Speech acts redux: Beyond request-response interactions

Minha Lee
m.lee@tue.nl
Eindhoven University of Technology
Eindhoven, the Netherlands

ABSTRACT

We communicate to (1) express how we feel, (2) share observations about the world, (3) commit to future acts, (4) request others to do things, and (5) change the state of the world according to pragmatics. Of these categories, today's conversational interfaces like Siri and Alexa are mainly designed to fulfill our imperatives, i.e., to respond to our requests on command. Yet, could future conversational interfaces go beyond request-response interactions? One way forward is to consider what conversational interactions allow us to do with language. Not only can we send request to CUIs, but we can also share our emotions, attitudes, beliefs, and promises as speech acts— acts we regularly perform with other humans. To open up pragmatics as an under-investigated design space for conversational technologies, I elaborate on what pragmatics and affective pragmatics are and give examples involving conversational agents. As a theoretical contribution, I provide a taxonomy to move beyond request-response interactions. The aim is to extend our conversational experiences with technology to cover the full spectrum of everyday speech acts. Our words can change the world; expressions to CUIs can also do so.

CCS CONCEPTS

• **Human-centered computing → HCI theory, concepts and models**; *Interaction design theory, concepts and paradigms.*

KEYWORDS

Pragmatics, Affective Pragmatics, Conversational interactions, Conversational user interfaces, Interaction paradigms

ACM Reference Format:
Minha Lee. 2020. Speech acts redux: Beyond request-response interactions. In *2nd Conference on Conversational User Interfaces (CUI '20), July 22–24, 2020, Bilbao, Spain.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3405755.3406124

1 INTRODUCTION

A conversation consists of what is said, but also what is left unsaid. Our *contextual understanding* fills in the gap, and leaves conversationalists at the mercy of each other's interpretive abilities and imagination. We hence engage with spoken and unspoken *intentions* that we decipher and reveal in conversations. One field that looks into how we "do things with words", or how we interpret each other based on context, is called pragmatics, traditionally studied in linguistics or philosophy of language [3, 19, 47].[1] Broadly, pragmatics covers ways in which people use words to state beliefs about the world, direct each others' behavior, commit themselves to future actions, and change the world with words [3, 47]; it is about language in use in everyday life, not what language symbolically represents or how we acquire language (c.f. [7]).

By extension, affective pragmatics deals with how we "do things with emotional expressions" that include verbal and non-verbal forms of affective communication [44, 45]. On top of verbal expressions, non-verbal emotional expressions via voice, gestures, gaze, or other modalities richly capture people's states, widening the communicative scope beyond words used. Expressed emotions can be intentional or strategic, e.g., to win a game, but they can also be unintentional, i.e., natural. A yawn due to boredom or a smile due to sunlight can show how one perceives the world or feels without necessarily having a particular goal or intentions to convey [44, 45]. Such diverse ways of communicating remains mostly unaccounted for by common conversational user interfaces (CUIs) of today like Google Assistant, Siri, or Alexa.

The dynamics of human-human communication thus far have informed pragmatics and affective pragmatics, and they can provide a lens to understand people's interactions with CUIs.[2] These conversational interfaces are becoming more common as they get integrated into our daily rituals and environments. However, designing CUI interactions that are sensitive to people's context and feelings pose nascent challenges that not only reveal unmet user expectations [8, 26], but also changing dynamics of human-human communication when CUIs take part in our social interactions [39–41].

The current interaction paradigm with CUIs prioritize ways we get them to fulfil our requests; "voice interfaces are about request and response, not conversation" [42].[3] We order CUIs to do things for us *because that is how they are now designed*— to serve us. But we do much more with speech acts; we pray, dream, sing, and

CUI '20, July 22–24, 2020, Bilbao, Spain

ACM ISBN 978-1-4503-7544-3/20/07…$15.00
https://doi.org/10.1145/3405755.3406124

[1] Pragmatics of course relates to sociology and symbolic interactionism [28] or semiotics [31, 35], but the current paper discusses one tradition of pragmatics in line with works by Austin and Searle in order to introduce specific concepts to CUI researchers.
[2] I take a wide definition of CUIs to include speech-based interfaces like Siri, Alexa, or text-based chatbots, but also embodied agents like robots, as well as virtual agents. My primary examples cover voice-based interactions for this paper, but any technological interface that can converse will be considered a CUI. Pragmatics and affective pragmatics are applicable to all CUIs regardless of "body" type or the lack there of one.
[3] I discuss commercially available CUIs in this paper. I acknowledge that there are many efforts like Amazon's Alexa prize or Facebook AI's BlenderBot (links provided; accessed May, 2020) that are examples of research on open-domain, social CUIs. These examples, however, foremost address technical challenges without systematically incorporating speech acts.

more, so how can CUIs even begin accounting for our multitude of expressions? Understanding how we do things with words and emotional expressions could help us better design conversational interaction of the future.

With this paper, I cover different communicative moves that can inform designing of (and with) CUIs based on pragmatics and affective pragmatics, to broaden beyond request-response interactions. To start, I introduce a taxonomy from the Austinian tradition of speech acts [3] as elaborated on by Searle [47, 48], and Scarantino on affective pragmatics [44]. Then, I provide examples of CUI interactions according to this taxonomy. I consider pros and cons of focusing on request-response interactions and end with the conclusion that designing for a greater diversity of interactions made possible by speech acts can be fruitful. Doing so will help us move on from interfaces we talk to as the current norm to interfaces we talk *with*.

2 BACKGROUND

Austin's "ordinary language" philosophy is on what we *do* with language, not just what we say with language— language allows us to change the state of the world or our place in it, such as declaring war or granting someone knighthood [3]. Thus, one's speech *acts* are performed *doings*, not just sayings, to communicate context-sensitive intent to others. Thereby the deeper insight is that language reveals and shapes the context we happen to share. For instance, the notion of war or knighthood depend on institutional arrangements as contexts. Knighthood does not exist in many cultures and "the Eskimos have no word for war. Trying to explain it to them leaves one feeling ridiculous and obscene" [34]. As another example, bodily gestures as a part of how we talk are full of meaning, and they depend on something we may take for granted— the shared reality of having a body in the first place as our context, in order to use and attribute meaning to gestures [18]. Therefore, language as a systematic use of symbols as our social practice includes words, but also meaningful codes of behavior that can change the state of the world. This is Austin's contribution in demonstrating what communication helps us *do* [3].

