Chapter 9

Speech acts

Stephen C. Levinson

Abstract

The essential insight of speech act theory was that when we use language, we perform actions – in a more modern parlance, core language use in interaction is a form of joint action. Over the last thirty years, speech acts have been relatively neglected in linguistic pragmatics, although important work has been done especially in conversation analysis. Here we review the core issues – the identifying characteristics, the degree of universality, the problem of multiple functions, and the puzzle of speech act recognition. Special attention is drawn to the role of conversation structure, probabilistic linguistic cues and plan or sequence inference in speech act recognition, and to the centrality of deep recursive structures in sequences of speech acts in conversation.

1. Introduction

The concept of speech act is one of the most important notions in pragmatics. The term denotes the sense in which utterances are not mere meaning-bearers, but rather in a very real sense do things, that is, perform actions. This is clear from a number of simple observations:

(i) utterances in conversation (and that is the only kind considered in this article) respond not to the shape or meaning of what was said, but to the underlying 'point' or action performed by the prior turn at talk, which might have been expressed in any number of ways;

(ii) utterances often have non-verbal counterparts (cf. waving to saying hello; bidding at auction by hand or voice);

(iii) utterances interdigitate with non-verbal actions in action sequences (cf. ordering a sandwich in a service encounter);

(iv) utterances have real-world consequences just like non-verbal actions (a $1,000 bid at an auction commits you to paying; saying you have nothing to declare in an airport can get you a big fine).

These actions are on a different ontological plane than the actions of the vocal organs in speech, which of course activate the motor cortex just as much as reaching for a glass – speech acts are more like moves in chess, whose meanings are circumscribed by rules and expectations. Trying to understand how utterances can have these abstract action-like properties, how they are coded linguistically and how we recognize them are some of the core issues in this domain.

Despite the fact that speech acts are clearly central to an understanding of language use, they have been largely off the linguistics agenda since the 1980s. As is often the case in science, research on speech acts boomed for a little over a decade (in the 1970s and 1980s), and then went out of fashion without the most fundamental issues being resolved at all. Amongst these unanswered questions are: How many types are there, and are they universal or culturally specific? How are they expressed in language? And how are they recognized or attributed in actual language use? These questions are addressed in sections 3-9 below.

2. A brief history of the concepts leading to the current state of the art

In philosophy of language during the 1930s and 40s the picture theory of meaning, and the broader correspondence theory of truth, began to be challenged by theories of language use being developed by the later Ludwig Wittgenstein at Cambridge and the 'ordinary language' philosophers like Gilbert Ryle and J.L. Austin in Oxford. It is Austin (Austin 1962) who is usually credited with the first developed theory of speech acts, although his influential lectures 'How to do things with words' were not published until 1962 after his death. Austin took the view that philosophy of language had wrongly concentrated on statements, or even just propositions, and in doing so had lost track of what language is mostly used for. Rather, he claimed, utterances attempt to do things, and just like other actions can fail for a range of reasons. He catalogued the kinds of actions performed, by noting that most speech acts (however colloquially expressed) can be paraphrased in the normal form "I hereby $V_{performative}$" where a delimited set of verb like *order, promise, warn, congratulate* could appear. He also classified the reasons for success or failure of speech acts, dubbed 'felicity conditions', noting that they often require appropriate subjective states (later called 'sincerity conditions' by Searle) as well as appropriate circumstances (Searle's 'preparatory conditions') . In this sort of way all the reasons for my bid at Christie's for a Picasso not succeeding (I am not a registered bidder, lack the funds, don't succeed in getting the attention of the auctioneer, etc.) can be spelled out. Speech acts can be understood on the analogy of ceremonies, like marriage or toasting the monarch's health – in the same sort of way they are conventional arrangements for creating new states of affairs, and consequently are in principle open-ended in kind. Austin went on to notice that these success conditions not only parallel truth-conditions, but actually subsume them; statements are therefore just a special class of speech acts with sincerity conditions of belief and presuppositions or preparatory conditions that must also be met. He also went to some pains to clarify all the different senses in which actions could be said to be performed by utterances: the 'locutionary act' is the saying of the words with the intended meanings, the 'illocutionary act (or force)' is the speech act proper (ordering, advising, warning, etc.), and the 'perlocutionary act' is the further act or consequences that are context-specific and not part of the specific conventions

invoked (e.g. by asking your advice I might flatter you). Austin also developed a number of notions whose importance was not immediately realized – for example the concept of 'uptake' (the ratified receipt and recognition by a recipient).

Austin's work was influentially systematized by John Searle, (Searle 1969) who connected the theory to sociology and jurisprudence on the one hand (speech acts are built as constitutive rules, whereby doing X counts as constituting a new state of affairs, like scoring a goal, or being guilty of a specific crime), and to linguistics on the other hand. Noting, following Hare (1952) , that the same propositional content could occur across speech acts (as in 'Pass the exam', 'Did you pass the exam?', 'Good luck with the exam'), he added a 'propositional content condition', so that the felicity conditions together now effectively defined the speech act. He went on to suggest that an exhaustive typology of speech acts could be arrived at by clustering types of felicity conditions, so that there can be seen to be just five main types: representatives (statements and the like), directives (questions, requests, orders), commissives (threats, promises, offers), expressives (thanking, apologizing, congratulating, etc.), and declarations (like christening, declaring war, firing, etc. which rely on elaborate institutional backgrounds). Searle's theory was well articulated and proved attractive to linguists, as recounted below.

Meanwhile, other philosophers took a more psychological view of language use, chief among them Grice and Strawson, who both thought that speech acts should be thought about as specific classes of intention, e.g. intentions to cause beliefs in addresses, or intentions to get them to do things. Grice (1957; 1975) reconstructed the notion of meaning along these lines, and characterized the use of language in conversation as guided by rational action between partners. Although he never laid this out in print, it is clear that he thought that felicity conditions simply follow from the specific classes of intention: If I want to get you to pass the water by saying "Could you pass the water?", it would simply be irrational if I didn't want the water, if the water is not in your reach, if you are deaf or otherwise preoccupied. This intentional perspective was followed up by work in natural language processing that related speech act recognition to plan recognition (see section 7 below.).

