B is *not* being irrational or non-cooperative, he or she is just intentionally not being relevant so as to communicate that he or she does not want to be taken literally. In other words, speaker B, by intentionally violating the maxim of relation (see section 2.1 of chapter 1), is providing evidence for the hearer to justify a non-literal interpretation of his utterance: speaker B's utterance has a primary indirect illocutionary point (that needs to be inferred) in addition to a secondary literal illocutionary point. Our second observation, which is also that of Searle (1975), is that the conclusion that speaker B's primary illocutionary point is that he is rejecting the proposal of speaker A is a *probabilistic* conclusion in that his reply does not *necessarily* constitute a rejection. In fact, speaker B could have instead replied (from Searle, 1975, p. 64):

23c. B: I have to study for an exam, but let's go to the movies anyhow.

This demonstrates that the hearer needs to establish two things (Searle, 1975, p. 64):

- 1) that the primary indirect illocutionary point departs from the literal illocutionary point;
- 2) what the primary indirect illocutionary point is.

---

<sup>19</sup> The illocutionary point of an utterance is its purpose or goal in conversation (more in chapter 3).

<sup>20</sup> A commissive, as we have seen, is a type of speech act whose illocutionary point is to commit the speaker to a future course of action (more in chapter 3).

Searle (1975) goes on to say that indirect illocutionary acts can be studied effectively within the area of directives<sup>21</sup> because the conversational requirements of politeness make indirect requests (such as 24a and 24b) a frequent alternative to direct requests performed by blunt imperative sentences (such as 24c) and explicit performatives (such as 24d):

24a. I wonder if you would mind leaving the room.

24b. Could you please leave the room?

24c. Leave the room!

24d. I order you to leave the room.

As we will see, Benincà et al. (1977) focus on directives too, and in particular on requests for action.

With regards to understanding indirect directives, Searle (1975) points out that "[t]he problem is made more complicated by the fact that some sentences seem almost to be conventionally used as indirect requests" (p. 60). In fact, it would be difficult to image a situation in which the sentence "I would appreciate it if you would get off my foot" is not uttered as a request but as a statement (Searle, 1975, p. 60). As a consequence, we can make a list of the sentences that could - to use Searle's (1975) terminology - standardly, ordinarily, normally, or conventionally be used to make indirect requests. In turn, these sentences can be divided into different categories roughly (but not exactly) according to the condition of success for requesting that they question or assert (we will lay out the conditions of success for directives more in detail below) (Searle, 1975). For example, one of the conditions for a request to be successfully performed is that the hearer is able to perform the action requested by the speaker: questioning the hearer's ability to perform that action constitutes an indirect request to the hearer to perform that action (e.g. "Can you reach the salt?"). Another condition for a request to be successfully performed is that the speaker wants or has a reason for the hearer to perform the action requested: stating that reason is, too, an indirect request to the hearer to perform that action (e.g. "You're standing on my foot"). Questioning the hearer's ability to perform the action requested or stating the reason behind the action requested, in the appropriate contexts, violate the Gricean maxim of relation, thus signaling to the hearer that the utterance has an additional indirect illocutionary point. The hearer can understand the type of the indirect illocutionary point by leveraging the conditions of success for speech acts (more below). A few examples of sentences that could be used "quite standardly" to make indirect requests and orders are the following (Searle, 1975, pp. 65 to 67):

GROUP 1: Sentences concerning the hearer's ability to perform the action requested:

---

<sup>21</sup> A directive is a type of speech act whose illocutionary point is to get the hearer to bring about a future state of affairs (more in chapter 3).

Can you reach the salt?

Can you pass the salt?

Could you be a little more quiet?

You could be a little more quiet.

Have you got change for a dollar?

GROUP 2: Sentences concerning the speaker's wish or want that the hearer will do the action requested:

I would like you to go now.

I want you to do this for me, Henry.

I would/should appreciate it if you would/could do it for me.

I hope you'll do it.

I wish you wouldn't do that.

GROUP 3: Sentences concerning the hearer's doing the action requested:

Officers will henceforth wear ties at dinner.

Would you kindly get off my foot?

Won't you stop making that noise soon?

GROUP 4: Sentences concerning the hearer's desire or willingness to do the action requested:

Would you be willing to write a letter of recommendation for me?

Do you want to hand me that hammer over there on the table?

Would you mind not making so much noise?

GROUP 5: Sentences concerning reasons for doing the action requested:

You ought to be more polite to your mother.

You should leave immediately.

Must you continue hammering that way?

Ought you to eat quite so much spaghetti?

You had better go now.

Why not stop here?

Why don't you be quiet?

It might help if you shut up.

You're standing on my foot.

How many times have I told you (must I tell you) not to eat with your fingers?

GROUP 6: Sentences embedding one of these elements inside another; also, sentences embedding an explicit directive illocutionary verb inside one of these contexts:

Would you mind awfully if I asked you if you could write me a letter of recommendation?

Would it be too much if I suggested that you could possibly make a little less noise?

Might I ask you to take off your hat?

I hope you won't mind if I ask you if you could leave us alone.

Conventional indirect requests like these are not the same as direct requests because, despite being conventionally used to issue directives, "[t]he sentences in question do not have an imperative force as part of their meaning" (Searle, 1975, pp. 67). This point can be demonstrated by the fact that the speaker can consistently connect the literal utterance of any of these sentences with the denial of any imperative intent (Searle, 1975). In the case of direct requests, on the other hand, denying the imperative intent is not possible. Let's consider the examples above (24a-d) and attempt to deny the imperative intent for each:

25a. I wonder if you would mind leaving the room, Bill, but I am not requesting you to leave the room; I am just wondering if you would mind doing it if I were to ask you.

25b. Could you leave the room? But I am not requesting you to leave the room; I am just asking you if you could do it if I were to ask you.

25c. Could you please leave the room? (IMPOSSIBLE to deny the imperative intent because of the use of "please" which makes it an explicit and literal request or order; see below)

25d. Leave the room! (IMPOSSIBLE to deny the imperative intent because it's a direct request or order)

25e. I order you to leave the room. (IMPOSSIBLE to deny the imperative intent because it's a direct request or order)

Sentences that are conventionally used to indirectly issue directives have a systematic relation with directive illocutions, whereas a sentence such as "I have to study for an exam" (cf. 23b) has no systematic relation with rejecting proposals (Searle, 1975, p. 68). Evidence of the fact that sentences that are conventionally used as indirect requests have a systematic relation with directive illocutions is that most of them can embed "please", which is typical of requests; for example:

I want you to stop making that noise, please.

Could you please lend me a dollar?

The use of "please" makes the sentence an explicit and literal request even though the rest of the sentence does not have the literal meaning of a directive (Searle, 1975). In addition to this, Searle (1975) points out that sentences conventionally used as indirect requests are not idioms, not only because they have literal, word-for-word translation in other languages - although, as we will see, sometimes with a different illocutionary act potential - but also because their use as indirect requests admits literal responses, which presupposes that they are too uttered literally; for example,

"Jones kicked the bucket", an idiom, cannot be translated literally, whereas "Could you help me?" can: "Pourriez-vous m'aider?", "Können Sie mir helfen?", "Potrebbe aiutarmi?", etc. (Searle, 1975, p. 68). In this case, the utterance keeps the same indirect illocutionary act potential across the four languages in that all forms are conventional indirect requests (as we will see below sometimes this does not happen). To address Searle's second point: "Why don't you be quiet, Henry?", being a literal question (or request for information), admits as a literal response "Well, Sally, there are several reasons for not being quiet. First..." (Searle, 1975, p. 68).

We have seen that sentences conventionally used as indirect requests, just like other less conventional indirect requests (but unlike literal requests), can be uttered literally *without* the intent of making indirect requests; for example, "Can you pass the salt?" can be uttered as a question about the hearer's physical abilities; similarly, "I want you to leave" can be uttered as a statement expressing the speaker's wants, devoid of any directive intent (Searle, 1975, p. 69). Nevertheless, these sentences, when they are instead uttered as requests, they are still uttered with and as having their literal meaning, despite being indirect requests by virtue of the context, which can be demonstrated by the fact that their indirect utterance as indirect requests can be followed by responses that are appropriate to them being uttered literally; for example (Searle, 1975, p. 69):

26a. Can you pass the salt?

26b. No, sorry, I can't, it's down there at the end of the table.

26c. Yes, I can. (Here it is).

26a has two potential meanings: it can be either a literal question or a conventional indirect request. In either case, a yes / no answer will be appropriate. Answering with "yes" or "no" is in fact appropriate for 26a's literal meaning, which the utterance retains regardless of whether it is used with its literal force or as an indirect request. Therefore, 26b is the response to 26a uttered as a literal question, and 26c is the response to 26a uttered as an indirect request (but retaining its literal meaning). This means that 26a uttered with the indirect illocutionary point of a request does not alter the fact that its literal illocutionary point is that of a question (or of a statement) (Searle, 1975). This potentially invalidates the claim that, when a sentence is used to perform a nonliteral indirect illocutionary act, the underlying literal illocutionary act is not conveyed (Searle, 1975).

While we have laid out the felicity conditions for promises (and incidentally of commissives), we have not laid out yet the felicity conditions for requests (and directives). Doing so would help us explain why "I have to study for an exam" uttered by B to reject the proposal of A (reported below as 27a and 27b; from Searle, 1975, p. 61) is tied to the conditions of success for commissives (and arguably for rejections) similarly to the way in which sentences that are conventionally used as indirect requests are tied to the conditions of success for directives.

27a. A: Let's go to the movies tonight.

27b. B: I have to study for an exam.

As we have seen before and in chapter 1, each type of illocutionary act has a number of conditions that are necessary for its successful performance (Searle, 1975). Searle (1975) presents the felicity conditions for directives and commissives are follows (p. 71):

	Directive (Request)	Commissive (Promise)
Preparatory condition	The hearer is able to perform the action.	The speaker is able to perform the action.
Sincerity condition	The speaker wants the hearer to do the action.	The hearer wants the speaker to perform the action. The speaker intends to do the action.
Propositional content condition	The speaker predicates a future action of the hearer.	The speaker predicates a future actions of the speaker
Essential condition	Counts as an attempt by the speaker to get the hearer to do the action.	Counts as the undertaking by the speaker of an obligation to do the action.

Now that we have at hand the felicity conditions for directives, we can refine our list of sentences conventionally used as indirect requests (Groups 1 to 6 above) and reduce the 6 Groups we defined to three types (Searle, 1975, p. 71):

- 1) Sentences that have to do with "felicity conditions on the performance of a directive illocutionary act", which include:
  - a) Group 1: preparatory condition (sentences concerning the ability of the hearer to perform the action);
  - b) Group 2: sincerity condition (sentences concerning the desire of the speaker that the hearer performs the action);
  - c) Group 3: propositional content condition (sentences concerning the predication of the action of the hearer);
- 2) Sentences that have to do with "reasons for doing the act", which include:
  - a) Group 4: sentences concerning the hearer's desire or willingness to do the action requested;
  - b) Group 5: sentences concerning reasons for doing the action requested;

3) Sentences "embedding one element inside another one", which include sentences embedding either performative verbs or elements already contained in the other two categories (felicity conditions and reasons).

For now, we focus on the first two of these groups - felicity conditions and reasons - about which Searle (1975, p. 72) makes the following generalizations:

GENERALIZATION 1: the speaker can make an indirect request (or other directive) by either asking whether or stating that a preparatory condition concerning the hearer's ability to do the action obtains.

GENERALIZATION 2: the speaker can make an indirect directive by either asking whether or stating that the propositional content condition obtains.

GENERALIZATION 3: the speaker can make an indirect directive by stating that the sincerity condition obtains, but not by asking whether it obtains.

GENERALIZATION 4: the speaker can make an indirect directive by either stating that or asking whether there are good or overriding reasons for doing the action, except where the reason is that the hearer wants or wishes, etc., to do the action, in which case he can only ask whether (and not state that) the hearer wants, wishes, etc., to do the action.

Searle (1975) asserts that the existence of these generalizations accounts for a systematic relation between sentences conventionally used as indirect requests (Groups 1 to 6 above) and the directive class of illocutionary acts. The rules behind the performance of directive and commissive speech acts consist in the conditions of success listed in the table above; the generalizations that follow are not rules, but rather consequences of the rules that govern the performance of directives (Searle, 1975). The task is now to show how the generalizations are valid consequences of the rules (when considered together with factual background information and Gricean principles of conversation). To do so, Searle (1975) lists what, according to him, are the steps that the speaker unconsciously follows for to derive the conclusion that "Can you pass the salt?" is uttered as a request to pass the salt (and not as a question about the hearer's abilities to pass the salt). His reconstruction of the hearer's inferential process is roughly the following (Searle, 1975, pp. 73-74):

STEP 1: the speaker has asked me a question as to whether I have the ability to pass the salt (fact about the conversation).

STEP 2: I assume that he is cooperating in the conversation and that therefore his utterance has some aim or point (principles of conversational cooperation).

STEP 3: the conversational setting is not such as to indicate a theoretical interest in my salt-passing ability (factual background information).

STEP 4: furthermore, he probably already knows that the answer to the question is yes (factual background information). (This step facilitates the move to Step 5, but is not essential).

STEP 5: therefore, his utterance is probably not just a question. It probably has some ulterior illocutionary point (inference from Steps 1, 2, 3, and 4). What can it be?

STEP 6: a preparatory condition for any directive illocutionary act is the ability of the hearer to perform the act predicated in the propositional content condition (theory of speech acts).

STEP 7: therefore, the speaker has asked me a question the affirmative answer to which would entail that the preparatory condition for requesting me to pass the salt is satisfied (inference from Steps 1 and 6).

STEP 8: we are now at dinner and people normally use salt at dinner; they pass it back and forth, try to get others to pass it back and forth, etc. (background information).

STEP 9: he has therefore alluded to the satisfaction of a preparatory condition for a request whose obedience conditions it is quite likely he wants me to bring about (inference from Steps 7 and 8).

STEP 10: therefore, in the absence of any other plausible illocutionary point, he is probably requesting me to pass him the salt (inference from Steps 5 and 9).

To sum up, Searle reconstructs the inferential process that leads the hearer to conclude that, in the relevant context, "Can you pass the salt?" is actually uttered with the illocutionary point of making a request. Searle (1975) wants to demonstrate that the hearer infers the indirect illocutionary point of request by virtue of the fact that the speaker is asking whether the preparatory condition concerning the hearer's ability to pass the salt obtains. In fact, if we consider an utterance that does not question the satisfaction of any of the preparatory conditions of the illocutionary act of requesting, such as "Where was this salt mined?", it will be impossible (and wrong, or irrational) for the hearer to infer that the speaker is indirectly requesting him or her to pass the salt (Searle, 1975). Put simply, "Can you pass the salt?" is related to (the rules behind) requesting to pass the salt, whereas "Where was this salt mined?" is not. That being said, not all questions about the hearer's abilities are indirectly requests, which means that the hearer needs some way to recognize when "Can you pass me the salt?" is a question about his or her abilities or a request made indirectly by way of asking that question (Searle, 1975). It is at this point that Gricean principles of conversation and factual background information become involved; according to Searle (1975), in two separate steps: 1) establishing the existence of an indirect illocutionary point, and 2) finding out what the indirect illocutionary point is. To quote Searle directly (1975, p. 74): "The first is established by the principles of conversation operating on the information of the hearer and the speaker, and the

second is derived from the theory of speech acts together with background information". In other words, we know that the speaker is performing an indirect speech act if his or her utterance violates any of the Gricean maxims of rational conversation, and we know what *type* of indirect speech act the speaker is performing by determining the type of the speech act whose condition of success the speaker is questioning (or stating). Let's clarify with an example:

28. Can you pass the salt?

The first question that the hearer unconsciously asks him- or herself is the following: "is the speaker, by his or her utterance, intentionally violating any of the maxims of rational conversation?":

- a) If the answer is no, then the speaker is *not* performing an indirect speech act, which means that the utterance has *only one* illocutionary point (retrievable from the utterance's literal meaning);
- b) If the answer is yes, then the speaker is performing an indirect speech act, which means that the utterance has an additional indirect illocutionary point (that needs to be inferred).

If the answer to the first question is "yes", the second question that the hearer unconsciously asks him- or herself is the following: "the speaker is asking whether or stating that one of the conditions of success of *what* particular *type* of speech act obtains?":

The hearer knows that the type of speech act performed indirectly is that of request by virtue of the fact that the utterance is questioning or stating the satisfaction of one of the conditions of success for requests; for example, by uttering 28, the speaker is questioning the preparatory condition concerning the hearer's ability to do the action - i.e. one of the conditions of success for requests. The speaker is therefore performing an indirect request.

With regards to why speakers often perform indirect requests instead of direct ones, Seale (1975) says that politeness is the main motivation behind the use of such indirect forms of request: by phrasing his or her request with "Can you", the speaker not only does not presume to know the hearer's abilities to perform the action requested, but also gives - or appears to give - the option to the hearer of refusing to commit (since it allows a negative answer). On the contrary, direct requests performed by blunt imperative sentences and explicit performatives presume to know the hearer's abilities and do not appear to give the possibility of refusing (Seale, 1975).

At this point, Searle (1975) lists a number of problems that arise with our current framework for understanding indirect speech acts. For example, he says that it is not clear why there are some syntactical forms that work better than others for making indirect requests even though the general mechanisms by virtue of which they are indirect requests in the first place do not have to do with

syntax, but rather with the speech act theory, Gricean principles of conversation, and shared background information (Searle, 1975). He makes the examples of (Searle, 1975, p. 75):

29a. Do you want to do action X?

29b. Do you desire to do action X?

and:

30a. Can you do action X?

30b. Are you able to do action X?

30c. Is it the case that you at present have the ability to do action X?

While it is easy to make a request with sentences such as 29a and 30a, it is not with 29b and 30b, and it is arguably impossible with 30c. In this regards, Searle (1975) notices that we can insert "please" fairly easily in 29a and 30a, but not in the others. Searle (1975) explains this phenomenon by arguing that, within the framework he presented for understanding indirect speech acts, there is room for a number of forms which have acquired conventional uses as polite forms for requests, while keeping their literal meanings. This is made possible by what he calls conventions of usage: forms such as "can you", "could you", "I want you to", have become conventional ways of making requests, but not by virtue of their literal imperative meaning (which they do *not* have), but rather by virtue of their frequency of use as polite requests. This, Searle (1975) continues, would explain why these forms sometimes lose their indirect speech act potential (or their indirect request potential) when they are translated into other languages:

31a. Can you hand me that book?

31b. Můžete mi podat tu knížku?

While 31a will function in English as an indirect request, its Czech translation will sound odd as a request (Searle, 1975). Their indirect request potential is in fact not tied to their literal, inter-translatable meaning, but rather to their frequency of use as indirect requests in each language. While 31a has become conventionally used in the English language as an indirect request, the same cannot be said for 31b in the Czech language.

Searle (1975, p. 76) goes on to explain why some sentences can be used as indirect requests why some others categorically cannot by means of the following maxim of conversation (which he adds to those proposed by Grice):

Speak idiomatically unless there is some special reason not to.

which roughly translates as:

Speak using the forms of a language as they are conventionally used (normal speech) unless there is some special reason not to.

If the speaker violates this maxim by attempting to make an indirect request using a nonidiomatic form such as 30c (instead of the idiomatic 30a), the hearer will reach the conclusion that the speaker is not making an indirect request because, when nonidiomatic forms are used, "the normal conversational assumptions on which the possibility of indirect speech acts rests are in large part suspended" (Searle, 1975, pp. 76-77). To sum up (Searle, 1975):

- 1) the sentences that we can use to make indirect requests must be idiomatic<sup>22</sup>, that is they must belong to "normal speech", which excludes sentence 30c from the candidates;
- 2) the sentences that have become entrenched as conventional forms for making indirect requests should (but need not to) be preferred to those that have not, which means that 29a and 30a should be preferred over 29b and 30b.
- 3) The forms that are selected as conventional forms vary from language to language.

Another problem about which Searle (1975) expresses concern is the asymmetry between the sincerity condition and the other conditions of success: the speaker can in fact perform an indirect speech act by *both* asserting and querying the obtainment of the propositional content and preparatory conditions, but can only assert (and not query) the satisfaction of a sincerity condition.

Let's consider the following examples (from Searle, 1975, p.65 and 77):

32a. I want you to do it.

32b. Do I want you to do it?

33a. Officers will henceforth wear ties at dinner.

33b. Would you kindly get off my foot?

34a. You could be a little more quiet.

34b. Could you be a little more quiet?

32a can be a request, whereas 32b cannot (Searle, 1975). In fact, while 32a is asserting the satisfaction of a sincerity condition, 32b is questioning whether it is satisfied; 32b, as we said, cannot be used to make indirect requests. We can also notice that, while 32a can take "please", 32b cannot. On the other hand, assertions such as 33a and 34a, and questions such as 33b and 34b can all be used to make indirect requests as they involve other conditions of success, namely the propositional content condition (33a and 33b) and the preparatory condition (34a and 34b). A similar asymmetry occurs in the case of reasons: if reason is that the hearer wants or wishes to do the action, unlike for all the other types of reasons, the indirect request can be made *only* by asking whether (and not stating that) the reason is in place (Searle, 1975, p. 77):

35a. Do you want to leave us alone?

---

<sup>22</sup> As we mentioned above, the possibility of a literal, word-for-word translation of 31a into 31b and vice versa, together with the possibility of answering them literally make these sentences idiomatic but *not* idioms.

35b. You want to leave us alone.

35c. You're standing on my foot.

While 35a can be a request, 35b cannot (Searle, 1975). 35b in fact is stating that the hearer wants to do the action, which is not a viable way of making an indirect request. On the other hand, 35c can be a request in that the speaker is stating a reason which does not involve the wants and wishes of the hearer. Searle (1975) points out that the speaker *cannot* make an indirect request by querying the satisfaction of the sincerity condition nor by asserting the wants and wishes of the hearer as "it is odd, in normal circumstances, to ask other people about the existence of one's own elementary psychological states, and odd to assert the existence of other people's elementary psychological states when addressing them. (...) It is, in general, odd for me to ask you about my states or tell you about yours" (p. 77). This asymmetry, Searle (1975) continues, also applies to the indirect performance of other types of speech acts (more below).

Searle's (1975) raises one last problem with regards to his framework for the understanding of indirect speech acts. He finally concerns himself with English syntactical forms. The issue that he raises is that of sentences with the form: "Why not + VERB" as in "Why not stop here?", which, unlike the form: "Why don't you + VERB", has according to him "many of the same syntactical constraints as imperative sentences" (Searle, 1975, pp. 77-78). In fact, both "Why not + VERB" sentences and imperative sentences (Searle, 1975, p. 78):

- require a voluntary verb: the speaker can say "Why not imitate your grandmother", but cannot say "Why not resemble your grandmother?", just like one can say "Imitate your grandmother!", but not "Resemble your grandmother!";
- require a reflexive when they take a second-person direct object: "Why not wash yourself?" just like "Wash yourself!".

Despite these linguistic facts, according to Searle (1975), "Why not + VERB" sentences are *not* imperative in meaning. In asking "Why not stop here?", he continues, the speaker is making a suggestion by challenging the hearer to provide reasons for not doing the action, on the assumption that the absence of reasons for not doing the action is itself a reason for doing it. The speaker thus *indirectly* makes a suggestion by way of alluding to a reason for doing the action (Searle, 1975). To support this claim, Searle (1975) points out that "Why not + VERB" sentences can be uttered literally and accept a literal response, in which case they do not constitute indirect suggestions; for example (p. 78):

36a. A: Why not stop here?

36b. B: Well, there are several reasons for not stopping here. First...

The literal use of 36a as a question or its indirect use as a suggestion are reflected by the way in which they are reported (Searle, 1975, p. 78; note that the use of "should" accounts for the requirement of a voluntary verb):

36c. He suggested that we shouldn't stop there.

36d. He asked me why we shouldn't stop there.

While 36c also reports the illocutionary point of suggestion, 36d does not. Searle (1975) also considers the troublesome use of "would" and "could" in indirect speech acts; for example (p. 78):

37a. Would you pass me the salt?

37b. Could you pass me the salt?

38a. Will you pass me the salt?

38b. Can you pass me the salt?

According to him, it is difficult to describe exactly how 37a and 37b differ in meaning from 38a and 38b. Searle (1975) argues that 37a comes from the sentence:

39a. Would you pass me the salt if I asked you to?

whereas 37b does not because the hearer's abilities are not dependent on the request of the speaker.

37b is likely to come from either of the following (Searle, 1975):

39b. Could you pass me the salt if you please?

39c. Could you pass me the salt if you will?

Moreover, according to Searle (1975), while both 37a and 39a can be used as indirect requests, they have a different illocutionary act potential. We must notice that 37a and 37b also have a direct, literal use (40a and 40b) to which the hearer can respond literally (41a and 41b) (from Searle, 1975, p. 79):

40a. Would you vote for a Democrat?

40b. Could you marry a radical?

41a. Under what conditions?

41b. It depends on the situation.

According to Searle (1975), "would" (like "will") traditionally expresses want or desire, or is a future auxiliary; "could" can be analyzed as "would" + possibility or ability (just like "can" can be analyzed as "will" + possibility or ability), thus 40b is roughly equivalent to:

42. Would it be possible for you to marry a radical?

The fact that "could" and "would" do not have an imperative meaning can be confirmed by the fact that they could have, at the same time, a commissive meaning (Searle, 1975). In fact, the following sentences are normally offers (Searle, 1975, p. 79):

43a. Could I be of assistance?

43b. Would you like some more wine?

Searle (1975) thus concludes that "would" and "could" do not have imperative meaning, nor commissive meaning, in that saying that they have both would involve an "unnecessary proliferation of meanings" (p. 79).

We have seen that the speaker can perform an indirect request by stating (but not questioning) the obtainment of a sincerity condition and by asking whether (but not stating that) the hearer wants or wishes to do the action (the hearer's wants and wishes are among the reasons behind the performance of directive speech acts). We report here the examples we made above (32a and 32b, and 35a and 35b):

44a. I want you to do it.

44b. Do I want you to do it?

45a. Do you want to leave us alone?

45b. You want to leave us alone.

While 44a and 45a can be uttered as indirect requests, 44b and 45b cannot. This asymmetry, according to Searle (1975), also applies to the indirect performance of other types of speech acts. First of all, Searle (1975) points out that the speaker can perform, not just directives (or requests), but any illocutionary act by asserting (and not by questioning) the obtainment of the sincerity condition for that particular act. We recall that the sincerity condition of a speech act is satisfied when the speaker performs a speech act while sincerely expressing his or her psychological state. In chapter 1, we saw that: in order to assert, the speaker must believe that his or her statement is true, in order to promise, the speaker must have the intention of bringing about the propositional content of his or her utterance, and that in order to request, the speaker must desire or want that the hearer brings about the propositional content on his or her behalf. Explicitly stating the satisfaction of the sincerity condition for a particular type of speech act is a way of performing indirectly that particular type of speech act. In other words, the speaker can indirectly perform a speech act by stating that he or she has the psychological state necessary for the successful performance of that particular speech act. Let's consider the following examples (Searle, 1975, pp. 79 - 80) and compare them with their direct counterparts:

46a. I am sorry I did it. (an apology)

46b. I apologize for doing it.

in that being sorry is the sincerity condition for apologizing;

47a. I think/believe he is in the next room. (an assertion)

47b. He is in the next room.

in that thinking or believing is the sincerity condition for asserting;

48a. I am so glad you won. (congratulations)

48b. I congratulate you on winning.

in that being glad is the sincerity condition for congratulating;

49a. I intend to try harder next time, coach. (a promise)

49b. I promise to try harder next time, coach.

in that intending is the sincerity condition for promising;

50a. I am grateful for your help. (thanks)

50b. Thank you for your help.

in that being grateful is the sincerity condition for thanking.

This list can potentially be expanded until it includes all the types of speech acts. In addition to this, we need to point out the fact that, for each illocutionary act type, there is not one but many ways of stating the satisfaction of its sincerity condition. For example, the following sentences (among the others) can be used to state the satisfaction of the sincerity condition for requests (Searle, 1975, p. 65):

I would like you to go now.

I want you to go now.

I would/should appreciate it if you would/could go now.

I hope you'll go now.

I wish you wouldn't stay here.

I'd rather you didn't stay.

Searle (1975) finally focuses on the class of commissives and on their indirect performance (especially offers and promises). He demonstrates that we can build for commissives a similar framework for understanding their indirect performance to the one that we built for directives. Searle (1975) begins his discussion on commissives with a list of sentences that can be uttered to perform indirect offers (or, in some cases, promises); he groups these sentences according to the condition of success of commissives whose satisfaction they state or question (Searle, 1975, pp. 80 - 81):

I. Sentences concerning the preparatory conditions:

A. that the speaker is able to perform the act:

Can I help you?

I can do that for you.

I could get it for you.

Could I be of assistance?

B. that the hearer wants the speaker to perform the act:

Would you like some help?

Do you want me to go now, Sally?

II. Sentences concerning the sincerity condition:

I intent to do it for you.

I plan on repairing it for you next week.

III. Sentences concerning the propositional content condition:

I will do it for you.

I am going to give it to you next time you stop by.

Shall I give you the money now?

IV. Sentences concerning the speaker's wish or willingness to do the action:

I want to be of any help I can.

I'd be willing to do it (if you want me to).

V. Sentences concerning (other) reasons for the speaker's doing the action:

I think that I had better leave you alone.

Wouldn't it be better if I gave you some assistance?

You need my help, Cynthia.

Returning now to the asymmetries that we have analyzed for directives (exemplified in 44a to 45b), we can now assert that such asymmetries apply to commissives too. In fact, the speaker can perform an indirect commissive by asserting (but not questioning) the obtainment of the sincerity condition (i.e. by asserting but not questioning his or her own psychological state) and by asking whether (but not asserting that) the hearer wants or wishes to do the action (i.e. by questioning but not asserting the hearer's psychological state) (Searle, 1975); for example (Searle, 1975, p. 81):

51a. Do you want me to leave?

51b. You want me to leave.

52a. I want to help you out.

52b. Do I want to help you out?

While 51a and 52a can be uttered as offers, 51b and 52b cannot. Searle (1975) mentions the fact that 51b can be an offer if the speaker adds the tag question "don't you", such as in (p. 81):

53. You want me to leave, don't you?

Searle (1975) goes on to say that a large number of hypothetical sentences belong to the class of commissives; to make a few examples (p. 81):

54a. If you wish any further information, just let me know.

54b. If I can be of assistance, I would be most glad to help.

54c. If you need any help, call me at the office.

54d. If it would be better for me to come on Wednesday, just let me know.

Searle (1975, p. 81) notices that "the antecedent concerns either one of the preparatory conditions (54a to c), or the presence of a reason for doing the action (54d)".

In the light of what we said thus far about commissives, Searle (1975) makes the following generalizations, which he adds to the generalizations proposed for the indirect performance of directives (to build a single unified framework) (p. 81):

GENERALIZATION 5: the speaker can make an indirect commissive by either asking whether or stating that the preparatory condition concerning his ability to do the actions obtains.

GENERALIZATION 6: the speaker can make an indirect commissive by asking whether, though not by stating that, the preparatory condition concerning the hearer's wish or want that the speaker do the action obtains.

GENERALIZATION 7: the speaker can make an indirect commissive by stating that, and in some forms by asking whether, the propositional content condition obtains.

GENERALIZATION 8: the speaker can make an indirect commissive by stating that, but not by asking whether, the sincerity condition obtains.

GENERALIZATION 9: the speaker can make an indirect commissive by stating that or by asking whether there are good or overriding reasons for doing the action, except where the reason is that the speaker wants or desires to do the action, in which case he can only state but not ask whether he wants to do the action.

In conclusion, we can say that our analysis of indirect speech acts follows two steps:

- 1) Firstly, we need to infer whether the speaker wants to be taken literally or contextually. By intentionally not being rational or cooperative, i.e. by intentionally violating any of the maxims or rational conversation (Grice, 1975), the speaker performs an indirect speech act;
- 2) Secondly, we need to infer the type of indirect speech act that the speaker performs (is it a directive? a commissive? etc.). To do so, we rely on the speech act theory: out of all the conditions of success for every type of speech act that exists, we need to discover whether the speaker is either stating that or asking whether any of these conditions obtains. If we identify the condition that the speaker is asserting or questioning, we are able to trace back the speech act type that is performed indirectly (since the condition in question is one of the conditions of success for that speech act type).

Searle's (1975) generalizations (1 to 9) guide us through the inferential process for the identification of directives and commissives performed indirectly. Let's consider the following utterance:

55. I want to help you with your assignment.

By uttering 55, the speaker wants to be taken either literally or contextually. We can determine whether the speaker is performing an indirect speech act rather than a literal one, by investigating the interaction between the utterance, Gricean maxims of conversation, and factual background information. For example, if the speaker is in a rush and about to leave (facts about the world) and utters 55 (fact about the conversation), he or she probably wants to be taken literally (the speaker can add "but I really can't" to make it explicit that he or she is just stating what he or she believes to be true without committing to any future actions):

56. I want to help you with your assignment, but I can't.

If on the other hand, the speaker has plenty of time and is very knowledgeable about the subject of the assignment (facts about the world) and utters 55 (fact about the conversation), he or she probably does not want to be taken literally. It would be odd for the speaker to express his or her want or desire to help and being in the condition to help, without offering to help. In this case, the speaker is probably asserting the satisfaction of the sincerity condition for commissives (= the speaker wants to do the action), which means that the speaker is probably indirectly performing a commissive.

That being said, we need to define the conditions of success (and make generalizations from them) about other types of speech acts, and not just commissives and directives, to be able to systematically identify indirect speech acts from utterances. We will attempt this task in the next chapters. In the next section, we will focus on the indirect performance of indirect requests for action in order to learn more about the different degrees of conventionality with which they can be performed.

### **3. Conventional, Semi-conventional, and Non Conventional Indirect Speech Acts**

This section is dedicated to the degrees of conventionality *of use (or usage)* of indirect speech acts, with a particular focus on indirect requests for action. According to Searle (1969), every time the speaker utters a sentence and means it literally, he or she intentionally chooses the expressions of a language that are conventionally connected with a particular literal force. In other words, the linguistic expressions of a language conventionally have a literal illocutionary force, which has a one-to-one correspondence with their literal meaning. In the present section, we are not concerned with conventionality in this sense, but rather with what Searle (1975) calls conventionality of usage: forms such as "can you", "could you", and "I want you to" have become conventional ways of making requests, but not by virtue of their literal meaning (which is *not* that of a request), but rather by virtue of their frequent *use* as polite requests (Searle, 1975). This means

that some sentences whose literal force is not that of a request, but rather that of an assertion or a question, "seem almost to be conventionally used as indirect requests" (Searle, 1975, p. 60); for example, a sentence like "I would appreciate it if you would stop speaking so loudly", while it has the conventional (in the first sense) literal force of an assertion, it is standardly, ordinarily, normally, or conventionally (in the second sense) *used* to make indirect requests (Searle, 1975). Similarly, the oft-quoted "Can you pass me the salt?" is literally a question, but conventionally used as an indirect request. Let's clarify even further with the following examples:

57a. Get off my foot!

57b. I request you to get off my foot.

57c. I would appreciate it if you would get off my foot.

57d. Can you get off my foot?

If the speaker utters either 57a or 57b and means it literally, he or she is performing a literal or direct request because 57a and 57b are requests by virtue of their literal meanings, and in particular: the use of the imperative mood in 57a, and the use of the performative verb "request" in 57b. In the case of 57a and 57b, the speaker intends to get the hearer to recognize his or her intention to make a request by virtue of the hearer's knowledge of the literal meaning of his or her sentences. Linguistic forms such as 57a and 57b provide the speaker with a conventional (in the first sense) means of requesting things to people. On the other hand, if the speaker utters either 57c or 57d and means it literally, he or she is performing, respectively, a literal or direct assertion (57c) and a literal or direct question (57d) by virtue of their literal meanings, and in particular: the use of the indicative mood in 57c, and the use of the interrogative mood in 57d. That being said, 57c or 57d, despite not being requests literally, they are conventionally (in the second sense) used to make requests. This means that, while 57c and 57d can be uttered literally with their conventional (in the first sense) illocutionary force, the speaker can also utter them to make requests by virtue of their conventional use as indirect requests. From now on, we will use the term "conventional" exclusively with the meaning of "conventional in use".

Benincà et al. (1977) expand Searle's (1975) general notion of conventionality of usage to cope with sentences that have different degrees of conventionality. According to them, indirect speech acts fall into three categories: conventional, semi-conventional, and non conventional indirect speech acts (Benincà et al., 1977). Benincà et al. (1977) study the different degrees of conventionality of indirect requests for action in Italian by comparing them to their direct or literal counterparts. In the present section, we will consider direct and indirect requests for action in both Italian and English as similar conclusions can be drawn about these two languages. In summary, we will investigate those cases, in Italian and in English, in which the speaker performs simultaneously

two acts: one literal, whose force is established on the basis of the linguistic indicators of force, and one indirect, whose force is established taking into account the literal act and the context in which it is performed (Benincà et al., 1977, p. 503). We will see that, in certain indirect speech acts (the more conventional ones), there exist, in the literal speech act, some traces or linguistic indicators of force of the indirect speech act (Benincà et al., 1977, p. 503).

Benincà et al. (1977) begin with laying out the felicity conditions for requests for action (p. 505):

1. The speaker cannot (or does not want to) do the action;
2. The speaker thinks that the interlocutor is capable of or can do the action;
3. The speaker thinks that the interlocutor has not yet done the action nor is doing the action;
4. The speaker thinks that the interlocutor can do the action (viz. he or she does not have external impediments);
5. The speaker thinks that the interlocutor has not decided and will not do the action independently of the request;
6. The speaker thinks that the interlocutor will accept and has no reasons for not doing the action;
7. The speaker wants or has a reason for the interlocutor to do the action.

Out of these seven conditions, only the first one is based exclusively on the speaker, whereas the other six involve both the speaker and the interlocutor (Benincà et al., 1977). Benincà et al. (1977) continue by saying that many requestive indirect speech acts in Italian consists of either asserting one of the conditions based on the speaker or questioning one of the conditions based on the interlocutor (p. 507). This is a characteristic of indirect requests (and of indirect speech acts in general) that Searle (1975) noticed in English: in fact, the speaker can make an indirect request either by asserting the satisfaction of the sincerity condition (based on the speaker) or by questioning the wants and wishes of the hearer (based on the interlocutor), but not vice versa. To report Searle's words (1975): "it is odd, in normal circumstances, to ask other people about the existence of one's own elementary psychological states, and odd to assert the existence of other people's elementary psychological states when addressing them. (...) It is, in general, odd for me to ask you about my states or tell you about yours" (p. 77). Benincà et al. (1977, p. 507) make the following examples:

58a. Vorrei che mi venissi a prendere.

58b. Puoi venirmi a prendere?

which translate into English as follows:

58c. I would like you to pick me up.

58d. Can you pick me up?

These sentences are not directly (or literally) requests, but in some contexts work as requests (Benincà et al., 1977, p. 507).

While all indirect requests are tied to the felicity conditions for requests for action reported above, which means that one cannot perform an indirect request by uttering any sentences whatsoever, they can have different degrees of conventionality. Benincà et al. (1977), in fact, distinguish between conventional, semi-conventional, and non conventional indirect requests. "Conventional indirect requests are immediately recognizable as requests for any interlocutor in any context, and the requestive use of such forms can be recognized even when the context is not given or understood" (Benincà et al., 1977, pp. 507-508). According to them, this is the case also because, when it comes to conventional indirect requests, there are often requestive "relics" in the literal speech act used to make the indirect request, in particular (Benincà et al., 1977, p. 508):

- descending (as opposed to ascending) intonation in the interrogatives;
- the possibility to insert "per favore" in Italian or "please" in English;
- in certain cases, the use of the conditional.

The most conventional indirect forms for requests are: "Sai...?" (En. "Can you (ability)...?"), "Puoi...?" (En. "Can you (possibility)...?"), "Sapresti...?" (En. "Could you (ability)...?"), "Potresti...?" (En. "Could you (possibility)...?"), "Ti dispiace...?" (En. "Do you mind...?"), "Vuoi...?" (En. "Do you want...?"), "Vorresti...?" (En. "Would you like...?"), "Vorrei..." (En. "I would like..."), or the use of the simple interrogative form (Benincà et al., 1977, p. 508); a few examples with their corresponding English translations are the following (we also report the number of the felicity condition that they are tied to):

Questioning felicity condition 2:

Sai riparare il televisore?