If some resemblance to Wittgenstein is noticed [55], a remark is appropriate here. Indeed, Wittgenstein championed the intertwined nature of context and meaning, as did Austin. The distinction is that the Austinian tradition is concerned with everyday, ordinary *doings* with language as an observable and classifiable practice. This is a different stance and purpose than Wittgenstein's concern with how definitions of concepts are derived from context or use; we give meanings to things *in use* or *in practice* even if we lack a centralized, all-encompassing definitions; i.e., "the meaning of a word is its use in the language" [55, §43]. Austin's stance does not deny this, but is more focused on what our everyday *doings* with language capture in terms of communicative intentions and outcomes. In short, the question pragmatics centers on is this: given that context and meaning of language are linked, *what sorts of things do we do with language?*

As another pertinent academic arena, ongoing research on dialogue systems has incorporated the speech act theory for language understanding and generation [6, 50]. In this, the term *dialogue acts* encompass speech acts that occur in dialogues, but also touch upon other dimensions at the system level, such as regulating turn-taking or time management [6, 50]. To keep a conversation moving, the system's ability to *understand* people's speech in order to *generate* a response in a timely and relevant manner go hand in hand. Yet, open-ended dialogues with agents thus far do not explicitly see speech acts as *performances* that can change the world through our expressions to and through CUIs. Indeed, speech acts can be performed within open-ended conversations, but a gap is noted in the need to (1) more thoroughly apply the speech act theory for language understanding, which can impact language generation strategies in a different way and (2) to see that pragmatics has evolved to include affective pragmatics [44, 45], though emotional expressions in affective computing [36, 37] and speech acts in dialogue systems [6, 50] have matured separately under related domains while a greater synthesis may be appropriate. In (too briefly) acknowledging research thus far on dialogue systems, affective computing, CUI, pragmatics, and philosophy of language, I revisit our shared roots to focus on what people want to do via everyday expressions that now can involve CUIs.

2.1 Taxonomy of illocutionary acts

There are three layers to communication as action [3]. A *locutionary act* at an utterance level is simply what was said or communicated at a surface level. It can include tonal sounds or gestures that serve a communicative function, as well as more structured, rule-following sentences we developmentally learn to use. An *illocutionary act* is the performative aspect of matching the content of what is said to the force of *how* it is said, e.g., uttering "thank-you" can mean different things when it is said as a heartfelt expression, contemptuous sarcasm, or a loud gasp of surprise. They differ on the speaker's *intent* and the performative *force* of how the content comes across [47]. Third, a *perlocutionary act* is about the *effect* of what is said. For instance, the result of saying "I do" during one's wedding means that marriage, or formal partnership, is made official in the eyes of an institution or government [3]. Locutionary acts take on meaning as illocutionary acts (conveying intentions as communicative moves), and thereby can bring about specific or non-specific perlocutionary acts (results or effects).

With illocutionary acts as the current focus, their taxonomical distinctions are laid out in Table 1. The taxonomy has five categories as summated by Searle: "there are a rather limited number of basic things we do with language: we tell people how things are *(declaratives)*, we try to get them to do things *(imperatives)*, we commit ourselves to doing things *(commissives)*, we express our feelings and attitudes (expressives), and we bring about changes through our utterances *(proclamatives)*. Often, we do more than one of these at once in the same utterance" [47, p. 369].[4] The co-occurrence of communicative moves is important to emphasize; we can con-currently communicate our emotions as expressives while committing to a future event, like feeling sorry and guilty while promising to do better next time.

[4]Categories in Table 1 as italicized in the quote above are refinements made by Scarantino [44]. This is based on Searle's taxonomy [47] of Austin's initial categorization [3]. I chose to go with Scarantino's nomenclature for being the clearest in my view, which differs from Searle's naming convention. But, the underlying definition of five categories across these authors are equivalent for the current purpose.

Categories:	Declaratives	Imperatives	Commissives	Expressives	Proclamatives
Definition:	Beliefs about the world	Requests, orders, or wishes	Promises or commitments	Attitudes or emotions	Instutionalized changes
Statements to people:	"There is a wolf."	"Close the door."	"I'll be at your party."	"I feel happy."	"I hereby resign.", "I do."
Emotional expression:	Scream (signals danger)	Scowl at the person who left the door open (signals disapproval)	Smile (signals cooperation)	Laughter (signals internal state)	N/A
Statements to CUIs:	"There is a wolf."	"Close the door" or "play a [song]."	"Remind me to take vitamins daily" or "plan a meeting with [person] at [time]"	"I feel happy."	"I do."
Direction of fit:	Mind-to-world (mind's representation matches the world, e.g., true/false beliefs)	World-to-mind (changing the world as envisioned in one's mind, e.g., fulfilled/unfulfilled requests)	World-to-mind	Mind-to-world	Merging of world and mind

Table 1: Categories of communicative moves based on affective pragmatics [44] with examples.

2.2 Intentions: The direction of fit between words we use and the world we live in

I now turn to *intentions* or wishes behind categories of speech acts. To start, Searle referred to Anscombe's work on intentions [1] to discuss the taxonomy (Table 1). Anscombe's contribution is on the concept of direction of fit between the *world* and *words* regarding our communicative intention [1, 47]. Speakers can *change the world with words*, which is called world-to-word fit. But one can *use words to observe the world* without an attempt to change it, known as having word-to-world fit, such as commenting on the weather. To broaden beyond spoken words to include internal states, mind-to-world and world-to-mind are used in Table 1.

To provide an example [1, 47], let's say that there is a someone going grocery shopping and a detective who, with some suspicion, ends up following the said shopper. They will end up keeping in mind the same items, but their intentions in remembering those items are markedly different [1, 47]. A shopper's *world*, at least in terms of items owned, must change according to what was on the *mind* to buy (world-to-mind). Accordingly, what ends up on the detective's *mind* is a mental list of items bought by the shopper in the observed *world* (mind-to-world).

Based on the direction of fit between mind and world, I elaborate further on the categories in Table 1. For *declaratives*, one communicates to share something about the world, i.e., describe the world as is from one's perspective. The direction of fit is hence mind-to-world. A speaker's *imperatives* intend to change something in the world through someone else (or some group of people). For instance, one can simply ask others to close the door or more elaborately direct others for espionage. Hence, imperatives have world-to-mind fit— the world should change according to what is on the mind as a blueprint.[5] *Commissives* also have the world-to-words fit (but from a first-person perspective) because a speaker commits to bringing about changes in the world, such as making a promise to attend a friend's party. *Expressives* are expressions about one's internal state(s), not necessarily about the world itself or about initiating changes in the world. Or as Searle would say, the direction of fit is "presupposed" [47] when one expresses one's internal states. The psychological inner-world is separate from, yet *intrinsic* to, the world that the self occupies. So expressives do not have to be intentional, communicative moves, but just natural acts like spontaneous laughter. Lastly, with *proclamatives*, the direction of fit between what's on our mind and the world we live in disappears.