During the period of Generative Semantics, linguists became increasingly interested in language usage and how sentences might encode aspects of the contexts in which they are used. Searle and other theorists had not concentrated on the actualities of speech act coding, presuming instead that illocutionary force is coded in the major sentence-types (imperatives, interrogatives and declaratives) and in the explicit performative verbs when so used – these would be the 'literal illocutionary forces' of utterances. But as any practical grammarian of English or other languages knows, in fact one has to learn idiomatic means of expressing speech acts. Gordon & Lakoff (1971) noted for the first time that 'indirect speech acts' could also routinely be expressed by querying or stating a felicity condition: "Do you need that pencil?", "Could I have that pencil", "Is that your pencil?", "I'd like that pencil" all query or state a precondition on requesting. They also noted that adverbials like *please* or *frankly* might force a particular speech act reading (as in "Please could we begin on time?"). There followed a large literature on indirect speech acts, investigating the forms used especially for requests across cultures, the psychological processing (indirect speech acts seemed to be processed without any complex detour through a literal meaning) and the politeness reasons for the mismatch between direct and indirect speech act coding. By the end of the 1980s, however, linguistic interests had moved largely elsewhere.

Meanwhile, a completely different approach, unrelated to the linguistic and philosophical traditions, was being taken in sociology, where the empirical study of conversation was being born in the late 1960s and early 1970s. Unencumbered by theory, the conversation analysts (Harvey Sacks, Manny Schegloff and Gail Jefferson especially) were observing all sorts of fundamental organizations for interactive language use: turn-taking, repair and sequence organization (see e.g. Schegloff & Sacks 1973, Schegloff 2007, and this volume). In doing so they were finding speech acts that had no vernacular names, no associated performative verbs or (it seems) special markings, for example pre-closings (e.g. the exchange of *well*s before goodbyes in phone calls), assessments (evaluations of shared events or things), repair-initiators (like *excuse me?*), pre-invitations (*What are you doing on Friday night?*), and so forth. Such *actions* (as the conversational analysts call them, treated here as equivalent to the notion of speech act) can only be understood against the background of sequential

position – that is, where they come with relation to prior or following turns. Despite the fact that many observations have now accrued about the sets of actions and their sequential placement, little systematic theory about actions has emerged from this work (for a survey see Levinson 2013a, Drew 2013).

Although this brief review cannot do justice to the extensive work that has been done in the different disciplines interested in speech acts (linguistics, psychology, conversation analysis) (see Levinson 1983, 2013a), it serves as a pointer to the state of the art. There is general acceptance of the importance of the subject, but little recent research that advances our understanding of the fundamental questions.

3. The essential insight and the leading issues

In contrast to the emphasis in modern linguistics on language as a device for an endless sound-meaning correspondence, J.L. Austin's core insight was that the central function of language is not to deliver meanings but to deliver speech acts. For the core ecological niche for language, and still its primary use and the locus of its acquisition, is conversation. Each of us produces on average perhaps 16,000 words and 1200 turns at talk a day – and each turn delivers a speech act: all in all we are participating in exchanges with something like 5,000 speech act moves a day. In order to respond on time (within the c. 200 ms allowed by the turn-taking system; Stivers et al. 2009) we need to decode or attribute speech acts at lightning speed, because it is the illocutionary force, not the meaning, that we primarily respond to. One of the central puzzles is that speech acts are not for the most part simply or directly coded in the linguistic form: for example, *Where are you going?* could be an idle question, or a challenge, or a reprimand, or a prelude (a pre-) to a request for a ride or to an offer to give you a ride, and the relevant response depends on the correct attribution. How then are speech acts recognized in the tight time-frame allowed? Is there a finite list of possible action types, or can they be created *de novo*? Further, as just illustrated, an utterance or turn can perform more than

one action simultaneously: in asking a question (*Where are you going?*) the speaker could also be transparently performing a pre-request in such a way that the addressee can make an offer in next turn (*Downtown, would you like a ride?*). How many acts can be performed at once?

These then are the central puzzles in this area, to be taken up below. Faced with these difficulties, to which current research yields no definitive answers, it is tempting for linguistic theory to simply hand over the can of worms to some other discipline (conversation analysis for example) as e.g. Bierwisch (1980) recommended. However, as discussed below (section 8.), there is a substantial intersection of speech acts and linguistic structure, which makes the topic of central importance for e.g. the study of syntax. Usage and structure in fact go hand in hand.

4. The nature of the beast: identifying speech acts

In this section we consider the problem of identifying and cataloguing speech acts given some problematic properties, like their implicit character and non-one-to-one mapping onto utterances.

There are four (three basic and one related) approaches to identifying or characterizing speech acts. First, one could rely on natural metalanguage, as in English *offer, request, invitation, greeting,* and so on. Austin's own tack here, recollect, was to do the lexicography of performative verbs (*I hereby declare/ choose/ delegate/ promise/ undertake/ bequeath …*). But there are many reasons to distrust natural metalanguage. Many speech acts have no vernacular names (such actions as pre-invitations, continuers, repair-initiators and the like), as discovered by the conversational analysts. In addition, while written languages often have large metalanguage resources of this kind, unwritten ones often do not, and they may have speech acts alien to us. So natural language terms are a poor guide.

A second approach is the use of felicity conditions to characterize speech acts, as in classical speech act theory. A problem here is that taken as necessary conditions jointly sufficient to define speech

acts, it is hard to specify them right. Thus the conditions for genuine information-seeking questions, exam questions, questions checking facts, and questions used in repair will all be subtly different – they form a loose family of speech act types not easily captured by a definitive checklist of conditions.