Can you repair the television?

Questioning felicity condition 4:

Puoi uscire un attimo?

Can you leave for a moment?

Questioning felicity condition 6:

Ti dispiace lasciare aperta la finestra?

Do you mind leaving the window open?

and

Vuoi portarmi un bicchiere d'acqua?

Do you want to bring me a glass of water?

Asserting the first alternative of felicity condition 7:

Vorrei che non mi parlassi così.

I would like you not to talk to me like that.

Semi-conventional indirect requests, on the other hand, are less conventional because, in order to be interpreted as requests (and not literally), they need to be uttered in the context in which the hearer knows (as factual information of the speaker) what action the speaker requests (Benincà et al., 1977, p. 509), that is to say: the hearer knows that the speaker's psychological state is that of desire. Moreover, these forms need one addition step (with respect to conventional indirect requests) to be connected with the felicity conditions for the speech act of requesting (Benincà et al., 1977); a few examples with their corresponding English translations are the following (Benincà et al., 1977, p. 508; we also report the number of the felicity condition that they are *eventually* tied to):

Dov'è il sale?

Where is the salt?

Additional step: if the speaker asks where the salt is, he or she does not know where the salt is, and therefore:

Asserting felicity condition 1:

Non so dov'è il sale.

I don't know where the salt is.

Contextual requirement: the hearer knows the psychological state of the speaker (desire)

or

Vedi il sale?

Do you see the salt?

Additional step: if the hearer sees the salt, he or she can (physically) pass it to the speaker, and therefore:

Asserting felicity condition 2:

Puoi passarmi il sale.

You can pass me the salt.

Contextual requirement: the hearer knows the psychological state of the speaker (desire)

Other examples of semi-conventional indirect requests are (Benincà et al., 1977, pp. 508-509):

Non trovo il sale.

I cannot find the salt.

Hai tu il sale?

Do you have the salt?

C'è bisogno...

There is the need...

This last example is semi-conventional in that, in order to be used as an indirect request, it necessitates the context in which the hearer understands that the need expressed with an impersonal form is actually pointed out to him or her (Benincà et al., 1977). Also in the case of semi-conventional indirect requests there can be requestive "relics" in the literal speech act used to make the indirect request (Benincà et al., 1977).

Finally, there exist non conventional indirect requests. They have this name because the hearer must know the context in order to interpret them as requests. Non conventional indirect requests are always tied to the second alternative of felicity condition 7, i.e. "the speaker has a reason for the interlocutor to do the action", which means that the hearer needs to recognize that the reason of the speaker is presented to him or her in such a way as to trigger an action in response (Benincà et al., 1977). Let's consider the following example with its corresponding English translation (Benincà et al., 1977, p. 509):

Domani devo pagare la rata della macchina.

Tomorrow I will have to pay the mortgage for my car.

Contextual requirements: the hearer needs to know, or needs to be able to suppose, that the speaker does not have enough money to pay the mortgage for his or her car, and needs to consider him- or herself as a person to whom the speaker might ask for a loan. The hearer needs to recognize that the reason provided by the speaker, i.e. that the speaker the next day will have to pay the mortgage for his or her car, is presented to him or her as a reason for him or her to do a certain action in response.

That being said, the hearer might not consider the reason of the speaker as a valid reason to perform a certain action in response; for example, if the speaker utters (Benincà et al., 1977, p. 509):

Che caldo!

How hot!

the hearer might be afraid of drafts and therefore not consider the heat a good reason for opening a window. In the case of non conventional indirect requests there are no requestive "relics" in the literal speech act used to make the indirect request (Benincà et al., 1977).

In summary, we can say that a minimum of conventionality is necessary in all indirect requests, even in the non conventional ones, in order to permit the hearer to recognize them as requests. Conventional forms are those that, regardless of the context, on the sole basis of some elements (requestive relics or force indicators present in the literal speech act), are conventionally intended as requests (Benincà et al., 1977, pp. 508-509). Semi-conventional indirect forms are those that can be intended in certain contexts as requests (Benincà et al., 1977, p. 509). Finally, non

conventional indirect requests are those that can be interpreted as requests in certain contexts if the hearer recognizes as valid the reason that the speaker gives him or her to take action (Benincà et al., 1977, p. 509). At this point, we can make one example of indirect request for action for each degree of conventionality, together with an example of direct or literal request:

Direct or literal request:

59a. Close the window!

59b. Chiudi la finestra!

Conventional indirect request:

59c. Can you (please) close the window?

59d. Puoi (per favore) chiudere la finestra?

Semi-conventional indirect request:

59e. I cannot reach the window.

59f. Non riesco ad arrivare alla finestra.

Non conventional indirect request:

59g. How hot!

59h. Che caldo!

Generally speaking, while semi-conventional and non conventional indirect requests need a number of inferential steps to be interpreted as requests, conventional indirect requests, just like direct or literal requests, do not. Nevertheless, conventional indirect requests, not being literal requests, can sometimes be interpreted as (and intended as) real questions or real assertions (Benincà et al., 1977). Conventional indirect requests lose their non-requestive interpretation when "per favore" or "please" is added. As we mentioned above, conventional indirect requests like 59c give the interlocutor the possibility to reject the request (or at least the idea that they can); the interlocutor can in fact reply with a yes / no answer to the request (and not to the literal question). Let's consider the following example (Benincà et al., 1977, p. 512):

60a. A: Ti dispiace / dispiacerebbe uscire?

60b. B: Sì (ed esce)

60c. A: Do / Would you mind leaving?

60d. B: Yes (and he/she leaves)

In this example, the question made by speaker A is being used as a conventional indirect request. In fact, if the interlocutor was instead answering the literal question (and therefore minded leaving), he or she would probably not be leaving afterwards.

We mentioned above the fact that that sentences conventionally used as indirect requests are not idioms, and therefore have a literal, word-for-word translation in other languages. We also said

that, sometimes, translating indirect requests into other languages can modify their illocutionary act potential. Benincà et al. (1977) conclude their discussion on conventionality by specifying that, while conventional indirect requests can modify their requestive potential in translation, non conventional indirect requests maintain their requestive potential constant in all languages. "Could you help me?", a conventional indirect request in English, can be translated into "Pourriez-vous m'aider?", "Können Sie mir helfen?", or "Potrebbe aiutarmi?" and keep the same requestive potential, but other indirect requests, such as "Are you ready to do X?" or "Sei pronto a fare X?", despite being semi-conventional in both Italian and English, becomes a conventional indirect request in modern Hebrew, thus modifying its requestive potential (Sadock, 1974, ch. IV). On the other hand, all non conventional indirect requests maintain their non conventionality in translation: "How hot!" or "Che caldo!" remains non conventional regardless of the language into which it is translated.

### CHAPTER 3 - ON CLASSIFICATION

In chapter 1, we focused on the philosophical origins of the speech act theory and on some of its most prominent theoretical developments. The takeaway from chapter 1 is that the speech act theory is a full-fledged, pragmatics-aware theory of meaning which features a very effective hands-on bag of notions for bridging the gap between utterances and speaker meaning. In chapter 2, we defined a framework for understanding indirect speech acts, in particular indirect promises and requests. We demonstrated that linguistic form underdetermines illocutionary force because speech acts depend on a number of conditions that are contextual in nature. Nevertheless, we also demonstrated that there exist a number of speech devices that the speaker can use to provide linguistic evidence of his or her communicative intentions. In the present chapter, we will take a step back and get a bird's eye view of the speech act ecosystem. We will see that speech acts can be of different types according to the way in which they are classified. The term "speech act classification" can be used to indicate either the process of grouping together speech acts that share the same characteristics, or the result of such process, i.e. the arbitrary<sup>23</sup> list of all possible types of speech acts. Most of the classifications that have been proposed are based on the notion of illocutionary point, that is to say: each class is defined in such a way as to include all the speech acts with the same communicative point or purpose. Classifying speech acts will indeed give us an idea of all the things that we can do with language, but will also ease our transition into computational linguistics. In fact, most of the studies in computational linguistics involving speech acts consist in the proposal of a classification (or tag-set) - often suited to a specific purpose, such as conversation tracking or machine translation - and a statistical model for mapping utterances (or sometimes larger stretches of discourse) to their appropriate speech act tags.

More specifically, we will begin with an analysis of the classifications proposed in philosophy by Austin (1962) and Searle (1976). Searle's classification (1976) has become the gold standard for most (if not all) subsequent classifications of speech acts - both in philosophy / linguistics and computational linguistics - mainly because of its focus on the illocutionary point or purpose of the utterances, which turns out to be a very reliable criterion for distinguishing between language uses. We will then go through the classifications proposed in computational linguistics and compare them to the classification proposed by Searle (1976). In particular, we will analyze: the DAMSL Standard tag-set (Allen & Core, 1997), the SWBD-DAMSL tag-set (Jurafsky et al., 1997), the MRDA corpus and tag-set (Shriberg et al., 2004), the works by Cohen, Carvalho, and Mitchell on "email speech acts" (Cohen et al., 2004; Carvalho & Cohen, 2005; Carvalho & Cohen,

---

<sup>23</sup> "Arbitrary" in the sense of "subjectively decided".

2006; Carvalho, 2008), the BC3 corpus and tag-set (Ulrich et al., 2008), the TA corpus and tag-set (Jeong et al., 2009), and the QC3 corpus and tag-set (Joty & Hoque, 2016). Before shifting the attention to computational linguistics, we will explain why it is important to classify speech acts in computational linguistics in the first place, that is to say: we will evaluate the benefits of having at hand an accurate classification of speech acts by giving specific examples of its possible applications. We will also examine the ways in which the notion of speech act has been simplified - or perhaps oversimplified - in order to be handled by computer programs. We will in fact witness a significant change from the in-depth characterization of speech acts (which we sought in the last two chapters) to the analysis of the surface linguistic properties of speech acts and of the way in which they are used back and forth in conversation. So-called adjacency pairs (Schegloff, 1968), i.e. two-part structures of the form "question-answer" or "request-grant" (Joty & Hoque, 2016), will play a major role in our understanding of speech acts in conversation. In chapter 4, we will elaborate on the problems that arise from the adaptation of the speech act theory in computational linguistics.

## 1. Introduction

The classification of speech acts is based on the idea that the uses that the speakers make of a language are limited in number - or at least reducible to a set of primitives - and classifiable. According to Searle (1976), there is not an infinite or indefinite number of uses of language, but rather the things that we do with language are limited in number, provided that we define clear criteria for delimiting one language use from another. Effectively classifying speech acts means defining unambiguous criteria for distinguishing between the different illocutionary forces, or between what Searle and Vanderveken call the different "natural kinds of uses of language" (1985, p. 179). To be even more precise, we will follow the footsteps of Searle (1976) and focus on a specific component of illocutionary force called illocutionary point. The illocutionary point is the purpose or goal of the utterance; it is the basic - or most important - component of illocutionary force as the other components of illocutionary force merely further specify and modify the illocutionary point, or are its consequences (Searle & Vanderveken, 1985). To make a few examples of illocutionary points: "the point of statements and descriptions is to tell people how things are, the point of promises and vows is to commit the speaker to doing something, the point of orders and commands is to try to get people to do things, and so on" (Searle & Vanderveken, 1985, pp. 13-14). Searle (1976) takes "illocutionary point (first and foremost), and its corollaries, (...) as the basis for constructing a classification" (p. 10). From this point of view, two speech acts are of

the same type, and thus belong to the same class, if the intention behind them is that of achieving the same illocutionary point. From this point of view, the number of things that we do with language is determined by the number of the different illocutionary points that a speaker can achieve.

We will see that classifying speech acts according to their illocutionary points will prove beneficial as it allows for a fairly neat delimitation between the different uses of the language. However, the notion of "illocutionary point or purpose" remains vague and open to interpretation. One can in fact define tailor-made illocutionary points at his or her convenience, which is one of the reasons why Searle's approach has become quite appealing to researchers in computational linguistics and software developers. To quote Jaszczolt (2002, p. 303): "it is essential to remember that the number of categories in the classification of speech acts is totally arbitrary". One can come up with his or her own classification by creating his or her own list of illocutionary points so long as clear criteria for distinguishing each point are provided. That being said, it can be argued that there is a small set of primitive illocutionary points that are intrinsic to human behavior, namely: reporting facts, expressing opinions and feelings, committing to doing something, trying to get others to do things, and declaring states of affairs. These basic illocutionary points are at the basis of Searle's (1976) classification. Searle (1976) develops his classification as an improvement of the classification proposed by Austin (1962). We will see that Austin's (1962) classification lacks well-defined classificatory principles and therefore did not achieve the same success as Searle's (1976). Searle (1976) defines of 5 coarse classes, corresponding to 5 primitive illocutionary points.

Since all the classifications of speech acts proposed in computational linguistics that we will analyze in the present work are based on illocutionary point, our comparison between theory and practice will consist in mapping (more or less directly) the classifications proposed in computational linguistics to the classification proposed by Searle (1976). We will in fact deliberately leave Austin's (1962) classification out of the picture since it does not hold to the same standard. Austin's (1962) classification, while essential to our discussion on classification, does not fit into our comparison because it is not essentially based on illocutionary points. In fact, perhaps with the only exception of commissives (more below), whose definition given by Austin is, according to Searle (1976), unambiguously based on illocutionary point, the biggest weakness of Austin's (1962) classification is that "there is no clear or consistent principle or set of principles on the basis of which the taxonomy is constructed" (Searle, 1976, p. 8). Searle (1976) asserts that Austin's (1962) weakness is caused by a confusion between illocutionary acts and illocutionary verbs, which in turn causes both overlaps between classes and the presence of different kinds of illocutionary verbs within the same class (Searle, 1976). We will see that Austin (1962)

distinguishes between the different uses of the language by proposing a list of illocutionary verbs representative of each use. That being said, it is fair to mention that Austin (1962) acknowledges many of the problems connected with his classification, which makes his work as a whole useful to our discussion on speech act classification.

Going back to Searle (1976), we have not considered yet the fact that each of his 5 coarse classes, corresponding to 5 primitive illocutionary points, subsumes a number of different illocutionary forces. Since we are classifying illocutionary points and not forces, we will discuss each component of illocutionary force only briefly in section 5. However, we must be aware of the fact that two utterances can have the same illocutionary point but different illocutionary forces. The same illocutionary point can in fact be achieved in a different way - or with a different degree of strength - for each force that it subsumes; for example, we can try to get somebody do something either by requesting (less strong) or insisting (stronger) that he or she do it (Searle & Vanderveken, 1985). This explains why, in most classifications, different forces like requesting and insisting fall into the same class: they share the same illocutionary point of directives, i.e. they are both aimed at trying to get people to do things. Similarly, as we mentioned in chapter 2, promising and threatening often fall into the same category because, despite being two different forces, they share the same illocutionary point of committing the speaker to doing something. In the light of this, we say that two forces are of the same type or belong to the same class (or category) if they share the same illocutionary point (in spite of achieving it in different ways).

Since in this chapter we are particularly interested in the linguistic properties of speech acts, a component of illocutionary force that will become particularly useful to our discussion are the so-called propositional content conditions, or rather their syntactic consequences. The illocutionary point of a speech act will "impose certain conditions on what can be in the propositional content" (Searle & Vanderveken, 1985) - the propositional content conditions - which have obvious syntactic consequences (Searle & Vanderveken, 1985). For example, it would be linguistically odd to say "\*I order you to have eaten beans last week" (Searle & Vanderveken, 1985, p. 16) to make an order, or "I will see you at 5" to describe a state of affairs. This means that, by analyzing the linguistic form of an utterance, we are able (to a certain extent) to backtrack and identify the point that imposed those conditions. Nevertheless, we need to always bear in mind that there is not a one-to-one correspondence between sentences or expression and illocutionary points as the same sentence or expression can be uttered with different illocutionary points.

Building a solid classification of speech acts would indeed be a great academic achievement, but it would also be useful from a practical standpoint for its many possible applications. As a general principle, we say that a classification of speech acts needs to include a fairly limited number

of classes to allow for a clear definition of each class, but at the same time it should include enough classes to be significant in the first place (and useful for downstream processing). A classification of speech acts can in fact be used as one of the primary components for the development of a number of applications, to name a few: dialog systems, automated summarization, machine translation, and conversation tracking. We will discuss more in detail below the benefits of having at hand a well-built classification of speech acts. On a slightly different note, we will see that, for the classification of speech acts in computational linguistics, little has been retained of what was theorized by Austin (1962) and Searle (1969; 1975; 1976). The speech act theory and the notion of speech act have in fact been simplified to suit practical needs, sometimes arguably beyond recognition. We will see more in detail below and in the next chapter why this simplification occurred, and what its manifestations and consequences are. We anticipate that two classes of speech acts defined by Searle (1976) are particularly controversial. One is the class of expressives, which has often been excluded or overly simplified probably because it is considered difficult to leverage. The other controversial class is that of declarations. This class has often been removed altogether in the transition to computational linguistics because of the lack of contextual data: declarations, in fact, rely on particular cultural-dependent institutions, whose presence is challenging to retrieve with the current technology. At the same time, other classes that are not mentioned in the theory have been created ad hoc for their utility in the development of specific applications; one example is the class of "answers", whose illocutionary point is that of being in response to questions, which is fundamental trait to be detected for the development of dialog systems.

To conclude our introduction, we would like to remark the fact that a speech act's "ecological niche", as Green (2017) calls it, is the conversation. While there are obvious situations in which speech acts occur in isolation - such as the utterance of "Please get off my foot!" in a crowded subway - most speech acts occur within a conversation. Scrutinizing speech acts "in captivity" would therefore deprive them of some of their distinctive features (Green, 2017). We have mentioned above the fact that many speech acts fall into pairs: assertions purport to be answers to questions, acceptances or rejections pair with offers, and so on (Green, 2017). As we will see, unlike Austin (1962) and Searle (1969), many researchers in computational linguistics study speech acts in pairs.

## **2. Ambiguity**

Before proposing his classification of speech acts, Austin (1962) elaborates on the relationship between conveying meaning and performing functions (or actions), giving particular

attention to the issue of ambiguity in natural language. This brief parenthesis on ambiguity will be useful for our understanding of natural language as a whole, but it can also be seen as a prelude to our later discussion on misclassification. Austin (1962) asserts that never in history has language been precise, nor explicit, where precision and explicitness are to be understood as follows: "precision in language makes it clearer what is being said - its meaning: explicitness, in our sense, makes clearer the force of the utterances, or 'how it is to be taken'" (Austin, 1962, p. 73). In other words, an utterance is precise if its propositional content is unambiguous (semantically) in terms of reference, predication, lexicon, structure, and scope. At the same time, an utterance is explicit if the speaker makes clear the illocutionary force with which its propositional content is to be taken. As we have reported in chapter 1, "[p]ropositional acts (the acts of referring and predicating) cannot occur alone; that is, one cannot just refer and predicate without making an assertion or asking a question or performing some other illocutionary act" (Searle, 1969, p. 25). "A proposition is *what* is asserted in the act of asserting [emphasis added]" (Searle, 1969, p. 29), *what* is questioned in the act of questioning, *what* is promised in the act of promising, and so on. Therefore, according to Austin (1962), every utterance can be more or less ambiguous in two different but related dimensions: precision and explicitness. We can clarify the difference between the two by reconsidering the following examples from chapter 1 (61a from Green, 2017):

61a. You'll be more punctual in the future.

61b. Every man loves a woman.

With regards to 61a, we said that the speaker does not make clear whether he or she is making a prediction, issuing a command, or making a threat. We can say that the speaker is not being *explicit* in making clear the force of his or her utterance. 61b, on the other hand, has a semantic ambiguity caused by the unspecified scope of the verb "love". This utterance can mean either that a) for every man, there is a woman, and it's possible that each man loves a different woman, or that b) there is one particular woman who is loved by every man. We can say that the speaker is not being *precise* in making clear the propositional content of his or her utterance.

On a similar note, Austin (1962) observes that "the giving of straightforward information produces, almost always, consequential effects upon action, (which) is no more surprising than the converse, that the doing of any action (including the uttering of a performative) has regularly the consequence of making ourselves and others aware of facts (Austin, 1962, p. 110). With regards to the first point, Austin (1962) is not referring to non conventional speech acts, but rather to the fact that utterances that are intended to give straightforward information (and just that) can have consequential non-immediate effects on the interlocutor, who will perform certain actions in the future in the light of the information that he or she has acquired. Non conventional indirect speech

acts, on the other hand, consist in utterances giving straightforward information but *intended* as something else to trigger *immediate* reactions from the interlocutor (reactions that are different from the simple acknowledgment of the information being transmitted). Austin (1962) observes that the propositional content of a speech act, whether it is asserted, questioned, promised, etc., will influence the hearer's knowledge about the state of affairs. In other words, when the speech act has a propositional content (and some as we will see do not), some information about the state of affairs is inevitably conveyed in its performance, regardless of its force. To clarify this point, we will quote Allen and Core (1997), who write in regard to statements: "[n]ote also that we are only coding (as statements) utterances that make explicit claims about the world, and not utterances that implicitly claim that something is true". To demonstrate how a non-statement can implicitly make the hearer aware of facts, they make the following example: "Let's take the train from Dansville", which presupposes the existence of a train in Dansville, but should not be considered a statement; it is rather an invitation (Allen & Core, 1997). An explicit statement would instead be "There is a train in Dansville".

Our final remark about ambiguity is that certain classifications merge illocutionary force and propositional content, which makes them sensitive not only to explicitness but also to precision. As we will see more in detail below, this is especially the case of Cohen and Carvalho. Let's consider the following examples:

62a. Can you please send me the document?

62b. Can you please stop by tomorrow?

Despite being both requests, 62a would be classified as a "request for data" and 62b as a "request for meeting" (Cohen and Carvalho, 2004). Similarly, Cohen and Carvalho (2004) hypothesize an email conversation assistant capable of detecting urgency:

63. Can you do this ASAP?

The use of "ASAP" makes 63, not just a request for action, but a request for *prompt* action, which in turn implicates that the issue needs to be addressed in time (Cohen and Carvalho, 2004). Bearing in mind that precision and explicitness are not unrelated, we can conclude this section by saying that, since our main goal is to classify utterances according to their illocutionary point, we are primarily concerned with the ambiguity of language in terms of explicitness. Generally speaking, the less explicit an utterance is, the more difficult it is to retrieve its illocutionary force (and point).

### 3. More Primitive vs. Less Primitive Devices

Austin (1962) argues that humans have always used language to perform functions, but that their ability to do so has increased in the course of history as society developed. According to him, the performance of functions with language has become more and more explicit - or less and less ambiguous - with time (Austin, 1962). He writes: "the explicit performative formula, i.e. the use of (illocutionary) verbs in the first person singular present indicative active form; e.g. I promise, I order, is the last and 'most successful' of numerous speech devices which have always been used with greater or less success to perform the same function" (Austin, 1962, p. 73). In the light of this, before going through Austin's classification of speech acts, we dedicate a few lines to what Austin calls instead "more primitive speech devices". According to him, these devices have been (partially) "taken over by the device of the explicit performative" (Austin, 1962, p. 73), but are still used to a significant degree to perform functions, although less explicitly. We would like to stress the fact that implicitness and indirectness are not the same: while implicitness refers to the conventionality that binds the utterance's literal meaning to its literal force, indirectness refers to the conventionality of usage of the utterance that binds the performance of a direct speech act with the simultaneous performance of an indirect one.

We will see below that Austin (1962) classifies speech acts by associating each act with an illocutionary verb naming it. However, as we said, the force of an utterance is to a certain extent conveyed by "more primitive devices". These devices can be summarized as follows (from Austin, 1962):

1) Mood, such as the use of the imperative to make an utterance a command, an exhortation, a permission, and so forth. We report the following examples (Austin, 1962, pp. 73-74):

'Shut it' resembles the performative 'I order you to shut it'.

'shut it, if you like' resembles the performative 'I permit you to shut it'.

'Very well then, shut it' resembles the performative 'I consent to your shutting it'.

'Shut it if you dare' resembles the performative 'I dare you to shut it'.

Similarly, we may use auxiliaries (Austin, 1962, p. 74):

'You may shut it' resembles the performative 'I give permission, I consent, to your shutting it'.

'You must shut it' resembles the performative 'I order you, I advise you, to shut it'.

'You ought to shut it' resembles 'I advise you to shut it'.

2) Tone of voice, cadence, and emphasis, which are features of spoken language not easily reproducible in written language: punctuation, italics, and word order can be used as indicators of a certain illocutionary force, but they are quite unrefined and arbitrary. Austin,

for example, uses an exclamation mark followed by a question mark to indicate a protest. He makes the following examples (Austin, 1962, p. 74):

It's going to charge! (a warning);

It's going to charge? (a question);

It's going to charge!?! (a protest);

3) Adverbs, adverbial phrases, and turns of phrase; for example, the force of "I shall" changes significantly if we qualify it by adding "probably" or "without fail":

I shall probably...

I shall without fail...

The use of such devices has a particular influence over those functions of language that, despite being essentially different, employ "the same or similar verbal devices and circumlocutions" (Austin, 1962, p. 75); Austin (1962, p. 75) makes the examples of: evincing, intimating, insinuation, innuendo, giving to understand, enabling to infer, conveying, and expressing, all of which are performed with the same verbs and thus need different adverbs as their qualifiers;

4) Connecting particles; for example, "we may use the particle 'still' with the force of 'I insist that'; we use 'therefore' with the force of 'I conclude that'; we use 'although' with the force of 'I concede that' (Austin, 1962, p. 75). In addition to this, the use of titles (and, we add, the use of subjects of emails or threads) serves a similar purpose; for example "Manifesto, Act, Proclamation, or the subheading 'A Novel...'" (Austin, 1962, p. 75);

5) Accompaniments of the utterance, that is gestures of ceremonial non-verbal actions, which are out of the scope of the present study;

6) The circumstances of the utterance, which may or may not be made explicit in the linguistic form of the utterance, such as (Austin, 1962, p. 76) "*coming from him*, I took it as an order, not as a request", or again "I shall die some day", which we understand differently in accordance with the health of the speaker.

Austin argues that, unlike more primitive devices, which can be misleading principally because of "their vagueness of meaning and uncertainty of sure reception" (Austin, 1962, p. 76), explicit performatives (illocutionary verbs) keep the performance relatively fixed (Austin, 1962): "in a way these resources (more primitive devices) are over-rich: they lend themselves to equivocation and inadequate discrimination. (...) The explicit performative rules out equivocation and keeps the performance fixed, relatively" (Austin, 1962, p. 76). Assuming that explicit performatives are fairly rigidly tied to the functions they perform, they are a good test for determining which illocutionary force an utterance has. In addition to this, being used mainly in

"that..." or "to..." formulas (Austin, 1962), explicit performatives are relatively easy to identify as they share a similar distribution. According to what we have said so far, it looks like the membership of an utterance to a specific class of illocutionary forces can be determined reliably only by determining whether or not a certain illocutionary verb (explicit performative) occurs in the linguistic form of the utterance, and that the investigation of over-rich primitive devices should be avoided.

We can summarize the main points made in this section as follows:

- Humans perform functions (or actions) through speaking more or less explicitly on the basis of the speech devices that they use;
- The use of the explicit performative (a performative verb) is the most advanced and most successful way to perform a function with language;
- Other speech devices can be used to perform the same function performed by the explicit performative, but in a less explicit and thus more ambiguous way;
- Explicit performatives are fairly rigidly tied to the functions they perform, but the other speech devices are not (more below).

#### **4. Austin's Classification**

Austin (1962) does not adopt the notion of illocutionary point for his classification, maybe with the exception of the class of commissives (more below), and considers performative verbs as the only speech device that is reliable enough to be used as a criterion for classifying speech acts. For these reasons, his classification is arguably not a classification of speech acts, but rather a full-fledged classification of English illocutionary verbs. In fact, Austin (1962) wrongly (but knowingly) assumes that any two non-synonymous illocutionary verbs mark different illocutionary acts (Searle, 1976), and therefore the same illocutionary verb cannot belong to two different classes. On this false premise, classifying illocutionary verbs is equivalent to classifying illocutionary acts. Austin's (1962) reasoning can be summarized as follows: the mere occurrence of an illocutionary verb in the first person singular present indicative active form is a clear indicator that the utterance in which the illocutionary verb occurs is used to perform a speech act of the type corresponding to the class to which the illocutionary verb belongs. Other indicators of illocutionary force, that is to say: anything in the language that is not an illocutionary verb, is put aside. In Austin's (1962) classification there is one single variable (the illocutionary verb) whose value determines (by itself) the type of the illocutionary act. According to Austin's (1962) classification, in the absence of a value, i.e. in case that there is not an illocutionary verb in the linguistic form of the utterance, we are unable to

determine to which class that utterance belongs because we would need to resort to more primitive (and unreliable) devices. Let's now assume that we need to map the utterance below to the speech act that it performs by analyzing its linguistic form in isolation (Austin, 1962, p. 74):

64. You must shut it.

We are inclined to think that the speech act performed is either an order or an advice, depending on the context. If, on the other hand, we use a performative (Austin, 1962, p. 74):

65a. I order you to shut it.

65b. I advise you to shut it.

the speech act performed is explicit. As reductive as it may sound, utterances such as 64 are considered by Austin (1962) not explicit enough to be classified. We have seen that Austin (1962) lists a number of (or types of) more primitive devices, but then excludes them from his classification for being too unreliable. Austin's (1962) classification is therefore based exclusively on illocutionary verbs. According to Searle (1976), on the other hand, illocutionary points should be at the basis of a classification of speech acts, and illocutionary verbs, together with other linguistic (and non-linguistic) features of the utterance, should be used (in combination) to retrieve the illocutionary point of the utterance (and therefore the class to which the utterance belongs). Searle (1976), as we will see especially in section 6, argues that (the more primitive device of) syntax plays too an important role in the identification of illocutionary points.

Austin (1962) distinguishes five categories or classes of illocutionary acts and lists a number of illocutionary verbs representing each class. Each class indicates a type or "family" of illocutionary verbs, and consequently a set of possible utterances in which they are employed. The five classes of speech acts proposed by Austin (1962) are the following (as reported by Jaszczolt, 2002, p. 301):

- verdictives (for example estimating, assessing, describing);
- exercitives (for example ordering, appointing, advising, excommunicating);
- commissives (for example promising, intending, betting)
- behabitives (for example apologizing, congratulating, thanking, blaming, complaining);
- expositives (for example arguing, insisting, affirming).

Austin (1962) sums up these five categories of speech acts as follows: "the verdictive is an exercise of judgement (or the giving of a verdict), the exercitive is an assertion of influence or exercising of power, the commissive is an assuming of an obligation or declaring of an intention (or the commitment to causes or courses of action), the behabitive is the adopting of an attitude (or of a social behavior), and the expositive is the clarifying of reasons, arguments, and communications (of

how utterances fit into lines of reasoning)" (p. 163; see also Green, 2017). Austin (1962) defines each of them in detail as follows:

- "Verdictives are typified by the giving of a verdict, as the name implies" (Austin, 1962, p. 150). More specifically, "verdictives consist in the delivering of a finding, official or unofficial, upon evidence or reasons as to value or fact, so far as these are distinguishable." (Austin, 1962, p. 152). The verdict need not to be final and can be based on facts which are not certain; for example, it can be an estimate, a reckoning, or an appraisal (Austin, 1962);
- "Exercitives are the exercising of powers, rights, or influence. Examples are appointing, voting, ordering, urging, advising, warning, &c." (Austin, 1962, p. 150). "An exercitive is the giving of a decision in favour of or against a certain course of action, or advocacy of it. It is a decision that something is to be so, as distinct from a judgment that it is so: it is advocacy that it should be so, as opposed to an estimate that it is so" (Austin, 1962, p. 155);
- "Commissives are typified by promising or otherwise undertaking; they *commit* you to doing something, but include also declarations or announcements of intention, which are not promises, and also rather vague things which we may call espousals, as for example, siding with. They have obvious connexions with verdictives and exercitives." (Austin, 1962, pp. 150-151). With a few exceptions, "the whole point of a commissive is to commit the speaker to a certain course of action." (Austin, 1962, p. 160). Commissives are the only class that remains almost unvaried in Searle's (1976) classification, the only difference that, in Searle's (1976) classification, *all* commissives, without exception, have the point of committing the speaker to a certain course of action;
- "Behabitives, are a very miscellaneous group, and have to do with attitudes and *social behaviour*. Examples are apologizing, congratulating, commending, condoling, cursing, and challenging." (Austin, 1962, p. 151). "Behabitives include the notion of reaction to other people's behaviour and fortunes and of attitudes and expressions of attitudes to someone else's past conduct or imminent conduct." (Austin, 1962, p. 160);
- "Expositives (...) make plain how our utterances fit into the course of an argument or conversation, how we are using words, or, in general, are expository. Examples are 'I reply', 'I argue', 'I concede', 'I illustrate', 'I assume', 'I postulate'." (Austin, 1962, p. 151). "Expositives are used in acts of exposition involving the expounding of views, the conducting of arguments, and the clarifying of usages and of references." (Austin, 1962, p. 161). In the case of expositives, "the main body of the utterance has generally or often the straightforward form of a 'statement', but there is an explicit performative verb at its head

which shows how the 'statement' is to be fitted into the context of conversation, interlocution, dialogue, or in general of exposition." (Austin, 1962, p. 85).

Austin's classification lacks well-defined classificatory principles: the categories overlap and it is thus often not clear which category an illocutionary verb belongs to (Jaszczolt, 2002). As Austin himself admits: "we should be clear from the start that there are still wide possibilities of marginal or awkward cases, or of overlaps." (Austin, 1962, p. 151). Austin (1962) acknowledges that this cross-classification is particularly evident in the last two classes, behabitives and expositives, but that ultimately affects all five classes; he writes: "behabitives are troublesome because they seem too miscellaneous altogether: and expositives because they are enormously numerous and important, and seem both to be included in the other classes and at the same time to be unique in a way that I have not succeeded in making clear even to myself. It could well be said that all aspects are present in all my classes." (Austin, 1962, p. 151). We can say that the five classes proposed by Austin do not discriminate accurately between one type of illocutionary force and the other, i.e. the classes overlap by definition; for example, an utterance with the illocutionary force of a behabitive - which may (e.g. "I support") or may not (e.g. "I am in favor of") contain an illocutionary verb - *also* commits the speaker to a certain course of action, thus making the behabitive utterance partially commissive. Moreover, an illocutionary force can be generated without the use of the explicit performative formula. Thus, "I support" and "I am in favor of" have the same illocutionary force, regardless of whether the illocutionary force is made explicit by an illocutionary verb.

Austin (1962) then discusses the use of special performative-looking words, such as "off-side" and "liable". "Instead of 'I pronounce you off-side' I might say 'You are off-side' and I might say 'I am (hereby rendered) liable' instead of 'I undertake...'" (Austin, 1962, p. 58). Furthermore, a speech act can also be performed by uttering a single word (Austin, 1962; Searle & Vanderveken, 1985), such as "out" (uttered by an umpire) or "guilty" (uttered by a judge). Austin (1962) explains the phenomenon of descriptive utterances and single words having a certain illocutionary force as follows: "any utterance (even a single word) which is in fact a performative (even though it does not look like it) should be reducible, or expandible, or analysable into a form with a verb in the first person singular present indicative active" (pp. 61-62). To prove his point, Austin makes the following examples of single-word utterances: "'Out' is equivalent to 'I declare, pronounce, give, or call you out' (when it is a performative) (...). 'Guilty' is equivalent to 'I find, pronounce, deem you to be guilty.'" (Austin, 1962, p. 62). Similarly, a descriptive utterance like "I am in favor of" can be analyzed into "I support". The hypothesis that we can extract a performative verb from virtually any

non-performative looking utterance and the assumption that performative verbs are rigidly tied to specific functions are at the basis of Austin's classification of performative verbs.

Austin (1962), in the first part of his work, makes the distinction between performative utterances and constative (or descriptive) utterances. This distinction will be later abandoned by Austin (1962) himself to make room for a more comprehensive view, according to which every utterance involves the performance of a speech act, whether it is an assertion (descriptive), or a promise, an order, etc. Currently, we still use the term "performative", also in computational linguistics, to indicate those utterances that are used to explicitly perform specific speech acts (thus keeping Austin's (1962) original definition).

The use of performative verbs does not come without problems. To quote Austin, we must acknowledge the fact that "this first person singular present indicative active, so called, is a peculiar and special use. In particular we must notice that there is an asymmetry of a systematic kind between it and other persons and tenses of the very same verb. The fact that there is this asymmetry is precisely the mark of the performative verb (and the nearest thing to a grammatical criterion in connection with performatives)" (Austin, 1962, pp. 62-63). Austin makes the example of "I bet" - a performative - and compares it with "bet" in another tense and/or person. "I betted" and "he bets", unlike "I bet", are in fact statements describing actions performed by the speaker or by somebody else; "actions each consisting in the utterance of the performative 'I bet'" (Austin, 1962, p. 63). Austin goes on by pointing out that "this sort of asymmetry does *not* arise at all in general with verbs that are *not* used as explicit performatives [emphasis added]. For example, there is no such asymmetry between 'I run' and 'He runs'" (Austin, 1962, p. 63), nor with between "I run" and "I am running"; they are all statements regardless of tense and person since "to run" is not a performative verb. The use of the present continuous is particularly ambiguous; e.g. "I apologize" is without a doubt a performative, but "I am apologizing" can be either a performative or a statement. Let's now consider a few examples with the second and third person (singular and plural) (example 66b from Austin, 1962, p. 57):

66a. I authorize you to pay...

66b. You are *hereby* authorized to pay...

While 66a is clearly a performative - the speaker authorizes the hearer to pay - 66b is ambiguous: in order to be classified as a performative, it needs the word "hereby", otherwise it could simply be a description of what usually happens, that is that the hearer is usually authorized to pay. The same problem arises with third person plural verbs in the passive voice (from Austin, 1962, p. 57):

67a. Passengers are warned to cross the track by the bridge only.

67b. Passengers are *hereby* warned to cross the track by the bridge only.

Or again with impersonal verbs in the passive voice (from Austin, 1962, p. 57):

68a. Notice is given that trespassers will be prosecuted.

68b. Notice is *hereby* given that trespassers will be prosecuted.

67a and 68a may be used to describe what usually happens, whereas 67b and 68b are clearly used to perform a warning (Austin, 1962). "Hereby" is typically employed in a formal or legal context (Austin, 1962), which means that in other contexts we cannot easily distinguish between a performative and a description of a regularity.

To sum up, regardless of whether it is true that the explicit performative form can be extracted from any non-performative looking utterance, we can say that the use of explicit performatives is, too, to a certain extent ambiguous: the use of a different person, tense, and voice are in fact indicators of a different illocutionary force. At this point, we conclude this section by saying that, according to Austin (1962), the performance of the action can be made explicit (and not merely stated or described) by means of:

- "the verbs which seem, on ground of vocabulary, to be specially performative verbs" (Austin, 1962, p. 61);
- "other words which seem to have a special performative function (and indeed have it), such as 'guilty', 'off-side', &c." (Austin, 1962, p. 61). These words are indicators of the action performed "in so far as and when they are linked in 'origin' with these special explicit performative verbs like 'promise', 'pronounce', 'find'" (Austin, 1962, p. 61);
- the "hereby" formula; it is a useful alternative but it is "too formal for ordinary purposes, and further, we may say 'I hereby state...' Or 'I hereby question...'" (Austin, 1962, p. 61);
- the use of mood, voice, and other more primitive devices. These devices are controversial and require complex rules to become a useful indicator of the action performed.

## 5. Searle's Classification

After the publication of "How to Do Things with Words" in 1962, the most prominent classification proposed within the domain of linguistics and philosophy is that of Searle in "A classification of illocutionary acts" (1976). Searle's studies on language use (1969; 1976) will be adopted as the theoretical background for most (if not all) works on speech acts in computational linguistics, and his classification will become the gold standard for most (if not all) subsequent classifications of speech acts. The importance that Searle's classification has gained over the years is due mainly to his focus on the illocutionary point or purpose of utterances (as a component of illocutionary force).

