Artificial Speech and Its Authors

Philip J. Nickel

[Forthcoming in *Minds and Machines*, 2013. Final version available at www.springerlink.com]

Abstract: Some of the systems used in natural language generation (NLG), a branch of applied computational linguistics, have the capacity to create or assemble somewhat original messages adapted to new contexts. In this paper, taking Bernard Williams' account of assertion by machines as a starting point, I argue that NLG systems meet the criteria for being *speech actants* to a substantial degree. They are capable of authoring original messages, and can even simulate illocutionary force and speaker meaning. Background intelligence embedded in their datasets enhances these speech capacities. Although there is an open question about who is ultimately responsible for their speech, if anybody, we can settle this question by using the notion of *proxy speech*, in which responsibility for artificial speech acts is assigned legally or conventionally to an entity separate from the speech actant.

I. Introduction

Technological advances over the last fifty years have enabled machines to simulate human speech and writing. These systems convert non-verbal information into a vocal production or a printed text. This is called *natural language generation* (NLG). Consider a program that gives a physician a morning update about the condition of an infant in a pediatric intensive care unit (PICU), based on the patient's history and instrument observations, and delivered in natural language adapted to the vocabulary used by clinicians in the PICU (Portet et al. 2009). Such an event is remarkable because it is more than just human speech delivered via technology. It appears to have an original linguistic content generated by an artifact. In this paper I ask whether NLG technology can really speak, i.e., whether it can produce speech acts with the semantic meaning and force of human speech. I argue that to a certain extent it can, and in particular that it can do so sufficiently to serve as a proxy speaker for natural and legal persons (e.g., corporations).

I coin the term *speech actants* to refer to an original source or author of a linguistic message in a way that is neutral between human speakers and other original sources of linguistic messages.[1] We usually think of communication technology as a medium for human utterances, implicitly contrasting the medium and the message. A pure communicative medium is an instrument for transmitting messages formulated independently of the medium. Examples of relatively pure media are telephones, recording-and-playback devices, and printed books. The message, by contrast, is an original linguistic item used to express something in context. Sometimes the content of messages is impacted by the nature of the medium. For example, televised news constrains the length and detail of sentences, and some kinds of social media messages composed and read on portable electronic devices must be very short and thus depend heavily on the common knowledge of the recipient(s). Here the medium is impure: this is the kernel of truth in McLuhan's dictum "the medium is the message" (1964). In these cases, though, we can still identify a human as the sole author of the message. In order for a communication technology to become an author of a message, it must do more than shape human messages. Speech actants are not merely media for the transmission of preformulated messages, but create their own messages with original meanings, or substantially contribute to such messages.

II. Criteria for being a speech actant

In this section I spell out the notion of a speech actant in more detail. In his 1637 work *Discourse on the Method*, René Descartes considers the extent to which meaningful speech can be produced by a machine without a soul (a classification which, for him, included non-human animals):

> [W]e can certainly conceive of a machine so constructed that it utters words, and even utters words which correspond to bodily actions causing a change in its organs (e.g., if you touch it in one spot it asks what you want of it, if you touch it in another spot it cries out that you are hurting it, and so on). But it is not conceivable that such a machine should produce different arrangements of words so as to give an

[1] The term "actant" is due to Latour (1999), who uses it to explain phenomena in terms not reducible to individuals, social groups, institutions or artefacts on their own. My usage is much more limited and is not intended to rule out such a reduction.

> appropriately meaningful answer to whatever is said in its presence, as the dullest of men can do (Descartes 1985 [1637], 140).[2]

In this passage Descartes appeals to the linguistically productive complexity characteristic of an intelligent, authorial voice, making clear that we need an account of the ability to "give an appropriately meaningful answer to whatever is said in its presence" in order to explain and clarify the notion of a speech actant. There is no sharp dividing line between Descartes' push-button assertions and those with more authorial complexity, but I will give a rough account. It is a matter of degree the extent to which a given entity satisfies the conditions.