A third approach favoured by conversation analysts is to use the character of responses to identify prior actions. For example, if a range of utterances X-Y-Z are all immediately responded to by fellow interactants passing the speaker something, then *prima facie* X-Y-Z are requests. The observation is that many speech acts come in pairs ('adjacency pairs'), with an initiating action having a characteristic response, as in greetings followed by greetings, offers by acceptances (or declinings), questions by answers, and so forth (Schegloff 2007, Stivers 2013). Thus if one can independently characterize the responding action, one can type the eliciting action. Conversation analysts argue that this is how we check that we are understood – we expect a response of a certain type. Consider, the following example, where the response marked by *thanks* and excuses suggests that *for B*, A's turn appears to have been an offer, though that is not obvious from its structure or content:

<1> [Levinson 1983:335]

 A: She says you might want that dress I bought, I don't know whether you do

 B: Oh thanks, well, let me see I really have lots of dresses

A fourth, related approach is to appreciate that an utterance gets parts of its identity from the sequential position it occupies. Consider the following tokens of the utterance *Okay*, each doing entirely different things (labeled here with the action codings used in conversation analysis – see Schegloff 2007):

<2> (a) *N: Y'wanna drink?*
        C: Yeah
        N: Okay (← Sequence-closing third)

   (b) *C: Okay* (← Preclosing)
       R: Okay [Bye
       C: Bye

   (c) *B: How are you?*
        A: Okay (← Answer)

One aspect of speech acts thus highlighted is that they are necessarily interactional in character. Consider a proposal (say about going for a walk together) – for success, the action depends on the uptake: it takes two to tango. This is a fundamental aspect of speech acts neglected in Searlian analysis – almost all speech acts are joint actions (Clark 1996).[1]

Most analysis actually makes use of all four of these different kinds of identifying properties, trading on our vernacular terminology, trying to tighten it up by defining criteria, considering how participants themselves respond to utterances, and noting how utterances play different roles depending on their positioning vis-à-vis other speech acts.

5.   The inventory and its universality

A natural question is how many kinds of speech acts are there? The question presumes a level of abstraction away from the specific propositional content, which may of course be unique: it's a question about how many types of illocutionary force exist. Austin suggested an open-ended list, convention-based, so cultural in nature. In contrast Grice (in unpublished work (Grice 1973); see also Schiffer (1972)) had suggested that complex speech act types could be built up from the two propositional attitudes of wanting and judging. His target was the 'moods' expressed in the major sentence types, namely declaratives, imperatives and interrogatives. Most languages grammatically code at least two of these, which could be taken as a hint of a cross-cultural core of basic speech

---

[1] A possible exception are 'outlouds' or 'response cries' like private exclamations (E. Goffman 1978), which may be produced with or without an audience, but by definition without an addressee.

acts. However it is moot whether these forms really code speech acts since they are in practice used for diverse action types, while other minor sentence types like English expressives more directly code for force (see section 6.) But the idea that speech acts fall into *classes of intention* is persistent (see e.g. Tomasello 2008).

Searle, taking an intermediate position, has argued that there are in fact just five large classes of things one can do with language – five major speech act types. The classification uses three parameters: the 'essential conditions' (Searle's term for the intentional goal), the sincerity conditions, and 'direction of fit' (whether the words copy the world as in statements or the world copies the words as in promises). Searle's classes are *representatives* (assertion-like), *directives* (questioning, requesting, etc.), *commissives* (promising, threatening, offering), expressives (thanking, apologizing, etc.) and *declarations* (blessing, christening, etc., which rely on special institutional backgrounds).

Searle's classification cannot however be exhaustive. First it fails to accommodate many of the actions noted by the conversation analysts (e.g. the continuer *hmhm*, the pre-s, the repair initiators and the repair responses, and so forth). Second, it is culture-bound. Consider the following exchange simplified and in translation from the language Yélî Dnye (Levinson 2005):

<3> *A: He's yelling into a bit of bush-knife*

   B: He's yelling under a mangrove tree

This is an adjacency pair of a special kind peculiar to this matrilineal Papuan culture, in which men make jokes by alluding to some unfortunate accident or event that befell the other man's father-in-law, to which the response must be immediate and in kind (B's father-in-law killed his wife and then himself with a bush-knife, while A's father-in-law died falling from a mangrove tree; they are ostensibly commenting on a man yelling down a megaphone). These utterances are paired father-in-law jokes and they don't describe states of affairs or express the feelings of the speakers or otherwise fall within Searle's taxonomy. In addition, Searle's classification is of course a higher-order

grouping of types, so it will not help us understand the specifics of action and response in conversation.

Austin or Searle's arm-chair classifications are based on intuitions about salient types of speech acts. These are nearly always first parts of (base) adjacency pairs (see Schegloff 2007, this volume) – that is the initiating actions (like questions, offers, invitations) to which responses are due (even then, many such initiatory actions have proved relatively unavailable to intuition, like repair initiators, continuers, assessments and the like). But the actions that lead in to these initiators (e.g. pre-announcements, pre-closings) or the responses themselves (e.g. answers, agreements, continuers, counter-offers), or the actions that interpose between first part and second (e.g. clarification questions) escape proper treatment in classical speech act theory. Consider (with arrowed action labeling):

<4> (Terasaki 1976]

1. D: Didju hear the terrible news?   ← pre-pre-announcement
2. R: No. What   ← answer + go-ahead
3. D: Y'know your Grandpa Bill's brother Dan? ← pre-announcement
4. R: He died.   ← guess
5. D: Yeah   ← confirmation

Describing line (1) as a question would miss its basic function, namely to check whether a news announcement should be made; line (2) makes clear it should (note the *what*); line (3) sets up the topic of the announcement in such a way than no announcement proves necessary, for the recipient guesses in line (4). Thus although (1) and (2) could be said to be questions that is not their main function, which is as preliminaries to an announcement (see Levinson 1983:345ff and Schegloff 2007 for more on pre-s). Recollect as mentioned above that conversation analysts have emphasized that it is the character of the response, or the locus in a sequence, that plays a major role in giving speech acts their identities.

To return to the central questions of this section: Is there a finite set of speech act types, and if so how big is it? The answers are that we really don't know. Is the set universal in character? Not in the

sense that all speech acts are pan-cultural (witness Yélî Dnye father-in-law jokes, or any of the institutionally circumscribed acts like finding guilty, proposing toasts, declaring war, etc.), but it is an open question as to whether there is a pan-cultural core with such plausibly general functions as telling, questioning, requesting, greeting, agreeing, or initiating repair.