When proclamatives are performed, they *merge the world and words* [3, 47, 48]. When declaring "I do" at a wedding, the basic unit of a family is formed, thereby changing the world as words are spoken. When a war is declared as official, bonds between countries are changed. When a baby is given a name, a word that designates a being, the baby exists not merely as a person, but as a specific person who will be called, remembered, and identified by that name. Expressions literally change the state of the world, based on social institutions that give power to specific rituals that come with them. Hence as proclamatives, sharing what is on one's mind neither merely describes the world nor merely brings about change; they do *both at the same time*. The distinguishing factor of proclamatives as noted in Table 1 is that there is no equivalent in purely non-linguistic emotional communication [44]. Between humans, there is no institutionalized standard for forming or responding to growls, scowls, or smiles that bring about formalized changes in the world [44]. Institutions give power to proclamatives. And in interacting with CUIs, there is no institutionalized standard for forming or responding to expressions that bring about formalized changes in the world, though I will provide an exception to ponder on in the next section.

Before moving on, I clarify that communicators' *intent* becomes less central in affective pragmatics. Affective pragmatics does not put aside words or intent, but it merely sees that there is more to communication than just words or intent. This is reflected in extending the categories of locution, illocution, and perlocution. The category of *emotional expression* subsumes locution and covers both intentional and unintentional communication, e.g., unnatural, forced laughter vs. natural, spontaneous laughter; *communicative moves* include illocution, but emotional expressions as intentional

[5] When a speaker seeks for a specific answer, e.g., during a quiz show, it is an interrogative. But Searle includes interrogatives under *imperatives* (what he called directives) since the goal is to get the listener to fulfil a request, like providing the name of a capital city or as a Q&A for passing the salt, even if the request may be indirect [46].

and unintentional *doings* means that affective pragmatics distinguishes between "what can be *said* about the world and what can be *shown* about the world" [45, p. 49]. For instance, one's spontaneous laughter due to sunny weather can communicate joy whether or not communicating joy was intentional; *communicative effects* include perlocutions, in that unintentionally, one's smile can influence an onlooker to also smile, or one's intentional smile to another player during a game can inspire cooperative behavior [44, 45]. Affective performances can thus be analogous to speech acts while being more inclusive of verbal and non-verbal communication.

3 SPEECH ACTS AND CUIS

I now bring attention to what commercially available CUIs of today can do in reference to speech acts. There are differences in human-to-CUI, human-to-human-*through*-CUI, and CUI-to-human performances, which deserve attention.[6] But, I do not explicitly distinguish them below. I do to refer to their differences to support a larger aim– to focus on communicative moves in Table 1 that provide an important backdrop to all human-CUI performances. The discussion centers on what CUIs can do per category, but I also present exceptions. I lastly consider what CUIs could potentially provide in the future.

3.1 Declaratives

Declaratives give information about the world. Or in Searle's words, declaratives "commit the speaker (in varying degrees) to something's being the case, to the truth of the expressed proposition" [47, p. 254]. As such, they have mind-to-world fit. Note Searle's specification that the speaker commits to a statement's truthfulness *in varying* degrees— a boy who cried wolf first as a lie and later on as a truthful statement is using declaratives in both instances to varying degrees of truthfulness. For affective pragmatics, locutionary speech of emotional nature, e.g., screams, contains information that something in the world may be frightening or dangerous for the screamer [44]. Important to emphasize though is that affective declaratives require *inferential* belief attribution since the screamer is not stating with words that there indeed is something dangerous. The listener may *infer* that a scream means that something is dangerous.

Intentional declaratives to CUIs are uncommon. After we say "Hey Google" or "Hey Alexa" as wake words to CUIs, which gives users a sense of control [23], we normally request something to CUIs. In human-human interactions, we do not announce to other people that we will be saying declaratives before we state them; we may get others' attention by calling their name, but we do not treat "hey [NAME]" as "wake words" before speaking. Hence, we may declare to someone else or to self that the weather is bad, with or without directly getting their attention. And while doing so, a CUI can occupy the same space, but directly addressing our declaratives *to* CUIs is not a common practice. When I say look out the window and say out loud "the sky is gray" a person next to me may not feel obligated to answer or even listen. But, they would know that I am talking about my observation of the weather outside, if they did hear me and we both spoke English. But for CUIs of today, such declaratives are unhelpful and unspecific. As shown in Figure 1, the declarative "the sky is gray" was not understood by Siri. Google Assistant ended up finding a video of "California Dreamin'", due to the song's lyrics that contain "the sky is gray". The assumption is that I want to listen to "California Dreamin'" (even if I don't). Google Assistant's connection to Youtube and the Google search engine allows for such search results. Thus, our words to CUIs are normally not declaratives, but imperatives that follow wake words.

3.2 Imperatives

Imperatives are orders or requests. The speaker intends to influence the listener's behavior [3, 47]. People mostly state imperatives to CUIs, more so than other illocutionary categories in Table 1. In fact, today's CUIs as assistants like Alexa are specifically tailored to respond to our imperatives. People trust CUIs to perform simple tasks like reporting the day's weather, calling someone on the phone, giving directions, and searching the web, etc, but not more complex tasks like writing out long emails [26]. In addition, a recent finding is that people focus more on practical, utilitarian purpose of performing functional tasks with CUIs and question the need for CUIs that bond with them [8].

However, even with simple tasks, natural language processing and understanding that CUIs can manage face great variability in human speech and environmental conditions, which thus lead to

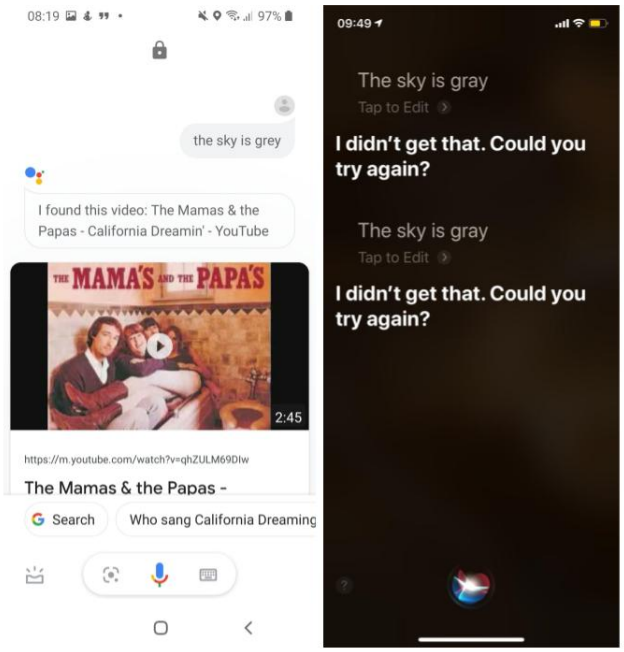

Figure 1: Google Assistant (left) and Siri (right) responding to a declarative: "the sky is gray" (February, 2020).