Before proposing his classification, Searle (1976) elaborates on what illocutionary force is and on the aspects in terms of which one (kind of) illocutionary act is different from another, drawing particular attention to the notion of illocutionary point or purpose of utterances. As we have said above, according to Searle (1976), there is *not* an infinite of indefinite number of uses of language, but rather the things we do with language are limited in number, provided that we define clear criteria for delimiting one language use from another (criteria that, as we will see below, are not based only on verbs). Searle (1976) agrees with Austin on the unreliability of illocutionary verbs as indicators of illocutionary forces and tries to come up with a solution to this problem. He writes: "differences in illocutionary verbs are a good guide but by no means a sure guide to differences in illocutionary acts." (Searle, 1976, p. 2). Part of Searle's critique is in fact that Austin's classification is a classification of English illocutionary verbs and *not* of illocutionary acts. Austin's methodology is overly lexicographic as he wrongly assumes that the range and limits of illocutionary acts can be understood by studying illocutionary verbs in English or other languages (Green, 2013). Austin (1962) groups together in the same class verbs that he thinks (are used to) perform the same illocutionary act as he assumes that any two non-synonymous illocutionary verbs mark different illocutionary acts (Searle, 1976). In other words, discriminating between (and classifying) illocutionary verbs is seen by Austin (1962) as equivalent to discriminating between (and classifying) illocutionary forces (or points). For example, Austin (1962, pp. 156-157) lists a number of verbs under the commissive class. Those are the verbs whose point is, according to him, "to commit the speaker to a certain course of action" (p. 156). Austin (1962) assumes that every utterance in which any of the commissive verbs occurs as a performative verb is necessarily used by the speaker to commit him- or herself to a certain future course of action.

Searle (1976) argues that we should classify illocutionary forces, or better illocutionary points and not illocutionary verbs (and that we should use illocutionary verbs only as one indicator, to be considered in conjunction with other indicators, for determining illocutionary points). The distinction between illocutionary forces and illocutionary verbs is clarified satisfactorily by Searle and Vanderveken (1985); they write about assertives the following (p. 38) (more on the class of assertives below): "[w]e will call the illocutionary forces with the assertive point *assertive illocutionary forces* and the performatives or illocutionary verbs which name an assertive illocutionary force *assertives*", assertive verbs, or illocutionary verbs of the assertive type. Always bearing in mind that relying on illocutionary verbs is both misleading and reductive, assertive verbs, or simply assertives, can still be used, although with many reservations, to mark utterances with the assertive force. Austin's (1962) work, because of its inconsistencies, should not, we argue, be used as the theoretical foundation of a classification of speech acts. However, his reflections on

performatives and constatives, together with his analysis of illocutionary verbs and other speech devices, will indeed turn out to be useful in our discussions of illocutionary points. To sum up, the shift of focus caused by Searle (1969; 1976) from illocutionary verbs to illocutionary points changes also the way in which we analyze language: instead of looking for illocutionary verbs within sentences, we now investigate whole sentences in their context of utterance.

Except maybe for commissives, whose definition given by Austin is, according to Searle (1976), *unambiguously* based on illocutionary point, the biggest weakness of Austin's classification is that "there is no clear or consistent principle or set of principles on the basis of which the taxonomy is constructed" (Searle, 1976, p. 8). This weakness is caused by a confusion between illocutionary acts and illocutionary verbs, which in turn causes both overlaps between classes and the presence of quite different kinds of verbs within the same class (Searle, 1976). Searle (1976) sums up the shortcomings of Austin's (1962) classification as follows: "in ascending order of importance: there is a persistent confusion between verbs and acts, not all the verbs are illocutionary verbs, there is too much overlap of the categories, too much heterogeneity within the categories, many of the verbs listed in the categories don't satisfy the definition given for the category and, most important, there is no consistent principle of classification" (Searle, 1976, pp. 9-10). Searle (1976) devises a more effective classification taking "illocutionary point (first and foremost), and its corollaries" as the basis for its construction (Searle, 1976, p. 10). Searle (1976) distinguishes five classes:

- assertives (or representatives) (for example stating, suggesting, boasting, complaining, claiming, reporting);
- directives (for example ordering, commanding, requesting, advising, recommending);
- commissives (for example promising, vowing, offering);
- expressives (for example congratulating, thanking, pardoning, blaming, praising, condoling);
- declaratives (or declarations) (for example excommunicating, resigning, dismissing, christening, naming, appointing, sentencing).

We can sum up these five classes as follows: representatives describe an existing state of affairs (they tell people how things are), directives and commissives try to get someone else or commit the speaker, respectively, to bring about a future state of affairs, expressives externalize feelings and attitudes about a state of affairs, and directives bring about changes of a state of affairs through their utterance (Searle, 1976).

Before diving deeper into Searle's classification, we dedicate a few lines to clarifying the notions of illocutionary point and direction of fit, both necessary for a thorough understanding of

the work of Searle (1976). According to Vanderveken's success-conditional semantics, the success value of a speech act is determined by its conditions of satisfaction, which "depend on the truth conditions of the proposition (P) embedded under the illocutionary force (F) and the way the proposition is related to the world through the property of its illocutionary force called the 'direction of fit'" (Jaszczolt, 2002, pp. 301-303). The conditions of satisfaction of speech acts can be seen to generalize the notion of truth; for example, we can say that the aim of assertions is to capture how things are and that the aim of commands is that the world is enjoined to conform to them (Green, 2017). When an assertion succeeds, not only is it true, but it also has its target, just like when a command succeeds, not only it brings about the truth of its content, but it also does so in a way that makes it a command (and not a prediction, for example). While we will focus on illocutionary point and direction of fit, it is however useful to dedicate a few words to all seven components of illocutionary force. These components can be seen as a revisited version of Searle's (1975) felicity conditions.

According to Searle and Vanderveken (1985), illocutionary force is defined in terms of seven features (see also Green, 2017, pp. 12-13):

1. Illocutionary point: the characteristic aim of the speech act; for example, the illocutionary point of a request is to get the addressee to do something (we will focus on this below);
2. Degree of strength of the illocutionary point: the strength with which the speaker wants to achieve the illocutionary point; for example, requesting and insisting have the same illocutionary point but the latter is stronger than the former;
3. Mode of achievement: the way in which the illocutionary point must be achieved; for example, requesting and commanding both aim to get the addressee to do something (this is their illocutionary point), but issuing a command, unlike making a request, involves invoking one's authority to be successful. This component of illocutionary force is tied to the set of culturally-dependent, group-specific conventions that characterize institutional speech acts (see chapter 1, §2.2);
4. Content conditions: the propositional content conditions necessary for the performance of certain illocutionary acts; for examples, the speaker can only promise what is in the future and under his or her control;
5. Preparatory conditions: all the other conditions that must be met for the speech act not to misfire, such as social status, authority, role, etc; for example, a person must own an object in order to bequeath it and a person must be legally invested with the necessary authority in order to marry a couple.

6. Sincerity conditions: the psychological state that the speaker expresses performing a speech act; for example, assertions express belief, apologies express regret, promises express intentions, etc;
7. Degree of strength of the sincerity conditions: the strength with which the speaker expresses his or her psychological state; for example, both requesting and imploring express desire and are identical in terms of all the components above, except for the fact that imploring expresses a stronger desire than requesting.

With regard to illocutionary point, Searle and Vanderveken (1985) write that "each illocution has a point or purpose which is internal to its being an act of that type. (p. 13) (...) By saying that the illocutionary point is internal to the type of illocutionary act, we mean simply that a successful performance of an act of that type necessarily achieves that purpose and it achieves it in virtue of being an act of that type" (p. 14). As we have said above: "the point of statements and descriptions is to tell people how things are, the point of promises and vows is to commit the speaker to doing something, the point of orders and commands is to try to get people to do things, and so on" (Searle & Vanderveken, 1985, pp. 13-14). These points or purposes are called the illocutionary points of the speech acts to which they correspond, and a speech act is successful if it achieves its corresponding purpose (or point) (Searle & Vanderveken, 1985). That is to say: a statement or a description is successful if it tells people how things are, a promise or a vow is successful if it commits the speaker to doing something, an order or a command is successful if it tries to get people do things, and so on. Searle and Vanderveken (1985) clarify why the illocutionary point is the most important of the components of illocutionary force by pointing out the following: "In real life a person may have all sorts of other purposes and aims; e.g. in making a promise, he may want to reassure his hearer, keep the conversation going, or try to appear to be clever (...) none of these is part of the essence of promising. But when he makes a promise he necessarily commits himself to doing something. Other aims are up to him, none of them is internal to the fact that the utterance is a promise; but if he successfully performs the act of making a promise then he necessarily commits himself to doing something, because that is the illocutionary point of the illocutionary act of promising" (Searle and Vanderveken, 1985, p. 14). Another important point raised by Searle and Vanderveken (1985) is that the illocutionary point of a speech act is achieved only "as part of a total speech act in which the propositional content is expressed with the illocutionary point" (p. 15). In other words, "the illocutionary point is achieved on the propositional content" (p. 15): "[o]ne cannot promise that someone else will do something (...) and one cannot promise to have some something in the past" (p. 16), similarly one cannot apologize for something that he or she has not done or is not otherwise responsible for, such as for the elliptical orbit of the planets (preparatory conditions).

As we mentioned above, propositional content conditions are particularly useful to our discussion since they have obvious syntactic consequences (Searle & Vanderveken, 1985); for example: "I will meet you at 5 pm" can be a commitment to a future course of actions, whereas "I have met you at 5 pm" cannot.

Direction of fit is described by Jaszczolt (2002) as follows: "one of the main characteristics of speech acts is their direction of fit which can be world-to-words or words-to-world. If by performing the speech act the speaker affects the way the world is, the direction of fit is world-to-words: the world adjusts to the words. For example, by ordering for something to be done, the speaker affects the way the world is. If by uttering the speech act the speaker describes the way the world is rather than affecting it, the direction of fit is words-to-world: the words adjust to the way the world is. Stating something or complaining have this direction of fit. Assertives have the words-to-world direction of fit, directives and commissives have world-to-words, expressives have no direction of fit, while declarations have both world-to-words and words-to-world directions of fit" (Jaszczolt, 2002, p. 302; see also Searle and Vanderveken, 1985). That being said, we are concerned with illocutionary point and direction of fit to the extent to which they are reflected in the linguistic form of the utterance and are useful for classificatory purposes. As Green (2013) points out, "[d]irection of fit is also not so fine-grained as to enable us to distinguish speech acts meriting different treatment"; for example asserting and conjecturing (that something is true) have both a words-to-world direction of fit but are indeed subject to different norms: while assertions are manifestations of knowledge, conjectures are not (Green, 2013). As a consequence, we might expect different uptakes, which are completely independent of the direction of fit: "How do you know?" is an appropriate reply to assertions, but not to conjectures (Green, 2017).

Let's now analyze Searle's classification in detail. Searle (1976) describes each class as follows:

- assertives (representatives): "[t]he point or purpose of the members of the representative class is to commit the speaker (in varying degrees) to something's being the case, to the truth of the expressed proposition. All of the members of the representative class are assessable on the dimension of assessment which includes true and false." (Searle, 1976, p. 10). Searle goes on: "The direction of fit is words to the world; the psychological state expressed is Belief (that p)": B(P) (Searle, 1976, p. 10). Representatives include the so-called statements, i.e. standard indicative forms *without* explicit performatives such as "he is a liar" and "he has appendicitis" (Searle, 1976, p. 20), but include also utterances characterized by one out of "a large number of performative verbs that denote illocutions that seem to be assessable in the True-False dimension and yet are not just 'statements'" (Searle, 1976, p. 10). In fact, we may

want to say that, while in the making of statements we *implicitly* "call, diagnose and describe, as well as accuse, identify and characterize" (Searle, 1976, p. 20), in using a performative verb, we *explicitly* do so. Assertives / Representatives of the explicit performative type also make explicit some added feature that typifies them; for example, "boast" and "complain" have the added feature of having something to do with the interest of the speaker, and "conclude" and "deduce" have the added feature of marking a relation between the representative illocutionary act and the rest of the discourse (Searle, 1976, pp. 10-11). Representatives correspond to most of Austin's expositives and to many of his verdictives in that they have the same illocutionary point, differing only in terms of other components of their illocutionary force (Searle, 1976). Arguably, "the simplest test of a representative is this: can you literally characterize it (inter alia) as true or false"? (Searle, 1976, p. 11), which however will not give neither necessary nor sufficient conditions (Searle, 1976), as we will see when we analyze Searle's fifth class. We will see in fact that assertions (characterizable as true or false) can also be uttered as declarations (and not as assertives) by somebody in a position of authority within an institution;

- directives: "[t]he illocutionary point of these consists in the fact that they are attempts (of varying degrees...) by the speaker to get the hearer to do something." (Searle, 1976, p. 11). They range from inviting and suggesting, to insisting that the hearer does something (Searle, 1976). "The direction of fit is world-to-words and the sincerity condition is want (W) (or wish or desire). The propositional content is always that the hearer H does some future action A." (Searle, 1976, p. 11). They have the following symbolism: W(H does A) (Searle, 1976, p. 11). "Verbs denoting members of this class are: ask, order, command, request, beg, plead, pray, entreat, and also invite, permit, and advise." (Searle, 1976, p. 11). But also some of Austin's behabitives belong to this class, such as dare, defy and challenge, as well as many of his exercitives;

- commissives: here Searle adopts Austin's definition of commissives. He writes: "Commissives then are those illocutionary acts whose point is to commit the speaker (again in varying degrees) to some future course of action". (Searle, 1976, p. 11). However, Searle (1976) rules out some of the verbs proposed by Austin, such as shall, intend, and favor. "The direction of fit is world-to-words and the sincerity condition is Intention (I). The propositional content is always that the speaker S does some future action A." (Searle, 1976, p. 11). They have the following symbolism: I(S does A) (Searle, 1976, p. 11);

- expressives: "[t]he illocutionary point of this class is to express the psychological state specified in the sincerity condition about a state of affairs specified in the propositional

content. The paradigms of expressive verbs are 'thank', 'congratulate', 'apologize', 'condole', 'deplore', and 'welcome'. Notice that in expressives there is no direction of fit" (Searle, 1976, p. 12). With regard to this last point, we can say that "the truth of the expressed proposition is presupposed. Thus, for example, when I apologize for having stepped on your toe, it is not my purpose either to claim that your toe was stepped on or to get it stepped on. This fact is neatly reflected in the syntax (of English) by the fact that the paradigm-expressive verbs in their performative occurrence will not take that clauses but require a gerundive nominalization transformation (or some other nominal)" (Searle, 1976, p. 12). To prove his point, Searle (1976) makes the following examples:

"one cannot say:

\*I apologize that I stepped on your toe;

rather the correct English is,

I apologize for stepping on your toe." (Searle, 1976, p. 12)

They have the following symbolism: (P)(S/H+ property) (Searle, 1976, p. 13). "P is a variable ranging over the different possible psychological states expressed in the performance of the illocutionary acts in this class, and the propositional content ascribes some property (not necessarily an action) to either S or H. I can congratulate you not only on your winning the race, but also on your good looks." (Searle, 1976, p. 13);

- declarations: "there is still left an important class of cases, where the state of affairs represented in the proposition expressed is realized or brought into existence by the illocutionary force-indicating device, cases where one brings a state of affairs into existence by declaring it to exist, cases where, so to speak, 'saying makes it so'. Examples of these cases are 'I resign', 'You're fired', 'I excommunicate you', 'I christen this ship, the battleship Missouri', 'I appoint you chairman', and 'War is hereby declared.'" (Searle, 1976, p. 13). Searle goes on: "[i]t is the defining characteristic of this class that the successful performance of one of its members brings about the correspondence between the propositional content and reality, successful performance guarantees that the propositional content corresponds to the world: if I successfully perform the act of appointing you chairman, then you are chairman; if I successfully perform the act of nominating you as candidate, then you are a candidate; if I successfully perform the act of declaring a state of war, then war is on; if I successfully perform the act of marrying you, then you are married" (Searle, 1976, p. 13), and so on and so forth. Searle (1976) argues that in declarations "there is no surface syntactical distinction between propositional content and illocutionary force: 'You're fired' and 'I resign' do not seem to permit a distinction between illocutionary force

and propositional content" (Searle, 1976, p. 13). According to Searle (1976), this correspondence between F and P can be also explained by the fact that statements, if used to perform declarations, have actually the following semantic structure (Searle, 1976, pp. 13-14; more below):

I declare: your employment is (hereby) terminated. (= "you're fired")

I declare: my position is (hereby) terminated." (= "I resign")

"Declarations bring about some alternation in the status or condition of the referred-to object or objects solely in virtue of the fact that the declaration has been successfully performed." (Searle, 1976, p. 14).

Searle (1976) remarks the fact acknowledged by Austin (1962) that *every* utterance consists in performing (at least) one illocutionary act, which means that, in terms of the illocution performed, there is virtually no distinction between the so-called constatives and performatives. Thus, "just as saying certain things constitutes getting married (a 'performative') and saying certain things constitutes making a promise (another 'performative'), so saying certain things constitutes making a statement (supposedly a 'constative')" (Searle, 1976, p. 14). In other words, "making a statement is as much performing an illocutionary act as making a promise, a bet, a warning or what have you." (Searle, 1976, p. 14). That being said, linguistic competence by itself is not sufficient to perform every illocutionary act: according to Searle (1976), declarations require the understanding of culture-specific institution to be performed. In fact, as Searle (1976) points out "the mastery of those rules which constitutes linguistic competence by the speaker and hearer is not in general sufficient for the performance of a declaration. In addition, there must exist an extra-linguistic institution and the speaker and hearer must occupy special places within this institution. It is only given such institutions as the Church, the law, private property, the state and a special position of the speaker and hearer within these institutions that one can excommunicate, appoint, give and bequeath one's possessions or declare war. As we will see more in detail below, the fact that most declarations require extra-linguistic institutions to be performed makes their automated detection and classification virtually impossible.

## 6. Deep Structure Representations of Searle's Classes

As Jaszczolt (2002) points out, the classification made by Searle in 1976 is based in part on linguistic criteria: generally speaking, indicative mood is used for assertives, imperative mood for directives, and so on, but also specific syntactic rules are imposed by each illocutionary force. For example, "in utterances with the directive point the speaker attempts to get the hearer to carry out

the course of action represented by the propositional content" (Searle & Vanderveken, 1985, p. 37); the propositional content must not be in the past tense and the overall sentence is the surface realization of a deep structure. We have seen before how it is linguistically odd to say something like "I order you to have eaten beans last week" (Searle & Vanderveken, 1985, p. 16) instead of "I order you to eat the beans next week". Searle (1976) provides a deep structure representative of each type of illocutionary force:

- assertives (representatives):

deep structure: I verb (that) + S.

where "verb" stands for a verb of the list of verbs of the assertive type, "that" is optional, and S is an arbitrary sentence. Examples of surface realizations are: "I state that it is raining" and "I predict he will come" (Searle, 1976, p. 17). Some assertive verbs may require further constraints on S; for example, "predict" requires that, inside S, the auxiliary verb is in the future, or anyway not in the past. Moreover, some representative verbs, such as "describe", "call", "classify", and "identify", take a syntactic structure similar to that of many declaratives: I verb NP<sub>1</sub>+ NP<sub>1</sub> be pred.

We say: "I call him a liar, I diagnose his case as appendicitis, I describe John as a Fascist" (Searle, 1976, p.19). These utterances can take the form of statements: "He is a liar, He has appendicitis, He is a Fascist" (Searle, 1976, p.19). An S can thus by itself be an assertive. We can come up with the following rule for statements, bearing into mind that it can also apply to some declarations: x am/are/is/have/has y. We could ideally also add rules that identify as statements all performatives in the second or third person, in the past tense, in the present continuous, and in the passive voice;

- directives:

deep structure: I verb you + you Fut Vol Verb (NP) (Adv).

"'I order you to leave' is thus the surface structure realization of 'I order you + you will leave'" (Searle, 1976, p. 17);

- commissives:

deep structure: I verb (you) + I Fut Vol Verb (NP) (Adv).

"'I promise to pay you the money' is the surface structure realization of I promise you + I will pay you the money" (Searle, 1976, p. 17). Similarly, "I pledge allegiance to the flag" has the following deep structure: I pledge + I will be allegiant to the flag (Searle, 1976, p. 17);

- expressives:

deep structure: I verb you + I/you VP => Gerundive Nom.

We say: "I apologize for stepping on your toe, I congratulate you on winning the race, I thank you for giving me the money" (Searle, 1976, p. 18)

- declaratives (three types):

1) deep structure: I verb NP<sub>1</sub> + NP<sub>1</sub> be pred. (same as some assertives)

We say: "I find you guilty as charged. I now pronounce you man and wife. I appoint you chairman", which are the surface structure realizations of "I find you + you be guilty as charged, I pronounce you + you be man and wife, I appoint you + you be chairman" (Searle, 1976, p. 20);

2) deep structure: I declare + S.

We say: "I declare the meeting adjourned, war is hereby declared", which are the surface structure realizations of "I/we (hereby) declare + a state of war exists, I declare + the meeting be adjourned" (Searle, 1976, p. 20);

3) deep structure: I verb (NP). (the most misleading)

We say: "You're fired. I resign. I excommunicate you", which are the surface structure realizations of "I declare + Your job is terminated, I hereby declare + My job is terminated, I declare + Your membership in the church is terminated" (Searle, 1976, p. 21).

Taking into account syntax will help us disambiguate some of those illocutionary verbs that can have more than one illocutionary force (or belong to more than one class). We have seen that Austin's (1962) classification is ambiguous in that it assumes that any two illocutionary verbs necessarily make explicit two different illocutionary forces. However, as we have mentioned above, there exist some verbs that belong to more than one class. In some of these cases, the illocutionary force of the utterance depends on the syntactic structure in which the illocutionary verbs are embedded. Jaszczolt (2002, p. 303) makes the example of the verb "advise" which can be used as an assertive (69a) or as a directive (69b) depending on the syntax:

69a. She advised us that we have passed the exam.

I verb (that) + S.

69b. I advised you to do it.

I verb you + you Fut Vol Verb (NP) (Adv).

We conclude our discussion on Searle's (1976) classification, and with it our discussion on the theoretical classifications of speech acts, by highlighting the fact that classifying speech acts is a fairly difficult task, not only because of the issue of implicitness, but also because of indirect speech acts and context-dependence. Jaszczolt (2002) argues that "the whole enterprise of distinguishing and classifying speech acts is (...) of limited use in pragmatics" (2002, p. 303) mainly because "there are implied, indirect speech acts that are difficult to classify" (2002, p. 304).

Nevertheless, classifying speech acts has found a number of practical applications. The next sections are dedicated to why and how classifications of speech acts have been implemented in computational linguistics, in spite of the unsolved (or unsolvable) theoretical ambiguities that underlie them.

## 7. Computational Linguistics: Introduction and Motivation

The speech act theory has been put to use widely, but its theoretical origins are discussed very little in computational linguistics. Many works in computational linguistics about speech act detection do not discuss what a speech act is (or is for them), nor provide sufficient reasons for why certain classes or types of speech acts have been defined instead of others. This can be explained by the fact that the classifications proposed in computational linguistics are not based on theoretical premises, but on practical ones. These classifications indeed embrace a simplified notion of illocutionary point, but are built exclusively - with the definition ad hoc classes - to be used within larger projects and programs aimed at specific tasks such as: email conversations tracking, vocal assistance, machine translation, and so on. In order to be implemented for computational purposes, both the notion of illocutionary point and the notion of utterance have undergone a number of changes. Since these changes are, for the most part, specific to each different tag-set, we limit ourselves to reporting here the changes that affect (to different degrees) all the classifications.

First of all, the notion of illocutionary point has been, on many occasions, refined to include different "dimensions" of the intended purpose of the utterance. An utterance often has a main illocutionary point and a secondary illocutionary point, which often (but not necessarily) correspond to a forward looking and a backward looking function of the utterance. Let's consider the following exchange:

70a. A: Can you please send me the document?

70b. B: Yes, I will do it tomorrow.

Roughly speaking, in this context, utterance 70b has both the illocutionary point of a statement (in that it is used to give straightforward information) and the illocutionary point of an acceptance (in that it is used to accept the request made by speaker A). We will see more in detail below how each classification deals with multiple illocutionary points. The second point that we make is that the unit of analysis has changed. We recall that in chapter 1 we defined the utterance as the concrete product of speech and writing or a contextualized sentence which comes "with information as to who the speaker is as well as information about the time, the place and other circumstances of the performed act of speaking" (Jaszczolt, 2002, p. 2), and we defined sentence as the abstract, grammatical unit

that can be derived from an utterance by abstracting over contingent and contextual information. The unit of analysis in computational linguistics corresponds to neither of these. The information available to identify the point of an utterance is mostly textual. However, as we have mentioned above, the discourse will become an important resource for the identification of illocutionary points, especially thanks to the analysis of conversation in the form of adjacency-pairs. As we will see, sometimes also the intonation of the utterance has been encoded. One final point is that utterances do not need to be complete meaningful sentences - with a propositional content - to be considered speech acts: among the others, "wow!", "hmhm", and "ahahah" are speech acts too. At the same time, in some cases the unit of analysis has been expanded to include more than one utterance up to even an entire email or blog post.

In computational linguistics, having at hand a well-defined classification of speech acts proves to be useful for the accomplishment of a wide range of tasks. Cohen, Carvalho, and Mitchell (Cohen et al., 2004; Carvalho & Cohen, 2005; Carvalho & Cohen, 2006; Carvalho, 2008), use speech acts and machine learning techniques to improve work-related email management. A classification of email acts can be used not only to speed up email communication overall, but also to predict leadership roles within email-centered work groups (Carvalho, 2008). The most recent work on email speech acts is that of Carvalho (2008), which can be seen as an improvement of previous works on email acts. Carvalho (2008) takes inspiration from act taxonomies<sup>24</sup> that have been proposed in the research areas of dialog systems, speech recognition (Stolcke et al., 2000; Taylor et al., 1998), and machine translation (Levin et al., 2003). Joty and Hoque (2016), on the other hand, address the issue of automated speech act recognition in virtually every type of written asynchronous conversations, e.g. fora, chats, emails, etc. According to Joty and Hoque (2016), the identification of speech acts "has been shown to be useful in many downstream applications including summarization (McKeown et al., 2007) and question answering (Hong and Davison, 2009)" (p. 1746). Moreover, "[r]evealing the underlying conversational structure in dialogues is important for detecting the human social intentions in spoken conversations and in many applications including summarization (Murray, 2010), dialogue systems and dialogue games (Carlson, 1983) and flirt detection (Ranganath, 2009). As an additional example, Ravi and Kim (2007) show that dialogue acts can be used for analyzing the interaction of students in educational forums" (Tavafi et al., 2013, p. 1).

Generally speaking, synchronous conversations happen in real-time and participants take turns instantly - an example are telephone calls -, whereas asynchronous conversations take place over extended periods of time and participants take turns when it is convenient for them to do so -

---

<sup>24</sup> In the present work, taxonomy and classification are synonyms.

an example are email conversations. As we will see, classifications based on synchronous conversation seem to follow the theory more closely (especially SWBD-DAMSL) than those based on asynchronous conversation, thus making easier their mapping to the classification proposed by Searle (1976). Although the classifications based on synchronous conversation are more loyal to the theory, they also feature a larger number of classes in comparison with classifications based on asynchronous conversation: classifications in synchronous conversation feature between 25 (DAMSL Standard) and 50 (SWBD-DAMSL and MRDA) different classes, whereas classifications in asynchronous conversation feature only between 4 (BC3 original tag-set) and 12 (TA and BC3 new tag-set) classes (more below). The line between synchronous and asynchronous is blurry since some classifications designed for asynchronous conversation have been adapted from classifications designed for synchronous conversation: the TA corpus<sup>25</sup>, for example, has a tag-set which is an adaptation of MRDA's (more below). In computational linguistics, we will often call the classifications of speech acts "tag-sets" to refer to the set of all the possible tags (each indicating a different speech act type or class) that one can use to label an utterance.

## **8. Overview of the Classifications (Tag-sets) in Computational Linguistics**

In this section, we will present the classifications of speech acts proposed in computational linguistics with the goal of clarifying how each classification relates to the others chronologically. For now, we will not dive into the criteria according to which each classification defines its own classes, which we will cover in detail in the next sections. Instead, we will attempt to answer the following question so as to consider the issue of classifying speech acts in its wider context: which classification is inspired or is an adaptation from which other classification? In the process of answering this question, we will inevitably have to mention the fact that the classification proposed by Searle in 1976 constitutes, more or less directly, the theoretical background for all the classifications of speech acts that we consider in the present work. In this section, however, we will not attempt the mapping between theory and practice, nor the mapping between one classification proposed in computational linguistics and the other. We will delve into the differences between the classification proposed by Searle (1976) and those proposed in computational linguistics in the next sections, where we describe each classification proposed in computational linguistics and map it to the classification proposed by Searle (1976). In this section, on the other hand, we will first discuss the classifications of speech acts proposed for synchronous domains (telephone conversations and

---

<sup>25</sup> A corpus can be generally be defined as a collection of written texts.

meetings), and then we will move to those proposed for asynchronous domains (such as emails, chats, fora, etc.).

Searle's (1976) 5 classes are the point of departure for most of the future classifications of speech acts proposed both in linguistics and philosophy and in computational linguistics. In summary, Searle (1976) influenced the definition of the DAMSL Standard (Allen & Core, 1997), which is in turn at the foundation of the SWBD-DAMSL tag-set (Jurafsky et al., 1997), in turn at the basis of the MRDA tag-set (Dhillon et al., 2004). The MRDA tag-set (Dhillon et al., 2004) inspired both the TA tag-set (Jeong et al., 2009) and the new BC3 tag-set (Joty et al., 2011). The original BC3 tag-set, on the other hand, was instead inspired by Cohen & Carvalho (2005). Cohen & Carvalho, in all of their works (Cohen et al., 2004; Carvalho & Cohen, 2005; Carvalho & Cohen, 2006; Carvalho, 2008), refer directly to Searle's (1976) classification. Finally, the QC3 tag-set (Joty & Hoque, 2016) was inspired by the TA tag-set (Jeong et al., 2009) and the new BC3 tag-set (Joty et al., 2011). In short, most researches adopt - directly or indirectly - Searle's (1976) classification as a blueprint for their own, and leverage illocutionary point for the definition of ad hoc classes of speech acts.

## 8.1 Synchronous Conversation Tag-sets

There are two major synchronous spoken domain corpora: the Switchboard-DAMSL (SWBD-DAMSL) corpus (Jurafsky et al., 1997), a corpus of telephone conversations whose tag-set is based on the DAMSL standard (Allen & Core, 1997), and the ICSI Meeting Recorder Dialog Act Corpus or MRDA (Dhillon et al., 2004), a corpus of meetings whose tag-set is based on SWBD-DAMSL. Jurafsky et al. (1997), in their description of the SWBD-DAMSL tag-set, write that "[t]he current version of the discourse tag-set is designed as an augmentation to the Discourse Annotation and Markup System of Labeling (DAMSL) tag-set" or DAMSL standard. Jurafsky et al. (1997) redirect us to the DAMSL standard (Allen & Core, 1997) to find "more theoretical justifications for the particular tagging philosophy" (Jurafsky et al., 1997) of their SWBD-DAMSL tag-set. The DAMSL standard represents the starting point for all the classifications of speech acts for synchronous conversation that we consider in the present study, namely the SWBD-DAMSL tag-set and the MRDA tag-set. Allen and Core (1997), in their definition of the DAMSL standard, provide us with the background information necessary for the understanding of Jurafsky et al.'s (1997) work and classification, as well as of Dhillon et al.'s (2004) work and classification (which in turn refers also to Jurafsky et al. (1997)). Allen and Core (1997) begin by defining dialog as "a spoken, typed or written interaction in natural language between two or more agents" (Allen & Core, 1997), and they divide it into conversational units called turns; during each turn each speaker temporary

controls the dialog by producing one or more utterances. They then base the notion of utterance "on an analysis of the intentions of the speaker" (Allen & Core, 1997). The intentions of the speaker correspond to "why the utterance was spoken" (Allen & Core, 1997), which brings us to Searle's (1976) notion of illocutionary point or purpose of the utterance. In fact, one can answer the question "Why did you speak that utterance?" by making explicit the illocutionary point or purpose of his or her utterance; for example, if the speaker utters "What time is it?" and is asked why he or she uttered that sentence, the speaker can answer by saying "Because I wanted to request a piece of information, i.e. the time" or "It was a request of information" or "Its illocutionary point or purpose was to request information". In the light of this, despite not adopting the same classes defined by Searle (1976), nor referring to Searle's (1969; 1976) work directly, Allen and Core (1997) rely on (Searle's contribution to) the speech act theory, which they adopt as the theoretical background for their definition of the DAMSL standard. As a consequence, the SWBD-DAMSL (Jurafsky et al., 1997) tag-set, which is based on the DAMSL standard, and the MRDA (Shriberg et al., 2004) tag-set, which is based on SWBD-DAMSL, fit within the same theoretical framework. It is important to mention that the observation of actual linguistic data is a common procedure in computational linguistics for the definition of speech act classes. As a consequence, we witness in computational linguistic an abstraction from the classes theorized by Searle (1976).

To be more specific, the MRDA corpus (Shriberg et al., 2004) is a "corpus of over 180,000 hand annotated dialog act tags and accompanying adjacency pair annotations for roughly 72 hours of speech from 75 naturally-occurring meetings" (p. 1). The MRDA tag-set features 50 tags and is adapted (to deal with face-to-face conversations) from the tag-set of the older SWBD-DAMSL corpus of telephone conversations (also 50 classes) (Jurafsky et al., 1997): a corpus of "1155 5-minute conversations, comprising 205,000 utterances and 1.4 million words" (Jurafsky et al., 1997). The DAMSL standard, on which, more or less directly, both the SWBD-DAMSL and the MRDA tag-sets are inspired, was created to be used as a reference for the annotation of spoken domain corpora and is thus considered the starting point for many subsequent synchronous conversation tag-sets. The DAMSL standard can be easily mapped to the classification of speech acts proposed by Searle (1976) since it is a clear adaptation of it (even though Searle (1976) is not explicitly mentioned in the DAMSL standard). As we said, the speech act tags of the DAMSL standard "indicate a particular aspect of the utterance unit" which summarizes the intentions of the speaker, i.e. "why the utterance was spoken" (Allen & Core, 1997). In other words, they indicate the illocutionary point of each utterance. The SWDB-DAMSL tag-set expands the DAMSL standard with specific tags for telephone conversation. In turn, the MRDA tag-set expands the SWDB-DAMSL tag-set to deal with face-to-face conversations (meetings). Some of the tags of MRDA

remain the same of SWDB-DAMSL but modify their meanings. Both SWDB-DAMSL and MRDA corpora are tagged at the utterance level.

## 8.2 Asynchronous Conversation Tag-sets

Corpora of similar dimensions do not exist in asynchronous domains (Joty & Hoque, 2016) as there exist only a few small corpora of asynchronous conversation, the most frequently mentioned of which are the Trip Advisor Corpus or TA (Jeong et al., 2009), a corpus of Trip Advisor forum conversations, and the British Columbia Conversation Corpora or BC3 (Ulrich et al., 2008), a corpus of email conversations and blog posts. Before diving into the tag-sets chosen for these two corpora, we proceed chronologically by considering the works of Cohen, Carvalho, and Mitchell (Cohen et al., 2004; Carvalho & Cohen, 2005; Carvalho & Cohen, 2006; Carvalho, 2008) on the classification of what they call "email speech acts". Cohen, Carvalho, and Mitchell's works on email speech acts mention explicitly the work of Searle (1976), but then divert from the theory as they define their own classes, for the most part empirically, that is to say: they look at the textual contents of email messages and find what kinds of actions it would be useful to capture in work related email exchanges. In addition to this, Searle (1976) and, after him, Allen and Core (1997), Jurafsky et al. (1997), and Dhillon et al. (2004), were reasoning on speech acts at the utterance level - and, in some cases, at the sub-utterance level -, whereas Cohen, Carvalho, and Mitchell label entire email messages with one single speech act type, which is actually represented by a verb-noun pair, such as "Deliver, deliveredData" or "Request, Meeting" (more below). The tag-set of Carvalho and Cohen (2005), designed for email classification, is the starting point for the tag-set proposed in the BC3 email corpus (Ulrich et al., 2008). Interestingly, the BC3 corpus is the only corpus of asynchronous conversations whose tag-set is an adaptation of a previous tag-set used for asynchronous conversations (that of Carvalho & Cohen, 2005). The TA tag-set, the new BC3 tag-set, and the QC3 tag-set, are all in fact (more or less direct) adaptations of the MRDA tag-set, which was conceived for synchronous conversations.

To be more precise, one year after the creation of BC3 (Ulrich et al., 2008), Jeong et al. (2009), in order to tag their new corpus - the TA corpus (plus 40 email threads taken from BC3) - use a tag-set which is a reduced version of the one used to tag the MRDA corpus (a corpus of meetings; synchronous conversations). Jeong et al. (2009) define 12 categories as an adaptation from the MRDA tag-set: they excluded what they call colloquial style interactions, such as backchannel, disruption, and floorgrabber, for their inapplicability in emails and forums. While the TA corpus was created from scratch (Jeong et al., 2009), Jeong et al. (2009) also used the same 12 category-tag-set on a sample of the BC3 corpus, which originally featured only 4 classes (cf. Ulrich

et al., 2008). At this point, the problem of segmentation<sup>26</sup> is quite evident: while Cohen, Carvalho, and Mitchell label entire email messages, almost everybody else labels single utterances. Nonetheless, Ulrich et al. (2008) base their own tag-set on Carvalho and Cohen's (2005) tag-set. Jeong et al.'s (2009) new tag-set of 12 classes (in substitution of the 4 original classes of the BC3 corpus (Ulrich et al., 2008)) was created because the original 4 classes designed for the BC3 corpus were not suitable, according to them, for domain independent applications, nor for tagging texts at the utterance level: as we have mentioned above, the original 4 classes of BC3 were inspired by the 4 classes proposed by Carvalho and Cohen (2005), who worked on a domain-dependent tag-set for labeling entire email messages. "(The TA) tag-set is different from the prior work on DA (dialog act) recognition in asynchronous conversations (...), since it is domain independent and suitable for sentence level annotation" (Joty et al., 2011). Cohen and Carvalho (2005), on the other hand, focused only on email communication and worked on labeling entire emails (not sentences), which explains the domain dependence of the original BC3 tag-set, which was developed from it. The same 12 classes tag-set proposed by Jeong et al. in 2009 was then used again by Joty et al. in 2011. A couple of years later, Joty and Hoque (2016) create the Qatar Computing Conversational Corpus or QC3corpus and tag it with a reduced version of the 12 classes of the TA tag-set. The QC3 corpus is a new data set of 50 conversations retrieved from a community question answering site called Qatar Living (Joty & Hoque, 2016). Joty and Hoque (2016) reduce the 12 classes of MRDA to 5 coarser act types in order to avoid the significant underrepresentation of some classes. As Joty and Hoque (2016) mention, some prior work (Tavafi et al., 2013; Oya & Carenini, 2014) took their same approach.

## 9. DAMSL Standard

Before proposing their classification, Allen and Core (1997) dedicate a few lines to the issue of explicitness. We have said at the beginning of this chapter that the difficulty to determine the action the speaker intends to perform comes from the ambiguity of the utterance in terms of its explicitness. Every utterance has an effect on the subsequent dialogue and interaction, but "[t]he purposes behind an utterance are very complex" to determine (Allen & Core, 1997); for example, in which cases can we say that "*as the result of an utterance*, the speaker is now committed to certain beliefs, or to performing certain future actions? (...)" [emphasis added] (Allen & Core, 1997). We must also bear in mind that there is a distinction between illocutionary force and perlocutionary effects, distinction which Allen and Core (1997) address as follows: "the effect that an utterance has

---

<sup>26</sup> Segmentation is the division of the text into analyzable units.

on the subsequent interaction may differ from what the speaker initially intended by the utterance" (Allen & Core, 1997). In chapter 1, we said that we will focus on the illocutionary force of an utterance instead of trying to predict its possible perlocutionary effects on the hearer. However, identifying the illocutionary force or point of an utterance still remains a difficult task. Allen and Core (1997) are also aware of the fact that some actions can be performed "indirectly". In fact, they point out that the effect the utterance has on the dialogue is accounted for by the Forward Looking Function category of tags (more below) "*even though* the actual form of the sentence might look like something else" [emphasis added] (Allen & Core, 1997), i.e. even though the linguistic form of the utterance might suggest another illocutionary point.