6. The multiple action problem

One particularly troubling feature of the mapping of speech acts onto utterances is that such a mapping is not necessarily, or even mostly, 1:1. Sometimes turns at talk have more than one constructional component, and each part can perform an action, as in <4> above and (5) below:

<5> *A: How are you=*
    B: =Fine. How are you?   ← answer and question

But often a single constructional unit (whether or not it exhausts the turn) can do more than one action (as in <4> where *Didju hear the terrible news?* might be said to be a question, but carries with it the obligation to tell the news, conditional on the answer 'no'). Consider the following example from a verbal tussle between a mother and her 14year old daughter Virginia wanting more allowance or pocket money:

<6> Virginia
    VIR: But- you know, you have to have enough mo:ney¿
           I think ten dollars'ud be good.   ←Proposal
           (0.4)
    MOM: ˙hhh Ten dollahs a week?   ← Repair-I, Q, Pre-challenge
    VIR: Mm hm.   ← Repair,   A, Go-ahead
    MOM: Just to throw away?   ← Repair-I, Q , Challenge and Pre- Rejection
           (0.5)
    VIR: Not to throw away, to spe:nd.   ←Repair,   A, defense

Viriginia's proposal is responded to by a question-like response, which has the form of an other-initiator of repair or OIR (i.e. is initiated by the responder, seeking repair on the prior turn). But it is a

prosodically incredulous OIR, adumbrating an upcoming challenge (call it a pre-challenge), which after a go-ahead, is duly delivered (*Just to throw away?*) but again in the form of a question inviting repair. That extreme-formulation of the question in turn pre-figures a rejection (call the turn then a pre-rejection), and gets a defense. And so forth. But now notice we have multiple layers of function for each turn – up to four actions packed into the one sub-clausal turn in *Just to throw away*!

The question that arises is whether there is any limit to the number of actions that a single turn can bear. Notice that some of these might merely be a matter of granularity of description, e.g. a special kind of question is often used to ask for repair. But that is not the kind of relation between the question and say the challenge: notice how the response deals with both. The literature acknowledges the existence of turns performing two actions: on one account, a 'literal speech act' is used to deliver an 'indirect speech act' (Searle 1975), and conversation analysts talk about one action being the vehicle for one other action (Schegloff 2007). But there is no explanation for turns that perform three or more actions (see however the suggestions in terms of plan-reconstruction at the end of the next section).

7. Bottom-up and top-down inference in speech act recognition and attribution

Speech acts, it has been suggested, are not easy to individuate or identify, are not known to come from a finite or universal set, and can be laminated one on top of another. These are problematic properties. But an even greater problem is how they are recognized (more properly attributed[2]) under the extraordinary time pressures of spoken conversation (or any other interactional use of language). Here we concentrate on the comprehension problem. As already mentioned, on average across languages the gaps between turns are on the order of 200-300 ms (Stivers et al. 2009, Levinson & Torreira in press). Given that the fastest response from conception to word takes 600+ ms (Levelt 1989; responses of any complexity, e.g. three or more words, take 900-1500 ms or more

---

[2] 'Recognition' presupposes correct attribution that matches speaker intent, but since we are interested in the comprehension process which will include occasional misattributions, 'attribution' is the more accurate term.

to prepare), it is clear that speakers in conversation predict the end of the incoming turn in order to launch their own response on time. But that response must 'type' the incoming turn as, e.g. a question, request, statement, before it has finished in order to compose the relevant response and launch it so it comes out on time. Probably this is done on average about half way through the incoming turn(see Magyari et al., 2014).

This makes the speed at which speech acts are attributed appear quite miraculous. For, as already made clear, the coding of speech acts is for the most part not directly marked: Most syntactic forms, even whole constructions like *Why don't you ....*, are multi-duty (*why don't you* turns out to code proposals, advice, invitations, and complaints, while *Do you want* codes requests, invitations, offers, and so forth; Couper–Kuhlen 2010).

Speech act recognition is similar to any perception problem, where pattern has to be discerned and categorized out of noise. Both 'bottom up' information (in the signal) and 'top down' information (expected categories) are usually involved, and the noisier the channel the greater the role for 'top down' factors. Let us consider them in turn. Bottom-up information is whatever clues to speech act type can be found directly coded or cued in the signal, by lexical choice, construction, or prosody. Given the turn-taking facts, it is clear that signals early on in a turn are going to be more important than signals at the end of turns, since by then the choice of response must have already been made. This suggests that effective cues will be 'front loaded', coming early in the turn (see Levinson 2013a). Here the cross-linguistic facts are curious. Take the grammar of interrogatives, associated (though not exclusively) with the illocutionary force of questioning. First, *wh-* or content interrogatives are only grammatically initial in about one third of languages (Dryer 2011b); however, this is the dominant single strategy since the alternative positions are various, and Dryer notes that only "a few languages exhibit at least a weak tendency to place interrogative phrases at the end of sentences" (he mentions two out of a sample of 900 languages). These facts are in line with the 'front loading' prediction from the psycholinguistic facts, but only as a tendency. The prediction would be that

languages with late (right-located) *wh-* words would have developed compensatory cues like prosody or particles positioned earlier in the clause.

Second, take polar (yes/no) questions (Dryer 2011a). The commonest coding strategy (60% of languages) is by particle, and of these about 30% are in initial or second position; however the commonest position of particles is final (50% of all particle types). It is worth noting however that 30% of languages have no lexical or morphosyntactic coding at all for polar questions, relying solely on intonation or prosody. These facts do not seem to be in line with the 'front loading' expectation. Further light is thrown on these issues by studies of usage in corpora. In a study of 10 languages, we found that those sentence-final particles are omitted or absent 40% of the time in Lao, and 70% in Korean (Enfield et al. 2010); two of the languages lacked any coding (including prosodic); and morphosyntactic coding as in English inversion is also mostly omitted. One can conclude that polar-question marking must carry low functional load, wherever it is located. These usage studies also showed that interrogatives (whether content or polar) only perform the function of seeking new information about 30% of the time; around 40% of the them are involved in repair or checking or confirming just-given information, and the remaining 30% perform many different functions including offers, requests and so on.

To summarize so far: there is no one-to-one match of form to function. Even where apparently dedicated morphosyntactic machinery exists to code speech acts (as in interrogatives), the coding may be omitted: about 60-70% (in various corpora) of English polar questions are unmarked declaratives in form, and do not carry rising intonation (Geluykens 1988). Crosslinguistically the tendency is for two thirds or more of all questions (in a broad sense) to be polar questions (unpublished data from Stivers et al. 2009). Even though *wh-* or content questions would seem to require a *wh-*form, this is not necessarily true; many languages have indefinite quantifiers that double as interrogative words, and many allow gaps to code the variable (as in *John is going to _?* instead of *Where is John going?*).