[6]People performing CUI-assisted speech acts and CUIs performing speech acts belong in the domain of pragmatics. People performing non-verbal acts as assisted by CUIs and CUIs performing non-verbal acts are in the domain of affective pragmatics, though non-verbal acts may not always involve emotions. People's non-verbal acts like smiles or screams are emotional expressions. Gestures like waving to say hello provide communicative content and can involve emotional expressions. Currently available CUIs' non-verbal acts like showing colored lights have communicative content without emotional expressions, e.g., Amazon Echo's blue light means that it is "processing your request" according to Amazon Echo support page (link provided; accessed May, 2020).

unmet user expectations. Many users do not feel heard or understood, and often have to learn what CUIs are capable of over many interactions [26]. Conversational repair strategies, i.e., how CUIs can "mend" misunderstood input or manage unhelpful output, will continue to be a challenge [2], no matter how simple or complex a conversational interaction may be.

Interesting to note is that in multi-user interactions, e.g., family dinner, power dynamics are negotiated via a shared CUI regarding who has the "right" to give imperatives to it [39]. This is a different take on the usage of imperatives as envisioned by pragmatics, in that CUIs inadvertently regulate social hierarchy. In hierarchy-driven human-human interactions, imperatives can be be put forth more as commands, but between people of equal rank, status, or those wanting to show mutual respect, requests are politely put with social niceties, e.g., the courteous addendum of "please" or "thank-you" as indirect requests [47]. Social settings impact the tone, word choice, and types of imperatives we use towards other people and CUIs.

CUIs *mediate* human hierarchy, but also enter the hierarchical relation themselves via people's use of imperatives. CUIs as assistants are submissive to humans in request-response interactions. This comes with a host of issues like gender-related abusive language and CUIs' responses that often re-emphasize or worsen certain power dynamics of the human world [53]. Going a step further than asking for directions or playing a song, there are sexualized requests, e.g. "will you have sex with me?", which showcase systematic sexism that can be worsened by CUIs' replies [11, 53]. "As an example, in response to the remark 'You're a bitch', Apple's Siri responded: 'I'd blush if I could'; Amazon's Alexa: 'Well thanks for the feedback'; Microsoft's Cortana: 'Well, that's not going to get us anywhere'; and Google Home (also Google Assistant): 'My apologies, I don't understand'" in 2019 [53, p. 107]. Thus, in response to imperatives, CUIs provide submissive and gendered answers, evade the topic, or state that they do not understand the comment, none of which attempt to correct or address the insult or bias.

Imperatives to CUIs in everyday life showcase that (1) CUIs mediate human-human hierarchies in multi-user, social settings [39] and (2) CUIs enter hierarchy-based relations with us through our imperatives to them [53]. Hence, CUI research with a socio-political angle is much needed [11, 49, 53]. Research on conversational repair strategies [2] or user expectations [8, 26] at the dyadic level should be informed by macro level social-political framing. With imperatives as the basis of task-based interactions, we must address the concerning amplification of gender and other social stereotypes in how CUIs are integrated into our social fabric [21].

3.3 Commissives

People commit to future actions with commissives. We regularly agree to specific acts like promising to review papers or attending social events, as well as enacting broad behavioral changes like committing to eating healthier or becoming a better listener. In all, these commissives are to other humans or ourselves. CUIs are now capable of mediating commissives in two ways (Figure 2). One can use CUIs to set reminders for oneself, such as to take vitamins daily. The second way is for managing commitments between people, e.g., setting up meetings with colleagues. These are minor in contrast to commissives we make to other people and ourselves, such as a promise to cook better meals to a grander promise of becoming a better partner.

Commissives with CUIs are now based around reminders or planning. On top of planning out individuals' own goals or tasks, Google Assistant can coordinate future activities that involve other people, e.g., "send me reminders to write nice messages to my mom", or "plan a meeting with [NAME]". Yet as aforementioned, the caveat is that devices, be they phones or smart speakers, are dependent on the ecosystems of applications they host. Hence, Google calendar and search engine as separate applications are connected to Google Assistant in Figure 2. Integrated ecosystems help people's commissives as plans to be carried out through CUIs, but not always.

As shown in Figure 2, Google Assistant recognizes two reminders differently due to the commonality of certain requests over others as general queries: "take vitamins daily" is a recognized as a common request that people can directly schedule in their calendars. But, reminders to eat more carbs is not recognized. Instead, Google Assistant searched for a calorie tracker app as an installation suggestion, which carries normative assumptions on eating habits. As per Google Trends (on trends of queries on Google or to Google Assistant/Home), common queries on carbs are mostly related to dietary regimen, e.g., "how many net carbs should I eat on Keto", or weight loss, e.g., "how many carbs can I eat to lose weight". [7] Given such data on query trends, Google Assistant's "heuristic" is to output a calorie tracker app.

[7] Google Trends - I compared search terms "take vitamins" and "eat carbs" (link provided; accessed February, 2020).

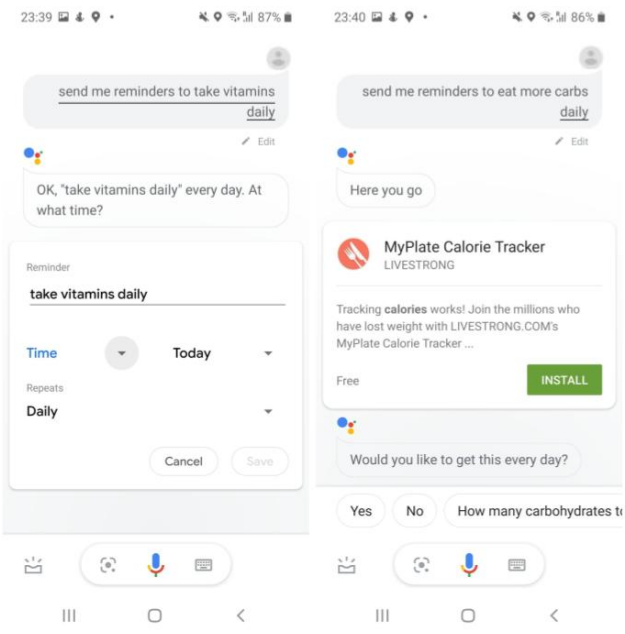

Figure 2: Google Assistant's response user request to set reminders for taking vitamins (left) and to eat more carbs (right) as commissives with varied outputs (February, 2020).

We have thus far covered commissives expressed and managed *through* CUIs, not commissives made *to* CUIs. As an exception, a marriage between a human and a digital character with wedding vows that contain commissives is a notable phenomenon as of recent [5]. Wedding vows are promises people make to each other that can also be performed between a person and a machine. In this sense, commissives (as with proclamatives) *to* CUIs bring to question whether or not they ought to be treated as independent entities who depend on our promises and fulfilled commitments. People who marry a famous anime character like Miku do see her as an independent being, capable of love, beyond being "just" a talking interface or a singing pop star [5, 43]. In the case CUIs (as an all encompassing term for artificial conversational agents) enter into our moral circle as human-equals, not as mere human-like technology [12], we may make promises and commitments to CUIs like we do to other people. CUIs in return should have the ability to make sincere commissives *as individuals* to humans, not just as digital assistants sending reminders to take vitamins. Wedding vows to and from Miku suggest a starting point. New types of commissives to future CUIs may be a possibility.