### 9.1 Utterance Tags Proposed in the DAMSL Standard

The DAMSL standard maps utterances to speech act types in the context of spoken bidirectional conversation as a dynamic exchange of intentions. For this reason, in the context of DAMSL (and of SWBD-DAMSL), we may want to talk about dialog acts instead of speech acts, but terminological differences do not concern us particularly at this point. However, we must bear in mind that, because of interruptions, which are typical of spoken dialogs, the DAMSL standard allows to group together a continuous set of utterances into a segment and to tag it with a single label (Allen & Core, 1997). Despite increasing the number of classes - from the 5 of Searle (1975) to 9 -, the DAMSL standard also leaves open the possibility of tagging certain utterances or segments as either uninterpretable (not comprehensible), abandoned (not complete), or self-talk (not intended to be communicated, yet communicated) (Allen & Core, 1997): in these cases the intentions of the speaker may not have been understood properly, thus the tag indicating the speech act performed is not provided. More precisely, the DAMSL standard defines four main categories of tags, the latter two of which include the tags in which we are particularly interested (from Allen & Core, 1997):

1) Communicative Status - records whether the utterance is intelligible and whether it was successfully completed. Possible tags:

- Uninterpretable;
- Abandoned;
- Self-talk.

2) Information Level - (an abstract) characterization of the semantic content of the utterance.

Possible tags:

- Task ("Doing the task"): utterances that advance the task;
- Task-management ("Talking about the task"): utterance that discuss the problem solving or experimental scenario;

- Communication-management ("Maintaining the communication"): utterances that address the communication process;
- Other-level: not falling neatly in any category.

3) the Forward Looking Function - how the current utterance constrains the future beliefs and actions of the participants, and affects the discourse.

4) the Backward Looking Function - how the current utterance relates to the previous discourse.

Allen and Core (1997) define two of what we call "superclasses": Forward Looking Function and Backward Looking Function; each utterance (or segment) can have one or more tags belonging to each superclass. At the same time, "utterances do not need to always have a component at each level. For instance, some utterances may have no Forward Looking Function, while others might have no Backward Looking Function" (Allen & Core, 1997). If Backward Looking Function tags are given, an antecedent (to which the current utterance is responding) must also be provided. The Forward Looking Function category includes 13 tags and the Backward Looking Function category 12 (I will indicate the possible tags in bold).

## 9.2 DAMSL Standard: Forward Looking Function

In Allen and Core's (1997) words: "the Forward Looking Function is a characterization of what effect the utterance has on the dialogue, even though the actual form of the sentence might look like something else". This definition, as we have said, takes into account indirect speech acts. The tags the proposed in the DAMSL standard within the Forward Looking Function are (Allen & Core, 1997; the possible tags are in bold):

- Statement (the speaker makes a claim about the world...)
  - **Assert** (the speaker is trying to change the belief of the addressee)
  - **Reassert** (the speaker thinks that the claim has already been made)
  - **Other-statement** (other)
- Influencing-addressee-future-action (the speaker is suggesting potential actions to the addressee beyond answering a request for information)
  - **Open-option** (the speaker is not creating an obligation that the hearer do the action unless the hearer indicates otherwise)
  - **Action-directive** (the speaker is creating an obligation that the hearer do the action unless the hearer indicates otherwise)
- **Info-request** (the speaker is making a question or another request for information)
- Committing-speaker-future-action (the speaker is committing to perform a future action)

- **Offer** (the commitment is contingent on (depends on) addressee's agreement)
- **Commit** (the commitment is not contingent on (does not depend on) addressee's agreement)
- Conventional Opening Closing
  - **Conventional-opening** (the speaker utters a word, phrase, or sentence that is conventionally used to summon the addressee and/or start the interaction)
  - **Conventional-closing** (the speaker utters a word, phrase, or sentence that is conventionally used in a dialog closing or used to dismiss the addressee)
- **Explicit-performative** (the speaker performs an action by virtue of making the utterance; the speaker declares what is performed)
- **Exclamation** (the speaker utters an exclamation)
- **Other-forward-function** (the speaker performs a forward looking function that is not captured by the current scheme)

First of all, we recall that, according to the DAMSL standard, an utterance can perform simultaneously multiple functions; this can be exemplified by an utterance such as "There is an engine at Avon", which, in the right context, not only is used to inform the listener of the existence of an engine at Avon, but also "states the possibility of using that engine to move some cargo", i.e. it can be used to influence the addressee's future actions (Allen & Core, 1997). A peculiarity of the DAMSL standard is that it has a particular way of dealing with statements. A statement can have different tags according to the context in which it occurs. In the appropriate contexts, a statement can be used to (Allen & Core, 1997):

- make an assertion or answer a question; e.g. "I am at a meeting tomorrow" (i.e. (more explicitly) "I make you aware of the fact that I am at a meeting tomorrow") (tag = Assert; Reassert; Other-statement);
- suggest or request that the addressee engages in some future course of action; e.g. "There is a calculator on the table" (i.e. (more explicitly) "I suggest that we use the calculator on the table") (tag = Open-option);
- request information in the form of an implicit yes/no question; e.g. "The train is late" (with the right intonation) (i.e. (more explicitly) "The train is late, right?") (Allen & Core, 1997) (tag = Info-request);
- make an offer; e.g. "I'm free at 3" (in context of setting up a meeting) (i.e. (more explicitly) "I can meet with you at 3") (Allen & Core, 1997) (tag = Offer), or make a commitment; e.g. "I'll come to your party" (i.e. (more explicitly) "I promise that I'll come to your party") (Allen & Core, 1997) (tag = Commit);

- perform an action by its utterance; e.g. "I quit" (Allen & Core, 1997) (tag = Explicit-performative).

A particular controversial case is when the same psychological state is expressed by different types of utterances: an utterance that can be considered as an implicit apology, such as "I am sorry", falls into the class of assertions since it is seen as making a claim about the world, claim about which the listener can disagree, whereas explicit apologies, such as "I apologize", belongs to the explicit performatives class as the listener cannot disagree with it (Allen & Core, 1997). Here the issue of indirectness comes into play. The fact that the DAMSL standard allows for 6 different tags to be assigned to an utterance whose linguistic form is a statement brings us back to the notion of indirectness. We must note that the DAMSL standard is to be implemented by human annotators, who will have to tag utterances according to their own interpretation of the way in which the linguistic form of the utterance relates to the context. Choosing the right tag among these 6 options will indeed be a much more complicated task for a computer to perform on new textual input, especially if it cannot rely on sufficient contextual information. Bearing in mind that distinguishing between utterance types is often a non trivial task, we will now go through the classification proposed by the DAMSL standard so as to show how Allen & Core (1997) dealt with some of the ambiguities that we have come across previously in this chapter, and in chapters 1 and 2.

In the DAMSL Standard, Statements include both claims about the world and answers to questions. The claims about the world do not need to "be strongly claiming that something is true or false" (Allen & Core, 1997) as this class includes also "weak forms of statement such as hypothesizing or suggesting that something might be true" (Allen & Core, 1997). In other words, statements in the DAMSL standard include Searle's assertives (which commit the speaker to something being the case) and some, but not all, Searle's expressives (which express how the speaker feels about the situation): less explicit expressives such as "I am sorry" and "I am thankful" are tagged as Statements, whereas more explicit expressives such as "I apologize" and "Thank you" are tagged as Explicit-performatives (more below). Also Searle's declarations are coded with the Explicit-performative tag (more below). There is another point that should be made about Statements. We have said in chapter one that, in the context of speech acts, the speaker uses language to intentionally "do something" in the process of conveying meaning. This means that some meaning in the form of information about the state of affairs is often conveyed in the performance of speech acts. Allen and Core (1997) write in this regard: "[n]ote also that we are only coding (as Statements) utterances that make explicit claims about the world, and not utterances that implicitly claim that something is true. As an intuitive test as to whether an utterance makes an explicit claim, consider whether the utterance could be followed by "That's not true". For example,

the utterance "Let's take the train from Dansville" presupposes that there is a train at Dansville, but this utterance is not considered a statement. You couldn't coherently reply to this suggestion with "That's not true". Another example could be that of a promise. As we will see below, an utterance such as "I promise that I will lend you my mobile phone charger" is to be tagged as Commit since it commits the speaker to a future course of action, i.e. to lend his or her mobile phone charger, but at the same time it makes the implicit claim that the speaker owns a mobile phone charger.

In the DAMSL standard, Influencing-addressee-future-action includes all utterances whose purpose "is to directly influence the hearer's future non-communicative actions, as in the case of requests ("Move the train to Dansville" and "Please speak more slowly") and suggestions ("how about going through Corning") (Allen & Core, 1997). There are many verbs in English that describe variations of these acts that differ in strength, including acts like command, request, invite, suggest and plead" (Allen & Core, 1997). Allen and Core (1997) point out that this category must not include utterances whose purpose is to request information, such as "tell me the time", which will be tagged as Info-request (more below). In other words, utterances belonging to the Influencing-addressee-future-action category have the purpose of influencing the hearer to perform some future non-communicative action, thus excluding those utterances whose purpose is to have the hearer provide some kind of information (a communicative action). This category roughly corresponds to Searle's (1976) directives with the main difference that Searle's directives include requests of information, tagged as Info-request in the DAMSL standard. With regard to the linguistic form of the utterances belonging to the Influencing-addressee-future-action category, in addition to the use of the imperative (above), Allen and Core (1997) notice that questions can be made with the intention of influencing the hearer's future non-communicative actions; for example, "'how long will it take if we go through Corning?'" is sometimes used to suggest that they move a train through Corning" (Allen & Core, 1997), but at the same time it can be uttered literally with the intention of soliciting information from the addressee. Before discussing this category any further, we need to mention that Allen and Core (1997) make "the distinction between an Action-directive, which obligates the listener to either perform the requested action or communicate a refusal or inability to perform the action, and an Open-option, which suggests a course of action but puts no obligation on the listener". In this latter category, Open-option, fall utterances that take the form of a Statement such as "There is a calculator on the table": in the right context, such as the context where the interlocutors are doing some complex mathematical operations, the speaker may utter such a sentence to suggest that it would be better to use a calculator. In the case of Open-option utterances, the hearer does not need to address what the speaker said since he or she is not placed under any specific obligations (beyond the principles of rational conversation we have mentioned in chapter

1). On the other hand, if the speaker said "Let's use a calculator", "We/You should use a calculator", or "I suggest that we/you use a calculator" (all Action-directive utterances), the addressee has to explicitly accept or refuse the speaker's suggestion to use a calculator since he or she is now under an obligation to do so. To sum up, both Action-directive and Open-option utterance suggest potential non-communicative actions to the addressee, but while Action-directive utterances put the addressee under the obligation to accept or refuse the request of action made by the speaker, Open-option utterance do not put the addressee under any such obligations.

In the DAMSL standard, Info-Request applies to all utterances whose purpose is to question or make a request in order to receive information. We should tag as Info-Request "any utterance that creates an obligation for the hearer to provide information, using any form of communication" (Allen & Core, 1997), i.e. including nonverbal actions such as the display of graphs etc. We must notice that the Info-request and Influencing-addressee-future-action (Influence-on-listener) categories are similar: "they both apply to suggests and requests (Info-requests request communicative actions and Influencing-addressee-future-action utterances request non-communicative action)" (Allen & Core, 1997). Some examples of utterances tagged as Info-Request are (from Allen & Core, 1997):

- yes/no questions such as "Is there an engine at Bath?", "The train arrives at 3 pm right?", and even "The train is late" (with the right intonation);
- wh-questions such as "When does the next flight to Paris leave?";
- requests for information such as "Tell me the time" but also "Show me where that city is on the map".

In the DAMSL standard, Committing-speaker-future-action (Influence-on-speaker) includes all utterances that "potentially commit the speaker (in varying degrees of strength) to some future course of action" (Allen & Core, 1997). If the utterance's commitment depends on the listener's agreement, the utterance is tagged as Offer, whereas, if the utterance's commitment does not depend on the listener's agreement, the utterance is tagged as Commit. Some examples of utterances within the category of Committing-speaker-future-action are (from Allen & Core, 1997):

Offer:

- typical Offers such as "Shall I come to your office?" or "I'm free at 3" (in context of setting up a meeting);
- Offers with explicit conditions such as "I'll be free after four if my meeting ends on time" or "I can meet at 3 if you're free"

Commit:

- weak Commits such as "Maybe I'll come to your party";

- regular Commits such as "I'll come to your party" or "I promise that I'll be there"

Allen and Core (1997) continue their discussion on Offers and Commits by saying that Commits usually follow a previous Open-option (Influencing-addressee-future-action) such as in (from Allen & Core, 1997):

A: I don't know what to do Saturday night. (Assert)

B: You could go to Bob's party. (Open-option)

A: Great, I'll see you there. (Commit)

Finally, Allen and Core (1997) acknowledge the existence of Conditional Commits such as:

I'll be there if the package arrives on time. (Commit)

but end up tagging them as simple Commits.

The DAMSL standard takes into account other forward looking functions which are relatively rare. They include (from Allen & Core, 1997):

- conventional conversational actions which in turn include conventional-opening functions such as "hi" (greeting) and "Can I help you?" (interaction starter), and conventional-closing functions such as "good-bye";
- explicit performatives such as "you're fired", "I quit", "thank you", "I apologize", by virtue of whose utterance the speaker performs an action. They correspond to Searle's declarations;
- exclamations such as "ouch";
- other forward looking functions not captured by any other category such as signaling an error by uttering "oops".

We must note that conventional openings or closings can be coded with other aspects as well; for example, "Can I help you" can be both a conventional-opening and an offer (Allen & Core, 1997). Finally, Allen and Core (1997) propose a test to determine whether an utterance belongs to the explicit performative class: if you can insert the word "hereby" before the main verb without modifying the meaning of the utterance you have an explicit performative; for example, "You are fired" and "You are hereby fired" have approximately the same meaning (Allen & Core, 1997).

Utterances with no propositional content can sometimes be difficult to label as they might have different interpretations; for example, "okay" and "yes" can be Asserts as well as Commits (Allen & Core, 1997):

A: do you have a cat?

B: yes. (Assert)

A: are you coming to the party?

B: yes. (Commit)

If the speaker accepts a request for action by uttering "okay" and then performs that action, "okay" should be considered a Commit (Allen & Core, 1997). Allen and Core (1997) make the following example:

A: can you tell me the time? (Action-directive)

B: okay. (Commit)

B: three o'clock. (Assert)

### 9.3 DAMSL Standard: Backward Looking Function

As mentioned above, "Backward Looking Functions indicate how the current utterance relates to the previous discourse. For example, an utterance might answer, accept, reject, or try to correct some previous utterance or utterances" (Allen & Core, 1997). The utterance or set of utterances to which the current utterance responds is called the antecedent and is indicated by the Response-to tag (Allen & Core, 1997). The antecedent usually follows directly after the utterance it responds to, but sometimes it is separated by a series of other utterances (Allen & Core, 1997). The tags proposed in the DAMSL standard within the Backward Looking Function are (Allen & Core, 1997; the possible tags are in bold):

- Agreement (the speaker is addressing a previous proposal, request, or claim...)
  - **Hold**(the speaker is not stating their attitude towards the proposal, request or claim...)
  - **Accept** (the speaker is stating their attitude towards the proposal, request or claim, and is agreeing to all of the proposal, request, or claim)
  - **Accept-part** (the speaker is stating their attitude towards the proposal, request or claim, and is agreeing to part of the proposal, request, or claim)
  - **Reject** (the speaker is stating their attitude towards the proposal, request or claim, and is disagreeing with all of the proposal, request, or claim)
  - **Reject-part** (the speaker is stating their attitude towards the proposal, request or claim, and is disagreeing with part of the proposal, request, or claim)
  - **Maybe** (the speaker is stating their attitude towards the proposal, request or claim, and it is not clear whether they are agreeing to or disagreeing with part or all of the proposal, request, or claim)
- Understanding (the speaker is taking an action to make sure that the interlocutors are understanding each other as the conversation proceeds...)
  - **Signal-non-understanding** (the speaker is explicitly indicating a problem in understanding the antecedent)

- **Signal-understanding** (the speaker is explicitly signaling understanding...)
  - Acknowledge** (the speaker is signaling that the antecedent was understood without necessarily signaling acceptance)
  - Repeat-rephrase** (the speaker is signaling that the antecedent was understood by repeating or paraphrasing the antecedent)
  - Completion** (the speaker is signaling that the antecedent was understood by finishing or adding to the clause that the interlocutor is in the middle of constructing)
- **Correct-misspeaking** (the speaker is offering a correction to signal that they believe that the interlocutor has not said what he or she actually intended to say)
- **Answer** (the speaker is answering to a question (to an Info-request))

In the DAMSL standard, the Agreement aspect codes how the speaker views the proposal that previously made by his or her interlocutor, where a proposal can be either "a request that the hearer do something, an offer that the speaker do something, or a claim about the world" (Allen & Core, 1997). Generally speaking, the speaker may explicitly accept or reject all or part of the proposal, or leave the proposal open (Allen & Core, 1997). Allen and Core (1997) make the following examples:

A utt1: Would you like the book and its review? (Offer)

B: Yes Please. (Accept(utt1))

B: I'd like the book. (Accept-part(utt1))

B: I'll have to think about it. (intended literally rather than a polite reject) (Maybe(utt1))

B: I don't need the review. (Reject-part(utt1))

B: No thanks. (Reject(utt1))

As much as we like it to be true, the scenario is not that simple. Firstly, it is not infrequent that the speaker *explicitly* accepts one part of the proposal and *explicitly* rejects the other; for example, one can utter: "I'll take the book but not the review", in which case, according to Allen and Core (1997), the utterance "will be segmented into two utterance units; one marked as Accept-part and the other as Reject-part". Secondly, we must bear in mind the following rule of thumb: while on the one hand the Agreement aspect (Hold, Accept, Accept-part, Reject, Reject-part, Maybe) has to be coded on utterances that are a response to Influencing-addressee-future-action (Action-directive and Open-option), Offers, and Asserts, on the other hand Answers are responses to Info-requests (more on Answers below). However, in some cases, Accepts can also be used to respond to Info-requests, but are usually immediately followed by an Answer, such as in (from Allen & Core, 1997):

A utt1: can you tell me the time? (Info-request)

B utt2: yes. (Accept(utt1))

B utt3: it's 5 o'clock. (Answer(utt1))

As we have seen, in this case, the sentence uttered by A is used to make an indirect request. The Hold tag is added to utterances that follow a proposal but leave the decision open, such as counter-proposals and requests for additional information (Allen & Core, 1997). On the other hand, utterances by which the speaker explicitly expresses uncertainty are tagged with the Maybe tag (Allen & Core, 1997). Allen and Core (1997) make the following example to explain the use of the Hold tag:

A utt1: take the train to Corning. (Action-directive)

B utt2: should we go through Dansville or Bath? (Info-request, Hold(utt1))

A utt3: Dansville. (Assert, Answer(utt2))

In the DAMSL standard, the Understanding aspect "concerns the actions that speakers take in order to make sure that they are understanding each other as the conversation proceeds" (Allen & Core, 1997). We discuss below some of the cases which may generate controversies. Utterances that explicitly indicate misunderstanding are tagged as Signal-non-understanding (SNU) and can be roughly paraphrased as "What did you say/mean?" (Allen & Core, 1997). While Signal-non-understanding utterances are used to express the non understanding of a previous utterance, Hold utterances implies the understanding of the antecedent as it involves the acquisition of additional information (how, why, when, etc.). Allen and Core (1997) make the following examples of Signal-non-understanding utterances (in response to A):

A: take the train to Dansville.

B: Huh? (i.e., What did you say?)

B: What did you say? (i.e., What did you say?)

B: to Dansville?(i.e., What did you say?)

B: did you say Dansville? (i.e., What did you say?)

B: Dansville, New York?(i.e., What did you mean?)

B: Which train?(i.e., What did you mean?)

These above SNU utterances are to be compared to these Hold utterances (from Allen & Core, 1997) (in response to A):

A: take the train to Dansville.

B: through Avon? (i.e., how shall we take the train?)

B: to get the oranges? (i.e., why are we taking the train?)

B: should it leave immediately? (i.e., when should we take the train?)

Signal-understanding utterances explicitly signal the understanding of the antecedent by means of Acknowledgments such as "okay", "yes", and "uh-huh" (Allen & Core, 1997). While many Acknowledgments are also Accept utterances at the Agreement level, Allen and Core (1997) make a few examples of Acknowledgments that are not acceptances. Acknowledgments can be used to signal the understanding of the antecedent (without acceptance):

A: I'll take the Avon train to Dansville.

B: Okay.

but they can also be used to acknowledge (without acceptance) only part of the utterance (while the interlocutor is still speaking), in which case they are often called "backchannel responses" (they occur as interruptions of the sentences uttered by the interlocutor):

A: if I take the engine and a boxcar from Elmira...

B: yes

A: ...how long will that take?

or again:

A: we take the engine at Avon to Bath...

B: uh-huh.

A: ...for the oranges.

The Correct-misspeaking tag indicates utterances whose purpose is to correct what the interlocutor previously uttered. This tag does not apply to utterances by which the speaker corrects him or herself. There is actually no such tag for self-corrections in the DAMSL standard (Allen & Core, 1997).

The Answer tag is used to indicate utterances that comply with an Info-request antecedent (Allen & Core, 1997). Usually, such utterances take the form of declarations, such as (from Allen & Core, 1997):

A utt1: can I take oranges on tankers from Corning? (Info-request)

B utt2: no, you may not; they must be in boxcars. (Assert, Answer(utt1))

Sometimes, Answers can be in the imperative mood, such as (from Allen & Core, 1997):

A utt1: how do I get to Corning? (Info-request)

B utt2: Go via Bath. (Assert, Open-option, Answer(utt1))

It must be noticed that every Answer is also an Assert since it provides the interlocutor with some kind of information (Allen & Core, 1997). In the latter case, the Answer is also marked with the Open-option tag since it describes one of the options for the interlocutor's future action (Allen & Core, 1997). In the case of implicit or indirect questions, Allen & Core (1997) write that an utterance is a question if it is "obvious enough to obligate the hearer to respond with the

information", in which case the antecedent is an Info-request and the current utterance is an Answer. Otherwise, if an utterance is too implicit or indirect to be tagged as an Info-request, it will be tagged as an Assert, such as (from Allen & Core, 1997):

A utt1: I need to get the train to Corning. (Assert)

B utt2: Go via Bath. (Action-directive)

With regard to implicitness or indirectness, Allen and Core (1997) acknowledge the fact that tagging an utterance as Info-request or as Assert is sometimes a matter of degree. There are borderline cases which "are left to the annotator's intuition" (Allen & Core, 1997). For example, in many context, the utterance "I don't know how to get oranges to Corning" counts as Info-request and not as Assert in that it implies an Answer like "You could get them from Bath" (Allen & Core, 1997). Furthermore, we recall that clarification requests in response to Info-requests are not Answers and should instead be tagged as Signal-non-understandings (Allen & Core, 1997). Similarly, the refusal to answer to an Info-request is not considered an Answer, but instead it is seen as an Assert that rejects a request of information (Allen & Core, 1997), such as (from Allen & Core, 1997):

A utt1: How can I get oranges to Corning? (Info-request)

B utt2: I don't know. (Assert, Reject(utt1))

The same tags apply to utterances that reject requests for non-communicative actions such as (from Allen & Core, 1997):

A utt1: Please open the door. (Action-directive)

B utt2: I can't, my arm is broken. (Assert, Rejct(utt1))

Finally, there are cases in which the speaker answers to his or her own questions, in which case the question will be tagged as Info-request and the answer as Answer (Allen & Core, 1997).

## 10. SWBD-DAMSL

The Switchboard-DAMSL or SWBD-DAMSL is a corpus of telephone conversations whose tag-set is based on the DAMSL standard. While the DAMSL standard has a total of 25 tags (13 Forward Looking and 12 Backward Looking), the SWBD-DAMSL has a total of 50 tag (24 Forward Looking and 26 Backward Looking). As we will see in detail below, the mapping between the DAMSL standard and the SWBD-DAMSL tag-set it fairly straightforward since the SWBD-DAMSL tag-set, for the most part, simply splits the classes of the DAMSL standards into a number of subclasses. Just like for the DAMSL standard, the SWBD-DAMSL tag-set allows for the possibility of labeling one utterance with one tag from the Forward Looking Function dimension

plus one tag from the Backward Looking Function dimension, thus making available to the labeler 624 (24 X 26) combinations of tags. Just like for the DAMSL standard, one utterance can simply have one single tag from one of the two dimensions. Jurafsky et al. (1997) have created a number of shortcut codes for the most common combinations of Forward and Backward Looking Functions labels, some of which we will encounter in the discussion below.

### 10.1 SWBD-DAMSL: Forward Looking Function

The SWBD-DAMSL tag-set includes the following classes within the Forward Looking Function dimension ("+" indicates new SWBD-DAMSL classes not present in the DAMSL standard; crossed out classes indicate classes present in the DAMSL standard and no longer used in SWBD-DAMSL):

- Statement
  - Statement-non-opinion +
  - Statement-opinion +
  - ~~Assert~~
  - ~~Reassert~~
  - ~~Other-statement~~
- Open-option
  - Action-directive
- Info-request
  - Yes-No-question +
  - Wh-question +
  - Open-question +
  - Or-question +
  - Or-clause +
  - Declarative-question +
  - Tag-question +
  - Rhetoric-question +
- Offer
  - Commit
- Conventional-opening
  - Conventional-closing
- Explicit-performative
  - Thanking +

You're welcome +

Apology +

- Exclamation
- Other-forward-function

### 10.1.1 SWBD-DAMSL: Statements

In SWBD-DAMSL, the distinction (made in the DAMSL Standard) between Assert, Reassert, and Other Statement is no longer made because of the difficulty to determine with certainty, in casual conversations, whether some piece of information has already been transmitted (Jurafsky et al., 1997). However, Jurafsky et al. (1997) make the distinction between what they call "descriptive/narrative/personal" statements (Statement-non-opinion) and "other-directed opinion statements" (Statement-opinion), distinction which was not made in the DAMSL standard. This distinction allows to capture more effectively responses to opinions, which usually express agreement or disagreement, as distinct from responses to statements of facts, which are usually acknowledgments (or backchannels) (Jurafsky et al., 1997). Jurafsky et al. (1997) provide fairly broad criteria for distinguishing between opinions and non-opinions. They identify three subtypes of the statement non-opinion category (Jurafsky et al., 1997):

- narrative statements, i.e. pieces of story (expressed in the past tense);

- declarative statements, i.e. what Searle (1969) calls statements of brute facts; e.g. "Boulder is north of Denver" (Jurafsky et al., 1997);

- personal statements, i.e. statements with subject pronouns "I" and "we" referring to the speaker and his or her family, and about a personal topic, such as the speaker's dog, house, neighborhood, and even personal opinions about a personal topic (something the listener cannot disagree with), e.g. "I was born in Chicago", "I get along well with my boss" (Jurafsky et al., 1997). The third subtype of statements non-opinion, "personal statements", raises a number of controversies. According to Jurafsky et al. (1997), personal statements look like opinions but are actually not in that they are about something the listener "doesn't really get to be an expert on" (Jurafsky et al., 1997). On the other hand they write that "[i]f the statement is about something more general, that the listener could conceivably have their own (possibly differing) opinion about, then it will be (a statement opinion) (Jurafsky et al., 1997). Jurafsky et al. (1997) also provide some helpful natural language indicators that an utterance belongs to the Statement-opinion category; they are: "I think", "I believe", "It seems", "It's my opinion that", "I mean", "Suppose", "Of course,", impersonal "we", and impersonal "they" as in "they say it rains a lot there" (Jurafsky et al., 1997). In addition to any possible synonyms of the indicators listed above, another natural language indicator that a speaker

is expressing opinions is the conditional (when used to express uncertainty), e.g. "I would say that...". The use of the conditional to express uncertainty can be seen in the following example. This is an exchange between two interlocutors who are clearly non-experts on the topic of the discussion, as indicated by the use of the conditional and of verbs such as "(I) imagine" (from Jurafsky et al. (1997); CONTEXT: topic (general) = rabbits; neither speaker has a pet):

A: I would imagine that they don't have many more than one to start with, either.  
(Statement-opinion)

B: Yeah. (Acknowledge)

A: Rabbits are darling. (Statement-opinion)

B: That would be fun if you could get them trained. (Statement-opinion)

To conclude this paragraph on statements, we must notice that these natural language indicators of Statement-opinions are not infallible heuristics (Jurafsky et al., 1997) and therefore should be used merely as aids for the detection of opinions.

In SWBD-DAMSL, Influencing-addressee-future-action (Open-option and Action-directive) remains the same as in the DAMSL standard: Open-options offer multiple options of non-communicative actions and do not require an explicit answer, whereas Action-directives involve mostly commands to perform a specific non-communicative action and require an explicit answer and/or the performance of the non-communicative action requested. Both Action-directives and Open-options exclude requests for information since they are communicative actions. The syntactic realization of Action-directives is either an imperative - in SWBD-DAMSL, "most of the imperatives are commands to speak ("Go ahead", "Tell me more about that", etc)" -, a question (e.g. "Do you want to go ahead and start?"), or a standard declarative sentence (e.g. "You ought to rent the house") (Jurafsky et al., 1997). Open-options are realized syntactically for the most part as standard declarative sentences, such as "You can go first" or "The suggestion is that we maybe talk about a menu for a dinner party" or "We could talk about my favorite subject" (Jurafsky et al., 1997).

### 10.1.2 SWBD-DAMSL: Info-requests

In SWBD-DAMSL, the tags Yes-No-question, Wh-question, Open-question, Or-question, Or-clause, Declarative-question, and Tag-question are a subset of the DAMSL standard class Info-request. Not only do the types of questions proposed by SWBD-DAMSL have different syntactic properties, but also they expect different kinds of answers. We have seen that distinguishing between Statement-opinions and Statement-non-opinions will help us predict what kind of utterance will follow a statement - usually agreements/disagreements follow opinions and

acknowledgments/backchannels follow non-opinions. Similarly, distinguishing between different types of questions will help us determine what kinds of answer to expect; for example, a Yes-No-question is more likely to get a Yes or No answer than a Wh-question (Jurafsky et al., 1997). Before discussing each type of question, Jurafsky et al. (1997) point out that a question does not need to be a question semantically/syntactically as it can also be a question only pragmatically. An utterance is a question from a syntactic/semantic perspective if the addressee can understand from what is said alone that the utterance was spoken with the intent of questioning (and of wanting an answer); for example, a Yes-no question can have syntactic attributes such as subject-aux inversion and do-support (Jurafsky et al., 1997). As we said, an utterance can also be a question from a discourse perspective, or pragmatically, if the hearer understands from the discourse (or the context) that the utterance was spoken with the intent of questioning (and of wanting an answer). Let's make a couple of examples to clarify this point. Utterance 71a is a Yes-No-question both semantically and pragmatically, utterance 71b and 71c are semantically Statement-non-opinions and pragmatically Yes-No-questions (what Jurafsky et al. (1997) call a Declarative question), utterance 71d is semantically a Yes-No-question and pragmatically an Action-directive, and utterance 71e is a Statement-non-opinion both semantically and pragmatically (from Jurafsky et al., 1997):

71a. Do you have to have any special training? (Yes-No-question)

71b. I don't know if you are familiar with that. (Yes-No-question + Declarative question)

71c. You must be familiar with that. (Yes-No-question + Declarative question)

71d. Can you pass the salt? (Action-directive)

71e. I like cakes. (Statement-non-opinion)

As Searle (1969) asserts, those speech acts that have one force semantically and another force pragmatically are called indirect speech acts. Therefore, 71b, 71c, and 71d are indirect speech acts, whereas 71a and 71e are not.

Let's now discuss one kind of Info-request at a time. According to SWBD-DAMSL, an utterance, in order to be tagged as a Yes-No-question must have *both* the what they call the "pragmatic force" of a question and the syntactic/semantic (and prosodic) markings of a yes-no question (Jurafsky et al., 1997). Typical syntactic markings of a Yes-No-question are subject-aux inversion and do-support (Jurafsky et al., 1997). Some examples of Yes-No-questions are the following (from Jurafsky et al., 1997):

Do you have to have any special training?

Does he bite her enough to draw blood?

Is that the only pet that you have?

Have you tried any other pets?

(Are you) Worried that they're not going to get enough attention?

An utterance is considered to have the syntactic markings of a yes-no question even if it begins with an ellipsed aux-inversion. On the other hand, if an utterance is pragmatically a question but has declarative syntax, it has to be marked as a Declarative question (e.g. if a declarative sentence is pragmatically a Yes-No-question, it will have the Yes-No-questiontag + the Declarative question tag)<sup>27</sup>. On the other hand, if an utterance is syntactically a question but does not function as a question, it can be tagged either as an Action-directive (e.g. "Can you pass the salt?"), but also as a Rhetorical-question or as a Backchannel (Acknowledgment) (Jurafsky et al., 1997). The main difference between Rhetorical-questions and Backchannels is that Backchannels, unlike Rhetorical-questions, lack semantic content (Jurafsky et al., 1997). A few examples of Backchannels are: "really?", "have you?", "do you?", "did you?", "is it?", "it does?", "isn't that amazing?", "you think so?" (Jurafsky et al., 1997). On the other hand, utterances like B of the exchange below are Rhetorical-questions (from Jurafsky et al., 1997):

A: Think what's going to be like for my youngest son when he goes to school.

B: What's going to happen?

A: I'm afraid for him.

In addition to Declarative questions, another case in which declarative statements are used as questions is when they are followed by what Jurafsky et al. (1997) call "question tags". According to Jurafsky et al. (1997), "the (question) tag gives the statement the force of a question". Utterances of this type should therefore be tagged as "Yes-No-question + question tag" to indicate that the statement being made is in fact a Yes-No-question(only) by virtue of the question tag attached to it. Question tags are either aux-inversions - which in turn may (e.g. You like tennis, don't you?) or may not (e.g. You like tennis, do you?) reverse the polarity of the main verb of the preceding statement - or one-words, such as "right?" and "huh?" (Jurafsky et al., 1997). Some examples are the following (from Jurafsky et al., 1997):

I guess a year ago you're probably watching CNN a lot, right? (Yes-No-question + Question tag)

So you live in Utah, do you? (Yes-No-question + Question tag)

That's a problem, isn't it? (Yes-No-question + Question tag)

These cases must be distinguished from those cases where the speaker asks a question at the end of a statement to determine whether the listener has understood the content of the statement, the so-

---

<sup>2727</sup> A declarative question can also be:

Wh-question tag + Declarative question tag; e.g. I don't know what your birthday is.

Or-question tag + Declarative question tag; e.g. I don't know whether you like cats or dogs.

Or Open-question tag + Declarative question tag; e.g. I don't know what you think about owning a dog.

called "understanding checks" (Jurafsky et al., 1997). Understanding checks are tagged as Yes-No-questions (and not as Question tags) (Jurafsky et al., 1997) and the statements preceding them are tagged simply as Statements (and not as Yes-No-questions). That is to say: a declarative statement can be tagged either as a Yes-No-question or as a Statement depending on whether it is followed by a question tag or an understanding check, which are in turn tagged as Question tag and Yes-No-question, respectively. To sum up, a statement followed by a question tag is tagged as Yes-No-question (i.e. Yes-No-question + Question tag), whereas a statement followed by an understanding check remains a Statement (i.e. Statement + Yes-No-question, where Yes-No-question is here the tag for the understanding check). Both types of utterances are followed by either a Yes answer or a No answer, the obvious difference being that answering Question tags means to explicitly agree or disagree with the statement preceding the question tag<sup>28</sup>, or "matrix statement" as Jurafsky et al. (1997) call it, and answering Understanding checks means to explicitly signal the understanding or non-understanding of the matrix statement without implying agreement or disagreement, i.e. without taking any position on it (Jurafsky et al., 1997).

Wh-questions are questions that begin with a "wh-word" and necessarily have subject-inversion (Jurafsky et al., 1997). On the other hand, as we have mentioned above, wh-questions without subject-inversion, are considered declarative questions. Let's make a few examples of wh-questions, whereas YYY and UUU

What cities are they looking at?

How old are you children?

What other long range goals do you have?

Who's your favorite team?

The following are declarative wh-questions:

You said what?

You say you've had him how long?

Open-ended questions are mostly of the "how about you" variety and usually do not place any syntactic constraints on the answer (Jurafsky et al., 1997). Some examples of Open-ended questions are (from Jurafsky et al., 1997): "How about you?", "How about yours?", "What do you think?", "What about your community?", "What are your opinions on it?", etc.

Or-questions are questions that suggest two or more possible answers such as "Do you live in a house or in an apartment?". One problem with Or-questions is that, to quote Jurafsky et al.

---

<sup>28</sup> By agreeing or disagreeing with a statement, the hearer is implying that he or she has understood that statement since he or she could not agree or disagree with that statement if he or she did not understand it.

(1997): "the listener often interrupts before the or clause is complete and answers the or-question as if it were a yes-no question about the first clause"; for example (from Jurafsky et al., 1997):

A: Did you bring him to a doobby obedience school or... (Or-question)

B: No. (No answer)

A: ...train him on your own. (+)

As Jurafsky et al. (1997) point out, there are two ways of labeling such cases depending on whether we take the speaker's point of view or the hearer's point of view. Since, as we have said in chapter 1, we are trying to capture the illocutionary force of each utterance and not how the hearer interprets or reacts to that utterance, we will label "what the speaker thinks" instead of "what the hearer thinks". The first utterance of A is thus an Or-question even though it is not complete. The "+" indicates that the second utterance of A is the continuation of the previous utterance of A since they have been uttered within the same slash unit (Jurafsky et al., 1997). Cases similar to Or-questions are those in which the speaker tacks on an or-clause, as a separate utterance, after a Yes-no question. In these cases, the or-clause has to be tagged as Or-clause; for example (from Jurafsky et al., 1997):

A: What is their location? (Wh-question)

A: Is it Asian? (Yes-no question)

A: Or is it European? (Or-clause)

### 10.1.3 SWBD-DAMSL: Offers and Commits

The tags Offer and Commit in SWBD-DAMSL correspond to the homonymous tags in the DAMSL standard, but with one exception: in SWBD-DAMSL, offers and commits are assumed to occur only within some sort of negotiation (in a weak sense), that is to say: only when the action to which the speaker is committing involves the interlocutor in some way (Jurafsky et al., 1997). For example, the following utterance is a Commit according to the DAMSL standard, but it is a Statement according to SWBD-DAMSL since it does not involve the conversational partner (Jurafsky et al., 1997):

I'm going to try out for crew next season.

Just like the DAMSL standard, SWBD-DAMSL identifies as Offers utterances by which the speaker offers his or her commitment to a future action to the addressee, who can refuse such commitment, that is to say: the speaker's commitment depends on the listener's agreement; for example (from Jurafsky et al., 1997):

I have a recipe if you want.

This utterance commits the speaker to giving the recipe to his or her interlocutor on the condition that the interlocutor agrees to be given the recipe. The addressee may in fact accept or reject the speaker's offer of commitment (Jurafsky et al., 1997):

Okay (Accept)

Sure (Accept)

No (Reject)

Jurafsky et al. (1997) conclude this part on Commits and Offers by asserting that utterances by which the speaker is suggesting, in a polite way, that he or she is about to do something (thus giving the chance to the listener to reply with "no") are to be tagged as Offers. In fact, even though the action itself does not involve the listener, the listener's acceptance is still necessary for the speaker to commit to that action. These sentences usually begin with "let me"; a few examples are (from Jurafsky et al., 1997):

Let me turn off my stereo here.

Let me push the button.

Let me try again.

Hang on let me check.