There are then distinct limits to the bottom-up coding and inference of speech act force. Nevertheless, some detailed studies suggest that underlying the apparent many-to-many correspondences between utterance forms and speech acts there might be a clockwork system. For example, in a study of requests in English telephone calls, it was found that the *Can you/Could you/Would you…* forms are used for requests where the speaker has clear rights or entitlements and knows what the request would involve; where the entitlements are low and the contingencies involved less clear, the *I wonder if* form is preferred (Drew and Curl 2008). This suggests that where multiple forms are available, they may each carry subtly different presuppositions about background conditions.

Nevertheless, it is more likely that the cues to illocutionary force are multiple and probabilistic in character. Indeed, there is now considerable work in natural language processing (NLP) that seems to show this. This work takes speech corpora, usually from task-oriented dialogues, and tags them by hand with a very constrained set of speech act categories that seems to reflect the functions in each particular corpus. Machine-learning algorithms are then trained on a sub-corpus, inducing the association between surface cues - lexical items, phrases or intonation - and the pre-coded tags. The algorithm is then let loose on the rest of the corpus to see how well it emulates the human tagging. So for example, it was found that 'assessments' (value judgements like "That was great" that usually call for a response in kind) have quite restricted elements (Goodwin 1996): *that* as subject in 80% of cases, intensifiers *really* or *pretty* and adjectives drawn from a short list including *great, good, nice, wonderful …* etc. (Jurafsky 2004). So a combination (an unstructured list) of surface cues may be a crude but very effective trigger for speech act categorization: the chances of being an assessment given just one cue like *really* might be low, but in combination with *that* and *great* may be greatly increased. This would be just the kind of low-level associative process that could rapidly deliver probabilities of speech act assignment in comprehension, and since these cues are distributed throughout the turn, an incoming turn could be incrementally classified with increasing certainty.

Turning to top-down information, this includes all the accumulated contextual and sequential information that forms the niche for the incoming turn. For example, in service encounters, the goals for speaker and addressee will be largely pre-set, so that an utterance like *Do you have coffee to go?* can be understood directly as a request. In free conversation, though, the context is usually more local. One factor of constant relevance is the current state of the common ground between participants. We noted earlier that polar questions in English and many other languages are typically unmarked, and thus have the shape and often the prosody of declaratives. How then can they be understood as questions? As Labov and Fanshel (1977) pointed out, the recognition is done on the basis of knowledge asymmetry: thus *You're hungry* is likely to be understood as a question, while *You're smart* is likely to be interpreted as a compliment. Statements about what the other knows best are candidate questions, and this explains how a fifth of languages can do without any lexical or morphosyntactic marking of polar questions (prosody may often help of course, but in some languages it seems never to play this role; see e.g. Levinson (2010) on Yélî Dnye*, or Dryer (2011a) on Chalcatongo Mixtec). Epistemic asymmetry or symmetry is such a strong indicator that it can over-rule interrogative marking: thus *Isn't it a beautiful day* is not likely to be interpreted as a question, since we can all be presumed to have access to the weather. Heritage (2012) argues that epistemic status trumps question marking in all cases.

A second always relevant factor is sequential location in the sequence of turns. The power of sequential location to map illocutionary force onto utterances can be appreciated from a number of angles. Consider as a limiting case silence, where there is literally no signal, yet the silence can imply a response, as in the following example where the two second silence is taken to imply "no" and functions to block a forthcoming request:

<7> [Levinson 1983:320]

    C: I was wondering would you be in your office on Monday (.) by any chance?

    (2.0) ←(Pre-Request won't go through)

    C: Probably not

The inference relies on the 'conditional relevance' of a second pair part and on the principle that dis-preferred responses are typically delayed or mitigated. Another way to appreciate the power of sequence to attribute speech act force is to consider cases where ambiguities arise, as in the following example <8> where the arrowed turn is ambiguous (Schegloff 1988). It could be a straight question, or it could be a pre-announcement, that is an offer to tell conditional on the recipient indicating that he doesn't know the indicated news. Note that the question force is not the 'literal force' (a question about knowledge), but a question about who is going. Pre-announcements often have this form (cf. *Do you know the joke about the plumber?*) and the pre-announcement reading is encouraged by the context, where Russ had produced a pre-announcement just before in the first line, and Mom could be reciprocating in kind. The ambiguity comes about because both readings are salient in the context.

<8>
Russ: I know where you're goin',
Mom: Where?
Russ: To the uh (eight grade)=
Mom: = Yeah. Right.
Mom: Do you know who's going to that meeting? ← (speech-act ambiguous turn)
Russ: Who?
Mom: I don't kno:w!
Russ: O::h probably Missiz McOwen en ....

A related type of high-level information can also be brought to bear on the interpretation of a turn, namely an assessment of how the turn fits into the likely goal-structure or plan of the speaker. For this is the inference schema we use to understand any sequence of actions: if you are sitting opposite and grasp your mug and lift it up, I'll expect you to put it to your mouth and take a drink. The sub-actions I see (grasping the mug, lifting it) are preconditions to the action I infer (taking a sip), and seeing the initial parts I can make the metonymic inference to the whole. Interestingly, the same pattern of inference works for speech acts. Consider the following service encounter (example <9>), where a precondition to buying pecan danish pastries is queried, and the seller responds both to the question and the underlying request.

<9> [Merritt 1976]
  C: *Do you have pecan danish today?*   ← Q + (pre-)Request
  S: *Yes we do.*   ← Answer
   Would you like one of those?   ← deals with request

Notice however that no request has been issued, so how exactly does this work? Consider the analysis sketched in <10>, in terms of customer C's plans and the seller S's reconstruction of them from the first utterance in the sequence. From *Do you have pecan danish today* the seller can infer that this is a precondition on asking for some, therefore the request is likely to follow – given which the seller can truncate the sequence as she does, by responding to the foreseeable forthcoming request (in dotted box). It is this projected request that gives *Do you have pecan danish today* its pre-request flavor; in this way speech acts can acquire multiple actions mapped onto one turn by virtue of projectable next actions.