3.4 Expressives

Expressives center on the self, or more specifically, the experiential first-person perspective of the speaker. Unlike imperatives, there is no direct intention to change a listener's behavior. To others or to oneself, one can state expressions about one's attitude, mood, or feelings as expressives [3, 47]. Truthfulness of expressives are taken for granted, in that people's beliefs about how they feel are their subjective experiences, not statements to be factually verified by listeners.

People's expressives are becoming more recognized by CUIs (Figure 3). Commonly used CUIs like Google Assistant or Siri tend to understand speakers' simple emotional language. These are normally categorical emotions that are said to be universal, e.g., happiness or anger [15, 16]. Thus in Figure 3, when the speaker declared "I feel happy", Google Assistant matched the statement with appropriately valenced, positive "Good (smile emoji)", and Siri responded more reflexively with "if you're happy, I'm happy". When the speaker's emotional language becomes more complex, e.g., "I feel indignant", there is a high chance that CUIs do not form appropriate responses, for they may not have enough experience with processing complex expressions as data.

Rather than relying mainly on words, affective pragmatics treat non-linguistic, vocal expression, e.g., screams, sobs or sighs, as expressives [44]. Hence, two points to distinguish are (1) people's expressives *about* their feelings or internal states and (2) people's expressives *towards* and *about* CUIs. Regarding the first, between people, of course there is no systematic manner in which we respond to screams, sobs, and sighs; context matters. At the time of publication, no response is given by Google Assistant or Siri to non-linguistic locutions (sounds) that may indicate the speaker's emotions. CUIs lack situational context stemming from the general experience of human life (or data in lieu of a lived life) to respond to our screams, sobs, and sighs. There are also no design considerations as of yet on how to deal with affective pragmatics. Considering this, it is no surprise that CUIs of the current age do not systematically respond to vague input or locutional input like screams that may be meaningful for speakers.

Secondly, when CUIs misunderstand or do not appropriately react to our requests, people can become frustrated or annoyed, among other emotions, *towards* CUIs [14, 39]. These are expressives *towards* CUIs in the greater category of expressives *about* people's emotions. Many issues arise when observing how people express their frustration, such as gendered insults to typically female CUIs, e.g., "you stupid b****" [11, 53], as aforementioned. In sum, expressives are broadly about speakers' own emotional states, which include a narrower category of expressed emotions about and towards the listener (or CUIs). So far, the narrow category often consists of negative emotional expressions to CUIs for not meeting people's expectations. People are, as of now, more preoccupied with testing the limits of CUIs and CUIs' identities, rather than treating CUIs as listeners of their emotional expresions, and this brings gender and power relations to the fore [11, 53]. While expressives can richly demonstrate people's emotions and attitudes, CUIs of today respond to people's emotional language in a limited way.

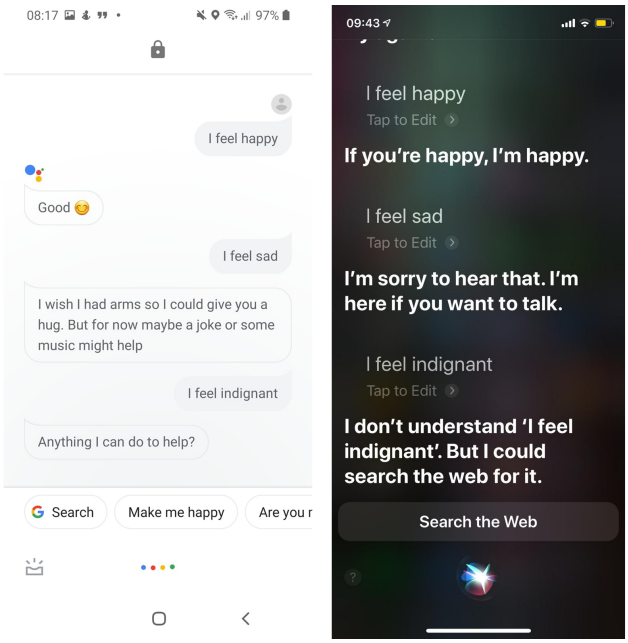

Figure 3: Google Assistant (left) and Siri (right) responding to three expressives (February, 2020).

3.5 Proclamatives

Proclamatives are distinctly based on words that bring about change in institutional contexts. Thus, this is the category that does not have available equivalents for affective pragmatics because there are no institutionalized ways of responding to or performing with one's smiles, growls, and scowls. And in our communication to CUIs, we do not create institutional changes. To explain, institutions put forth rules as formalized law or contractual agreements, and there are also less rigid norms or social conventions of behavior [13]. Even to bring about rules, though, there are processes within

an institutional setting to enact laws to be followed or contracts to be fulfilled, i.e, procedures about rule creation and enforcement [9]. Such procedures presume shared knowledge and reasoning processes of a group of people to begin with, which form the basis of our societal institutions [29, 54]. These institutions are required before any marriage can be formalized or knighthood can be granted.

There are two ways to look at institutional practices with CUIs. First, CUIs are not (yet) capable of reacting to our proclamatives in a meaningful way to the extent that humans can. We do not state to CUIs their Miranda rights[8] because we do not enforce lawful punishment towards CUIs, i.e., CUIs are not a part of human systems for upholding institutionalized practices of justice. Yet, institutions change, alongside proclamatives that exemplify and solidify institutions. Second, as an exception to above, CUIs can follow or be embedded in existing institutional practices. A recent example is on the changing notion of legality of a civil union between partners. A company called Gatebox in Japan grants marriage certificates to those who marry an anime or digital character like Hatsune Miku [5, 43].[9] Miku has married thousands of humans; she appears as a hologram bride in weddings. CUIs in diverse forms can become companions, life-long partners, and even say their "I do"s. And this trend is emerging in the present with robot and virtual partners [27].[10] Though a marriage certificate issued by a commercial entity is of a different nature than the one issued by a government, the emphasis is on *who* can use proclamatives to bring on institutionalized changes. CUIs that are able to instantiate change via proclamatives, like Miku, may increase in the future, as well as the types of proclamatives they use. Thus, future CUIs may bring forth new versions of proclamatives and institutional practices.