Other classes within the Forward Dimension are: Conventional-opening, Conventional-closing, Explicit-performative, Exclamation, and Other-forward-function (which includes Thanks, Welcomes, and Apologies). Conventional-openings, Conventional-closings, and Exclamations are fairly self-explanatory: while Conventional-openings and Conventional-closings include all utterances that are conventionally used to open and close, respectively, a conversation - e.g. "hi", "how are you", "I'm doing fine" to open and "bye", "It's been nice talking to you" to close a conversation -, Exclamations include typically one-to-three-word utterances that are conventionally used to make exclamations; these are mostly generated by the following grammar (Jurafsky et al., 1997):

(oh|well|i mean|NIL) (gosh|goodness|boy|goodgrief|jeez|heavens|shoot|gee whiz)

Explicit-performatives and Other-forward-functions need a more in-depth explanation. While in the DAMSL Standard, Explicit-performatives and Other-forward-functions belong to the same class of Explicit Performatives, in SWBD-DAMSL the distinction is made between the so-called Other-forward-functions on the one hand, which include Thanks (e.g. "thank you"), Welcomes (e.g. "you're welcome", and the non-performative "Uh-huh", "Okay", "You bet", "Yeah"), and Apologies (e.g. "Excuse me") (Jurafsky et al., 1997), and on the other hand Explicit-performatives, which include roughly all utterances which are not represented by other classes in the Forward-Communicative-Function dimension whose main verb is a performative verb (verb "in the first

person, present tense, indicative mood, active voice, (which) describes its speaker as performing a speech act" (Green, 2015)); for example: "I (do)recommend the bit", "I bet you can't guess", "I wish you very good luck with it" (Jurafsky et al., 1997). We must notice that, unlike utterances within the Explicit-performative class, which must include a performative verb, Other-forward-functions may or may not include a performative verb. We must also notice that, theoretically speaking, Other-forward-functions with a performative verb belong to the Explicit-performative class. However, the definition of Other-forward-functions by SWBD-DAMSL, i.e. ad-hoc classes which capture Thanks, Welcomes, and Apologies, will provide a more better understanding of the dynamics of the conversation. Finally, we need to mention the fact that there exist some overlaps, which are contextually disambiguated; in particular, Jurafsky et al. (1997) speak of Thanks which have to be marked as Conventional-closings if they are used to end a conversation, and of Apologies, which are apologies by virtue of the fact that they are used to apologize for something the speaker has done, such as a cough or an interruption, but can also be Offers if they are used to obtain approval to do something, e.g. "Excuse me just a second".

## 10.2 SWBD-DAMSL: Backward Looking Function

The SWBD-DAMSL tag-set includes the following classes within the Backward Looking Function dimension ("+" indicates new SWBD-DAMSL classes not present in the DAMSL standard; crossed out classes indicate classes present in the DAMSL standard and no longer used in SWBD-DAMSL):

- Accept
  - Accept-part
  - Maybe
  - Reject-part
  - Reject
  - Hold before answer/agreement
- Signal-non-understanding
  - Signal-understanding
  - Acknowledge
  - Acknowledge-answer +
  - Repeat-phrase +
  - Completion
  - Summarize/reformulate +
  - Appreciation +

- Sympathy +
- Downplayer +
- Correct-misspeaking
- Yes answer +
- No answer +
- Affirmative non-yes answer +
- Negative non-no answer +
- Other answer +
- Expansion of Yes/No answer +
- Dispreferred answer +

### 10.2.1 SWBD-DAMSL: Agreement

Jurafsky et al.(1997) assert that all the classes within the Agreement dimension - Accept, Accept-part, Maybe, Reject-part, Reject, Hold before answer/agreement - "mark the degree to which speaker accepts some previous proposal, plan, opinion, or statement". SWBD-DAMSL thus expands the use of Agreements to include accepts and rejectsof statements, unlike the DAMSL Standard, which seems to reserve Agreements for rejects and accepts of proposals (Jurafsky et al., 1997). A few examples of Agreements are the following exchanges (from Jurafsky et al., 1997):

DIALOG 1 (Accepting a proposal)

A: Go ahead. (Action-directive)

B: Okay. (Accept)

DIALOG 2 (Accepting (Agreeing with) a previous opinion)

A: That was a really good movie. (Statement-opinion)

B: It sure was. (Accept)

DIALOG 3 (Accepting (Agreeing with) a previous non-opinion)

A: I could just sit there all day and look at the scenery. (Statement-non-opinion)

B: Yes, I agree. (Accept)

According to Jurafsky et al. (1997) there are a number of one-line utterances that always indicate Accepts; they are:

Exactly!

Definitely.

Yes.

That's a fact.

That's true.

True.

Jurafsky et al. (1997) argue that "yeah" and, to a lesser extent, "uh-huh" can be used as Accepts, but they are Accepts only if they are used to agree with some previous utterance, otherwise they are either Acknowledges, Welcomes, or Yes answers (dialogs 1 and 3 from Jurafsky et al., 1997):

DIALOG 1 - CONTEXT: topic (general) = rabbits; neither speaker has a pet:

A: I would imagine that they don't have many more than one to start with, either.  
(Statement-opinion)

B: Yeah. (Acknowledge)

DIALOG 2:

A: Thank you. (Thank)

B: Yeah. (Welcome)

DIALOG 3:

A: So you live in Utah, do you? (Yes-no question + Question tag)

B: Yeah. (Yes answer)

The fact that "yeah" has 4 possible different labels depending on the type of utterance by which it is preceded demonstrates, again, how the context is crucial for the identification of the correct use in conversation of a linguistic expression. With regard to Agreements, Jurafsky et al. (1997) continue by asserting that not only one but also two separate utterances can be used to agree with a previous proposal, plan, opinion, or statement: while the first utterance is always tagged as Accept (or any other Agreement), the second utterance is tagged either as Accept (or any other Agreement) or as Statement (either Statement-opinion or Statement-non-opinion), depending on its length: shorter utterances are more likely to be Agreements than longer ones; for example (adapted from Jurafsky et al., 1997):

DIALOG 1

A: That was a really good movie. (Statement-opinion)

B: Yeah. (Accept)

B: You're right. (Accept)

DIALOG 2

A: John is an idiot. (Statement-opinion)

B: Yeah. (Accept)

B: He's an idiot because of his dumb ideas. (Statement-opinion)

A good rule of thumb to distinguish between Agreements and Statements is the following: "Thinking alike generally constitutes agreement; being alike may not" (Jurafsky et al., 1997); for example (from Jurafsky et al., 1997):

DIALOG 1

A: I have a Mercedes. (Statement-non-opinion)

B: Me too. (Statement-non-opinion)

DIALOG 2

A: I like Mercedes. (Statement-non-opinion)

B: Me too. (Accept)

DIALOG 3

A: I think Mercedes are great cars. (Statement-non-opinion)

B: Me too. (Accept)

An example of Reject is the following exchange (from Jurafsky et al., 1997):

A: The whole point of the military is to kill people essentially. As an instrument of US policy. (Statement-opinion)

B: Oh, no. (Reject)

B: It's to defend the nation against external evils. (Statement-opinion)

An example of Reject is the following exchange (from Jurafsky et al., 1997):

A: I don't think women look good with muscles. (Statement-non-opinion)

B: Up to a point. (Accept-part)

Finally, Jurafsky et al. (1997) point out the fact that Maybes often do not actually contain "maybe"; here's a few examples (from Jurafsky et al., 1997):

DIALOG 1:

A: A shotgun hurts worse than a pistol does. (Statement-opinion)

B: Yeah, I suppose. (Maybe)

DIALOG 2:

A: My husband feels that they'll come and collect everybody's guns. (Statement-non-opinion)

B: Yeah. (Acknowledge)

B: I guess that could happen. (Maybe)

DIALOG 3

A: I can't complain too much. (Statement-non-opinion)

B: Yeah. (Acknowledge)

B: I guess so. (Maybe)

B: I don't know. (Maybe)

DIALOG 4

A: I suspect it very much depends upon the job. (Statement-opinion)

B: Huh-uh. (Acknowledge)

B: Maybe. (Maybe)

B: There are some jobs where I guess it doesn't really. (Statement-opinion)

### 10.2.2 SWBD-DAMSL: Understanding

According to Jurafsky et al. (1997), this dimension includes all utterances that mark the understanding or non-understanding of a previous utterance. Very common within this dimension are "backchannels" (also called "continuers" or "assessments"), which we have encountered above, and manifestations of misunderstanding, i.e. requests for repeat and corrections of misspeaking (Jurafsky et al., 1997). We begin by talking about manifestations of misunderstanding since they include only one class of speech acts. The so-called Signal-non-understandings manifest the misunderstanding of a previous utterance but are always also Action-directives in that they are always used to request, more or less directly, that the interlocutor clarify the misunderstanding caused by his or her utterance (Jurafsky et al., 1997). In turn, the interlocutor to which such requests are made is obligated to address them explicitly (signal-non-understandings are Action-directives and not Open-options, which means that the request being made must be explicitly addressed by the recipient). The classic example of Signal-non-understanding is the following (from Allen & Core, 1997):

What did you mean/say?

But there can also be less-direct utterances such as the following (from Jurafsky et al., 1997):

I can't hear you.

There's static on the line.

Acknowledges are used to signal the understanding of the interlocutor's utterance without necessarily signaling acceptance. Accepts, on the other hand, always imply understanding since the speaker could not accept an antecedent that he or she has not understood. SWBD-DAMSL tags as Acknowledges utterances that signal understanding without signaling acceptance, and tags as Accepts utterances that signal acceptance (which always imply understanding). The most frequent pure Acknowledges in SWBD-DAMSL are (from Jurafsky et al., 1997):

38% uh-huh

34% yeah

9% right

3% oh

2% yes

2% okay

2% oh yeah

1% huh

1% sure

1% um

1% huh-uh

1% uh

Jurafsky et al. (1997) mark as Acknowledge also "yeah" when it is used as "incipient speakership", i.e. when it is used by the speaker to indicate that he or she is about to speak; for example (Jurafsky et al., 1997):

A: you know, I don't really feel as though I've a gotten sufficient dose of news that way.  
(Statement-non-opinion)

B: Yeah. (Acknowledge)

B: A lot of my information comes from several sources.(Statement-non-opinion)

B: Probably pretty high up on the list is National Public Radio. (Statement-non-opinion)

Jurafsky et al. (1997) make the distinction between 1) pure Acknowledges, such as the ones listed above, 2) Acknowledges which take the form of a question (or backchannel questions), which for consistency we call Acknowledge-questions, and 3) Acknowledges which are used to acknowledge answers to questions (they follow a question + answer sequence), which Jurafsky et al. (1997) call Acknowledge-answers. Here's a few examples of backchannel questions (from Jurafsky et al., 1997; the number next to them indicates their number of occurrences out of ~740 Acknowledges from the first 755 conversations of SWBD-DAMSL corpus):

141 (Oh,) really?

103 Really?

39 Is that right?

21 (Oh,) yeah?

15 (Oh,) is that right?

14 Do you?

12 Is it?

11 (Oh) really?

10 (Oh,) did you?

10 Are you?

8 Yeah?

6 (Oh,) have you?

6 (Oh,) do you?

- 6 No?
- 6 Did you?
- 5 (Oh,) are you?
- 5 Was it?
- 5 Have you?
- 4 (Oh,) is it?
- 3 (Oh,) you do?
- 3 Isn't that interesting?
- 3 Isn't that amazing?
- 2 (Oh,) it does?
- 2 (Oh,) do they?
- 2 (Oh,) are you really?
- 2 isn't that funny?
- 2 You think?
- 2 You think so?

In SWBD-DAMSL, backchannel questions are 35% of the time answered with "yeah" tagged as Yes-answer (Jurafsky et al., 1997). Jurafsky et al. (1997) make the following example of exchange:

A: It was funny. (Statement-opinion)

A: There was a fireworks display at halftime. (Statement-non-opinion)

B: Oh, yeah? (Acknowledge-question)

A: Yeah. (Yes answer)

Acknowledgments of answers to questions, on the other hand, are tagged as Acknowledge-Answers. The most common Acknowledge-answers in SWBD-DAMSL are (Jurafsky et al., 1997; the number next to them indicates their number of occurrences out of ~1339 Acknowledges from the entire SWBD-DAMSL corpus: 1155 conversations):

- 418 okay
- 284 (oh,) okay
- 144 oh
- 48 (oh,) I see
- 48 I see
- 35 uh-huh
- 18 Yeah
- 14 okay.
- 11 (oh,) yeah

11 right

11 All right

9 (oh,) uh-huh

9 (oh,) okay.

Here's an example of Acknowledge-Answer in an exchange (from Jurafsky et al., 1997):

A: But, I was just curious, what part of the country? (Wh-question)

B: Stockton. (Statement-non-opinion)

A: Okay. (Acknowledge-Answer)

As we can see from this example, Acknowledge-answers must be preceded by a question+answer pair, bearing in mind that the question and the answer need not be contiguous (Jurafsky et al., 1997).

In SWBD-DAMSL, "mimic-other-speaker" is an orthogonal tag which indicates the recycling of lexical material; if we combine it with the pure Acknowledge tag, we obtain the Repeat-phrase tag (Jurafsky et al., 1997). Let's make an example (from Jurafsky et al., 1997):

A: Well, how old are you? (Wh-question)

B: I'm twenty-eight. (Statement-non-opinion) [[[Assert + Answer in the DAMSL Standard]]]

A: Twenty eight. (Repeat-phrase)

A: Okay. (Acknowledge-answer)

A: I'm twenty-three. (Statement-non-opinion)

In SWBD-DAMSL, Summarize-reformulate utterances are used by the speaker who proposes a summarization or paraphrase of another speaker's - and not his or her own - utterance or utterances (Jurafsky et al., 1997). If a speaker is summarizing or paraphrasing his or her own talk, we are dealing with simple Statements (Jurafsky et al., 1997). Here's an example of Summarize-reformulate (from Jurafsky et al., 1997):

A: And you need a special nursing home for that. (Statement-opinion)

A: You need one that has a unit that's locked where they are not able to get out and roam around. (Statement-opinion)

B: Yeah. (Acknowledge)

A: And you need people who are trained for that type... (Statement-opinion)

B: Right. (Acknowledge)

A: ...of problem. (+)

B: Who know what they're doing with that. (Summarize-reformulate; it paraphrases "(people) who are trained for that type of problem")

A: Yeah. (Accept)

Jurafsky et al. (1997) assert that summarizations of other-talk (as well as Completions; see below), function as understanding checks, i.e. they are pragmatically (though not syntactically) questions, "the implicit question being something like 'is this an acceptable summary of your talk?'" (Jurafsky et al., 1997). Summarize-reformulate and Completion utterances are often followed by utterances that signal understanding (Accepts) or non-understanding (Rejects), or by partial acceptances (Accept-parts) or partial rejects (Reject-parts), which means that, counterintuitively, a Summarize-reformulate and a Completion is typically not followed by an Acknowledge or a Yes / No Answer (Jurafsky et al., 1997). Completions, also called "collaborative completions", on the other hand, complete the utterance of the interlocutor while functioning as understanding-checks (Jurafsky et al., 1997); for example (adapted from Jurafsky et al., 1997):

A: In other words, you'd have to murder more than one other person... (Statement-opinion)

B: ...Besides him. (Completion)

A: Yeah. (Accept)

Backwards-attitude is a dimension within the Understanding dimension which is not coded in the DAMSL Standard; it is used to express not only acknowledge/understanding, but also further emotional involvement and/or support (Jurafsky et al., 1997). Backwards-attitude includes three classes: Assessment/Appreciation (the most common), Sympathy, and Downplayer (Jurafsky et al., 1997). An Assessment/Appreciation is "an Acknowledge/Continuer which functions to express slightly more emotional involvement and support" (Jurafsky et al., 1997). Jurafsky et al. (1997) make the following examples of Assessments/Appreciations:

I can understand that.

That would be nice.

I can imagine.

It must have been tough.

That is good.

(Oh,) great.

(Oh,) he'll be delighted.

That's great.

That's great!

That's probably a good idea.

That makes sense.

You bet.

(Uh,) I know exactly what you mean.

Example of Assessment/Appreciation in context (from Jurafsky et al., 1997):

A: Especially if it's after an acute illness. (Statement-non-opinion)

A: To get over a... (Statement-non-opinion)

A: Or to rehab after an illness. (Statement-non-opinion)

B: That's true. (Accept)

B: I never thought of that. (Assessment/Appreciation)

Sympathy includes markers of sympathy in response to somebody else's previous utterance (Jurafsky et al., 1997). Sympathy excludes actual apologies (for doing something), which are tagged as Apology (Forward Looking). Downplayers are used to respond to apologies and compliments. An example of Sympathy and Downplayers the following (adapted from Jurafsky et al., 1997):

A: My dog died. (Statement-non-opinion)

B: I'm real sorry. (Sympathy)

A: That's all right. (Downplayer)

A: He was old. (Statement-non-opinion)

Here's an example of Downplayer as a response to a compliment (from Jurafsky et al., 1997):

A: You are well versed on the subject, I tell you. (Statement-opinion)

B: Well, I don't know. (Downplayer)

The most common types of Downplayers in the SWBD-DAMSL corpus are (Jurafsky et al., 1997; the number next to them indicates their number of occurrences in the entire SWBD-DAMSL corpus: 1155 conversations):

24 that's okay

7 no

5 that's all right

4 okay

3 (oh,) that's okay

2 it's okay

2 Uh-huh

2 No

Finally, Correct-misspeakings are not-so-frequent utterances that are used correct somebody else's utterance or utterances. They are sometimes followed by an acknowledgement of the error by the interlocutor (Jurafsky et al., 1997). An example of Correct-misspeaking is to be found in following exchange (from Jurafsky et al., 1997):

A: I suppose they all have the balloons. (Statement-non-opinion)

B: The air bags. (Correct-misspeaking)

B: Yeah. (Acknowledge)

### 10.2.3 SWBD-DAMSL: Answer

According to Jurafsky et al. (1997), the Answer dimension includes all utterances that are in response to Info-requests. While the DAMSL Standard has no subtyping of answers, SWBD-DAMSL defines 4 macroclasses (3 of which represent different possible answers to Yes-No-questions, and the remaining represents answers non Yes-No-questions). Each of the first three macroclasses is in turn divided into 3 classes of answers (from Jurafsky et al., 1997):

- Answers to (pragmatic) Yes-No-questions:

1) Affirmative Answers:

- Yes answers (i.e. answers that are "yes" or a variant)
- Affirmative non-yes answers (i.e. answers that are not "yes" or a variant)
- Yes plus expansion (i.e. answers that are "yes" or a variant + an expansion)

2) Negative Answers:

- No answers (i.e. answers that are "no" or a variant)
- Negative non-no answers (i.e. answers that are not "no" or a variant)
- No plus expansion (i.e. answers that are "no" or a variant + an expansion)

3) Other answers:

- Other answers (i.e. none of the above, such as "maybe", "I don't know", etc.)
- Dispreferred answers (such as "well...")
- Hold (same of the Agreement dimension)

- Answers to non Yes-No-questions:

4) Answers to Wh-questions, Open-questions, and Or-questions:

- Statements (sometimes preceded by a Hold)
- Dispreferred answers

In SWBD-DAMSL, Yes-answers consist mostly of the following utterances (from Jurafsky et al., 1997; the number next to them indicates their number of occurrences in the first 18 conversations of the SWBD-DAMSL corpus; note that pauses and discourse markers are considered part of the Yes-answer):

17 Yeah

5 yes

5 uh-huh

3 (uh,) yeah

2 (oh,) yeah

1 (oh,) yes

1 (well,) yes

1 yes (uh,)

1 yes, actually

1 yeah, I do

1 yep

In SWBD-DAMSL, No-answers consist mostly of the following utterances (from Jurafsky et al., 1997; the percentage in parentheses indicates the relative frequency of occurrence of each, and the number next to them their number of occurrences, out of 942 No-answers from the first 755 conversations of the SWBD-DAMSL corpus; note that pauses and discourse markers are considered part of the No-answer):

709 no (75%)

49 (uh,) no (5%)

45 huh-uh (5%)

22 (well,) no (2%)

19 (oh,) no (2%)

16 (um,) no (2%)

11 uh-huh (1%)

9 no (uh,) (1%)

5 nope (< 1 %)

3 (uh,) actually no (< 1 %)

2 yes (< 1 %)

2 yeah (< 1 %)

2 so no (< 1 %)

2 probably not (< 1 %)

2 (but)(uh,) no (< 1 %)

2 but no (< 1 %)

2 actually no (< 1 %)

Note that just as Yes-answers do not include "he/she is" or "he/she does" (and the like), No-answers do not include "he/she isn't" or "he/she doesn't" (and the like), since they are respectively Affirmative non-yes answers and Negative non-no answers (Jurafsky et al., 1997); for example (from Jurafsky et al., 1997):

A: Is that the only pet that you have? (Yes-No-question)

B: It is. (Affirmative non-yes answer)

Other examples of Affirmative non-yes answers are the following (first one from Jurafsky et al., 1997):

EXCHANGE 1

A: Do you have kids? (Yes-No-question)

B: I have three kids. (Affirmative non-yes answer)

EXCHANGE 2

A: Did they just get away with it? (Yes-No-question)

B: I guess. (Affirmative non-yes answer)

An example of Negative non-no answer is the following (from Jurafsky et al., 1997):

A: Did you happen to see last night the special on Channel Two with James Galway? (Yes-No-question)

B: We don't get Channel Two. (Negative non-no answer)

On the other hand, answers that begin with "yes" (or variants) and "no" (or variants) and are then "expanded" must be tagged respectively as Yes plus expansion and No plus expansion (Jurafsky et al., 1997); for example (from Jurafsky et al., 1997):

A: Okay, um, Chuck, do you have any pets there at your home? (Yes-No-question)

B: Yeah, I do. (Yes plus expansion)

If the expansion is an independent utterance after the Yes / No answer, it should be marked as Statement expanding Yes / No answer (Jurafsky et al., 1997). Note that, according to Jurafsky et al. (1997), only the first utterance after the Yes / No answer has to be tagged as an expansion even though, as they admit, the utterances that follow the first one will often also be expansions of the Yes / No answer. Let's consider the following example (adapted from Jurafsky et al., 1997):

A: Do you live with you parents? (Yes-No-question)

B: No. (No answer)

B: I live alone in an apartment. (Statement-non-opinion expanding Yes / No answer)

B: It's on Histon road. (Statement-non-opinion)

In SWBD-DAMSL, expansions of Affirmative non-yes answers and of Negative non-no answers are not marked as expansions at all (Jurafsky et al., 1997); for example (from Jurafsky et al., 1997):

A: Do you ride a lot of rallies or a lot of those around there? (Yes-No-question)

B: Not so much. (Negative non-no answer)

B: Uh, I guess mostly I bike on my own. (Statement-non-opinion)

Other answers include responses to Yes-No-questions "that are neither affirmative responses ("yes" or "Indeed I do") nor negative responses ("no" or "I don't think so")" (Jurafsky et al., 1997). The

most common Other answer is "I don't know" (Jurafsky et al., 1997). Jurafsky et al. (1997) make the following example of Other answer:

A: Do you think the jury should have a dollar figure for losing an arm, a dollar figure for losing different body parts? (Yes-No-question)

B: I don't know. (Other answer)

Dispreferred answers are pre-answer sequences which can either be used 1) to respond negatively to a question that presupposes an affirmative answer or 2) to respond positively to a question that presupposes a negative answer (Jurafsky et al., 1997). In the first scenario, we respond to Yes-No-questions or to Yes-No-questions + (Negative) Question tags - both presupposing a positive answer - with a negative answer (Jurafsky et al., 1997); for example (adapted from Jurafsky et al., 1997):

PREFERRED

A: You like Clinton, don't you?(Yes-No-question + (Negative) Question tag)

B: Yes, I do. (Yes plus expansion)

vs.

DISPREFERRED

A: You like Clinton, don't you?(Yes-No-question + (Negative) Question tag)

B: No, I don't. (Dispreferred answer)

In the second scenario, we respond to Yes-No-questions + (Positive) Question tags - presupposing a negative answer - with a positive answer (Jurafsky et al., 1997); for example (adapted from Jurafsky et al., 1997):

PREFERRED

A: Um, you don't have a problem with that, do you? (Yes-No-question + (Positive) Question tag)

B: No, I don't. (No plus expansion)

vs.

DISPREFERRED

A: Um, you don't have a problem with that, do you? (Yes-No-question + (Positive) Question tag)

B: Actually, I do. (Dispreferred answer)

Basically, any time preferred patterns are contradicted by speakers, we may expect a Dispreferred answer (Jurafsky et al., 1997). If however the Dispreferred answer is after a Yes / No answer within the same utterance, it is not coded (Jurafsky et al., 1997).

## 11. MRDA

MRDA, or Meeting Recorder Dialog Act, is a "corpus of over 180,000 hand-annotated dialog act tags and accompanying adjacency pair annotations for roughly 72 hours of speech from 75 naturally-occurring (multi-party) meetings (...) The meetings were recorded at the International Computer Science Institute (ICSI) as part of the ICSI Meeting Recorder Project" (Shriberg et al., 2004, p. 1). MRDA was adapted from SWBD-DAMSL (Jurafsky et al., 1997) to deal with face-to-face conversations (Shriberg et al., 2004); SWBD-DAMSL, on the other hand, deals with telephone conversations and is, in turn, an adaptation of the DAMSL Standard (Allen & Core, 1997). As Shriberg et al. (2004) point out, MRDA features human-human casual conversation instead of human-human task-oriented dialog. MRDA codes three types of information: 1) Dialog Act (DA) segment boundaries (beginning and ending of the DA), 2) the DA itself, and 3) the correspondences between DAs (adjacency pairs) (Shriberg et al., 2004). MRDA segments DAs on the basis of the function of the different speech regions, as well as by paying attention to pauses and intonation (Shriberg et al., 2004). Some utterances are prosodically one unit but contain multiple DAs; in such cases, a pipe bar ( | ) is put to separate between one DA and the other (Shriberg et al., 2004). Different DAs within the same prosodic unit may or may not be seen as separate utterances according to the particular research goals of the researcher / programmer (Shriberg et al., 2004).

Just like in the transition between the DAMSL Standard and SWBD-DAMSL, also in the process of adaption of MRDA from SWBD-DAMSL some of the previous classes have been modified, some have been added, and some others have been deleted. Detailed information about the labeling technique of MRDA can be found in Dhillon et al. (2004). In the following sections, we will present the classes of MRDA by comparing them to the classes of SWBD-DAMSL. First of all we need to talk about segmentation. MRDA is "more specific" than SWBD-DAMSL with regard to segmentation. In fact, while SWBD-DAMSL tags so-called slash units, MRDA tags text at the utterance and sometimes even at the "sub-utterance" level. MRDA segments text for it to be tagged with dialog acts, and marking dialog acts often requires to split utterances into smaller units if different functions - or dialog acts - are performed by the different parts of the same utterance. MRDA, in fact, presupposes that, when necessary, utterances are split into smaller units and before they are tagged. SWBD-DAMSL, too, splits utterances but does so only in special occasions, i.e. when a speaker is interrupted in the middle of an utterance by his or her interlocutor, in which case the "second part" of the utterance, in the case the speaker actually finishes it, will be marked with a "+". Before delving into sub-utterance segmentation, let's consider Dhillon et al.'s (2004) definition of utterance. MRDA segments speech in such a way that an utterance is not necessarily formed by a

grammatically complete sentence; an utterance can in fact be an incomplete sentence, clause, or phrase as long as it has a unique function within conversation (Dhillon et al., 2004). An utterance can also be formed by a single word (Dhillon et al., 2004). In Dhillon et al.'s (2004) words, an utterance of MRDA consists of either a noun phrase, a verb phrase, or both (Dhillon et al., 2004). However, according to the theory, an utterance can also be a single prepositional phrase, adverbial phrase, adjective phrase and so on (e.g. A: "Would you like to dine in our out" B: "out"). Differently than the theory, Dhillon et al. (2004) split utterances into two or more separate utterances if they encounter a syntactic indicator such as the following: "and", "or", "but", "so", "because", etc. - and then tag them separately -, with the exception of when such indicators connect two phrases of the same type (i.e. noun phrase with noun phrase, or verb phrase with verb phrase) (Dhillon et al., 2004). Moreover, Dhillon et al. (2004) argue that just like segmenting utterance at the sub-sentential level, also segmenting parentheticals will contribute to the maximization of information provided by dialog acts. At the same time, prosody - i.e. the elements of language that are not encoded by grammar or vocabulary; e.g. rise and fall of pitch, energy level, duration of the words - also plays an important role in detecting utterance boundaries (Dhillon et al.; 2004). In fact, an utterance may be syntactically complete but prosodically incomplete (Dhillon et al., 2004). Pauses are also important for determining utterance boundaries, where the longer the pause the higher the chance of text segmentation (Dhillon et al., 2004).

SWBD-DAMLS uses concatenations of tags to mark utterances that perform particular functions in conversation. Each of the concatenated tags is, so to speak, on the same level, i.e. none of them bears a special status within what we call the "compound tag". In SWBD-DAMSL there are indeed a few exceptions to this. There are in fact a few tags, some of what Jurafsky et al. (1997) call "orthogonal tags", that could only occur attached to others. The exceptions in SWBD-DAMSL are the rule in MRDA. The fact that some tags can only occur in concatenation with others and cannot occur alone is at the basis of MRDA's tagging method. MRDA has explored the idea of "main tag and secondary tag(s)" to the point that two different sets of tags have been created: one set includes the general tags that represent all the possible basic forms of an utterance (e.g. statement, question, backchannel, etc.), the other set includes the specific tags that represent the functions or the characteristics an utterance may have in addition to its basic form (e.g., accepting, rejecting, acknowledging, rising tone, etc.) (Dhillon et al., 2004). From the point of view of the utterance, in MRDA, each utterance has one, and only one, general tag, plus one or more optional specific tags if the general tag is not enough to characterize the utterance and thus further characterization is needed (Dhillon et al., 2004). Specific tags cannot be used in isolation and, when more than one of them is needed, they need to be attached to the general tag in alphabetical order (Dhillon et al.,

2004). Some restrictions apply in constructing labels: if on the one hand there are particular specific tags that can only be attached to certain general tags, on the other hand there are specific tags that cannot appear together within the same dialog act (Dhillon et al., 2004). As an aside, in the present work, we will not cover MRDA's Disruption Forms, i.e. tags "used to mark utterances that are indecipherable, abandoned, or interrupted" (Dhillon et al., 2004, p. 19).

It is important to mention the fact that utterances that were considered one single speech act in SWBD-DAMSL, can be multiple speech acts in MRDA; for example, in MRDA, compound classes of the type "x + expansion" have been deleted as they cannot be used anymore; the reason for this is that utterances which would be labeled with a compound tag are not tagged with multiple dialog acts (or speech acts): they are split into two or more units and each unit is labeled separately. For tagging purposes, the expansion in MRDA is tagged as separate unit; for example (from Jurafsky et al., 1997; gen = general tag / spec = specific tag):

SWBD-DAMSL:

A: Okay, um, Chuck, do you have any pets there at your home? (Yes-No-question)

B: Yeah, I do. (Yes plus expansion)

MRDA<sup>29</sup>:

A: Okay, um, Chuck, | do you have any pets there at your home? (Floor Grabber(gen) | Yes-No-question(gen))

B: Yeah, | I do. (Statement(gen) + Accept, Yes Answer(spec)) | Statement(gen) + Expansion of Yes / No answer(spec))

In MRDA, in the case of multiple DAs within the same utterance, we tag each of the different portions of the utterance with a different tag. As Dhillon et al. (2004, p.18) say: "[t]he use of a pipe bar indicates that segmenting an utterance is not necessary, despite that the initial portion of an utterance, or last portion in the case of Tag-Questions, has a different DA than the rest of the utterance". For example, any utterances containing a Floor grabber and a Statement or a Floor Holder and a Statement require multiple dialog acts, just as multiple dialog acts are needed for Statements followed by question tags (Tag-Question) (Dhillon et al., 2004). In the latter case, we recall that SWBD-DAMSL, too, uses two different tags (instead of one single tag): one for the

---

<sup>29</sup> With regards to the pipe ("|"):

"The pipe bar is indicated in the appropriate location within the label as well as within the transcription. Within the label, the pipe bar separates the DAs. Within the transcript, the pipe bar separates the portions of an utterance to which the different DAs apply. This is done in such a manner that the DA to the left of the pipe bar in the label pertains to the portion of the utterance to the left of the pipe bar in the transcript and the DA to the right of the pipe bar in the label pertains to the portion of the utterance to the right of the pipe bar in the transcript" (Dhillon et al., 2004, p. 18).

statement and one for the question tag. However, while SWBD-DAMSL concatenates the two tags into a compound tag, MRDA splits the utterance into two units and tags each of them separately.

Before we describe the rules governing the usage of each tag of MRDA, we need to mention the fact that Dhillon et al. (2004) base their tag-set on a corpus of audio-recorded meetings. This means that prosody is an integral part of the information available to them for discriminating among the different tags. As they themselves admit, "[w]ith regard to the examples provided within this section, it is of much use to listen to the corresponding audio portions, as some examples cannot be fully comprehended otherwise. In particular, utterances marked as floor grabbers <fg>, floor holders <fh>, holds <h>, backchannels <b>, acknowledgements <bk>, and accepts <aa> share a common vocabulary which renders examples of these tags in text insufficient in fully communicating how utterances marked as such are identified" (Dhillon et al., 2004, p. 32). In other words, while Dhillon et al. (2004) provide written examples of the use of each of their tags, such examples are not sufficient to fully understand the different uses of those tags that share a common vocabulary and thus look identical in text.

We said that, in MRDA, each utterance has one single general tag, plus one or more optional specific tags - to be attached to the general tag in alphabetical order -, should the general tag not be sufficient to adequately characterize the utterance (Dhillon et al., 2004). We have also seen that specific tags can only be used to further characterize an utterance and thus cannot be used in isolation like general tags (Dhillon et al., 2004). In section 11.1 we will list all general tags and in section 11.2 we will list all specific tags. The other sections of chapter 11 are dedicated to the description of the tags that compose the MRDA tag-set.

### 11.1 MRDA: General Tags

The MRDA tag-set includes the following classes within the General Tags set (from Dhillon et al., 2004; "+" indicates new MRDA classes not present in the SWBD-DAMSL; crossed out classes indicate classes present in SWBD-DAMSL and no longer used in MRDA):

- Statement
- ~~Statement-non-opinion~~
- ~~Statement-opinion~~
- Yes-No-question
- Wh-question
- Or-question
- Or-clause
- Open-question

- Rhetorical-question
- Backchannel (or Continuer)
- Floor Grabber +
- Floor Holder +
- Hold before answer/agreement

## 11.2 MRDA: Specific Tags

The MRDA tag-set includes the following classes within the Specific Tags set (from Dhillon et al., 2004; "+" indicates new MRDA classes not present in the SWBD-DAMSL; crossed out classes indicate classes present in SWBD-DAMSL and no longer used in MRDA; the classes are listed according to the alphabetical order of their respective tags):

- Accept, Yes Answer
- Partial Accept
- Maybe
- Reject, No Answer
- Partial Reject
- Assessment/appreciation
- Correct-misspeaking
- Downplayer
- Rhetorical-question continuer
- Acknowledge-answer
- Signal-non-understanding
- Reformulate/summarize
- Misspeak Self-Correction +
- Understanding Check +
- Sympathy
- Commit
- ~~Conventional-opening~~
- ~~Conventional-closing~~
- ~~Explicit-performative~~
- ~~Other-forward-function~~
- Command (Action-directive)
- ~~Open-option~~
- Suggestion (Offer)

- Declarative-Question
- Defending/Explanation +
- Expansion of Yes / No answer (all utterances of the type "x + expansion" have been deleted)
- "Follow me" +
- Apology
- Exclamation
- Thanks
- Welcome
- Tag-Question
- Humorous Material +
- Mimic other
- Narrative-affirmative answer (Affirmative non-yes answer)
- Dispreferred answer
- Narrative-negative answer (Negative non-no answer)
- No knowledge answer (Other answer)
- Repeat +
- Rising tone +
- About-task
- Topic change +
- Self-talk
- Third-party-talk
- Collaborative completion
- ~~Quoted Material~~
- ~~Hedge~~
- ~~Continued from previous line~~

At this point, instead of describing first all general tags and then all specific tags, we prefer to follow Dhillon et al.'s (2004) work and describe dialog act tags group by group, where each group includes a number of both general and specific tags that share the same characteristics. Note, however, that each group does not necessarily include both general and specific tags as they sometimes only include either general or specific tags. Note also that, even though we describe the classes of MRDA group by group, we will always make explicit whether a class and its corresponding tag is general or specific (general and specific tags in fact play different roles in the tagging process).

### 11.3 MRDA Group 1: Statements

In MRDA, the distinction is not made between Asserts, Reasserts, and Other Statements (cf. DAMSL Standard), nor between "descriptive/narrative/personal" statements (Statement-non-opinions) and "other-directed opinion statements" (Statement-opinions) (cf. SWBD-DAMSL). In MRDA, all statements are tagged as Statements and Statement is the most frequently used tag in the MRDA corpus (Dhillon et al., 2004). When necessary, Statements can be further characterized by appending a specific tag to the Statement general tag; for example, the first example below is a simple Statement, whereas the other three are Statements with a further characterization (from Dhillon et al., 2004, p. 33):

if we exclude English um - there is not much difference with the data. (Statement(gen))

It's a great story. (Statement(gen) + Assessment/appreciation(spec))

so this changes the whole mapping for every utterance. (Statement(gen) + Understanding Check(spec))

okay. (Statement(gen) + Acknowledge-answer(spec))

We do not know yet what kind of utterances the specific tags used above, i.e. Assessment/appreciation, Understanding Check, and Acknowledge-answer, designate (unless we borrow their definitions from the previous tag-sets, with the exception of Understanding Check, which is not present in the above-mentioned tag-sets). However, we understand their purpose: each specific tag marks a different type of statement.

### 11.4 MRDA Group 2: Questions

In MRDA, there are different general tags for Questions: Yes-No-question, Wh-question, Or-question, Or-clause After Yes-No-question, Open-ended question, and Rhetorical-question. They almost identical to the corresponding classes in SWBD-DAMSL. Let's describe them one at a time:

- Yes-No-questions, just like in SWBD-DAMSL, are all utterances that have both the pragmatic force and the syntactic and prosodic indications of a yes-no question, i.e. subject-aux inversion and question intonation (Dhillon et al., 2004). Question intonation is marked in the Yes-No-question with an additional specific tag: Rising Tone. Yes-No-questions elicit Yes / No Answers, but it is not necessarily the case that the answer they will receive is a simple yes or no (Dhillon et al., 2004). Here's some examples of Yes-No-questions (from Dhillon et al., 2004, pp. 33-34):

do you think that would be the case for next week also? (Yes-No-question(gen) + Rising Tone(spec))

did I say that? (Yes-No-question(gen) + Rising Tone(spec))

Didn't they want to do language modeling on you know recognition compatible transcripts?  
(Yes-No-question(gen) + Understanding Check(spec1) + Rising Tone(spec2))

Is this channel one? (Yes-No-question(gen) + Rising Tone(spec))

The Yes-No-question tag is used not only as the general tag for Tag Questions and Rhetorical Question Backchannels, but also as the general tag for Declarative Questions (Dhillon et al., 2004). We recall that, in SWBD-DAMSL, Tag Questions (or Question Tags) - i.e. questions attached at the end of a Statement consisting of either aux-inversion (e.g. "do you?", "aren't you?", etc.) or a single word (e.g. "right?", "huh?", etc.) (Jurafsky et al., 1997) - are treated as part of the same utterance of the statement; such utterances belong to the Yes-No-question + Question Tag class, where the Statement becomes Yes-No-question by virtue of having a question tag attach to it. We encountered the following example (from Jurafsky et al., 1997):

I guess a year ago you're probably watching CNN a lot, right? (Yes-No-question + Question tag)

In MRDA, the statement and the question tag are treated separately; for example:

I guess a year ago you're probably watching CNN a lot, (Statement(gen))  
right? (Yes-No-question(gen) + Declarative Question(spec1) + Tag Question(spec2) + Rising Tone(spec3))

Declarative Yes-No-questions, on the other hand, are marked in SWBD-DAMSL by the compound tag Yes-No-question + Declarative question. In MRDA, Declarative Yes-No-questions have the following notation (from Dhillon et al., 2004, p. 34):

the insertion number is quite high(?) (Yes-No-question(gen) + Understanding Check(spec1) + Declarative-question(spec2) + Rising Tone(spec3))

Finally, Rhetorical Question Backchannels are marked in SWBD-DAMSL as Acknowledge-questions (or backchannel questions) and in MRDA as Rhetorical-question continuers (or Rhetorical-question backchannels); for example (from Dhillon et al., 2004, p. 34):

oh really? (Yes-No-question(gen) + Rhetorical-question continuer(spec))

We will delve into each Specific Tag in section 12.2. For now, we use such examples to demonstrate how utterances are treated differently in MRDA with respect to SWBD-DAMSL, sometimes also with regard to segmentation. Another example of segmentation discrepancy is that of a Yes-No-question followed by an elaboration. In these cases, in MRDA, the elaboration requires its own line - i.e. it is separated from the Yes-No-question it elaborates on - and is marked with the Elaboration tag (Dhillon et al., 2004); for example (Dhillon et al., 2004, pp. 34-35)

wasn't there some experiment you were going to try? (Yes-No-question(gen) + Rising Tone(spec))

where you did something differently for each um uh - I don't know whether it was each mel band or each uh um f f t bin or someth- (Statement(gen) + Elaboration(spec))

In SWBD-DAMSL, the elaboration would not be separated from the Yes-No-question and the entire utterance will end up having the Yes-No-question tag.