<10> Plans underlying speech acts in <9>
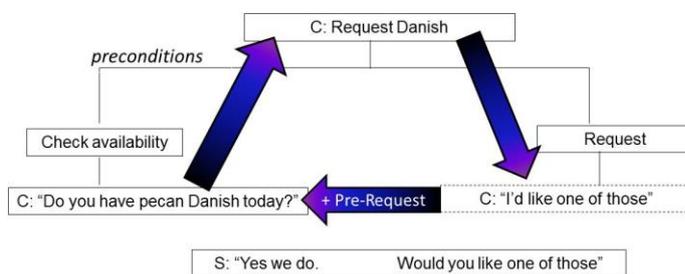
Notice this account explains why mentioning a felicity condition on a speech act is one way of performing that speech act (this is the classical theory of 'indirect speech acts', as in Searle (1975)). But it has much wider application. Consider the telephone exchange in <11>: the caller C in line 3 queries what the recipient is doing, which is a potential prequel to an invitation. The response in line 4 not only answers the query but at the same makes clear that there is no impediment to an invitation, thus projecting an acceptance. The lamination of actions throughout this sequence is straightforwardly explicable in terms of current action plus foreseeable next action, as sketched in <12>.

<11>

1. C: Hi
2. R: Hi
3. C: Whatcha doin'.   ← Q+ Pre-invitation
4. R: Not much.   ← A+ Go-ahead for invitation
5. C: Y'wanna drink?   ← Q+Invitation
6. R: Yeah   ←A+Acceptance
7. C: Okay.

<12> Plans underlying speech acts in <11>
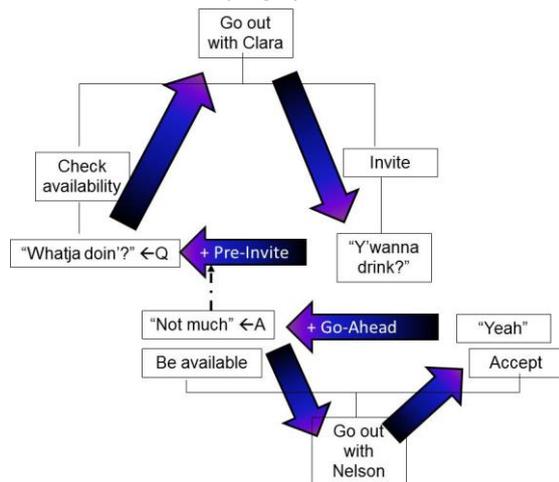
The virtues of this mode of analysis become especially clear when one considers cases like the following where the main actions are projected, but never actually performed.

<13> [Schegloff 2007:64]

D: 'hh My ca:r is sta::lled   ← Announcement of problem
((5 lines omitted))
 I don' know if it's po:ssible, but (0.2 hhh)   ← Unvoiced Request for ride
see I haveta open up the ba:nk.hh
(0.3) a:t uh: Brentwood?hh=
M: =Yeah:- en I know you want- (.)   ← Unvoiced Rejection
en I whoa- (.) en I would,
but I've gotta leave in about five min(h)utes. (hheh)

Here there is no feasible 'indirect speech act' in terms of classical felicity conditions: there is rather

an indication of a predicament which would have an obvious solution, while the recipient produces

an account for why the obvious solution cannot be performed. In the same sort of way, in example

<6>, Mom's *Just to throw away?* performs four actions, as question, repair-initiator, challenge, and

pre-rejection because it is transparent that Mom intends to resist Virginia's claim for more weekly

pocket money by countering Viriginia's every move. Neither indirect speech act theory nor the conversation analyst's notion of one action being the 'vehicle' for another (as in Schegloff (2007)) can explain this kind of quadruple depth of speech act lamination on a single turn.

Plan-reconstruction as an account of speech act comprehension was first advanced by Allen (1979), Cohen and Perrault (1979) and applied to the problem of indirect speech acts by Allen and Perrault (1980); (see also Clark 1979, Levinson 1981). These approaches in classical Artificial Intelligence style make use of the heavily intentional approach favored by Grice and reviewed in section 1, cranking through a calculus of desire and belief to arrive at a final 'indirect speech act' (Cohen et al. 1990). The insights can be understood, however, in a slightly different way, in terms of an utterance being designed to reveal, variously, the whole or part of the iceberg of underlying interactional goals, where projectable next turns serve to laminate one or more 'indirect speech acts' onto the current turn.

Both bottom-up cues, which may be just probabilistic associations of linguistic features and speech acts, together with top-down factors like the role of sequence, epistemic asymmetries and plan-attribution, almost certainly play a role together in speech act comprehension. Curiously, cases where interlocutors misunderstand one another as in <8> are vanishingly rare. But there is no complete model of how these various kinds of information come together in action attribution.

8. Syntax, sentence types and the grammar of speech acts

We return now to the grammar of speech acts. We've noted that in general there is no one-to-one mapping between form and function. This is especially true of the 'big three' sentence types, declarative, interrogative and imperative, which are probably best seen as carrying a very general semantics (e.g. a *wh*-interrogative expresses an open proposition with a blank constituent, which is why the same form may double as an indefinite expression in many languages). However, as discussed above under the rubric of cues, there can be many surface elements that will help to

narrow down an illocutionary force. There are for examples adverbs like *please* that unambiguously mark requests or pleadings, adverbs like *obviously* or *frankly* that mark statements (Gordon and Lakoff 1971), and interjections like *Wow, My God* that mark exclamations. In addition there are minor sentence types that are indeed specialized for illocutionary force (Sadok and Zwicky 1985). A classic case are exclamatives, where English has rich specialized constructional resources as in *What a beautiful day!, That it should come to this!, Why, if it isn't the trouble maker!, You and your linguistics!, Of all the stupid things to do!, To think I nearly won a medal!* (well described in grammars like Quirk et al. 1989). Exclamatives are a category of some typological interest (see Michaelis 2001, who defines them semantically and finds them often coded in quasi-interrogative or topic constructions or NP complements). Similarly English codes wishes as optatives (*If only I'd done it, May the best man win, Oh to be in England*), and suggestions or proposals in special forms (*How about joining us?, What if you came earlier?, Let's go, Why not have a drink?)*. Many other languages have their own specialized forms for warnings, blessings, and the like. Unfortunately, studies of the usages of these forms are still few and far between, so we cannot be sure they are as specialized in usage as the grammars suggest – but it is an important subject for future research.