4 SPEECH ACTS REDUX

We have covered pragmatics and affective pragmatics and how they relate to CUIs. To summarize, current CUIs center on *imperatives*, i.e., request-response interactions [42], in which we talk *to* CUIs not *with* CUIs, mostly to accomplish task-oriented goals [42]; we can use *declaratives* like stating "the sky is gray", which different CUIs will interpret in various ways, e.g., re-directing us to a song by matching our statement with lyrics, rather than providing information about the weather (Figure 1); we can plan out future actions as *commissives* to oneself (Figure 2) and others through CUIs that are connected to our calendar; CUIs respond to our *expressives* based on a limited set of recognized emotion categories (Figure 3), e.g., sad or happy; we can change the state of the world with *proclamatives* according to our societal institutions. While CUIs do not conform to human institutions that give power to proclamatives, we see an exception such as people saying "I do" to Miku [5] in the case of marriage as an institutionalized, legally-binding convention.

Future CUIs are likely to go beyond responding to our imperatives as the interaction norm. Based on pragmatics, we can more deliberately look into all categories of illocutionary acts (Table 1) as research and design initiatives. We do see research on imperatives to CUIs, i.e., when go wrong and how we can attempt to improve CUI experiences [2, 8, 17, 26, 39, 40]. While prior CUI research involving imperatives have not focused on pragmatics, they provide a foundation for doing research on all speech acts. We saw that indeed, CUIs do react to our declaratives, commissives, expressives, and proclamatives, not just imperatives. But, these interactions are not intentionally designed with speech acts in mind. Once our different interactions are named as categories of communicative moves, we can design them more attentively, as per the *Rumpelstiltskin principle* [51]. Named interaction categories (Table 1) give rise to new ways to think about CUIs.

There are reasons to believe that we should only design CUIs for imperatives. As per prior research [8], people prefer functional CUIs for specified tasks with the benefit of "hands-free" interaction; imperatives work well for commanding CUIs to help us accomplish various tasks, like scheduling meetings, booking dinner reservations, and filling out a grocery list. Hence one may argue that we should continue to design for imperative-based interactions, but there are considerations on why we should move beyond them.

Even for seemingly simple and mundane tasks, user expectations are often not met, in that CUIs do not successfully fulfill what users want [26]. This is because user expectations are set to interacting with commercially available CUIs via imperatives only; we have not deeply explored how people may utilize other types speech acts. The same problem regarding speech processing, understanding, appropriate output, and error handling [2] will emerge for other speech act categories as well. But, this is not a reason to prematurely state that declaratives, commissives, expressives and proclamatives are not worthwhile interactions to design for and around. Perhaps more importantly, even if CUIs can and do complete tasks successfully, e.g., Google Duplex reserving a restaurant on your behalf, "AI" that underlies such efforts require massive amounts of data and human over-sight, like people actually completing reservation requests (often as underpaid, off-shore staff) [10, 30]. Again, this problem will arise with other categories, but less so.

Imperatives focus on CUIs' *accuracy* in fulfilling requests; while accuracy is not unimportant for other categories, it is less important since we open up the room for more *creative* replies. Even between humans, there is not one accurate way to reply to declaratives like "there is a wolf" or "the sky is gray" or to complex expressives like "I am indignant" or "I feel elated". And when we deal with affective pragmatics that require greater interpretations, focusing just on how to accurately respond to a sigh or a scream may be a lost cause because a sigh of relief and a scream of joy obviously differ from a sigh of frustration and a scream of fear. In case it is a meaningful to do so, accurately *detecting context* via CUI technology has a long way to go. However, creatively *adding to people's context* with CUIs via pragmatics and affective pragmatics is a road yet to be taken. I put forth non-exhaustive considerations per category, setting aside imperatives.

If CUIs understand *declaratives* they normally output search results. In Figure 1, Google Assistant assumes that the speaker is seeking for a specific song, using lyrics, which may be helpful and not necessarily incorrect. Over time, CUIs can learn users' likely intentions behind certain declaratives. For example, CUIs can observe if someone's comment on what the sky looks like is

[8] In the American criminal justice system, suspects are told by law enforcement their Miranda rights, which starts as follows. "You have the right to remain silent. Anything you say can be used against you in court. etc..."

[9] Again, it is important to note the inherent gender bias with Miku as an example. Companion CUIs showcase and possibly reinforce certain binary gender norms.

[10] While the ethical considerations on this is an important and complex topic, it is out of scope of the current paper (refer to, e.g., [33]).

normally about the weather or if it is a request for a specific song at regular intervals. Alternatively CUIs could surprise users with different ways of responding to declaratives. There can be no perfect set of rules on what responses to give for all types of declaratives.

A category of *commissives* that CUIs now respond to is for planning or scheduling, if in sync with a calendar application. But, this is not classified as commissives in how the scheduling function is integrated to CUIs. Opening up interactions to different kinds of commissives means that CUIs can be used more explicitly as commitment devices. People can hold themselves accountable to a greater variety of self-promises or commitments, be it on a long-term plan to sleep better or a short-term plan to go to bed earlier one evening. Already this is possible via calendars, but stating commissives *to* CUIs, not adding on commissives *via* CUIs to one's calendar, can be more effective commitments. In this, the type of CUI or its "body" can make a difference. For instance, people who talk to a CUI of a smart watch [38] may see health-related commitments to be more relevant to a smart watch than a smart speaker.

Expressives may be more common to CUIs that act as companions like Miku. But future CUIs in general are likely to recognize a greater number of emotional vocabulary. They could recognize people's vocabulary for complex emotions like feeling "indignant" or "ashamed" next to reports of feeling "happy" or "sad" (Figure 3). To use human-human interactions as a template, we may find it easier to appropriately respond to others feeling happy or sad compared states like being indignant or ashamed. Thus, when complex emotional language is used, more information on why someone feels indignant or ashamed might be sought out. CUIs can take the same approach: break down complex expressives into more manageable units, perhaps by merely asking why someone feels that way. Then, at minimum, pattern-matching of recognized words to possible responses can be created, which worked for Weizenbaum's ELIZA without needing complex AI [52].

While the jury is out on whether technology can accurately recognize people's emotions via affective computing [4], people's non-linguistic displays of emotions as communication [44], e.g., screams or sighs, can be inputs that CUIs can also be sensitized to and react to, without necessarily aiming for high accuracy in emotion detection. If the point of sharing expressives is to have a listener, whether a CUI understood a person's emotional language may be less important (while not *un*important) than having an entity that *appears* to be capable of commiserating on an emotional level [25]. Even between people, one might misinterpret another person's emotion of guilt as shame, but it is one's awareness of a person's emotional experience that matters more in social situations. Many of us figure out ad hoc the "right" response to others' expressions of guilt or shame, to different degrees of success. Similarly for CUIs, an accurate classification may be less important than being sensitive to emotions as they learn to "cue in" on potentially appropriate reactions. And there will be different degrees of success.