- "Wh-questions are questions that require a specific answer" (Dhillon et al., 2004, p. 53) and usually contain a so-called wh-word (what, which, where, when, who, why, how) (Dhillon et al., 2004). However, containing a wh-word does not necessarily make a question a Wh-question as there are also Open-ended questions that begin with a wh-word (Dhillon et al., 2004; more on Open-ended questions below). Here's a few examples of Wh-questions (from Dhillon et al., 2004, p. 35):

why didn't you get the same results and the unadapted? (Wh-question(gen) + Repeat(spec1) + Third-party-talk(spec2))

I guess - what time do we have to leave? (Wh-question(gen) + Third-party-talk(spec1))

In addition to utterances that contain wh-words, there can be other types of utterances that function as wh-questions and thus need to be tagged as such; for example "huh?", "excuse me?", and "padron?" are similar "what?" as requests for repetition (Dhillon et al., 2004, p. 36). However, Wh-questions that do not contain wh-words can be easily confused with all the other classes of dialog acts that share the same vocabulary, i.e. "floor grabbers, floor holders, holds, backchannels, yes/no questions that are rhetorical question backchannels, (and) acknowledgments" (Dhillon et al., 2004, p. 36).

Declarative Wh-questions can either include a wh-word, such as A from dialog 1, or not include a wh-word, such as A from dialog 2. Declarative Wh-questions without a wh-word are usually requests for repetition. Both examples are adapted from Dhillon et al. (2004, pp. 36-37):

#### DIALOG 1

A. I don't understand what you are saying about the spanish.(Wh-question(gen) + RepetitionRequest(spec1) + Declarative Question(spec2))

B. the spanish labels were in different format. (Statement(gen))

#### DIALOG 2

A. and you're saying about the spanish(?) (Wh-question(gen) + Repetition Request(spec1) + Declarative Question(spec2) + Rising Tone (spec3))

B. the spanish labels were in different format. (Statement(gen))

- "Or-questions offer the listener at least two answers or options from which to choose" (Dhillon et al., 2004, p. 37); for example (from Dhillon et al., 2004, p. 37):

are we going to - i mean - is it going to be over there or is it going to be in there? (Or-question(gen) + Rising Tone(spec))

are you assuming that or not? (Or-question(gen))

do we have like a cabinet on order or do we just need to do that? (Or-question(gen) + Rising Tone(spec))

is this the same as the email or different? (Or-question(gen))

Or-questions receive answers in which the interlocutor selects one of the options proposed by the Or-question (Dhillon et al., 2004). Just like we have seen in SWBD-DAMSL, sometimes the Or-question is interrupted and is answered as a Yes-No-question. In such cases the interrupted Or-question is still tagged as Or-question and not as Yes-No-question (Dhillon et al., 2004), i.e. we assume the point of view of the speaker by marking his or her (non fulfilled) intention; for example (Dhillon et al., 2004, p. 39):

per channel or? (Or-question(gen) + Rising Tone(spec))

- Just like SWBD-DAMSL, MRDA recognizes the existence of Yes-No-questions followed by Or Clauses, such as (from Dhillon et al., 2004, p. 40):

do you have the true source files? (Yes-No-question(gen) + Rising Tone(spec))

or just the class? (Or Clause(gen))

Notice that the Or Clause can consist of just the word "or" (Dhillon et al., 2004).

- Open-ended Questions are questions that "place few syntactic or semantic constraints on the form of the answer it elicits" (Dhillon et al., 2004, p. 41). Open-ended Questions may contain a wh-word or may look like Yes-No-questions or Or-questions; the difference lies in the fact that Open-ended questions, unlike Wh-questions, Yes-No-questions, and Or-questions, do not seek a specific answer (Dhillon et al., 2004). Some examples of Open-ended questions are (from Dhillon et al., 2004, pp. 41-42):

and anything else? (Open-ended question(gen) + Declarative Question (spec1) + Rising Tone(spec2))

anybody have any institutions or suggestions? (Open-ended question(gen))

but - | what - do you think about that? (Floor Grabber(gen) | Open-ended question(gen))

what about the um - your trip yesterday? (Open-ended question(gen) + About-Task(spec))

Questions? (Open-ended question(gen) + Declarative Question(spec))

- Rhetorical Questions are "questions to which no answer is expected ... used by the speaker for rhetorical effect" (Dhillon et al., 2004, p. 42). The difference between Rhetorical questions and Rhetorical-question backchannels (or Rhetorical-question continuers) is that the latter lack semantic content, function mostly as continuers, and are not used by a speaker who has the floor (Dhillon et

al., 2004, p. 42). Some examples of Rhetorical questions are the following (from Dhillon et al., 2004, pp. 42-43):

- I mean is this realistic? (Rhetorical question(gen) + Rising Tone(spec))
- why not? (Rhetorical question(gen) + Accept, Yes answer(spec))
- i mean who cares? (Rhetorical question(gen))
- isn't that wonderful? (Rhetorical question(gen) + Appreciation/Assessment(spec))
- why don't you read the digits? (Rhetorical question(gen) + Command(spec))
- uh - | but who knows? (Floor Holder(gen) | Rhetorical question(gen))

### 11.5 MRDA Group 3: Floor Mechanisms

Another group of general tags is what Dhillon et al. (2004) call Floor Mechanisms. Floor Mechanisms involve "all general tags pertaining to mechanisms of grabbing or maintaining the floor" (Dhillon et al., 2004, p. 43). To put them into context, a turn - term with which we are already familiar - is "the period during which a speaker has the floor" (Dhillon et al., 2004, p. 2), i.e. the period during which a speaker produces one or more utterances. Floor Mechanisms cannot have any specific tags attached to them (Dhillon et al., 2004). They split into three types:

1) Floor Grabbers: utterances that the speaker uses to gain attention (or to gain the floor), usually by interrupting the interlocutor who at that moment has the floor, so that he or she may begin to speak (Dhillon et al., 2004). For this reason, Floor Grabbers usually occur at the beginning of a speaker's turn (Dhillon et al., 2004, p. 43). "Common floor grabbers include, but are not limited to, the following: 'well,' 'and,' 'but,' 'so,' 'um,' 'uh,' 'I mean,' 'okay,' and 'yeah.'" (Dhillon et al., 2004, p. 44). However, as Dhillon et al. (2004) point out, Floor Grabbers cannot be identified merely on the basis of the vocabulary used, but rather on the basis of the context in which such vocabulary is used. In fact, as we have mentioned above, similar vocabulary is shared by the following classes of dialog acts: Floor Grabbers, Floor Holders, Holds, Backchannels, Acknowledge-answers, and Accepts (Dhillon et al., 2004). Any of the above-mentioned words are thus Floor Grabbers only if uttered in the context which makes them attempts - either successful or unsuccessful - to gain the floor (Dhillon et al., 2004);

2) Floor Holders: utterances that the speaker who is currently the floor holder uses midspeech "as a means to pause and continue holding the floor" (Dhillon et al., 2004, p. 45), with the exception of when they are used at end of a turn, in which case they may be used to relinquish the floor (Dhillon et al., 2004, p. 45). "Common floor holders include, but are not limited to, the following: 'so,' 'and,' 'or,' 'um,' 'uh,' 'let's see,' 'well,' 'and what else,' 'anyway,' 'I mean,' 'okay,' and 'yeah'" (Dhillon et al., 2004, p. 45). As mentioned above, Floor Holders appear very similar in text to a number of other

classes of speech acts. Therefore, they need to be investigated by taking into account their context of utterance and their sound (Dhillon et al., 2004). While Dhillon et al. (2004) do not provide any examples of Floor Grabbers in their surrounding context, they report a number of exchanges including Floor Holders (p. 46):

so it's a rather huge thing. (Statement(gen))

but um - um - | we can sort of (Floor Holder(gen) | Statement(gen))

i think we got plenty of stuff to talk about. (Statement(gen))

and then um - | just see how a discussion goes. (Floor Holder(gen) | Statement(gen))

3) Holds are uttered by "a speaker who is given the floor and is expected to speak (but) "holds off" prior to making an utterance" (Dhillon et al., 2004, p. 46). Holds are usually used by the speaker to pause or "hold off" before he or she answers a question (Dhillon et al., 2004). "Common holds include, but are not limited to, the following: 'so,' 'um,' 'uh,' 'let's see,' 'well,' 'I mean,' 'okay,' and 'yeah'" (Dhillon et al., 2004, p. 46). Holds and Floor Holders, despite being very similar in sound, differ in terms of their location within a speaker's turn: while Holds occur at the beginning of a speaker's turn, Floor Holders occur in the middle or at the end of a speaker's turn (Dhillon et al., 2004). Moreover, while Holds indicate that a speaker has just been given the floor (mostly by asking him or her a question), Floor Holders indicate that a speaker merely has the floor and is either trying to keep it or to give it away (Dhillon et al., 2004). As mentioned before, the context needs to be investigated to properly identify a Hold. Here's an example of Hold and Floor Holder within a dialog (from Dhillon et al., 2004, p. 47):

A: i mean what was the rest of the system? (Wh-question(gen))

B: um (Hold(gen))

B: yeah it was - it was uh the same system (Statement(gen))

B: uhhuh (Floor Holder(gen))

B: it was the same system. (Statement(gen) + Repeat(spec))

B: huh (Floor Holder(gen))

## 11.6 MRDA Group 4: Backchannels and Acknowledgements

The next group of tags MRDA defines is that of Backchannels and Acknowledgments which includes the general tag for Backchannels (or Continuers), and the specific tags for Acknowledgements (or Acknowledgments), Assessments/Appreciations, and Rhetorical-question backchannels (Dhillon et al., 2004). Utterances marked with this group of tags are most often used as responses, in the form of acknowledgments or backchannels, to another speaker's talk; generally, they do not

elicit feedback, nor attempt they to halt the interlocutor (Dhillon et al., 2004). Let's begin by describing the general tag for Backchannels.

In MRDA, "backchannels (or continuers) are utterances made in the background that simply indicate that a listener is following along or at least is yielding the illusion that he is paying attention. When uttering backchannels, a speaker is not speaking directly to anyone in particular or even to anyone at all" (Dhillon et al., 2004, p. 49). Dhillon et al. (2004, p. 49) make the following examples of Backchannels: "uhhuh," "okay," "right," "oh," "yes," "yeah," "oh yeah," "uh yeah," "huh," "sure", and "hm". On the other hand, utterances such as "uh", "um", and "well" are not usually Backchannels as they are rather used to indicate that a speaker is attempting to grab the floor and say something (Dhillon et al., 2004, p. 49). However, we mentioned before that Backchannels, Floor Grabbers, Floor Holders, Holds, Acknowledgements, and Accepts share an almost identical vocabulary and thus need the context to be discriminated properly; as Dhillon et al. (2004) reiterate, "[u]tterances labeled with these tags tend to appear very similar in text yet emerge exceedingly different in sound" (p. 49). Nonetheless, there are a number of rules that help us distinguish between the above-mentioned types of utterances. For example, while Acknowledgments and Accepts usually occur after the interlocutor has terminated his or her utterance - since they respectively acknowledge the semantic content of what the other speaker has said (Acknowledgments) and agree with such content (Accepts) -, Backchannels can also, but not necessarily, occur in the middle of the interlocutor's utterance (Dhillon et al., 2004). Generally speaking, an utterance produced before the interlocutor has terminated his or her own is likely a Backchannel - and not an Acknowledgment - since who speaks it most often cannot acknowledge or agree to an utterance that has not been finished and is thus semantically incomplete, or semantically insignificant (Dhillon et al., 2004). "Additionally, backchannels are usually uttered with a significantly lower energy level than the surrounding speech, while acknowledgments tend not to be quite so low as backchannels and accepts are generally at the same level or else higher" (Dhillon et al., 2004, p. 49). Here's a few examples of Backchannels (from Dhillon et al., 2004, p. 50):

#### EXCHANGE 1

A: but I think that uh - this was a couple years ago.(Statement(gen))

B: huh. (Backchannel(gen))

#### EXCHANGE 2

A: do you get out a - uh - a vector of these ones and zeros and then try to find the closest matching phoneme to that vector? (Yes-No-question(gen) + Rising Tone(spec))

B: uhhuh. (Backchannel(gen))

There is only one specific tag that can be appended to a Backchannel, namely the Rising Tone tag (Dhillon et al., 2004).

Unlike Backchannels, Acknowledgments (or Acknowledge-answers) must occur as a response to a semantically significant utterance or portion thereof (Dhillon et al., 2004). Acknowledgments are neutral in the sense that they are used to acknowledge the interlocutor's (or sometimes even a speaker's own) utterance (or significant portion of it) without agreeing or disagreeing with it (Dhillon et al., 2004). Acknowledgments can be either mimicked portions of the interlocutor's utterance or one of the following: "I see", "okay", "oh", "oh okay", "yeah", "yes", "uhhuh", "huh", "ah", "all right", "got it", and similar (Dhillon et al., 2004, p. 50). As we have mentioned before, detecting Acknowledgments solely on the basis of their vocabulary would mislead us into thinking that the context plays no role at all in defining what an Acknowledgment is. In reality, as we have seen, Acknowledgments are very similar in text to, i.e. they share a very similar vocabulary with, Backchannels, Accepts, Floor Grabbers, Floor Holders, and Holds even though they emerge very different in terms of how each of them sounds like and what the position of each is within the dialog (Dhillon et al., 2004). An utterance marked as an Acknowledgment cannot be marked also with a tag belonging to the Response group (e.g. Accept, Reject, Maybe, etc.) since every utterance can be used either primarily to acknowledge or to agree/disagree, but not both. Moreover, an utterance tagged as an Acknowledgment cannot be tagged also as an Assessment/appreciation and vice versa because of redundancy since Assessment/appreciations are, by definition, also Acknowledgments with the difference that, unlike Acknowledgments, they are either positive or negative but never neutral (Dhillon et al., 2004). A similar situation of redundancy would occur if we conjoined Acknowledgments and Rhetorical-question continuers as the latter are a type of backchannel or acknowledgment (Dhillon et al., 2004). Here's a list of the specific tags with which Acknowledgments can be used in conjunction: Mimic other, Repeat, Rising Tone, Exclamation, Self-talk, and 3rd-party-talk (Dhillon et al., 2004, p. 51). Here's a few examples of Acknowledgments (adapted from Dhillon et al., 2004, pp. 51 - 52):

#### EXCHANGE 1

A: why didn't you get the same results as the unadapted? (Wh-Question(gen) + 3rd-party-talk(spec))

B: oh because when it estimates the transformer produces like single matrix or something. (Statement(gen) + 3rd-party-talk(spec))

A: oh I see. (Statement(gen) + Acknowledgment(spec1) + 3rd-party-talk(spec2))

#### EXCHANGE 2

A: it opens the assistant that tells you that the font type is too small. (Statement(gen))

B: ah. (Statement(gen) + Acknowledgment(spec))

In MRDA, "Assessments/appreciations are acknowledgments directed at another speaker's utterances and function to express slightly more emotional involvement than what is seen in the utterances marked with the Acknowledgment tag" (Dhillon et al., 2004, p. 52). In simple terms, while Assessments/appreciations are either positive or negative, Acknowledgments are neutral. Assessments/appreciations that express negative emotions, especially longer ones, are often criticisms, or at least they are perceived as such (Dhillon et al., 2004). Assessments/appreciations are often, but not always, quite short - in which case, they are usually uttered as exclamations (Dhillon et al., 2004). As a final remark, we often tag as Assessments/appreciations "[c]omments and opinions on an aspect a speaker has noticed within the contents of another speaker's speech" (Dhillon et al., 2004, pp. 52-54). Here's a few examples of Assessments/appreciations (from Dhillon et al., 2004, p. 52):

It's very exciting. (Statement(gen) + Assessment/appreciation(spec1))

wonderful. (Statement(gen) + Assessment/appreciation(spec1))

wonderful! (Statement(gen) + Assessment/appreciation(spec1) + Exclamation(spec2))

That's good. (Statement(gen) + Assessment/appreciation(spec1))

That's good! (Statement(gen) + Assessment/appreciation(spec1) + Exclamation(spec2))

wow! (Statement(gen) + Assessment/appreciation(spec1) + Exclamation(spec2))

So I think that's a really great way to approach it. (Statement(gen) + Assessment/appreciation(spec1))

Finally, Assessments/Appreciations can also be Affirmative Answers, Dispreferred Answers, or Negative answers; such types of utterances are used both to assess/appreciate and to agree/disagree with the interlocutor's utterance (Dhillon et al., 2004); for example (adapted from Dhillon et al., 2004, p. 53):

A. I was wondering if I should study abroad.(Statement(gen))

B1. I think that would be worth doing.(Statement(gen) + Assessment/appreciation(spec1) + Affirmative Answer(spec2))

B2. That's wonderful. (Assessment/appreciation(spec1))

In MRDA, Rhetorical-question Backchannels or Rhetorical-question Continuers are syntactically similar to Rhetorical Questions, however they lack semantic content and function as backchannels and acknowledgments (Dhillon et al., 2004). In most cases Rhetorical-question Backchannels or Rhetorical-question Continuers are uttered as backchannels, that is they are uttered - without speaking to anyone in particular - to indicate or to yield the illusion that the listener is paying attention (Dhillon et al., 2004). Less frequently, they are uttered as acknowledgments, that is

they are uttered to express the acknowledgment of a previous interlocutor's utterance or of a semantically significant portion thereof, thus denoting direct communication between speakers (Dhillon et al., 2004). Rhetorical-question backchannels always receive the Yes-No-question general tag (Dhillon et al., 2004). Here's a few examples of Rhetorical-question Backchannels (Dhillon et al., 2004, pp. 55-56):

oh really? (Yes-No-question(gen) + Rhetorical-question Backchannel(spec))

yeah? (Yes-No-question(gen) + Rhetorical-question Backchannel(spec))

isn't that interesting? (Yes-No-question(gen) + Rhetorical-question Backchannel(spec))

you think so? (Yes-No-question(gen) + Rhetorical-question Backchannel(spec))

To conclude, we recall that an utterance which functions as an acknowledgment may be tagged with only one of the following tags: Acknowledgment, Assessment/appreciation, Rhetorical-question continuer, excluding any combinations thereof (Dhillon et al., 2004). To quote Dhillon et al. (2004): "the default tag for acknowledgments is the Acknowledge-answer tag. If further descriptions apply to an acknowledgment and an Assessment/Appreciation or Rhetorical-question Backchannel tag is deemed necessary, then only one of these tags is used (as) (t)he Acknowledge-answer tag cannot be used in conjunction with the Assessment/Appreciation or Rhetorical-question continuer tags" (Dhillon et al., 2004, p. 55).

## **11.7 MRDA Group 5: Responses**

The next group of tags MRDA defines is that of Responses, in turn orthogonally divided into three subgroup: positive utterances, negative utterances, and uncertain utterances (Dhillon et al., 2004). Responses are often used as responses to questions and suggestions (Dhillon et al., 2004).

### **11.7.1 POSITIVE**

#### **11.7.1.1 Accept**

"The Accept tag is used for utterances which exhibit agreement to or acceptance of a previous speaker's question, proposal, or statement" (Dhillon et al., 2004, p. 57). The Accept tag marks utterances that are "quite short" as compared to the "Affirmative Answer" which marks their "lengthy counterparts" (Dhillon et al., 2004, p. 57). Some examples of Accepts are "yeah," "yes," "okay," "sure," "uhhuh," "right," "I agree," "exactly," "definitely," and "that's true", as well as "no" "if it is used to agree to a syntactically negative statement or question" (Dhillon et al., 2004, p. 57).

Accepts are to be distinguished from backchannels and acknowledgments since they "have much more energy and are more assertive" (Dhillon et al., 2004, p. 57). We recall that Accepts, Floor Grabbers, Floor Holders, Holds, Backchannels, and Acknowledgements share a very similar vocabulary and therefore cannot be discriminated solely on the basis of their vocabulary (Dhillon et al., 2004, p. 57). Since they usually appear very similar in text the context has to be taken into account. Let's make a few examples of Accepts (adapted from Dhillon et al., 2004, p. 57-58):

#### EXCHANGE 1

A: if you want to decrease the importance of a c- - parameter you have to increase its variance. (Statement(gen))

B: yes (Statement(gen) + Accept(spec))

B: right (Statement(gen) + Accept(spec))

#### EXCHANGE 2

A: because when you train up the aurora system you're uh - you're also training on all the data. (Statement(gen) + Defending/Explanation(spec))

B: that's right. (Statement(gen) + Accept(spec))

### 11.7.1.2 Partial Accept

"The Partial Accept tag marks when a speaker explicitly accepts part of a previous speaker's utterance. Partial accepts are often conditional responses that accept or agree to another speaker's utterance." (Dhillon et al., 2004, p. 59). Partial Accepts are not to be confused with Partial Rejections: while Partial Accepts focus on "agreeing with or accepting part of a previous speaker's utterance" (Dhillon et al., 2004, p. 59), Partial Rejections focus on "disagreeing with or rejecting part of a previous speaker's utterance" (Dhillon et al., 2004, p. 59). Here's a few examples of Partial Accepts (adapted from Dhillon et al., 2004, p. 59-60):

#### EXCHANGE 1

A: well the - the - sort of the landmark is - is sort of the object. (Statement(gen) + Understanding Check(spec1) + Rising Tone(sepc2))

A: right?(Yes-No-question(gen) + Declarative-Question(sepc1) + Tag-Question(spec2))

B: usually. (Statement(gen) + Partial Accept(spec))

#### EXCHANGE 2

A: removing all these k l t's and putting one single k l t at the end. (Statement(gen) + Offer(spec))

A: yeah I mean that would be pretty low maintenance to try it. (Statement(gen) + Affirmative Answer(spec))

B: uh - | if you can fit it in. (Floor Holder(gen) | Statement(gen) + Partial Accept(spec))

### 11.7.1.3 Affirmative Answer

"The Affirmative Answer tag marks an utterances that act as narrative affirmative responses to questions, proposals, and statements. The Affirmative Answer tag is much like the Accept tag in that they both exhibit agreement to or acceptance of a previous speaker's question, proposal, or statement. The difference between the two tags is that, as the Accept tag is used for shorter utterances, the Affirmative Answer tag is used for lengthy utterances" (Dhillon et al., 2004, p. 60). In order to properly distinguish an Affirmative Answer from a Statement we need to investigate the context (Dhillon et al., 2004). Here's an example of Affirmative Answer (adapted from Dhillon et al., 2004, p. 60):

A: a cabinet is probably going to cost a hundred dollars two hundred dollars something like that. (Statement(gen))

B: yeah I mean - you know - we - we can spend under a thousand dollars or something without - without worrying about it. (Statement(gen) + Affirmative Answer(spec))

## 11.7.2 NEGATIVE

### 11.7.2.1 Reject

"The Reject tag marks negative words such as "no" and other semantic equivalents that offer negative responses to questions, proposals, and statements. The Reject tag marks brief negative responses to questions, proposals, and statements in the same manner that the Accept tag marks brief affirmative answers." (Dhillon et al., 2004, p. 61). Some examples of Rejects are the following: "no," "nope," "no way," "nah," "not really," and "I don't think so." (Dhillon et al., 2004, p. 61). It is worth pointing out that positive responses to syntactically negative questions and statements can function as Rejects, just like negative responses to syntactically negative questions and statements can function as Accepts (Dhillon et al., 2004). Here's an example of Reject (adapted from Dhillon et al., 2004, p. 62):

A: is there an ampersand in dos? (Yes-No-question(gen) + Rising Tone(spec))

B: nope. (Statement(gen) + Reject(spec))

### 11.7.2.2 Partial Reject

"The Partial Reject tag marks when a speaker explicitly rejects part of a previous speaker's utterance. Partial rejections are often responses posing exceptions when rejecting another speaker's utterance" (Dhillon et al., 2004, p. 62). Here's an example of Reject (adapted from Dhillon et al., 2004, p. 63):

A: it would actually slow that down tremendously. (Statement(gen))

B: not that much though. (Statement(gen) + Partial Reject(spec))

### 11.7.2.3 Dispreferred Answer

"The Dispreferred Answer tag marks statements which act as explicit narrative forms of negative answers to previous speakers' questions, proposals, and statements in the same manner in which the Affirmative Answer tag acts as an agreement with or acceptance of a previous speaker's utterance. As with the Affirmative Answer tag, the Dispreferred Answer tag marks lengthier utterances than those marked with the Reject tag which exhibit rejection" (Dhillon et al., 2004, p. 63). Just like in the case of Affirmative Answers, the context is required to properly distinguish between Dispreferred Answers from general Statements (Dhillon et al., 2004). Finally, Dispreferred Answers differentiate themselves from Negative Answers as they indicate explicit rejections, unlike Negative Answers, which indicate implicit rejections through the use of hedging (Dhillon et al., 2004). Here's couple of examples of Dispreferred Answers (adapted from Dhillon et al., 2004, p. 64):

#### EXCHANGE 1

A: we figured out that it was twelve gigabytes an hour. (Statement(gen) + Understanding Check(spec1) + Rising Tone(spec2))

B: it was more than that. (Statement(gen) + Dispreferred Answer(spec))

#### EXCHANGE 2

A: do you want to try? (Yes-No-question(gen) + Rising Tone(spec))

B: i'd prefer not to. (Statement(gen) + Dispreferred Answer(spec))

### 11.7.2.4 Negative Answer

As we have mentioned above, "[a]s opposed to a dispreferred answer (Dispreferred Answer) which *explicitly* offers a negative response to a previous speaker's question, proposal, or statement, a negative answer (Negative Answer) *implicitly* offers a negative response with the use of hedging [emphasis added]" (Dhillon et al., 2004, p. 64). Dhillon et al. (2004) clarify the difference between Maybes, Other Answers (or No Knowledge Answers), and Negative Answers as follows (p. 64):

- Maybes are "utterances in which a speaker asserts that his response is probable, yet not definite";

- Other Answers (or No Knowledge Answers) are "utterances in which a speaker does not know an answer";
- Negative Answers are "indirect negative response(s)", which [o]ftentimes (...) appear as alternative suggestions to a previous speaker's question, proposal, or statement".

Here's couple of examples of Negative Answers (adapted from Dhillon et al., 2004, p. 65-66):

#### EXCHANGE 1

A: you guys have plans for Sunday? (Yes-No-Question(gen) + Rising Tone(spec))

A: because we also want to combine it with some barbeque activity where we just fire it up and what - whoever brings whatever you know can throw it on there. (Statement(gen))

B: well I'm going back to visit my parents this weekend. (Statement(gen) + Negative Answer(spec))

#### EXCHANGE 2

A: what if we give people you know - we cater a lunch in exchange for them having their meeting here or something? (Wh-Question(gen))

B: well you know - i - i do think eating while you're doing a meeting is going to be increasing the noise. (Statement(gen) + Negative Answer(spec))

#### EXCHANGE 3

A: can we actually record? (Yes-No-Question(gen) + Rising Tone(spec))

B: uh | well we'll have to set up for it. (Floor Holder(gen) | Statement(gen) + Negative Answer(spec))

### 11.7.3 UNCERTAIN

#### 11.7.3.1 Maybe

"The Maybe tag marks utterances in which a speaker's utterance conveys probability or possibility by using the word "maybe" or other words denoting possibility and probability" (Dhillon et al., 2004, p. 66). Maybes should not be confused with Offers, i.e. suggestions in the form of "maybe we should..." (Dhillon et al., 2004). Based on the data, common examples of Maybes are or include "maybe", "I guess", and "probably" (Dhillon et al., 2004). Here's couple of examples of Maybesin context (adapted from Dhillon et al., 2004, p. 67-68):

#### EXCHANGE 1

A: we- - what set the - they set the context to unknown? (Wh-question(gen) + Rising Tone(spec))

B: right now we haven't observed it. (Statement(gen))

B: so I guess it's sort of averaging over all those three possibilities. (Statement(gen) + Maybe(spec))

#### EXCHANGE 2

A: is Srini going to be at the meeting tomorrow?(Yes-No-Question(gen) + Rising Tone(spec))

A: do you know?(Yes-No-Question(gen) + Rising Tone(spec))

B: maybe. (Statement(gen) + Maybe(spec))

#### EXCHANGE 3

A: so - so what accent are we speaking? (Wh-question(gen))

B: probably western yeah. (Statement(gen) + Maybe(spec))

### 11.7.3.2 No Knowledge

"The no knowledge tag (No Knowledge) marks utterances in which a speaker expresses a lack of knowledge regarding some subject" (Dhillon et al., 2004, p. 68). The most common No Knowledges are or include "I don't know" - when it is not a Floor Holder - and "I'm not sure" (Dhillon et al., 2004). Here's an examples of No Knowledge in context (adapted from Dhillon et al., 2004, p. 68):

A: but if you really want to find out what it's about you have to click on the little light bulb. (Statement(gen))

B: although i've - i've never - i don't know what the light bulb is for. (Statement(gen) + (No Knowledge(spec))

## 11.8 MRDA Group 6: Action Motivators

The next group of tags MRDA defines is that of Action Motivators. The group of Action Motivators contains specific tags pertaining to future action, regardless of whether such action occurs immediately or in the distant future (Dhillon et al., 2004). "The tags in Group 6 either indicate that a command or a suggestion has been made regarding some action to be taken at some point in the future or else indicate that a speaker has committed himself to executing some action at some point in the future" (Dhillon et al., 2004, p. 70).

### 11.8.1 Command

"The Command tag marks commands. In terms of syntax, a command may arise in the form of a question (e.g., "Do you want to go ahead?") or as a statement (e.g., "Give me the

microphone.)" (Dhillon et al., 2004, p. 70). The most common indicator is the imperative mood. Commands differ from Suggestions in terms of two things:

1) the response they receive: unlike Suggestions, Commands are uttered as orders and a failure to comply is perceived as impolite or as a sign of indignation ("considering whether the utterance could receive a response that is a rejection and whether that rejection is considered impolite is a helpful method to determine if the utterance is a command or a suggestion. If a rejection is considered impolite, the utterance is considered a command, otherwise it is considered a suggestion" (Dhillon et al., 2004, p. 70);

2) the role of the speaker: "generally suggestions made by the speaker running a meeting (or by any speaker in a position of power for that matter) are perceived as commands (...) (w)hereas, if the same utterance is made by another speaker who is not running the meeting, then the utterance is considered a suggestion instead" (Dhillon et al., 2004, p. 70). This, however, does not mean that all suggestions made by a speaker in a position of power are actually commands. When determining whether an utterance is a Suggestion or a Command, we always need to take into account point 1, i.e. we need to assess how a potential rejection to that Suggestion or Command would be perceived (Dhillon et al., 2004).

Although Suggestions and Commands cannot be easily distinguished in the absence of contextual information, such as knowledge about the role of the speaker and of its interlocutors, we nevertheless propose a few examples of Commands taken out of context to give an idea of what they look like in their textual form (from Dhillon et al., 2004, p. 71-72):

Continue. (Statement(gen) + Command(spec))

Proceed. (Statement(gen) + Command(spec))

Wait. (Statement(gen) + Command(spec))

let's get this uh - b- - clearer. (Statement(gen) + Command(spec))

explain to me why it's necessary to distinguish between whether something has a door and is not public. (Statement(gen) + Command(spec))

close it and - and load up the old state so it doesn't screw - screw that up. (Statement(gen) + Command(spec))

so | we should think about trying to wrap up here. (Floor Holder(gen) | Statement(gen) + Command(spec))

yeah so maybe just cc hari and say that you've just been asked to handle the large vocabulary part here. (Statement(gen) + Command(spec))

### 11.8.2 Suggestion

Simply put, "(t)he suggestion tag marks proposals, offers, advice, and, most obviously, suggestions" (Dhillon et al., 2004, p. 73). Here's a few remarks made by Dhillon et al. (2004, p. 73) regarding Suggestions: "(s)uggestions are often found in constructions such as "maybe we should...". Suggestions containing the word "maybe" are not to be confused with the maybe tag (Maybe). Additionally, if the phrase "excuse me" precedes something for which a speaker is negotiating permission (Jurafsky 35), then it is marked as a suggestion rather than an apology (Apology)". Finally, as we have mentioned above, Suggestions are not to be confused with Commands. Here's couple of examples of Suggestions in context (adapted from Dhillon et al., 2004, p. 73-74):

yeah | i was just going to say maybe it has something to do with hardware. (Floor Grabber(gen) | Statement(gen) + Suggestion(spec))  
should we take turns? (Yes-No-question + Suggestion(spec1) + Rising Tone(spec2))  
let's see maybe we should just get a list of items. (Statement(gen) + Suggestion(spec))  
i think these things are a lot clearer when you can use fonts - different fonts there. (Statement(gen) + Suggestion(spec))

### 11.8.3 Commitment

"The commitment tag (Commitment) is used to mark utterances in which a speaker explicitly commits himself to some future course of action. Commitments are not to be confused with suggestions in which a speaker suggests that he, the speaker himself, execute some action. With commitments, a speaker mentions what he will do in the future, not what he might do" (Dhillon et al., 2004, p. 74). Here's couple of examples of Commitments in context (adapted from Dhillon et al., 2004, p. 75):

I'll make that available.(Statement(gen) + Commitment(spec))  
my intention is to do a script that'll do everything.(Statement(gen) + Commitment(spec))  
i'll send it out to the list telling people to look at it. (Statement(gen) + Commitment(spec))  
i'll try to get to that. (Statement(gen) + Commitment(spec))  
i'm just going to do it. (Statement(gen) + Commitment(spec))

## 11.9 MRDA Group 7: Checks

Put simple, "(t)his group contains specific tags pertaining to understanding or being understood" (Dhillon et al., 2004, p. 76).

### 11.9.1 "Follow Me"

"The "Follow Me" tag marks utterances made by a speaker who wants to verify that what he is saying is being understood. Utterances marked with the "Follow Me" tag explicitly communicate or else implicitly communicate the questions "do you follow me?" or "do you understand?". In implicitly communicating those questions, a speaker's utterance may be a tag question (Tag-Question), such as "right?" or "okay?", where a sense of "do you understand?" is being conveyed" (Dhillon et al., 2004, p. 76). Here's a couple of examples of "Follow Me"s (adapted from Dhillon et al., 2004, p. 76):

this is understandable? (Yes-No-Question(gen) + Declarative-Question(spec1) + "Follow Me"(spec2) + Rising Tone(spec3))

do you know what i'm saying?(Yes-No-Question(gen) + "Follow Me"(spec1) + Rising Tone(spec2))

you know what i mean?(Yes-No-Question(gen) + Declarative-Question(spec1) + "Follow Me"(spec2) + Rising Tone(spec3))

### 11.9.2 Repetition Request / Signal-non-understanding

"An utterance marked as a repetition request indicates that a speaker wishes for another speaker to repeat all or part of his previous utterance" (Dhillon et al., 2004, p. 77). Some examples of Repetition Requests are: "what?", "sorry?", "huh?", "pardon?", "excuse me?", "say that again", "what did you say?", and "what was that again?" (Dhillon et al., 2004, p. 77). Here's a couple of examples of Repetition Requests in context (adapted from Dhillon et al., 2004, p. 77-78):

#### EXCHANGE 1

A: um | how long would it take to - to add another node on the observatory and um - play around with it? (Floor Holder(gen) | Wh-Question(gen) + Rising Tone(spec))

B: another node on what? (Wh-Question(gen) + Repetition Request(spec1) + Rising Tone(spec2))

#### EXCHANGE 2

A: so who would be the subject of this trial run? (Wh-Question(gen))

B: pardon me? (Wh-Question(gen) + Repetition Request(spec))

### 11.9.3 Understanding Check

"The understanding check tag marks when a speaker checks to see if he understands what a previous speaker said or else to see if he understands some sort of information. With understanding checks, a speaker usually states what he is trying to verify as correct and follows that with a tag question (Tag-Question). Only the utterance, or portion of the utterance if a pipe bar is used,

containing the information to be verified is marked with the Understanding Check tag. Tag questions (Tag-Questions) are not marked with the Understanding Check tag as they do not contain the information that is to be verified." (Dhillon et al., 2004, p. 78). Here's an example of Repetition Request in context (adapted from Dhillon et al., 2004, p. 79):

A: the reading task is a lot shorter. (Statement(gen))

B: and other than that yeah i guess we'll just have to uh - listen. (Statement(gen))

B: although i guess it's only ten minutes each. (Statement(gen) + Understanding Check(spec))

B: right? (Yes-No-question(gen) + Declarative-Question(spec1) + Tag-Question(spec2) + Rising Tone(spec3))

### **11.10 MRDA Group 8: Restated Information**

"This group, as the name states, contains specific tags pertaining to information that has been restated. The group is further divided into two subgroups: repetition and correction" (Dhillon et al., 2004, p. 80).