9.   Conclusions – the centrality of speech acts

The central function of language, it has been argued, is to deliver speech acts (Searle 1972). The rest of the linguistic apparatus, with all of its complex syntax and propositional structure, is there to serve this purpose. For speech acts are the coin of conversation, and conversation the core niche for language use and acquisition. A retort might be that the central function of missiles is to target explosives, but this doesn't help one understand much about the inner complex engineering of a missile – the outer function can be remote from design details, partly because there may be innumerable different engineering solutions that would answer the same function. Linguistics then would be effectively autonomous from the study of speech acts. What has been argued here, however, is that such a disjunction is unlikely to be tenable. First, language design has to accommodate to the tight constraints of conversation, so that speech acts have to be decoded early

partly from bottom-up aspects of the signal – hence constructions of many different kinds serve this purpose, if often in a non-deterministic way. Second, the very clausal structure of language is almost certainly due to the tight turn constraints into which sentences must fit, where each turn must deliver at least one speech act. Third, whatever ones' views on the origin of language, short turns delivering speech acts was almost certainly a design feature of protolanguage – languages have evolved within this ecological niche, spinning complexity in the tight confines of the turn.

Another way to appreciate the centrality of speech acts in language design is to appreciate how many of the features we think of as most intimately connected to language structure are actually also exhibited in the sequential organization of speech acts. Consider recursion, argued by Chomsky (2007, 2010) to be the most central design feature exclusive to language. Now consider that the clearest type of recursion, namely center-embedding, is restricted in language to just two, occasionally three, levels of nesting. Karlsson (2007) found no examples of triple embedding in huge corpora, and just 13 in the whole history of Western literature; for spoken language, the limit is two. Since small numbers of centre-embeddings can easily be modeled with a finite state device, there is poor evidence for the need for phrase-structure grammars here. Yet center-embedding within discourse shows none of these limits, and is sufficiently multiple and routine to provide a much better basis for escalation to phrase structure grammars. Here is a simple example of one degree center-embedding:

<14> [Merritt 1976]

A: ⌈ May I have a bottle of Mich?
B: | ⌈ Are you twenty one?
A: | ⌊ $^{0.1}$ No
B: ⌊ No

Since this can be recursively elaborated, we could express the indefinite recursion by the rule: Q&A →Q (Q&A) A (Levinson 1981, 2006; Koschmann 2010). The following shows an example with

degree three internal embedding (each level numbered), a level exceeding all syntactic embedding in spoken languages (the speech acts, or adjacency pairs, here relevant are request+compliance, question+answer, and two repair-initiator+repairs).

<15> [Merritt 1976]

```
S: Next                              ←   Request to order
0C: Roast beef on rye                ←   Order
1S:   | Mustard or mayonnaise?       ←   Q1
2C:   |  | Excuse me?                ←   Repair Initiator (RI₁)
3S:   |  |  | What?                  ←   Repair on RI
3C:₁  |2 |  |3 Excuse me?
2:    |  |  | I didn't hear what you said   ←   RI₂
1S:   |  | Do you want mustard or mayonnaise/  ←  Q₁ = Repair
1C:   | Mustard please.             ←   A₁
0S:     ((provides))                 ←   Compliance with order
```

It is easy to show that degree six or more center-embedding occurs in spoken dialogue (see Levinson 2013b). When one finds a domain where a capacity is more evolved than in another domain, there is reason to assume that it has a longer evolutionary history. While short-term memory constraints are often invoked to explain our failure to produce center-embedding in syntax, that doesn't not seem to be a constraint in the interactive domain. This would suggest that linguistic recursion at least partly originates from this type of push-down stack in action sequencing, which as far as we know is universal in dialogue. Incidentally, it is also possible to show that cross-serial dependencies can be found in the sequential structure of speech acts, showing once again that complexity attributed to syntax may be more easily found in dialogue structure. All in all, a better case can be made for the need to climb the Chomsky hierarchy of grammars based on speech acts in dialogue than on syntactic structure.

For all the reasons outlined in this article, speech acts are a fundamentally important area of study in the language sciences. Work in this domain has been relatively, and inexplicably, neglected since the 1970s and 1980s, and it is time for a renaissance of work on speech acts and their use in dialogue.[3]

References

Allen, James F. 1979. *A plan-based approach to speech act recognition*. Toronto: University of Toronto.

Allen, James F. and Raymond C. Perrault. 1980. A plan-based analysis of indirect speech acts. *Journal of Computational Linguistics* 6: 167-82.

Austin, John L. 1962. *How to do things with words.* Oxford: Clarendon Press.

Bierwisch, Manfred. 1980. Semantic structure and illocutionary force. In *Speech act theory and pragmatics.* ed. John R. Searle, Ferenc Kiefer and Manfred Bierwisch, 1-35. Dordrecht: Reidel.

Chomsky, Noam. 2007. Of minds and language. *Biolinguistics* 1: 1009-1037.

—. 2010. Some simple evo devo theses. How true might they be for language? In *Approaches to the evolution of language*. ed. Richard K. Larson, Viviane M. Deprez and Hiroko Yamakido, 45-62. Cambridge: Cambridge University Press.

Clark, Herb H. 1979. Responding to Indirect Speech Acts. *Cognitive Psychology* 11: 430-477.

—. 1996. *Using language*. Cambridge: Cambridge University Press.

Cohen, Philip, Jerry Morgan and Martha Pollack. 1990. *Intentions in communication.* Cambridge, MA: MIT Press.

Cohen, Philip and C. Raymond Perrault. 1979. Elements of a plan-based theory of speech acts. *Cognitive Science* 3: 177-212.

Couper-Kuhlen, Elizabeth. 2010. 'Recognizing actions in interaction'. Lecture given at LAGB, Leeds, September.

Drew, Paul. 2013. Turn design. In *The Handbook of Conversation Analysis*. ed. Tanya Stivers and Jack

---

[3] My thanks to Penelope brown and Kobin Kendrick for helpful comments on the manuscript.

Sidnell, 131-149. Chichester: Wiley-Blackwell.

Drew, Paul and Tracy Curl. 2008. Contingency and Action: A Comparison of Two Forms of Requesting. *Research on Language & Social Interaction* 41: 129-153.