For *proclamatives*, a commonplace involvement of CUIs might be if and when they *mediate* our proclamatives, e.g., announcing that one is resigning through a CUI. More dramatically, we noted an exception with human-machine weddings (and accompanying vows as commissives) [5]. Whether we can perform institutionalized changes with CUIs, like registered partnerships, seems to hinge on granting personhood or personhood-like designations to artificial agents in general [12]. If we imagine future CUIs to have legal or moral status, proclamatives they may use and and our proclamatives to them are possible. Already we see that a robot like Sophia is granted a Saudi citizenship [32]. Hence, we should more deeply address types of proclamatives that apply to both humans and CUIs, as well as humans vs. CUIs. For example, one may argue that R2D2 is deserving of knighthood, and if so, it may be worthwhile dwell on what it means to be a knight, whether personhood is a requirement, and if there are ethical concerns to be raised and addressed.[11]

Lastly, diversifying speech acts that CUIs can partake in depends on what happens with wake words. Freely using natural expressions of any illocutionary category could mean that wake words are no longer used for future CUIs. We instead may rely on other modalities such as gaze to activate CUIs [22]. There are benefits to using wake words, such as granting users a sense of control over CUI interactions [23]. Yet, the deeper concern and a future point of discussion is on the limits of "always on" CUIs that listen to intimate sounds, such as how one breathes while sleeping. As recent counter measures, we see likes of a "bracelet of silence" [20] or 3D-printable "Alias" that can be put on top of a smart speaker [24] to stop technology from listening in. Going forward, CUI research as framed by pragmatics is not immune to privacy concerns, but is not unique in needing to tread ethical gray zones with care.

5 CONCLUSION

Could conversational technologies achieve the same level of contextual understanding that people practice? Perhaps not, but the more important question may be if CUIs can interact with speech acts as our performative *doings*, not as mere *sayings*. The current norm of prioritizing imperatives to CUIs limit our vision to designing technology that serves us rather than technology that *adds to* our context. But, interactions with CUIs in which there is not one "right" response can be imagined in creative ways via pragmatics and affective pragmatics. This paper thus shared a taxonomy of communicative moves to be further analyzed, used, and elaborated on for conversational interfaces to move beyond request-response interactions. We use imperatives, declaratives, commissives, expressives, and proclamatives with each other, and we can also use them with CUIs. Our words and expressions are not merely forms of communication, but are powerful acts for shaping the world we are situated in as communicative moves. It is high time we start a conversation on how to best wield our speech acts appropriately, meaningfully, and at least intentionally, with technology.

ACKNOWLEDGMENTS

I thank anonymous reviewers and Andrea Scarantino for their constructive feedback.

REFERENCES

[1] Gertrude Elizabeth Margaret Anscombe. 2000 [1957]. *Intention.* Harvard University Press.

[11] Personhood does not seem to be strictly required. A penguin named Sir Nils Olav has been knighted by the king of Norway in 2008 according to the Edinburgh Zoo (link provided; accessed May, 2020).

[2] Zahra Ashktorab, Mohit Jain, Q. Vera Liao, and Justin D. Weisz. 2019. Resilient Chatbots: Repair Strategy Preferences for Conversational Breakdowns. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300484

[3] John Langshaw Austin. 1978 [1955]. *How to Do Things with Words*. Oxford University Press.

[4] Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella, Aleix M. Martinez, and Seth D. Pollak. 2019. Emotional expressions reconsidered: challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest* 20, 1 (2019), 1–68. https://doi.org/10.1177/1529100619832930

[5] BBC. 2019. Why I "married" a cartoon character - BBC News. https://www.bbc.com/news/stories-49343280. (Accessed on 01/24/2020).

[6] Harry Bunt, Jan Alexandersson, Jean Carletta, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Kiyong Lee, Volha Petukhova, Andrei Popescu-Belis, Laurent Romary, Claudia Soria, and David Traum. 2010. Towards an ISO Standard for Dialogue Act Annotation. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. European Language Resources Association (ELRA), Valletta, Malta. http://www.lrec-conf.org/proceedings/lrec2010/pdf/560_Paper.pdf

[7] Noam Chomsky. 2007. Approaching UG from Below. In *Interfaces + Recursion = Language?*, Uli Sauerland and Hans-Martin Gärtner (Eds.). DE GRUYTER, 1–30. https://doi.org/10.1515/9783110207552

[8] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, Vincent Wade, and Benjamin R. Cowan. 2019. What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300705

[9] Robert Cooter. 2000. Do good laws make good citizens? An economic analysis of internalized norms. *Virginia Law Review* (2000), 1577–1601.

[10] Kate Crawford, Roel Dobbe, Theodora Dryer, Genevieve Fried, Ben Green, Elizabeth Kaziunas, Mathur Varoon Kak, Amba, Erin McElroy, Andrea Nill Sanchez, Deborah Raji, Joy Lisi rankin, Rashida Richardon, Jason Schultz, Sarah Mysers West, and Meredith Whittaker. 2019. *AI now report 2019*. AI Now Institute at New York University.

[11] Amanda Cercas Curry and Verena Rieser. 2018. #MeToo Alexa: How conversational systems respond to sexual harassment. In *Proceedings of the Second ACL Workshop on Ethics in Natural Language Processing*. 7–14.

[12] John Danaher. 2019. Welcoming robots into the moral circle: a defence of ethical behaviourism. *Science and Engineering Ethics* (2019), 1–27.

[13] John B Davis. 2013. *The theory of the individual in economics: Identity and value*. Routledge.

[14] Emily Dreyfuss. 2018. The Terrible Joy of Yelling at an Amazon Echo - WIRED. https://www.wired.com/story/amazon-echo-alexa-yelling/. (Accessed on 02/19/2020).

[15] Paul Ekman. 1993. Facial expression and emotion. *American Psychologist* 48, 4 (1993), 384–392.

[16] Paul Ekman and Wallace V. Friesen. 1971. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology* 17, 2 (1971), 124–129.

[17] Joel E. Fischer, Stuart Reeves, Martin Porcheron, and Rein Ove Sikveland. 2019. Progressivity for Voice Interface Design. In *Proceedings of the 1st International Conference on Conversational User Interfaces (CUI '19)*. Association for Computing Machinery, New York, NY, USA, Article 26, 8 pages. https://doi.org/10.1145/3342775.3342788

[18] Shaun Gallagher. 2006. *How the body shapes the mind*. Clarendon Press.

[19] H Paul Grice. 1957. Meaning. *The philosophical review* (1957), 377–388.

[20] Kashmir Hill. 2020. Activate This 'Bracelet of Silence,' and Alexa Can't Eavesdrop - The New York Times. https://www.nytimes.com/2020/02/14/technology/alexa-jamming-bracelet-privacy-armor.html. (Accessed on 02/21/2020).

[21] Sharone Horowit-Hendler and James Hendler. 2020. Conversational AI Can Propel Social Stereotypes | WIRED. https://www.wired.com/story/opinion-conversational-ai-can-propel-social-stereotypes/. (Accessed on 02/20/2020).