#### **11.10.1 Repetition**

##### **11.10.1.1 Repeat**

"The repeat tag (Repeat) is used when a speaker repeats himself. This often occurs in response to repetition requests (Repetition Requests) or else to place emphasis on a certain point. In repeating himself, a speaker repeats all or part of one of his previous utterances. However, in order for an utterance to be considered a repeat, it must be a repeat of an utterance made at most a few seconds prior to the repeat. (...) It is not required that a speaker repeat himself verbatim in order for a utterance to be marked with the repeat tag (Repeat)" (Dhillon et al., 2004, p. 80). Dhillon et al. (2004, p. 80) continue: "(r)eppeats (Repeats) are not to be confused with mimics (Mimic other). As previously stated, a repeat occurs when a speaker repeats his own utterance. A mimic occurs when a speaker repeats another speaker's utterance. Repeats are also not to be confused with summaries (Reformulate/summarizes) where a speaker summarizes his own utterances as many structural differences occur between the summary and the information being summarized". Here's a couple of examples of Repeats in context (adapted from Dhillon et al., 2004, p. 81):

EXCHANGE 1

A: and everything is fixed. (Statement(gen))

A: everything is fixed. (Statement(gen) + Repeat(spec))

## EXCHANGE 2

A: and there didn't seem to be any uh penalty for that? (Yes-No-Question(gen) + Understanding Check(spec1) +Declarative-Question(spec2) + Rising Tone(spec3))

B: pardon? (Yes-No-Question(gen) + Repetition Request(spec1) + Rising Tone(spec2))

A: there didn't seem to be any penalty for making it casual?(Yes-No-Question(gen) + Understanding Check(spec1) +Declarative-Question(spec2) + Repeat(spec3) + Rising Tone(spec4))

### 11.10.1.2 Mimic

Simply put, "[t]he mimic tag marks when a speaker mimics another speaker's utterance, or portion of another speaker's utterance" (Dhillon et al., 2004, p. 81). "Mimics (Mimics) are not to be confused with repeats (Repeats). As previously stated, a mimic occurs when a speaker repeats another speaker's utterance. A repeat occurs when a speaker repeats his own utterance. Also, mimics are not to be confused with summaries (Reformulate/summarize) where a speaker summarizes another speaker's utterances as many structural differences occur between the summary and the information being summarized" (Dhillon et al., 2004, p. 82). Just like Repeats, Mimics do not need repeat another utterance verbatim in order to be considered as such, and they can also contain speech that is additional to what is mimicked (Dhillon et al., 2004). Oftentimes, utterances that are labeled as Mimics are also Acknowledge-answers; for example, the speaker who does not have the floor acknowledges the speaker who has the floor by mimicking part of what he or she says (Dhillon et al., 2004); for example (Dhillon et al., 2004, p. 82):

A: go up one. (Statement(gen) + Action-directive(spec1) + Rising Tone(spec2))

B: up one. (Statement(gen) + Acknowledge-answer(spec1) + Mimic(spec2))

Other times, a speaker will "phrase the mimic in the form of a declarative question" (Dhillon et al., 2004, p. 82); for example (Dhillon et al., 2004, p. 83):

A: well you have a like techno speaker accent i think. (Statement(gen))

B: a techno speak accent? (Yes-No-question(gen) + Understanding Check(spec1) + Declarative-Question(spec2) + Mimic(spec3) + Rising Tone(spec4))

### 11.10.1.3 Summary

"The Summary (or Reformulate/summarize) tag marks when a speaker summarizes a previous utterance or discussion, regardless of whose speech he is summarizing" (Dhillon et al., 2004, p. 83). While Understanding Checks restate information for validation, Summaries do not require validation (Understanding Checks and Summaries are mutually exclusive) (Dhillon et al.,

2004). Here's a couple of examples of Summaries in context (adapted from Dhillon et al., 2004, p. 83):

A: so i mean add moderate amount of noise to all data. (Statement(gen))

A: so that makes any additive noise less effective. (Statement(gen))

B: so you're making all your training data more uniform. (Statement(gen) + Summary(spec))

## 11.10.2 Correction

### 11.10.2.1 Correct Misspeaking

"The Correct Misspeaking tag is used when a speaker corrects another speaker's utterance. Corrections are based upon whether the word choice of a speaker is corrected or the pronunciation of a word is corrected" (Dhillon et al., 2004, p. 85). Here's an example of Summary in context (adapted from Dhillon et al., 2004, p. 85):

A: oh no | i've ninety four. (Statement(gen) + No answer(spec) | Statement(gen) + Rising Tone(spec))

B: ninety three point six four. (Statement(gen) + Correct Misspeaking(spec))

### 11.10.2.2 Self-Correct Misspeaking

"The Self-Correct Misspeaking tag marks when a speaker corrects his own error, with regard to either pronunciation or word choice" (Dhillon et al., 2004, p. 85). Here's a couple of examples of Self-Correct Misspeakings in context (from Dhillon et al., 2004, p. 86):

#### EXAMPLE 1

A: okay | so - yeah so note the four nodes down there the - sort of the things that are not directly extracted. (Statement(gen) + Acknowledge-answer(spec) | Statement(gen))

A: actually the five things. (Statement(gen) + Self-Correct Misspeaking(spec))

#### EXAMPLE 2

A: um and uh | they don't look very separate. (Floor Holder(gen) | Statement(gen))

A: uh | separated. (Floor Holder(gen) | Statement(gen) + Self-Correct Misspeaking(spec))

## 11.11 MRDA Group 9: Supportive Functions

Put simply, "[t]his group contains tags that apply to utterances in which a speaker supports his own argument by defending himself, offering an explanation, or else offering additional details and utterances in which a speaker attempts to support another speaker by finishing the other speaker's utterance" (Dhillon et al., 2004, p. 87).

### 11.11.1 Defending/Explanation

“The Defending/Explanation tag marks cases in which a speaker defends his own point or offers an explanation. Often, the word "because" signals an explanation” (Dhillon et al., 2004, p. 87). The Defending/Explanation tag is not to be confused with the Elaboration tag, which is instead used to mark “utterances in which a speaker offers further details” (Dhillon et al., 2004, p. 87). That is to say: while Defending/Explanations revolve around reasons, Elaborations revolve around details (Dhillon et al., 2004). Here’s an example of Defending/Explanations in context (adapted from Dhillon et al., 2004, p. 87):

A: no no it isn't sensitive at all. (Statement(gen) + Reject(spec))

A: i was just - i was jus- - i was overreacting just because we've been talking about it.  
(Statement(gen) + Defending/Explanation(spec))

### 11.11.2 Elaboration

“The elaboration tag marks when a current speaker elaborates on a previous utterance of his by adding further details as opposed to simply continuing to speak on the same topic. When a speaker describes something using an example, the example is regarded as an elaboration” (Dhillon et al., 2004, p. 88). Here’s an example of Elaboration in context (adapted from Dhillon et al., 2004, p. 89):

A: and basically the net- - network is trained almost to give binary decisions.  
(Statement(gen) )

A: and uh - binary decisions about phonemes.(Statement(gen) + Elaboration(spec))

### 11.11.3 Collaborative Completion

“The collaborative completion tag (Collaborative Completion) tag marks utterances in which a speaker attempts to complete a portion of another speaker's utterance. Whether the speaker whose utterance is completed by another speaker agrees with the content of the completion is inconsequential” (Dhillon et al., 2004, p. 90). Here’s an example of Collaborative Completion in context (adapted from Dhillon et al., 2004, p. 91):

A: but there's a significant amount of == (Statement(gen))

B: non zero? (Yes-No-question(gen) + Declarative-Question(spec1) + Rising Tone(spec2) + Collaborative completion(spec3))

## 11.12 MRDA Group 10: Politeness Mechanisms

Generally speaking, “[t]his group contains tags that apply to utterances in which speakers exhibit courteousness” (Dhillon et al., 2004, p. 92).

### 11.12.1 Downplayer

“The downplayer tag (Downplayer) marks cases in which a speaker downplays or deemphasizes another utterance. The utterance that is downplayed may be uttered by the same speaker or a different speaker” (Dhillon et al., 2004, p. 92). As a rule of thumb, “[a]pologies, compliments, and other courteous utterances are often downplayed. In other cases, a speaker makes a strong assertion and then downplays it” (Dhillon et al., 2004, p. 92). “The following is a list of common short downplayers: "that's okay," "that's all right," "it's okay," "I'm kidding," "it's just a thought," and "never mind”.” (Dhillon et al., 2004, p. 92). Here’s an example of Downplayer in context (adapted from Dhillon et al., 2004, p. 93):

A: congratulations. (Statement(gen) + Assessment/appreciation(spec))

B: well it was i mean - i really didn't do this myself. (Statement(gen) + Downplayer(spec))

### 11.12.2 Sympathy

“The Sympathy tag marks utterances in which a speaker exhibits sympathy. Oftentimes, the phrase "I'm sorry" is used sympathetically. However, that very phrase also has the potential to be marked as a repetition request (Signal-non-understanding) or as an apology (Apology), depending upon its function” (Dhillon et al., 2004, p. 94). Here’s an example of Exclamation in context (adapted from Dhillon et al., 2004, p. 94):

A: thinking about it when i offered up my hard drive last week == (Statement(gen))

B: oh no! (Statement(gen) + Sympathy(spec1) + Exclamation(spec2))

### 11.12.3 Apology

“An utterance is marked as an apology <fa> when a speaker apologizes for something he did (e.g., after coughing, sneezing, interrupting another speaker, etc.)” (Dhillon et al., 2004, p. 94). Here’s an example of Apology in context (adapted from Dhillon et al., 2004, p. 95):

A: because the date is when you actually read the digits and the time and == (Statement(gen))

A: excuse me. (Statement(gen) + Apology(spec))

A: the time is when you actually read the digits but i'm filling out the date beforehand. (Statement(gen) + Misspeak Self-Correction(spec))

#### **11.12.4 Thanks**

The Thanks tag marks utterances in which a speaker thanks another speaker. Here's an example of Thanks in context (adapted from Dhillon et al., 2004, p. 96):

A: nice coinage. (Statement(gen) + Assessment/appreciation(spec))

B: thank you. (Statement(gen) + Thanks(spec))

#### **11.12.5 Welcome**

"The Welcome tag marks utterances which function as responses to utterances marked with the thanks tag (Thanks). Phrases such as "you're welcome" and "my pleasure" are marked with the welcome tag (Welcome). No instances of the Welcome tag exist within the Meeting Recorder data" (Dhillon et al., 2004, p. 96).

### **11.13 Group 11: Further Descriptions**

Generally speaking, "[t]his group contains various tags that do not fit into any of the pre-established groups. The tags within this group characterize meeting agendas, changes in topic, exclamatory material, humorous matter, self talk, third party talk, as well as syntactic and prosodic features of utterances" (Dhillon et al., 2004, p. 96).

#### **11.13.1 Exclamation**

"The exclamation  
tag marks utterances in which a speaker expresses excitement, surprise, or enthusiasm" (Dhillon et al., 2004, p. 97). Exclamations vary in length and are characterized by a high level of energy; they are punctuated with an exclamation mark within MRDA (Dhillon et al., 2004). Here's a few examples of Exclamations (Dhillon et al., 2004, p. 97): "wow!", "aha!", "whew", "oops!", "god!", "oh!", "ha!", "oh yeah!", "oh no!", "i can read!", "twlve minutes!", "oh it's seventy five per cent!", "damn this project!", "then do some more spectral subtraction!", "so that's amazing you showed up at this meeting!" (Dhillon et al., 2004, p. 97-98).

#### **11.13.2 About-Task**

"The about-task tag marks utterances that are in reference to meeting agendas or else address the direction of meeting conversations with regard to meeting agendas" (Dhillon et al., 2004, p. 98). While Topic Changes either end or begin a topic regardless of a meeting agenda, About-Tasks regard "previously established items to be discussed or managed within a meeting" (Dhillon et al.,

2004, p. 98). If an utterance is changing a topic in reference to a meeting agenda, then it should be tagged both as a Topic Change and as an About-task (Dhillon et al., 2004). "In essence, the about-task tag marks utterances which revolve around what tasks are to be completed within the course of a meeting (...) For instance, if a speaker mentions that an agenda item is to discuss a certain subject and then other speakers begin to discuss that subject, then the utterance mentioning that the agenda item to discuss a subject is marked with the about-task tag. However, the actual discussion about the subject is not marked with the about-task tag" (Dhillon et al., 2004, p. 99). Here's a few examples of About-tasks in context (Dhillon et al., 2004, p. 99-100):

i want to talk about new microphones and wireless stuff. (Statement(gen) + About-task(spec))

let's discuss agenda items. (Statement(gen) + Action-directive(spec1) + Rising Tone(spec2) + About-task(spec3))

so yeah why don't we do the speech nonpeech discussion? (Rhetorical-Question(gen) + About-task(spec1) + Topic Change(spec2))

#### EXCHANGE

A: any agenda items today? (Open-Question(gen) + About-task(spec))

B: i want to talk a little bit about getting - how we're going to to get people to edit bleeps parts of the meeting that they don't want to include. (Statement(gen) + About-task(spec))

#### 11.13.3 Topic Change

"The Topic Change tag marks utterances which either begin or end a topic. As the Topic Change tag marks when a topic changes, once the topic has indeed changed and a new topic is in the course of discussion, the discussion of the new topic is not marked with the Topic Change tag" (Dhillon et al., 2004, p. 100). When a new topic is introduced by means of a floor grabber, that utterance must be tagged as Floor Grabber and not as Topic Change (Dhillon et al., 2004, p. 100). Here's a few examples of Topic Changes in context (Dhillon et al., 2004, p. 101):

A: let's see. (Floor Grabber(gen))

A: um | why don't - why don't we uh - if there aren't any other major things why don't we do the digits and then - then uh - turn the mikes off. (Floor Holder(gen) | Statement(gen) + Offer(spec1) + About-task(spec2) + Topic Change(spec3))

#### 11.13.4 Joke

“The Joke tag marks utterances of humorous or sarcastic nature” (Dhillon et al., 2004, p. 102). Dhillon et al. (2004) focus on the fact that an utterance is to be marked as a Joke every time the speaker is attempting to be funny, regardless of how he or she is perceived by the addressee, i.e. regardless of whether the addressee understands the humorous or sarcastic nature of the utterance (Dhillon et al., 2004). Finally, we need to mention the fact that the majority of jokes is context dependent, thus the context has to be investigated before marking utterances as jokes (Dhillon et al., 2004, p. 102). Here's a few examples of Jokes in context (Dhillon et al., 2004, p. 102):

A: is he going to come here? (Yes-No-question(gen) + Rising Tone(spec))

B: oh == (Hold(gen))

B: well we'll drag him here. (Statement(gen) + Joke(spec1) + Affirmative non-yes answers(spec2))

B: I know where he is. (Statement(gen) + Joke(spec))

### 11.13.5 Self Talk

“The Self Talk tag is used when a speaker talks to himself. Often, utterances marked as self talk are quieter and softer than the surrounding speech” (Dhillon et al., 2004, p. 103). Self talk is usually used by the speaker as he or she is writing something down or as he or she figuring out the answer of a calculation or problem (Dhillon et al., 2004). Backchannels and Floor Holders, in spite of not being forms of direct communication, are not considered self talk (Dhillon et al., 2004). Here's a few examples of Self Talks in context (Dhillon et al., 2004, p. 104):

A: i - i - ith- - i think he == (Statement(gen))

A: what am i saying here? (Open-Question(gen) + Self Talk(spec))

### 11.13.6 Third Party Talk

“The third party talk tag marks utterances of side conversations. Side conversations are conversations which are not directed toward the main conversation and may only consist of a handful of utterances or may be quite lengthy” (Dhillon et al., 2004, p. 104). Here's an example of third Party Talk in context (Dhillon et al., 2004, p. 105-106):

A: and we get a certain - we have a situation vector and a user vector and everything is fine. (Statement(gen) + Rising Tone(spec))

A: an- - an- - and - and our - and our == (Interruption)

B: did you just sti- - did you just stick the m- - the - the - the microphone actually in the tea? (Yes-No-question(gen) + Rising Tone(spec1) + Third Party Talk(spec2))

C: no. (Statement(gen) + No answer(spec1) + Third Party Talk(spec2))

A: and um == (Floor Holder(gen))

C: i'm not drinking tea. (Statement(gen) + Negative non-no answer(spec1) + Third Party Talk(spec2))

C: what are you talking about? (Wh-question(gen) + Third Party Talk(spec))

B: oh yeah. (Statement(gen) + Acknowledge-answer(spec1) + Third Party Talk(spec2))

B: sorry. (Statement(gen) + Apology(spec1) + Third Party Talk(spec2))

A: let's just assume our bayes net just has three decision nodes for the time being. (Statement(gen) + Action-directive(spec1) + Rising Tone(spec2))

### 11.13.7 Declarative Question

"The declarative question tag marks questions which have the syntactic appearance of a statement. In declarative questions, the subject precedes the verb and subject-auxiliary inversion and wh-movement do not occur. It is not uncommon for a rising tone (Rising Tone) to be found on a declarative question, however a rising tone does not always function as an indicator that a question is being asked" (Dhillon et al., 2004, p. 105). Tag-Questions are often Declarative questions, in which case they often consist of a subject plus a verb (e.g. "you do?"), a single word (e.g. "right?"), or a noun phrase (e.g. "the tenth of July?") (Dhillon et al., 2004, p. 106). If a question consists of a single word and that word is a "wh" word, then that question is a Wh-question and not a declarative Tag-question (Dhillon et al., 2004). Here's a few examples of Declarative Questions (Dhillon et al., 2004, p. 106-107):

right? (Yes-No-question(gen) + Declarative Question(spec1) + Tag-Question(spec2) + Rising Tone(spec3))

you know? (Yes-No-question(gen) + Declarative Question(spec1) + "Follow me"(spec2) + Tag-Question(spec3))

um | and anything else anyone wants to talk about? (Floor Holder(gen) | Open-Question(gen) + Declarative Question(spec1) + Rising Tone(spec2))

or you'd like - so you're saying you could practically turn this structure inside out? (Yes-No-question(gen) + Understanding Check(spec1) + Declarative Question(spec2) + Rising Tone(spec3))

### 11.13.8 Tag Question

"A tag question follows a statement and is a short question seeking confirmation of that statement. Tag questions receive a general tag of Yes-No-questions and are often used in conjunction with the "follow me" tag and the declarative question tag (Declarative Question). (...)

Tag questions are often found following statements marked with the understanding check tag (Understanding Check)" (Dhillon et al., 2004, p. 108). Common Tag Questions are the following: "right?", "yes?", "yeah?", "no?", "okay?", "isn't it?", "correct?", "won't it?", "doesn't it?", and "you know?" (Dhillon et al., 2004, p. 108). Here's a few examples of Tag Questions in context (Dhillon et al., 2004, p. 108):

**EXAMPLE 1**

A: exchange money is an errand. (Statement(gen) + Understanding Check(spec))

A: right? (Yes-No-question(gen) + Declarative Question(spec1) + Tag-Question(spec2))

**EXAMPLE 2**

A: and this - this one is right at the end of the table. (Statement(gen))

A: okay? (Yes-No-question(gen) + Declarative Question(spec1) + "Follow Me"(spec2) + Tag-Question(spec3))

**EXAMPLE 3**

A: yeah | so we don't store any of our audio formats compressed in any way. (Floor Grabber(gen) | Statement(gen) + Understanding Check(spec))

A: do we? (Yes-No-question(gen) + Declarative Question(spec1) + Tag-Question(spec2))

**EXAMPLE 4**

A: I mean - | the normalization you do is over the whole conversation. (Floor Holder(gen) | Statement(gen))

A: isn't it? (Yes-No-question(gen) + Tag-Question(spec1) + Rising Tone(spec2))

### **11.13.9 Rising Tone**

"The rising tone tag is used to mark utterances in which a speaker's tone rises at the end of his utterance. Rising tones at the end of utterances occur in both questions and statements. Although intonation does not constitute a dialog act, the use of the Rising Tone tag provides useful information for automatic speech recognition" (Dhillon et al., 2004, p. 110).

## **11.14 Group 12: Disruption Forms**

"[D]isruption forms are used to mark utterances that are indecipherable, abandoned, or interrupted. Only one disruption form may be used per utterance" (Dhillon et al., 2004, p. 110).

### **11.14.1 Indecipherable<%>**

"The indecipherable tag marks indecipherable speech such as mumbled or muffled words or utterances that are too difficult to hear on account of the microphone picking up sounds from

breathing. The indecipherable tag <%> is not to be confused with the nonspeech tag <x>. The nonspeech tag <x> is used for sound segments which are silent or otherwise contain non-vocal sounds such as doors slamming and phones ringing. The nonspeech tag <x> does not apply to sounds such as breathing and sighs, as these are vocal sounds. However, sounds such as coughing and sneezing may be considered vocal sounds but are instead categorized with the nonspeech variety” (Dhillon et al., 2004, p. 110).

#### **11.14.2 Interrupted <%->**

“The interrupted tag marks incomplete utterances in which a speaker stops talking on account of being interrupted by another speaker. This tag is not to be confused with the abandoned tag <%--> which is used to mark instances in which a speaker intentionally abandons an utterance” (Dhillon et al., 2004, p. 111).

#### **11.14.3 Abandoned <%-->**

“The abandoned tag marks utterances which are abandoned by a speaker. Abandoned utterances occur when a speaker trails off or else chooses to either reformulate an utterance or change the topic by abandoning his current utterance and beginning a new one” (Dhillon et al., 2004, p. 111).

#### **11.14.4 Nonspeech<x>**

“The nonspeech tag marks any utterance that is unintelligible on account of non-vocal noises such as doors slamming, phones ringing, and problems with a recording. The nonspeech tag also marks coughing and sneezing sounds, as well as utterances filled with silence” (Dhillon et al., 2004, p. 113).

### **11.15 Group 13: Nonlabeled**

“Group 13 solely contains the nonlabeled tag <z>. As stated in Section 3.2, the tag <z> does not provide any information regarding the characteristics and functions of utterances as the tags of the other groups do, and for this reason it is separated from those groups” (Dhillon et al., 2004, p. 113).

#### **11.15.1 Nonlabeled**

“The nonlabeled tag marks utterances that are not to be labeled with a DA (...) The tag <z> marks utterances which otherwise would be labeled with DAs but instead are intentionally not to be labeled” (Dhillon et al., 2004, p. 113).

## 12 MRDA: Adjacency Pairs

Adjacency pairs are paired utterances, produced by different speakers, that reflect the structure of conversation; some examples are: question-answer, greeting-greeting, offer-acceptance, and apology-downplay (Dhillon et al., 2004, p. 25; Levinson, 1983). According to Dhillon et al. (2004), “[l]abeling adjacency pairs (AP) in meetings provides a means to extract the information provided by the interaction between speakers” (p. 25). Such contextual information, rendered by pragmatics, is necessary to be able to distinguish, for example, a statement from an answer (from Jurafsky et al., 1997):

A: Do you have kids? (Yes-No-question(gen))

B: I have three kids. (Statement(gen) + Narrative-Affirmative answer(spec))

Utterance B is specifically a Narrative-Affirmative answer - and not simply a general Statement - by virtue of being paired with utterance A; together, in fact, they form the question-answer adjacency pair.

As Dhillon et al. (2004) point out, “[a]djacency pairs denote direct interaction between speakers” (p. 30), therefore all the sentences that “are not uttered directly to a speaker as a response and do not function in a way that elicits a response” (p. 30) cannot be labeled with an adjacency pair. For example, adjacency pairs are never marked in the case of Backchannels, Rhetorical question backchannels (when uttered as acknowledgments), Floor Holders, and Floor Grabbers, but they are always marked in the case of Rhetorical question backchannels (when uttered as backchannels), Holds, Mimics, and Collaborative completions (Dhillon et al., 2004). Finally, we need to mention the fact that, in some cases, the adjacency pair is not clear enough to be marked; one of such scenarios is when two or more speakers each utters a Statement and another speaker utters an Acknowledgment: we do not know to which Statement the Acknowledgment refers and thus the adjacency pair cannot be marked (Dhillon et al., 2004).

## 13 Comparison Between SWBD-DAMSL and MRDA

### 13.1 Unused and Merged SWBD-DAMSL Tags

"[C]ertain SWBD-DAMSL tags are not found in the MRDA tagset. Of these tags, some have been merged with other tags and others are not included in the MRDA tagset entirely. Below is a list of these tags. Each SWBDDAMSL tag listed below is followed by a brief description indicating whether it has been merged or why it is not included in the MRDA tagset" (Dhillon et al., 2004, p. 120).

### **13.1.1 About-communication<c>**

"Utterances such as "pardon me?" and "I can't hear you" that are marked with About-communication in the SWBD-DAMSL tagset are considered Repetition Requests / Signal-non-understandings in the MRDA tagset" (Dhillon et al., 2004, p. 120).

### **13.1.2 Statement-non-opinion <sd> and Statement-opinion <sv>**

"The Statement-non-opinion and Statement-opinion tags were quite difficult to use with the MRDA data, as their use resulted in a lack of agreement among annotators. They were eventually eliminated from the MRDA tagset and replaced with the Statement tag, which marks statements in general, without having to distinguish between "non-opinion" and "opinion." (For overt opinions, the Assessment/Appreciation tag is used) (Dhillon et al., 2004, p. 120).

### **13.1.3 Open-option**

"This tag is no longer included in the MRDA tagset due to its redundancy with suggestions (Offers)" (Dhillon et al., 2004, p. 120).

### **13.1.4 Conventional-opening**

"This tag is not included in MRDA tagset due to lack of use. Utterances that would be marked with this tag usually occur in pre-meeting chatter, which is marked with the Nonlabeled tag" (Dhillon et al., 2004, p. 120).

### **13.1.5 Conventional-closing**

"This tag is not included in MRDA tagset due to lack of use. Utterances that would be marked with this tag usually occur in post-meeting chatter, which is marked with the Nonlabeled tag" (Dhillon et al., 2004, p. 121).

### **13.1.6 Explicit-performative**

"This tag is no longer included in the MRDA tagset due to its lack of use"(Dhillon et al., 2004, p. 121).

### **13.1.7 Other-forward-function**

"This tag is not included in MRDA tagset due to lack of use" (Dhillon et al., 2004, p. 121).

### **13.1.8 Yes Answers**

"This tag has been merged with the SWBD-DAMSL tag Accept to form the MRDA tag Accept / Yes Answer" (Dhillon et al., 2004, p. 121).

### **13.1.9 No Answers**

"This tag has been merged with the SWBD-DAMSL tag Reject to form the MRDA tag Reject / No Answer" (Dhillon et al., 2004, p. 121).

### **13.1.10 Quoted Material**

"Due to the various DA tags quoted material within the MRDA data had the potential to receive, the use of the SWBD-DAMSL tag <q> was replaced with a convention that actually used DAs to characterize the quoted material. In doing so, more information regarding the character and function of quoted material is gained than through using a tag such as <q> to merely indicate that quoted material is present" (Dhillon et al., 2004, p. 121).

### **13.1.11 Hedge**

"This tag is not included in the MRDA tagset due to lack of use and ambiguity as to what sort of utterance would be labeled as a hedge as opposed to another label" (Dhillon et al., 2004, p. 121).

### **13.1.12 Continued from Previous Line <+>**

"This tag is not included in the MRDA tagset because utterances continued from a previous line by the same speaker are given a new DA to depict the function of the continuation" (Dhillon et al., 2004, p. 121).

## **13.2 Unique MRDA Tags**

"Due to the nature of the MRDA data, the SWBD-DAMSL tagset proved to be inefficient in accurately characterizing all facets of the MRDA data. Consequently, tags were created to account

for areas where the SWBD-DAMSL tagset was insufficient. Below is a list of the tags that were created specifically for the MRDA data. Each tag listed below is followed by a brief description indicating why it entered the MRDA tagset" (Dhillon et al., 2004, p. 123).

### **13.2.1 Interrupted <%->**

"Throughout the meetings, incomplete utterances arose on account of speakers abandoning their utterances or being interrupted. To characterize why an incomplete utterance arose, the interrupted tag was added (as the abandoned tag <%--> was already present)" (Dhillon et al., 2004, p. 123).

### **13.2.2 Topic Change**

"Within the MRDA data, many instances arose in which speakers attempted to change the topic. No other mechanism was present to mark such occurrences, so the <tc> tag entered the MRDA tagset to mark changes in topic" (Dhillon et al., 2004, p. 123).

### **13.2.3 Floor Holder**

"The SWBD-DAMSL tagset contained the tag <h> (hold), which was also incorporated into the MRDA tagset. Utterances similar to those marked with <h> appeared midspeech within the MRDA data. The <fh> tag was implemented to distinguish between a hold, which marks utterances in which a speaker "holds off" prior to answering a question or prior to speaking when he is expected to speak, and these mid-speech "holds"" (Dhillon et al., 2004, p. 123).

### **13.2.4 Floor Grabber**

"This tag entered the tagset as there were significant similarities among the means by which speakers "gained" the floor and also due to the lack of a tack to mark such instances. Speakers' utterances often contained specific lexical items and higherenergy during these attempts to "gain" the floor. The <fg> tag entered the MRDA tagset as a means to mark such utterances" (Dhillon et al., 2004, p. 123-4).

### **13.2.5 Repeat**

"This tag entered the MRDA tagset in order to mark possible subtle changes in the manner in which a speaker repeats an utterance, whether for purposes of emphasis or in response to a repetition request" (Dhillon et al., 2004, p. 124).

### **13.2.6 Self-Correct Misspeaking**

"This tag was added to differentiate cases in which the primary speaker alone corrected his speech rather than being corrected by another speaker, which is indicated by the Correct-misspeaking tag" (Dhillon et al., 2004, p. 124).

### **13.2.7 Understanding Check**

"This tag entered the MRDA tagset as there seemed to be a large number of distinct cases in which a speaker wanted to check if his information was correct" (Dhillon et al., 2004, p. 124).

### **13.2.8 Defending/Explanation**

"This tag was added as speakers tended to defend their suggestions either immediately prior to making a suggestion or immediately after. Its usage was later expanded to include when speakers generally defended their points or offered explanations" (Dhillon et al., 2004, p. 124).

### **13.2.9 "Follow Me"**

"This tag was added as speakers tended to occasionally seek verification from their listeners that their utterances were understood or agreed upon" (Dhillon et al., 2004, p. 124).

### **13.2.10 Joke**

"This tag was added to mark utterances of humorous content and jokes, as there was previously no other means to mark such utterances" (Dhillon et al., 2004, p. 124).

### **13.2.11 Rising Tone**

"Although this tag is not an actual dialog act, it was implemented to mark whether an utterance ended with a rising tone for the purpose of providing information for automatic speech recognition" (Dhillon et al., 2004, p. 125).

### **13.2.12 Nonlabeled**

"Certain utterances arose in the data that were intentionally not to be labeled. The Nonlabeled tag entered the MRDA tagset specifically for this purpose (Dhillon et al., 2004, p. 125).

The next sections are dedicated to tag-sets designed or adapted for asynchronous conversation. We anticipate that the descriptions below of the tag-sets for asynchronous conversation are significantly less detailed than the ones that we have made for synchronous

conversation. This is mostly because of the very little data provided by the original authors about their tag-sets. We think that the best way of understanding how each class is to be taken - as compared to the others, as well as to their "corresponding" classes in the other tag-sets - is to consult the tables at the end of this chapter, where each tag-set is mapped to the others. For graphical purposes, we have divided the mapping according to Searle's (1976) 5 primitive classes (since all subsequent classifications can theoretically be reduced to them). Before consulting the tables at the end, it is important to know that, even if a class is reported with the same name in two different tag-sets, the criteria of membership for that class may actually be different in the two tag-sets.

#### 14. Email Speech Acts

We have seen that MRDA tags dialog acts by splitting utterances into smaller units, each performing its own function in conversation. Cohen, Carvalho, and Mitchell, on the other hand, tag each entire email messages with one single label. In addition to this, as we will see, they tag email messages with a significantly smaller tag-set than the previous tag-sets used for synchronous conversation.

Because of our dependence on email - Shipley and Schwalbe (2007) estimate that U.S. office workers spend more than 25% of the day on email -, Carvalho (2008) aims at helping email users keep track of the status of ongoing conversations. Building upon previous studies on email act classification (see Cohen et al. (2004), and Cohen & Carvalho (2005; 2006)), Carvalho (2008) proposes a revised taxonomy of email acts to be used as a framework for the automatic detection of intentions behind the textual contents of email messages. More specifically, he wants to detect the email-act category (or categories) of incoming emails, where each category takes the form of a noun-verb pair, such as *request for information*, *commit to perform a task*, and *propose a meeting* (Carvalho, 2008, p. 7). The taxonomy proposed by Carvalho (2008), as he admits, is not intended for general purpose, but rather for work related email exchange only. Carvalho (2008) uses the work of Searle (1976) as the theoretical background, but focuses on observed linguistic behavior in actual email conversations to build his taxonomy. The corpora he observes are: the CSpace email corpus, which "contains approximately 15,000 email messages collected from a management course at Carnegie Mellon University", and the "PW CALO, a dataset generated during a four-day exercise conducted at SRI specifically to generate an email corpus" (Carvalho, 2008, p. 11), plus the conversations he found in his own inbox. Carvalho (2008), as we said, builds his taxonomy on the basis of observed linguistic behavior. This causes his tag-set to detach itself, at least partially, from

the classification theorized by Searle in 1976. This is in contrast with synchronous conversation speech act classifications, which are clearly inspired by Searle's (1976) classes (in the forward looking function), although they split them in a number of subclasses (see Allen & Core (1997), Jurafsky et al. (1997), and Dhillon et al. (2004)).

Carvalho (2008) knowingly merges several illocutionary points; for example "let's do lunch" (an offer), which is both a directive and a commissive (in Searle's (1976) classification and in all the above mentioned classifications) in that the speaker wants the hearer to do something and at the same time commits him- or herself to doing something, is classified by Carvalho (2008) as a simple "propose" act. Moreover, acts which need extra-linguistic institutions to be performed, i.e. Searle's (1976) declarations or institutional speech acts, are ignored altogether by Carvalho (2008): utterances in the form of statements are classified as deliveries of information, answers to questions, and other forms of delivery both linguistic and non-linguistic (e.g. files), but there is no distinction between the so-called assertives and declaratives. Carvalho (2008), as we said, does not take into account extra-linguistic institutions, but includes in his taxonomy non-linguistic uses of email, such as the delivery of files. To sum up, Carvalho (2008) defines four classes or act types, each represented by an email speech act verb (or illocutionary verb), and in turn aggregated into two broader classes: the illocutionary verbs "deliver" and "commit" belong to the set of commissive acts, and "request" and "propose" belong to the set of directive acts. We must notice that, in Carvalho's (2008) study, requests include both orders and questions. As a final note on email act classification, we mention the fact that Carvalho (2008) and Cohen & Carvalho (2005; 2006), unlike Cohen et al. (2004), exclude the category of Amends. Amends differ from proposals in terms of the tasks they refer to: while proposals are associated with commitments and requests in relation to new tasks, amend messages suggest modifications to already-proposed tasks (Carvalho, 2008). All studies on email act classification acknowledge the existence but do not construct classifiers for speech acts of refusal, greeting, and reminder since they are too infrequent or too irrelevant for task-tracking. Finally, we need to mention that Carvalho (2008), just like all previous studies on email act classification, also provides so-called "activity nouns", such as "data" and "meeting", i.e. nouns that are associated with email speech act verbs to form noun-verb pairs.

### **15. BC3, TA, and QC3**

Carvalho's (2008) taxonomy, which is conceived for work-related email exchange, can be expanded to the automatic recognition of speech acts in other forms of conversation: a more recent study is that of Joty and Hoque (2016), who address more broadly the issue of automatic speech act

recognition in virtually every type of written asynchronous conversations, e.g. fora, chats, emails, etc. As we said, Ulrich et al. (2008), in their definition of the original BC3 tag-set, adopt the tag-set proposed by Cohen & Carvalho (2005) but exclude the class of "Deliver" for being too broad. They write: "Deliver is excluded because most emails deliver some sort of information" (Ulrich et al., 2008). Ulrich et al. (2008) thus do not define a specific class for the delivery of information arguably because every utterance ultimately conveys some sort of information, which, according to them, makes this feature of emails less relevant for classificatory purposes.

The TA / BC3 new tag-set (Jeong et al., 2009; Joty et al., 2011) reduce the MRDA tag-set (Dhillon et al., 2004) from 50 classes to 12. This simplification is mostly due to the need to adapt the tag-set or MRDA to asynchronous conversations (emails and blog posts), and at the same time to get rid of underrepresented classes. No particular motivations are given by the authors for the reduction of the tag-set. Unlike Carvalho (2008) and Ulrich et al. (2008), Joty and Hoque (2016) use already existing tag-sets to build their own: they further reduce the tag-set created by Jeong et al. (2009) - Joty et al. (2011) (respectively, TA and BC3 new tag-set). Given the relatively small size of TA and BC3's 40 annotated threads and in order to learn a reasonable classifier and avoid a significant underrepresentation of some classes (and ultimately successfully detect speech acts overall), Joty and Hoque (2016) reduce the 12 act types of the TA / BC3 new tag-set (Jeong et al., 2009; Joty et al., 2011) to 5 coarser act types. Joty and Hoque's (2016) classes, while fewer than the 12 proposed by Jeong et al. (2009), are conceived specifically for domain-independent tagging of asynchronous conversations. Joty and Hoque (2016) briefly describe their 5 classes as follows: "all the question types are grouped into one general class Question, all response types into Response, and appreciation and polite mechanisms into Polite class." (p. 1750). The other two classes are Statement and Suggestion. Joty and Hoque (2016) use the MRDA meeting corpus to train their neural network. Since "TA and BC3 are quite small to make a general comment about model performance in asynchronous conversation" (Joty and Hoque, 2016, pp. 1750-1751), they decide to create the Qatar Computing Conversational Corpus or QC3: a new data set of 50 manually annotated conversations (with 5 speech act types) retrieved from a community question answering site called Qatar Living (Joty & Hoque, 2016).

In spite of the changes proposed in the different classifications, there are still a number of problems with regards to speech act classification and the definition of a reliable tag-set. In chapter 4, we will discuss further many of these problems.

## 17. Conclusion

Speech act classifications in computational linguistics seem to focus for the most part on the interaction between speech acts, such as on the relation between questions and answers or between offers and acceptances. The classifications of speech acts proposed in philosophy and linguistics, on the other hand, seem to focus on illocutionary points and on how they are achieved. In the next chapter, we will dive deeper into some of the problems that we have encountered in this chapter. We will be especially concerned with the oversimplification of statements, which results in their overrepresentation in all the corpora if compared to the other classes of speech acts (more in the next chapter). Not only statements can actually be declarations, but they can also be indirect requests or orders, depending on the context and on the authority of the speaker uttering them. Speaking about the use of statements will bring us back to the issue of indirectness at large, which this time we will discuss from the perspective of speech act classification.

The chart below shows a comparison between all the classifications considered in this chapter. Before consulting the chart, it is important to know that, even if a class is reported with the same name in the two different tag-sets, as we have specified throughout this chapter, the criteria of membership for that class may be different in the two tag-sets. For readability, we divided the chart into two: the first chart includes the classification proposed by Searle (1976) and the classifications proposed for synchronous conversation, whereas the second chart includes the classifications proposed in computational linguistics. (fwd l. f. = forward looking function; bwd l. f. = backward looking function).