Dryer, Matthew.S. 2011a. Polar questions. In *The World Atlas of Language Structures Online*. ed. Matthew S. Dryer and Martin Haspelmath. Munich: Max Planck Digital Library. Available online at http://wals.info/feature/116A.

Dryer, Matthew S. 2011b. Position of Interrogative Phrases in Content Questions. In *The World Atlas of Language Structures Online*. ed. Matthew S. Dryer and Martin Haspelmath. Munich: Max Planck Digital Library, feature 93A. Available online at http://wals.info/feature/93A.

Enfield, Nick J., Tanya Stivers and Stephen C. Levinson. 2010. Question-response sequences in conversation across ten languages: An introduction. *Special issue of Journal of Pragmatics* 42(10): 2615-2619.

Geluykens, Ronald. 1988. On the myth of rising intonation in polar questions. *Journal of Pragmatics* 12: 467-485.

Goffman, Erving. 1978. Response cries. *Language* 54: 787-815.

Goodwin, Charles. 1996. Transparent vision. In *Interaction and grammar*. ed. Eleanor Ochs, Emmanuel Schegloff and Sandra A. Thompson, 370-404. Cambridge: Cambridge University Press.

Gordon, David and George Lakoff. 1971. Conversational postulates. Paper presented to the CLS, Chicago.

Grice, Herbert P. 1957. Meaning. *Philosophical Psychology* 67: 377-388.

—. 1973. Probability, defeasibility and mood operators. *Paper delivered at the Texas Conference on Perfomratives, presuppositions and Implicatures.* Austin, Texas.

—. 1975. Logic and conversation. In *Syntax and semantics: Speech acts*. ed. Peter Cole and Jerry Morgan, 41-58. New York: Academic Press.

Hare, Richard M. 1952. *The language of morals*. Oxford: Clarendon Press.

Heritage, John. 2012. Epistemics in Action: Action Formation and Territories of Knowledge. *Research on Language and Social Interaction* 45: 1-29.

Jurafsky, Dan. 2004. Pragmatics and computational linguistics. In *The Handbook of Pragmatics*. ed. Lawrence Horn and Gregory Ward, 578-604. Malden, MA: Blackwell.

Karlsson, Fred. 2007. Constraints on multiple center-embedding of clauses. *Journal of Linguistics* 43: 365-392.

Koschmann, Timothy. 2010. On the universality of recursion. *Lingua* 120: 2691–2694.

Labov, William and David Fanshel. 1977. *Therapeutic discourse: Psychotherapy as conversation*. New York: Academic Press.

Levelt, Willem J. M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levinson, Stephen C. 1981. Some Pre-Observations on the Modeling of Dialog. *Discourse Processes* 4: 93-116.

—. 1983. *Pragmatics.* Cambridge: Cambridge University Press.

—. 2005. Manny Schegloff's dangerous idea. *Discourse Studies* 7: 431-453.

—. 2006. On the human interactional engine. Roots of human sociality. In *Culture, cognition and human interaction*. ed. Nick J. Enfield and Stephen C. Levinson, 39-69. Oxford: Berg.

—. 2010. Questions and responses in Yélî Dnye, the Papuan language of Rossel Island. In *Question-response sequences in conversation across ten languages.* ed. Nick J. Enfield, Tanya Stivers and Stephen C. Levinson, 2741-2755. Amsterdam: Elsevier.

--.2013a. Action formation and ascription. In T. Stivers & J. Sidnell (Eds.), *The Handbook of Conversation Analysis* (pp. 103-130). Chichester: Wiley-Blackwell.

--.2013b. Recursion in pragmatics. *Language, 89*: 149-162.

Levinson, Stephen C. and Torreira, Francisco. Submitted. The precision of turn-taking and its implications for processing models of language. Frontiers in Psychology.

Magyari, L., Bastiaansen, M. C. M., De Ruiter, J. P., and Levinson, S. C. 2014. Early anticipation lies behind the speed of response in conversation. Journal of Cognitive Neuroscience, 26(11): 2530-2539. doi:10.1162/jocn_a_00673.

Merritt, Marilyn. 1976. On questions following questions (on service encounters). *Language in Society* 5: 315-357.

Michaelis, Laura. 2001. Exclamative constructions. In *Language Universals and Language Typology: An International Handbook.* ed. Martin Haspelmath, Ekkehard König and Wulf R. Österreicher, 1038-1050. Berlin: Walter de Gruyter.

Quirk, Randolph, Sydney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1989. *A comprehensive grammar of the English language*. London: Longman.

Sadock, Jerry M. and Arnold Zwicky. 1985. Speech act distinctions in syntax. In *Language typology and syntactic description: Clause structure*. ed. Timothy Shopen, 155-196. Cambridge: Cambridge University Press.

Schegloff, Emanuel A. 1988. Presequences and Indirection - Applying Speech Act Theory to Ordinary Conversation. *Journal of Pragmatics* 12: 55-62.

—. 2007. *Sequence organization in interaction: A primer in conversation analysis*. Cambridge: Cambridge University Press.

Schegloff, Emanuel A. and Harvey Sacks. 1973. Opening up closings. *Semiotica* 8: 289-327.

Schiffer, Stephen R. 1972. *Meaning*. Oxford: Basil Blackwell.

Searle, John. 1969. *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.

—. 1972. Chomsky's Revolution in Linguistics. *The New York Review of Books.* June 29, 1972.

—. 1975. Indirect speech acts. In *Syntax & semantics: Speech acts*. ed. Peter Cole and Jerry Morgan, 59-92. New York: Academic Press.

Stivers, Tanya. 2013. Sequence organization. In *The Handbook of Conversation Analysis*. ed. Tanya Stivers and Jack Sidnell, 191-209. Chichester: Wiley-Blackwell.

Stivers, Tanya, Nick J. Enfield, Penelope Brown, Christina Englert, Makoto Hayashi, Trine Heinemann, Gertie Hoymann, Federico Rossano, Jan Peter de Ruiter, Kyung–Eun Yoon and Stephen C. Levinson. 2009. Universals and cultural variation in turn-taking in conversation. *PNAS* 106: 10587-92.

Terasaki, Alene. 1976. *Pre-announcement sequences in conversation*. University of Irvine, Social

    Sciences 99.

Tomasello, M. 2008. *Origins of human communication.* Cambridge, MA: MIT Press.