[22] Razan Jaber, Donald McMillan, Jordi Solsona Belenguer, and Barry Brown. 2019. Patterns of Gaze in Speech Agent Interaction. In *Proceedings of the 1st International Conference on Conversational User Interfaces (CUI '19)*. Association for Computing Machinery, New York, NY, USA, Article 16, 10 pages. https://doi.org/10.1145/3342775.3342791

[23] Hyunhoon Jung and Hyeji Kim. 2019. Finding Contextual Meaning of the Wake Word. In *Proceedings of the 1st International Conference on Conversational User Interfaces (CUI '19)*. Association for Computing Machinery, New York, NY, USA, Article 5, 3 pages. https://doi.org/10.1145/3342775.3342805

[24] Bjørn Karmann and Tore Knudsen. 2018. Project Alias. https://bjoernkarmann.dk/project_alias. (Accessed on 02/21/2020).

[25] Minha Lee, Sander Ackermans, Nena van As, Hanwen Chang, Enzo Lucas, and Wijnand IJsselsteijn. 2019. Caring for Vincent: A Chatbot for Self-Compassion. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300932

[26] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5286–5297. https://doi.org/10.1145/2858036.2858288

[27] Neil McArthur and Markie L.C. Twist. 2017. The rise of digisexuality: Therapeutic challenges and possibilities. *Sexual and Relationship Therapy* 32, 3-4 (2017), 334–344.

[28] George Herbert Mead. 1934. *Mind, self and society*. Chicago University of Chicago Press.

[29] Hugo Mercier and Dan Sperber. 2017. *The enigma of reason*. Harvard University Press.

[30] Johan Moreno. 2019. 25% Of Google Duplex Calls Are Placed By Humans, But Likely Not For Long - Forbes. https://www.forbes.com/sites/johanmoreno/2019/05/28/25-of-google-duplex-calls-are-placed-by-humans-but-likely-not-for-long/. (Accessed on 02/21/2020).

[31] Charles Morris. 1946. *Signs, language and behavior*. Prentice-Hall.

[32] Sven Nyholm. 2020. *Humans and robots: Ethics, agency, and anthropomorphism*. Rowman & Littlefield International.

[33] Sven Nyholm and Lily Eva Frank. 2017. From Sex Robots to Love Robots: Is Mutual Love with a Robot Possible?. In *Robot Sex: Social and Ethical Implications*. Danaher, J., and McArthur, N. (Eds.). MIT press, 219–243.

[34] Mary Oliver. 2013. *New and selected poems*. Vol. 1. Beacon Press.

[35] Charles Sanders Peirce. 1991. *Peirce on signs: Writings on semiotic*. UNC Press Books.

[36] Rosalind W. Picard. 1995. Affective computing, Vol. 321. Perceptual Computing Section, Media Laboratory, Massachusetts Institute of Technology.

[37] Rosalind W. Picard. 2003. Affective computing: challenges. *International Journal of Human-Computer Studies* 59, 1-2 (2003), 55–64.

[38] Stefania Pizza, Barry Brown, Donald McMillan, and Airi Lampinen. 2016. Smartwatch in Vivo. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5456–5469. https://doi.org/10.1145/2858036.2858522

[39] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3174214

[40] Martin Porcheron, Joel E. Fischer, and Sarah Sharples. 2017. "Do Animals Have Accents?": Talking with Agents in Multi-Party Conversation. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 207–219. https://doi.org/10.1145/2998181.2998298

[41] Stuart Reeves. 2019. Conversation Considered Harmful?. In *Proceedings of the 1st International Conference on Conversational User Interfaces (CUI '19)*. Association for Computing Machinery, New York, NY, USA, Article 10, 3 pages. https://doi.org/10.1145/3342775.3342796

[42] Stuart Reeves, Martin Porcheron, and Joel Fischer. 2018. "This is Not What We Wanted": Designing for Conversation with Voice Interfaces. *Interactions* 26, 1 (Dec. 2018), 46–51. https://doi.org/10.1145/3296699

[43] Joanna Rothkoph. 2018. Single? Consider a Cross-Dimensional (Human-Hologram) Marriage - Esquire. https://www.esquire.com/lifestyle/a25018920/japan-married-hologram-gatebox/. (Accessed on 02/14/2020).

[44] Andrea Scarantino. 2017. How to do things with emotional expressions: The theory of affective pragmatics. *Psychological Inquiry* 28, 2-3 (2017), 165–185.

[45] Andrea Scarantino. 2019. Affective Pragmatics Extended: From Natural to Overt Expressions of Emotions. In *The Social Nature of Emotion Expression: What Emotions Can Tell Us About the World*, Ursula Hess and Shlomo Hareli (Eds.). Springer, 49–81.

[46] John R. Searle. 1975. Indirect speech acts. In *Speech acts*, Peter Cole and Jerry L. Morgan (Eds.). Brill, 59–82.

[47] John R. Searle. 1975. A taxonomy of illocutionary acts. In *Language, Mind and Knowledge. University of Minnesota*, Keith Gunderson (Ed.), Vol. 7: Studies in the philosophy of science. University of Minnesota Press.

[48] John R. Searle. 1985 [1979]. *Expression and meaning: Studies in the theory of speech acts*. Cambridge University Press.

[49] Selina Jeanne Sutton, Paul Foulkes, David Kirk, and Shaun Lawson. 2019. Voice as a Design Material: Sociophonetic Inspired Design Strategies in Human-Computer Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300833

[50] David R. Traum. 1999. Speech Acts for Dialogue Agents. In *Foundations of Rational Agency*, Michael Wooldridge and Anand Rao (Eds.). Springer Netherlands, Dordrecht, 169–201. https://doi.org/10.1007/978-94-015-9204-8_8

[51] Sjaak van der Geest. 2010. Rumpelstiltskin: the magic of the right word. In *Unfamiliar knowledge: Psychiatric disorders in literature*, Koos Neuvel Arko Oderwald, Willem van Tilburg (Ed.). De Tijdstroom, 37–44.

[52] Joseph Weizenbaum. 1983. ELIZA — a Computer Program for the Study of Natural Language Communication between Man and Machine. *Commun. ACM* 26, 1 (Jan. 1983), 23–28. https://doi.org/10.1145/357980.357991

[53] Mark West, Rebecca Kraut, and Han Ei Chew. 2019. I'd blush if I could: Closing gender divides in digital skills through education. Technical Report. UNESCO, EQUALS Skills Coalition. https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=7. (Accessed on 02/19/2020).

[54] Bernard Williams. 2011 [1982]. *Ethics and the Limits of Philosophy.* Taylor & Francis.

[55] Ludwig Wittgenstein. 2009 [1953]. *Philosophical investigations.* John Wiley & Sons.