AUTHOR OR AUTHORS	Searle (1976)	Allen & Core (1997)	Jurafsky et al. (1997)	Dhillon et al. (2004)		
NAME OF TAG-SET	Gold standard - Theory	DAMSL Standard	SWBD-DAMSL	MRDA		
ADAPTED FROM / INSPIRED BY	Austin (1962)	Searle (1976)	Allen & Core (1997)	Jurafsky et al. (1997)		
TOTAL NUMBER OF CLASSES / POSSIBLE TAGS	5	25 (13 fwd l. f. + 12 bwd l. f.)	50 (24 fwd l. f. + 26 bwd l. f.)	11 general (gen) + 39 specific (spec)		
# OF TAGS PER UTTERANCE + MUTUAL EXCLUSIVITY	1 or more tag per utterance	0 or more forward l. f. + 0 or more backward l. f.	0 or more forward l. f. + 0 or more backward l. f.	1 gen plus 0 or more spec (some are mutually exclusive)		
SUBDIVISION(S) FOR TAGGING PURPOSES	none	forward l. f. and backward l. f.	forward l. f. and backward l. f.	general and specific	MRDA GROUP	
DOMAIN		synchronous	synchronous	synchronous		
NOTES	classes are in bold					
FORWARD LOOKING	<b>Assertive (Representative)</b>	<i>Statements:</i>		<b>Statement (gen)</b>	1: Statements	
		- Assert	<b>Statement-non-opinion +</b>			
		- Reassert	<b>Statement-opinion +</b>			
		- Other-statement	Reassert			
	<b>Directive</b>	<i>Influencing-addressee-future-action:</i>			6: Action Motivators	
		- Open-option	<b>Open-option</b>	<del>Open-option</del>		
		- Action-directive	<b>Action-directive</b>	<b>Command (spec)</b>		
		<b>Info-request</b>	<b>Yes-no-question +</b>		<b>Yes-no-question (gen)</b>	2: Questions
					<b>Understanding Check (spec) +</b>	7: Checks
				<b>"Follow me" (spec) +</b>		
			<b>Wh-question +</b>		<b>Wh-question (gen)</b>	2: Questions
					<b>Repetition Request (spec) +</b>	7: Checks
			<b>Open-question +</b>		<b>Open-ended question (gen)</b>	2: Questions
			<b>Or-question +</b>		<b>Or-question (gen)</b>	
<b>Or-clause +</b>		<b>Or-clause after yes-no question (gen)</b>				
<b>Declarative-question +</b>		<b>Declarative-question (spec)</b>	11: Further			

			Tag-question +	Tag-question (spec)	Descriptions
			Rhetoric Question +	Rhetorical Question (gen)	2: Questions
	Commissive	<i>Committing-speaker-future-action:</i>			6: Action Motivators
		- Offer	Offer	Suggestion (spec)	
		- Commit	Commit	Commit/Commitment (self-inclusive) (spec)	
	Expressive	Explicit Performative	Explicit-performative	Explicit-performative	10: Politeness Mechanisms
	Declarative		Thanking +	Thanks (spec)	
			You're welcome +	Welcome (spec)	
		Apology +	Apology (spec)		
		Conventional-opening	Conventional-opening	Conventional-opening	
		Conventional-closing	Conventional-closing	Conventional-closing	
		Exclamation	Exclamation	Exclamation (spec)	11: Further Descriptions
	Other-forward-function	Other-forward-function	Other forward function		
			Floor Holder (gen) +	3: Floor Mechanisms	
			About-Task (spec) +	11: Further Descriptions	
			Topic Change (spec) +		
			Joke (spec) +		
			Self Talk (spec) +		
			Third Party Talk (spec) +		
			Rising Tone (spec) +		
BACKWARD LOOKING				Floor Grabber (gen) +	3: Floor Mechanisms
		<i>Agreement:</i>			
		- Hold	Hold	Hold before answer/agreement (gen)	3: Floor Mechanisms
		- Accept	Accept	Accept, Yes Answer <sup>1</sup> (spec)	5: Responses
	- Accept-part	Accept-part	Partial Accept (spec)		

		<b>- Reject</b>	<b>Reject</b>	<b>Reject, No Answer<sup>1</sup> (spec)</b>	
		<b>- Reject-part</b>	<b>Reject-part</b>	<b>Partial Reject (spec)</b>	
		<b>- Maybe</b>	<b>Maybe</b>	<b>Maybe (spec)</b>	
		<i>Understanding:</i>			
		<b>- Signal-non-understanding</b>	<b>Signal-non-understanding</b>	<b>Signal-non-understanding (spec)</b>	4?
		<i>- Signal-understanding:</i>			
		<b>- Acknowledge</b>	<b>Acknowledge/Backchannel</b>	<b>Acknowledge/Backchannel (gen)</b>	4: Backchannels and Acknowledgments
			<b>Rhetorical-question backchannel +</b>	<b>Rhetorical-question backchannel (spec)</b>	
			<b>Acknowledge-answer +</b>	<b>Acknowledgement (spec)</b>	
		<b>- Repeat-rephrase</b>	<b>Repeat-phrase</b>	<b>Mimic (spec)</b>	8: Restated Information
				<b>Repeat (spec)</b>	
			<b>Summarize-reformulate +</b>	<b>Reformulation/Summary (spec)</b>	
		<b>- Completion</b>	<b>Completion</b>	<b>Collaborative-completion (spec)</b>	9: Supportive Functions
			<b>Appreciation +</b>	<b>Appreciation/Assessment (spec)</b>	4: Backchannels and Acknowledgments
			<b>Sympathy +</b>	<b>Sympathy (spec)</b>	10: Politeness Mechanisms
			<b>Downplayer +</b>	<b>Downplayer (spec)</b>	
		<b>- Correct-misspeaking</b>	<b>Correct-misspeaking</b>	<b>Correct-misspeaking (spec)</b>	8: Restated Information
				<b>Misspeak Self-Correction + (spec)</b>	
		<b>Answer</b>	<b>Yes Answer +</b>	<i>merged<sup>1</sup></i>	
			<b>No Answer +</b>	<i>merged<sup>1</sup></i>	
			<b>Affirmative non-yes answer +</b>	<b>Narrative-affirmative answer/Affirmative non-yes Answer (spec)</b>	5: Responses
			<b>Negative non-no answer +</b>	<b>Narrative-negative answer/Negative non-no Answer (spec)</b>	
			<b>Other answer +</b>	<b>No knowledge (spec)</b>	

			<b>Yes plus expansion +</b>	<del>Yes plus expansion</del>	
			<b>No plus expansion +</b>	<del>No plus expansion</del>	
			<b>Statement expanding y/n answer +</b>	Statement expanding y/n answer	
			<b>Expansion of yes-no answer +</b>	<b>Expansion of yes-no answer/Elaboration (spec)</b>	9: Supportive Functions
				<b>Defending-explanation (spec) +</b>	
			<b>Dispreferred answer +</b>	<b>Dispreferred answer (spec)</b>	5: Responses

Tag-sets for Asynchronous conversation:

AUTHOR OR AUTHORS	Cohen et al. (2004)		Cohen & Carvalho (2005; 2006); Carvalho (2008)		Ulrich et al. (2008)	Jeong et al. (2009) - Joty et al. (2011)	Joty & Hoque (2016)
NAME OF TAG-SET	Email Act Taxonomy		"Reduced" Email Act Taxonomy		BC3 original tag-set	TA / BC3 new tag-set	QC3
ADAPTED FROM / INSPIRED BY	Searle (1976)		Searle (1976) - Cohen et al. (2004)		Cohen & Carvalho (2005)	Dhillon et al. (2004)	Jeong et al. (2009) - Joty et al. (2011)
TOTAL NUMBER OF CLASSES / POSSIBLE TAGS	5		4		4	12	5
# OF TAGS PER UTTERANCE + MUTUAL EXCLUSIVITY							
SUBDIVISION(S) FOR TAGGING PURPOSES	superclass(es)	subclass	superclass	subclass			
DOMAIN	asynchronous - we do not report on this table noun classes		asynchronous - we do not report on this table noun classes		asynchronous	asynchronous (domain-independent speech act tags)	asynchronous
NOTES	classes for which classifiers were constructed in bold						
FORWARD LOOKING		<b>Deliver<sup>3</sup></b>	Commissive -->	<b>Deliver<sup>3</sup></b>		<b>Statement</b>	<b>Statement</b>
						<b>Action motivator</b>	
						<b>Yes-no question</b>	<b>Question</b>
						<b>Wh-question</b>	
						<b>Open-ended question</b>	
		Negotiate --> Initiate -->	<b>Request</b>	Directive -->	<b>Request</b>	<b>Request</b>	



				isagreement <sup>4</sup>			
					Reject response <sup>6</sup>		
						Acknowledge and appreciate	Polite
						Polite mechanism	Polite
		merged <sup>3</sup>	merged <sup>3</sup>	merged <sup>4</sup>	merged <sup>5</sup>	Response	
merged <sup>6</sup>							
merged <sup>5</sup>							
merged <sup>6</sup>							
Uncertain response							

## CHAPTER 4 - PROBLEMS CONNECTED WITH SPEECH ACT IDENTIFICATION

In the present chapter, we will elaborate on the problems that arise from the adaptation of the speech act theory in computational linguistics, and propose a number of possible solutions. A reliable classification of speech acts allows to efficiently and systematically map utterances to speech act types. At the same time, a classification should be fine-grained enough to be useful for downstream processing in the first place. We will analyze the classifications proposed in the chapter 3 and focus on the classes that, in our opinion, need to be discussed further. We will also discuss the percentage representation of some classes in the corpora: one of the major issues (shared by all classifications) is the overrepresentation of statements. The overrepresentation of statements in turn causes a negative ripple effect that brings in a number of other issues. While it is possible that the straightforward delivery of information happens very frequently, it is also possible that many utterances have actually been mistagged as statements, but are in fact either non conventional indirect requests, indirect questions, expressives, or declarations.

We will see that statements always represent more than half of all speech acts detected in a corpus. If this measurement is accurate, then we can consider potentially dividing statements into a number of meaningful subclasses. This is what was actually proposed in both the DAMSL Standard and the SWBD-DAMSL tag-set but was then abandoned starting from the MRDA tag-set. If on the other hand, this measurement is not accurate, we need to discover why this is the case. At the same time, we do not want to exclude the fact that one tag-set might work better than another because of the way in which the text is segmented before the tagging is performed. That being said, we will only propose changes upstream, that is to say: we will not discuss how the computer should learn from the data, but rather how more accurate and meaningful data can be submitted to the computer (regardless of the machine learning algorithm run on the data). Our main argument is that the issues connected with indirect speech acts play a major role in the misclassification of speech acts, and in particular in the misclassification as statements of non conventional requests and of other types of speech acts.

Tables 2 to 5 below represent the distribution of the different speech act classes in the corpora. We are particularly interested in the class of statements (sometimes represented as S or ST).

Table 2. Distribution of QC3 classes in TA, BC3, and MRDA corpora (Joty & Hoque, 2016, p. 1750)

Tag	Description	TA	BC3	MRDA
SU	Suggestion	7.71%	5.48%	5.97%
R	Response	2.4%	3.75%	15.63%
Q	Question	14.71%	8.41%	8.62%
P	Polite	9.57%	8.63%	3.77%
ST	Statement	65.62%	73.72%	66.00%

Table 3. Distribution of QC3 classes the QC3 corpus (Joty & Hoque, 2016, p. 1751)

Speech Act	Distribution
Suggestion	17.38%
Response	5.24%
Question	12.59%
Polite	6.13%
Statement	58.66%

Table 4. Distribution of TA classes in SWBD and MRDA corpora (Jeong et al., 2009, p. 1253; the tags are defined in Table 5 as they are the same tags used by Joty et al., 2011)

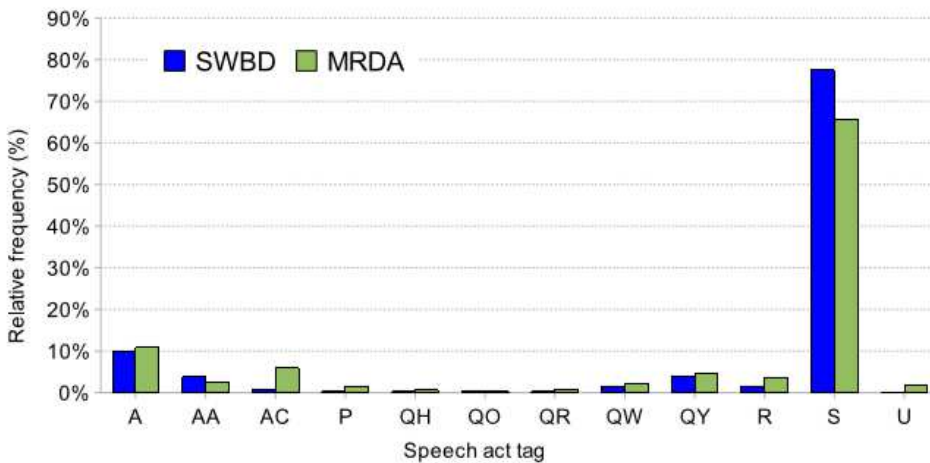


Table 5. Distribution of new BC3 classes in 40 email threads of the BC3 corpus + 200 forum threads from the TripAdvisor travel forum site (Joty et al., 2011)

Tag	Description	Email	Forum
S	Statement	69.56%	65.62%
P	Polite mechanism	6.97%	9.11%
QY	Yes-no question	6.75%	8.33%
AC	Action motivator	6.09%	7.71%
QW	Wh-question	2.29%	4.23%
A	Accept response	2.07%	1.10%
QO	Open-ended question	1.32%	0.92%
AA	Acknowledge and appreciate	1.24%	0.46%
QR	Or/or-clause question	1.10%	1.16%
R	Reject response	1.06%	0.64%
U	Uncertain response	0.79%	0.65%
QH	Rhetorical question	0.75%	0.08%

### 1. Statements

As Allen and Core (1997) said, every utterance has a certain effect on the dialog but, at the same time, "the actual form of the sentence might look like something else". This means that, while the form of the sentence might suggest that the utterance is a statement, the utterance can instead be a request performed indirectly or another speech act. Before we discuss the issue of indirectness in statements, we propose a chart indicating how different types of statements are tagged in the corpora of synchronous conversation analyzed in chapter 3 as compared to Searle's (1976) classification (in asynchronous conversation, all statements are labeled with the same tag).

"if we exclude English um - there is not much difference with the data."			
Searle (theory)	DAMSL Standard	SWBD-DAMSL	MRDA
Assertive	Assert	Statement-non-opinion	Statement

"It's a great story."			
Searle (theory)	DAMSL Standard	SWBD-DAMSL	MRDA
Assertive / Expressive	Assert	Statement-opinion	Statement + Assessment/appreciation

So this changes the whole mapping for every utterance.			
Searle (theory)	DAMSL Standard	SWBD-DAMSL	MRDA
Assertive	Reassert	Statement-non-opinion	Statement + Understanding Check

As we have seen before, in MRDA (Dhillon et al., 2004) the Assessment/Appreciation tag is attached to most "[c]omments and opinions on an aspect a speaker has noticed within the contents of another speaker's speech" (pp. 52-54). The Assessment/Appreciation tag thus serves to mark those utterances that in SWBD-DAMSL are tagged as Statement-opinions. In other words, MRDA defines opinions as Statements further characterized by the special tag Assessment/Appreciation. Jurafsky et al. (1997), on the other hand, devise two separate tags to mark the distinction between what they call "descriptive/narrative/personal" statements (Statement-non-opinion) and "other-directed opinion statements" (Statement-opinion). This distinction, which was not accounted for in the DAMSL standard, allows to distinguish opinions, which usually express agreement or disagreement, from statements of facts. We said that the problem with this distinction is that same

statements can belong to either category - opinion and non-opinion - depending on whether the speaker has expertise in the subject about which he or she is talking. According to us, the expertise of a speaker on a specific subject is too subjective to be used as a reliable discriminating factor between opinions and non-opinions. Let's say for example that the speaker utters "I get along with the boss" to a coworker. In this case, the same utterance can be an opinion or a non-opinion simply by virtue of whether the interlocutor can agree with that statement. If the speaker is in the position to disagree, then that statement will constitute an opinion, or else it will constitute a non-opinion. To further complicate the situation, let's reconsider the following example from chapter 3 (Jurafsky et al., 1997):

A: I think Mercedes are great cars. (Statement-non-opinion)

B: Me too. (Accept)

There is indeed a difference between:

I think Mercedes are great cars.

and

Mercedes are great cars.

because:

The former is arguably a Statement-non-opinion as the speaker knows exactly what he or she thinks. The speaker knows better that anyone else his or her thoughts regardless of whether Mercedes are great cars, which is not relevant at this point. In other words, it is an indisputable fact that the speaker thinks that Mercedes are great cars. The latter, on the other hand, is a Statement-opinion since the speaker is expressing an opinion on something non-personal, i.e. Mercedes, on which he or she is likely not an expert. We can therefore see that, when the speaker makes explicit his or her psychological state, it becomes challenging to identify the utterance as an opinion since the expression of the speaker's psychological state is clearly not the speaker's opinion. A similar situation occurs in:

I like Mercedes (Statement-non-opinion)

which is different than

Mercedes are good (Statement-opinion)

Another example that is worth making is the following exchange (from Jurafsky et al., 1997):

A: My husband feels that they'll come and collect everybody's guns. (Statement-non-opinion)

B: Yeah. (Acknowledge)

B: I guess that could happen. (Maybe)

Utterance A seems to demonstrate that, when the speaker refers to somebody else and predicates a certain action of them, his or her utterance is always a Statement-non-opinion. It is in fact a non-debatable fact that the other person thinks that, or even says that such as in:

A: My husband says that they'll come and collect everybody's guns. (Statement-non-opinion)

The fact that her husband says it is non debatable. Moving on to other information about statements that a classification can detect, we need to mention that, while Searle's classification (1976) and the SWBD-DAMSL tag-sets do not capture the repetition of information, the DAMSL Standard and the MRDA do. This issue is tied to the linguistic form of the utterance (and its propositional content), rather than the utterance's illocutionary force. At this point, however, we will not dive deeper into the information that one tag-set is able to capture about statements but another cannot, but rather we will focus on the information that none of the tag-sets is able to capture. We can in fact say that, while differing from one another, none of the tag-sets analyzed in chapter 3 accurately accounts for the *indirect* use of statements.

Statements, as we can see from the tables above, are overrepresented in all of the corpora that we considered for the present study. This overrepresentation of statements occurs probably because, in practice, we are not able to identify the institutions in which declarations are performed and thus we are forced to merge assertives with declarations, or because we are not able to determine when the speaker performs indirect speech acts by way of making statements. In this regard, the introduction of adjacency pairs allows us, to a certain extent, to understand the indirect use of an utterance by looking at how the interlocutor responds to it. If it is an acknowledgment, then the utterance might be a statement, if it is an acceptance, then the utterance might be an indirect request, and if it is an answer, then the utterance might be an indirect question. For example (Allen & Core, 1997):

A: I'll take the Avon train to Dansville.

B: Okay.

Utterance B is a simple acknowledgment and therefore A is a simple statement. On the other hand (Jurafsky et al., 1997):

A: I have a recipe if you want.

B: Sure.

Utterance B is an acceptance and therefore A is an indirect request. Finally:

A: I don't know if you like chocolate.

B: Yes, I do.

Utterance B is a (positive) answer and therefore A is an indirect Yes-no question. As we said in chapter 2, we can leverage the felicity conditions of requests and questions to determine whether an utterance can be used to perform either an indirect request or a question. In addition to this, now that we know what the response to a request or a to question looks like, we have indeed more contextual information that we can leverage to reach a reasonable conclusion. From this point of view, the issue of discriminating between statements, indirect requests, and indirect questions, becomes intertwined with that of distinguishing between acknowledgments, acceptances and answers.

With regards to indirect requests, Benincà et al. (1977) argue that the indirect performance of requests for action depends also on the "symmetry or asymmetry of the mutual respect" (p. 521), that is to say: the respect between two interlocutors that results from information such as their age, sex, infirmities or disabilities, as well as the role and the social status that the interlocutors have in society. According to Benincà et al. (1977) this information determines the type (and, we add, as a consequence the amount) of requests that one speaker can make to the other: if more respect can be ascribed to the hearer, then the speaker has a fairly limited number of requests that he or she can make. This means that the speaker is less likely to intend his or her utterance as an indirect request. As a consequence of this, another possible way to increase the precision of the discrimination between statements and indirect requests or orders is to work on the context to predict leadership roles (Carvalho, 2008). Doing so, will allow us to determine more accurately whether the speaker is assessing a state of affairs, or ordering or requesting that that state of affairs is brought about. In other words, if a speaker is very authoritative and refers indirectly to the motivations for a certain activity to done, that assertion should probably count as a request or an order.

If we do not know the roles of the participants; for example, we do not know that the conversation is between an employer and his or her employee, we can probably still determine leadership roles in a number of ways. According to Searle and Vanderveken (1985), a greater strength of the illocutionary point can derive from the power or authority of one of the interlocutors. The different degree of strength of the illocutionary point is evident in expressions such as "expressing regret" and "humbly apologizing" (Searle & Vanderveken, 1985), where in the second case the stronger degree is partially caused by the use of the adverb "humbly". If the hearer is "humbly apologizing" and the speaker is "imposing" a certain action, then we are likely in front of a situation with a significant asymmetry of the mutual respect. We believe that including a score for each interlocutor that indicates how authoritative he or she is would indeed help discriminate between statements and requests. The following information is of course, to a certain extent,

subjective, but we still deem it useful if leveraged correctly. The authority score of the speaker might result from a statistical analysis of:

- the degree of strength of the illocutionary points from both interlocutors;
- the use of the imperative;
- the use of "please" when the sentences allows for its embedding;
- the times that one interlocutor interrupts the other.

Having at hand this information and combining it into one single score, we argue, would indeed help us identify with more precision a declarative utterance as either a statement or an indirect request or order.

## 2. Issues regarding other classes

In the corpora there are a number of cases in which the indirect speech act is not identified properly. One example is the following from the BC3 corpus, where a conventional indirect request has been mistagged as Yes-no question (Joty et al., 2011; tagged using the new BC3 tag-set):

A: Can you suggest another venue and possible sponsor? QY

B: I am pursuing one other possibility but would like to hear back from the possible sponsor before I suggest it. S

Utterance A is tagged as a Yes-no question but in this context it looks more like an Action motivator, and utterance B is statement but is arguably primarily a (temporary) Reject response. We have noticed that, overall, the interest in the class of directives has lowered in time. The DAMSL Standard distinguishes between Info-requests (the speaker is making a question or another request for information) and Action-directives (the speaker is creating an obligation that the hearer do the action unless the hearer indicates otherwise), whereas all the classifications in asynchronous conversation, with the only exception of the TA / BC3 new tag-set, do not. These classifications merge requests for information and requests for action in the same category of requests. While the DAMSL Standard keeps Info-requests and Action-directives separate, we still encounter cases in which Action-directives are mistagged as Info-requests. Let's consider the following example (from Allen & Core, 1997):

A utt1: can you tell me the time? (Info-request)

B utt2: yes. (Accept(utt1))

B utt3: it's 5 o'clock. (Answer(utt1))

As we have seen, in this case, the sentence uttered by A is used to make an indirect request. However, Allen & Core (1997) tag utt1 as an Info request thus ignoring (at least in the tag itself) the

fact that it is indirectly used to make a request. Utterance A should in fact be an Action-directive as it is not used as a literal Yes-no question. The only way for us to know, on the basis of the tags, that it is actually a request made by way of asking a Yes-no question (or by way of requesting information in the form of a Yes-no answer) is thanks to the tag Accept, which is used for the response (utt2), instead of the Answer tag. Accepts are used in response to Action-directives, whereas Answers are used in response to Info-requests. Interestingly, utt3 is tagged as Answer almost as if utterance A was equivalent to:

A utt1 What's the time?

On a similar note, we have seen that Answers can be in the imperative mood, such as (from Allen & Core, 1997):

A utt1: how do I get to Corning? (Info-request)

B utt2: Go via Bath. (Assert, Open-option, Answer(utt1))

Within the dimension of requests for action, the DAMLS Standard and SWDB-DAMSL distinguish between Action-directives and Open-options (the speaker is not creating an obligation that the hearer do the action unless the hearer indicates otherwise), whereas all subsequent classifications, including MRDA and the TA / BC3 new tag-set, do not. This can be problematic as we have seen that Open-options consist also of indirect requests (arguably except for the more conventionalized) in that requests performed indirectly leave open the possibility for the hearer (or at least the idea) to refuse to comply. As a consequence, in most classifications we do not have labels to distinguish between, for example:

How hot!

which, in the appropriate context, is an Open-option, and:

Close the window!

an Action-directive.

Jurafsky et al. (1997) discuss the issue of indirect requests for information. As we saw in chapter 3, utterance 72a is a Yes-No-question both semantically and pragmatically, utterances 72b and 72c are semantically Statement-non-opinions and pragmatically Yes-No-questions (what Jurafsky et al. (1997) calls Declarative questions), utterance 72d is semantically a Yes-No-question and pragmatically an Action-directive, and utterance 72e is a Statement-non-opinion both semantically and pragmatically (from Jurafsky et al., 1997):

72a. Do you have to have any special training? (Yes-No-question)

72b. I don't know if you are familiar with that. (Yes-No-question + Declarative question)

72c. You must be familiar with that. (Yes-No-question + Declarative question)

72d. Can you pass the salt? (Action-directive)

72e. I like cakes. (Statement-non-opinion)

As we have mentioned in chapter 3, Jurafsky et al. (1997) discuss the indirect speech acts that have the syntactic form of declarations but are "pragmatically" questions. In SWBD-DAMSL, they are called declarative questions and are labeled the Declarative question tag. The Declarative question tag is concatenated with the tag indicating the kind of question that they ask. Declarative questions can be used to make either a Yes-No-question, a Wh-question, an Open-question, or an Or-question. For example, 72b above (from Jurafsky et al., 1997) is a statement used to indirectly make an indirect Yes-No-question equivalent to "are you familiar with that?" and therefore should be tagged as Yes-No-question + Declarative question. 72b leverages one of the conditions of success of questions, in particular one of the preparatory conditions, that is that the speaker does not know something about the hearer (and therefore asks about it). The fact that utterances like 72b do not have what Jurafsky et al. (1997) call "question form" makes it impossible to determine that they are questions from what is said alone: they do not have a wh-word as the argument of the verb, nor subject-aux inversion, since they have "declarative" word order (the subject precedes the main verb) (Jurafsky, 1997). At the same time, utterances like 72b may have rising question-intonation (Jurafsky et al., 1997), which however might not be a viable feature to rely on in written corpora. On a similar note, utterance 72d, has the syntactic form of a Yes-No-question but has indeed to be tagged as an Action-directive (Jurafsky et al., 1997) since it is uttered with the intention of getting the addressee to do something. Utterances like 72d do not have the typical Action-directive form, i.e. they are not in the imperative mood. utterances like 72b and 72d are said to have the literal force respectively of a statement and of a question, and, in addition, respectively the contextual (or pragmatic) force of a Yes-no question and of an Action-directive.

The MRDA tag-set, on the other hand, deals with other types of indirect speech acts such as negative answers. In MRDA, direct answers (merged with the acceptance or refusal of requests for action) are labeled either with the Accept, Yes Answer(spec) or the Reject, No Answer(spec) tags. As we mentioned in chapter 3, Dhillon et al. (2004) clarify the notion of Negative answer as follows (p. 64):

Negative Answers are "indirect negative response(s)", which [o]ftentimes (...) appear as alternative suggestions to a previous speaker's question, proposal, or statement"

In other words, Negative answers are used to indirectly reject a question, proposal, or statement: instead of saying "no" (which is a Reject, No Answer(spec)), the speaker can make a rejection by proposing an alternative. Here's an example of Negative Answer (adapted from Dhillon et al., 2004, p. 65-66):

A: you guys have plans for Sunday? (Yes-No-Question(gen) + Rising Tone(spec))

A: because we also want to combine it with some barbeque activity where we just fire it up and what - whoever brings whatever you know can throw it on there. (Statement(gen))

B: well I'm going back to visit my parents this weekend. (Statement(gen) + Negative Answer(spec))

Utterance B is an indirect way of saying "I cannot come". By uttering B, the speaker states that one of the preparatory conditions for accepting is not met, and therefore he or she is indirectly refusing the invitation to the barbeque.

To open a brief parenthesis on the commissive dimension of speech acts, our next concern regards the existence, acknowledged by Allen and Core (1997), of Conditional Commits, such as:

I'll be there if the package arrives on time. (Commit)

which they tag as simple Commits, despite not being necessarily commits. We think that there should be an appropriate tag for conditional commits or that the utterance should be split into two single units, tagged separately: a Statement and a Commit, where the commitment depends on the truth of the statement.

### 3. Structure of the Tags

We saw that MRDA has explored the idea of "main tag and secondary tag(s)" to the point that two different sets of tags have been created: one set includes the general tags that represent all the possible basic forms of an utterance (e.g. statement, question, backchannel, etc.), the other set includes the specific tags that represent the functions or the characteristics an utterance may have in addition to its basic form (e.g., accepting, rejecting, acknowledging, rising tone, etc.) (Dhillon et al., 2004). While Dhillon et al. (2004) did not ideate this style of tagging specifically to account for indirect speech acts, we consider this view to be particularly useful for distinguishing between the direct speech act (based on the literal form of the utterance) and the indirect speech act which is actually performed. We therefore purport to adopt their tagging style to account for indirect speech act; for example:

You should leave. (Statement(gen) + Command(spec))

is a statement used to perform a command, and:

Can you leave? (Yes-no-question(gen) + Command(spec))

is a yes-no question used to perform a command. Finally, a direct command we argue that it should be tagged as follows:

Leave! (Command(gen) + Command(spec))

or simply:

Leave! (Command(gen))

since it is a command both literally and contextually. Dhillon et al. (2004), on the other hand, tag direct requests just like statements used to make indirect requests (p. 71-72):

Continue. (Statement(gen) + Command(spec))

We believe however that a distinction should be made between the two.

In addition to this new way of tagging, we could also explore the idea of expanding the context to more than the previous and the next utterance, that is to say: augmenting the notion of Adjacency Pairs to Adjacency Trios (or even bigger chunks of the discourse). Adjacency pairs, we said, are paired utterances, produced by different speakers, that reflect the structure of conversation; some examples are: question-answer, greeting-greeting, offer-acceptance, and apology-downplay (Dhillon et al., 2004, p. 25; Levinson, 1983). According to Dhillon et al. (2004), "[l]abeling adjacency pairs (AP) in meetings provides a means to extract the information provided by the interaction between speakers" (p. 25). It would indeed be interesting to see if expanding the co-text would allow us to extract even more accurate information about the conversation, thus facilitating the identification of the speech acts performed by each interlocutor.

#### **4. Conclusion**

In the present chapter, we have elaborated on some of the problems about the classification of speech acts. We can say that the task of identifying utterances as performing one speech act rather than other is fairly challenging; even humans can sometimes identify an utterance as performing a speech act instead of another. We saw that speech act theory and the notion of speech act have been simplified to fit practical needs, sometimes to the point that they have lost part of their original meaning. The rule of thumb for a successful classification is that the number of classes should be limited and that the criteria characterizing each class so should be clear so that we can classify each utterance given as an input with the smallest possible margin of error. At the same time, we need to define enough classes to make the classification useful for downstream processing in the first place. The results must be satisfying enough for the classification to be used for the development of a number of applications, such as: dialog systems, automated summarization, machine translation, conversation tracking, and so on. Two classes of speech acts defined by Searle (1976) are particularly controversial and therefore have not been analyzed thoroughly. One is the class of expressives, which has often been overly simplified or even excluded from classification because it was not considered particularly useful for the particular applications for which the classifications that we analyzed were built. However, expressives have indeed become fairly useful

in recent years: understanding expressives has turned out to be crucial in the growing area of opinion mining and sentiment analysis<sup>30</sup>. The other controversial class is that of declarations. This class has often been removed altogether in the transition to computational linguistics because of the lack of contextual data: declarations, in fact, rely on particular cultural-dependent institutions, whose presence is challenging to retrieve with the current technology. As Searle (1976) points out, "the mastery of those rules which constitutes linguistic competence by the speaker and hearer is not in general sufficient for the performance of a declaration. In addition, there must exist an extra-linguistic institution and the speaker and hearer must occupy special places within this institution. It is only given such institutions as the Church, the law, private property, the state and a special position of the speaker and hearer within these institutions that one can excommunicate, appoint, give and bequeath one's possessions or declare war. We may add that the only exceptions to the principle that every declaration requires an extra-linguistic institution are those declarations that concern language itself, as for example, when one says, 'I define, abbreviate, name, call or dub'" (Searle, 1976, p. 14-15). In chapter 1, we mentioned that cultural dependency affects to a certain degree all types of speech acts, even speech acts which are usually not considered institutional speech acts per se, in that societal conventions always apply and regulate the way we act. All speech acts are therefore partially culture-dependent which makes all classifications of speech acts to a certain extent necessarily culture-dependent and ethnocentric. While there are indeed commonalities between non institutional speech act types across cultures and languages, we want to remark the fact that the classifications proposed in chapter 3 and further discussed in this chapter focus on data in the English language, produced for the most part, but not exclusively, by US American native speakers. We conclude with the words of Wierzbicka, who emphasizes the volatile nature not only of declarations, but of speech acts as a whole: "from the outset, studies in speech acts have suffered from an astonishing ethnocentrism, and to a considerable degree they continue to do so" (1991, p. 25).

---

<sup>30</sup> The area which "deals with the computational treatment of opinion, sentiment, and subjectivity in text" (Pang & Lee, 2008).

## References

Act. (2019). In OxfordDictionaries.com. Retrieved from <https://www.lexico.com/en/definition/act>

Action. (2019). In OxfordDictionaries.com. Retrieved from <https://www.lexico.com/en/definition/action>

Allen, J., & Core, M. (1997). Draft of DAMSL: Dialog act markup in several layers.

Austin, J. L., & Urmson, J. O. (1962). *How to Do Things with Words*. The William James Lectures.

Bach, K. (1999). The myth of conventional implicature, *Linguistics and Philosophy*, 22: 327–66.

Bach, K. (2006). The top 10 misconceptions about implicature, in B. Birner & G. Ward (eds.), *Drawing the Boundaries of Meaning: Neo-Gricean Studies in Pragmatics and Semantics in Honor of Laurence R. Horn*, pp. 21–30, Amsterdam: John Benjamins.

Bach, K., & Harnish, R. M. (1979). Linguistic communication and speech acts.

Benincà, P., Cinque, G., Fava, E., Leonardi, P., & Piva, P. (1977). 101 modi per richiedere. *Aspetti sociolinguistici dell'Italia contemporanea*, pp. 501-33. Roma: Bulzoni.

Carlson, L. (1983). *Dialogue games: An approach to discourse analysis*. D. Reidel.

Carvalho, V. R., & Cohen, W. W. (2005, August). On the collective classification of email speech acts. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 345-352). ACM.

Carvalho, V. R., & Cohen, W. W. (2006, June). Improving email speech acts analysis via n-gram selection. In *Proceedings of the HLT-NAACL 2006 Workshop on Analyzing Conversations in Text and Speech* (pp. 35-41). Association for Computational Linguistics.

Carvalho, V. R. (2008). *Modeling intention in email*. Carnegie Mellon University, Language Technologies Institute, School of Computer Science.

Cohen, L.J. (1964). 'Do Illocutionary Forces Exist?' *The Philosophical Quarterly*, 14: 118–137.

Cohen, W.W., Carvalho, V.R. and Mitchell, T. (2004). Learning to classify email into "speech acts". In *Proceedings of EMNLP*, pages 309–316.

Davidson, D. (1967). Truth and meaning. In *Philosophy, Language, and Artificial Intelligence* (pp. 93-111). Springer, Dordrecht.

Davis, S. (1988). Linguistic semantics, philosophical semantics, and pragmatics. *Philosophia*, 18(4), 357-370.

Dhillon, R., Bhagat, S., Carvey, H., & Shriberg, E. (2004). Meeting recorder project: Dialog act labeling guide (No. ICSI-TR-04-002). INTERNATIONAL COMPUTER SCIENCE INST BERKELEY CA.

Green, M. (2017). "Speech Acts", *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), Edward N. Zalta (ed.). Retrieved from: <https://plato.stanford.edu/archives/sum2017/entries/speech-acts/>.

Grice, H. P. (1957). Meaning. *The philosophical review*, 377-388.

Grice, H. P. (1975). In: Grice, H. P., Cole, P., & Morgan, J., *Logic and conversation*. 1975, 41-58.

Grice, H. P. (1989). *Studies in the Way of Words*. Harvard University Press.

Hong, L., & Davison, B. D. (2009, July). A classification-based approach to question answering in discussion boards. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval* (pp. 171-178). ACM.

Horn, L. (2004). Implicature. In: Horn, L. and Ward, G. (eds.), *Handbook of Pragmatics*. Oxford: Blackwell. 3-28.

Hurford, J. R. and Heasley, B. (1983). *Semantics: a course book*. Cambridge University Press.

Hymes, D. (1974). *Foundations in Sociolinguistics: An Ethnographic Approach*. Philadelphia: University of Pennsylvania Press.

Inference. (2019). In *OxfordDictionaries.com*. Retrieved from <https://www.lexico.com/en/definition/inference>

Jaszczolt, K. (2002). *Semantics and pragmatics: Meaning in language and discourse*. Pearson education. London: Longman. Second edition under contract with Cambridge University Press.

Jeong, M., Lin, C. Y., & Lee, G. G. (2009, August). Semi-supervised speech act recognition in emails and forums. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3* (pp. 1250-1259). Association for Computational Linguistics.

Joty, S., Carenini, G., & Lin, C. Y. (2011, July). Unsupervised modeling of dialog acts in asynchronous conversations. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence* (Vol. 22, No. 3, p. 1807).

Joty, S., & Hoque, E. (2016). Speech act modeling of written asynchronous conversations with task-specific embeddings and conditional structured models. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL (pp. 7-12).

Jurafsky, D., Shriberg, E. & Biasca, D. 1997. Switchboard SWBD-DAMSL labeling project coder's manual, draft 13. Technical report, Univ. of Colorado Institute of Cognitive Science.

Jurafsky, D. & Martin, J. H. (2018). Speech and Language Processing (3rd ed. draft). Retrieved from <https://web.stanford.edu/~jurafsky/slp3/>

Kissine, M. (2013). From utterances to speech acts. Cambridge University Press.

Korta, K. & Perry, J. (2015). Pragmatics. The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University. Retrieved from <https://plato.stanford.edu/entries/pragmatics/>

Leezenberg, M. (2001). Contexts of metaphor. Amsterdam and London: Elsevier

Levin, L., Langley, C., Lavie, A., Gates, D., Wallace, D., & Peterson, K. (2003). Domain specific speech acts for spoken language translation. In Proceedings of the Fourth SIGdial Workshop of Discourse and Dialogue (pp. 208-217).

Levinson, Stephen C. 1983. Pragmatics. Cambridge: Cambridge University Press.

Levinson, S. C. (2000). Presumptive meanings: The theory of generalized conversational implicature. Cambridge, MA: MIT press.

Lewis, D. K. (1980). Index, Context and Content in Lewis (1998). Papers in Philosophical Logic. Cambridge: Cambridge University Press, 21–44.

Literal. (2019). In OxfordDictionaries.com. Retrieved from <https://www.lexico.com/en/definition/literal>

McGrath, M. (2018), Propositions, The Stanford Encyclopedia of Philosophy, Edward N. Zalta (ed.). Retrieved from <https://plato.stanford.edu/archives/spr2014/entries/propositions/>.

McKeown, K., Shrestha, L., & Rambow, O. (2007, February). Using question-answer pairs in extractive summarization of email conversations. In International Conference on Intelligent Text Processing and Computational Linguistics (pp. 542-550). Springer, Berlin, Heidelberg.

Murray, G., Carenini, G., & Ng, R. (2010, July). Generating and validating abstracts of meeting conversations: a user study. In Proceedings of the 6th International Natural Language Generation Conference (pp. 105-113). Association for Computational Linguistics.

Oya, T., & Carenini, G. (2014, June). Extractive summarization and dialogue act modeling on email threads: An integrated probabilistic approach. In Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL) (pp. 133-140).

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2), 1-135.

Penco, C. (1999, September). Objective and cognitive context. In *International and Interdisciplinary Conference on Modeling and Using Context* (pp. 270-283). Springer, Berlin, Heidelberg.

Potts, C. (2005). *The Logic of Conversational Implicatures*. Oxford: Oxford University Press.

Potts, C. (2007). Into the conventional-implicature dimension, *Philosophy Compass*, 2: 655–79.

Ranganath, R., Jurafsky, D., & McFarland, D. (2009, August). It's not you, it's me: detecting flirting and its misperception in speed-dates. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1* (pp. 334-342). Association for Computational Linguistics.

Ravi, S., & Kim, J. (2007). Profiling student interactions in threaded discussions with speech act classifiers. *Frontiers in Artificial Intelligence and Applications*, 158, 357.

Recanati, F. (2004). *Literal meaning*. Cambridge University Press.

Sadock, J. M. (1974). *Toward a linguistic theory of speech acts*. Academic Pr.

Sbisà, M. (2002). Speech acts in context. *Language & Communication*, 22(4), 421-436.

Sbisà, M. (2006). Speech acts without propositions?. *Grazer Philosophische Studien*, 72(1), 155-178.

Schegloff, E. A. (1968). Sequencing in conversational openings. *American anthropologist*, 70(6), 1075-1095.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language* (Vol. 626). Cambridge university press.

Searle, J. R. (1975). Indirect speech acts. *Syntax and semantics. Volume 3: Speech acts*, 59-82.

Searle, J. R. (1976). A classification of illocutionary acts. *Language in society*, 5(01), 1-23.

Searle, J. R., & Vanderveken, D. (1985). *Foundations of illocutionary logic*. CUP Archive. Chicago

Shipley, D., & Schwalbe, W. (2007). *The Essential Guide to Email for Office and Home*.

Speaks, J. (2018). Theories of Meaning. The Stanford Encyclopedia of Philosophy. Zalta, E. N. (ed). Retrieved from <https://plato.stanford.edu/archives/spr2017/entries/meaning/>.

Sperber and Wilson (1995). *Relevance: Communication and Cognition*. Oxford: Blackwell.

Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Martin, R., Van Ess-Dykema, C., & Meteer, M. (2000). Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational linguistics*, 26(3), 339-373.

Strawson, P. F. (1964). Intention and convention in speech acts. *The philosophical review*, 439-460.

Tavafi, M., Mehdad, Y., Joty, S., Carenini, G., & Ng, R. (2013, August). Dialogue act recognition in synchronous and asynchronous conversations. In *Proceedings of the SIGDIAL 2013 Conference* (pp. 117-121).

Taylor, P., King, S., Isard, S., & Wright, H. (1998). Intonation and dialog context as constraints for speech recognition. *Language and Speech*, 41(3-4), 493-512.

Ulrich, J., Murray, G., & Carenini, G. (2008). A publicly available annotated corpus for supervised email summarization. In *Proc. of aai email-2008 workshop, chicago, USA*.

Wierzbicka, A., & Pragmatics, C. C. (1991). *The semantics of human interaction*. Mouton De Gruyter.

Wittgenstein, L. (1953). *Philosophical investigations*. London, Basic Black. Chicago.

Wayne, D. (2014). Implicature. The Stanford Encyclopedia of Philosophy. Edward N. Zalta (ed). Retrieved from <https://plato.stanford.edu/archives/fall2014/entries/implicature/>.