The criterion for being a speech actant should not be that of successfully imitating human speech. Early efforts to create speaking machines tried to meet a Turing Test criterion of intelligence, attempting to make it difficult for a listener to distinguish artificial interlocutors from human ones (Weizenbaum 1966). But after a period of extraordinary optimism about the prospects of automated machine translation and other forms of computer-generated natural language interface, there were setbacks. The initial ambitions for artificial intelligence employing natural language were scaled back or given a longer timeline (Madsen 2009). At the same time, it was accepted that NLG technology could be useful without passing the Turing Test, marking a shift from an anthropomorphic to a pragmatic standard. Current NLG focuses on providing useful information in specific contexts, and speech outputs are evaluated by how well they enable human task-performance, or by other subjective measures of output quality, rather than by human-likeness (Reiter & Belz 2009). Although *naturalness* is mentioned as an important anthropocentric dimension of quality, it is also noted that "[t]here is no guarantee that a human-produced description is a good one" (Dale *et al.*, 2005).

Bernard Williams, in an essay on the nature of belief, describes an articulate machine with four main capacities necessary (but not sufficient) for belief and assertion (1973, 145).

[2] This passage was brought to my attention by Salinga & Wuttig 2011.

First, it can make statements and distinguish them from mere hypotheses. Second, it embodies inferential processes, representing and responding to rational relations between propositions. Third, the machine gathers information and makes observations about its environment, adapting the statements it is prepared to assert accordingly. And finally, it is capable of being insincere. These are meant as criteria for being the kind of thing that really makes assertions, not as criteria for a given utterance to count as an assertion. Individual assertions could deviate substantially from the criteria, being false, irrelevant, etc., and still count as genuine assertions because they are made by an entity that satisfied these criteria generally.

For Williams, distinguishing between statements and hypotheses shows that the machine is sensitive to the truth-evaluable nature of assertion: an assertion, whatever the evidence for it, must be evaluated as satisfactory in a very important respect if it is true and unsatisfactory if it is false, and this is not what we would say about hypotheses, which are often satisfactory even if they turn out to be false.[3] At the crudest level, sensitivity to the truth-evaluable nature of assertion simply means that the machine produces mostly true assertions. But it can also be exhibited by systematic co-variations between assertions and the state of the world. This is compatible with making claims that are false — so long as they are not randomly false, but show rational co-variation with the world and the available evidence. At the same time, the ability to draw inferences and make observations about the environment is necessary for the capacity to have beliefs and make assertions because it is important that beliefs and assertions be adjusted to available, salient evidence, as well as to the total present collection of logically related assertable statements currently "held" by the machine. In addition, Williams argues that a machine does not have the capacity to be

[3] This is not to say that hypotheses are not truth-evaluable, only that their implicit purpose is not always undermined if they turn out to be false. Some hypotheses (such as those putting forward a claim to be investigated) need only be *plausible* in order to be satisfactory. Others, such as those used in proving a *reductio ad absurdum*, need not even be plausible in order to be satisfactory.

insincere if the connection between its representational state and what it "says" is unmediated:

> [T]here is a direct route from the state that [the machine] is in to what it prints out; … if something goes wrong on this route, it goes mechanically wrong, that is, if something interrupts the connexion between the normal inner state for asserting that *p* and its asserting that *p*, and it comes out with something else, this is merely a case of breakdown. It is not a case of insincere assertion, of its trying to get you to believe that *p* when really all the time it itself believes that not-*p*: we have not given it any way of doing that. … [W]hen I said this machine made assertions, I should have actually put that in heavy scare-quotes; 'assertion' itself has got to be understood in an impoverished sense here, because our very concept of assertion is tied to the notion of deciding to say something which does or does not mirror what you believe (1973, 145–6).

For Williams, it would take a machine with intentional states to achieve genuine insincerity. So for him the possibility of computerized speech actants is highly restricted.

Although Williams' conditions are a good starting point, they are too restrictive as an account of speech actants. In part, this is because they are not meant as an account of speech actants in general, but only for those whose aim is to produce the speech act of assertion, not other speech acts such as questions or commands. Even in the case of assertion, truth is not the only thing we care about when evaluating its intelligibility. We care just as much about salience, consistency and informativeness. Furthermore, sensitivity to the truth-evaluable nature of assertion and to evidence might be exhibited in other ways besides being able to make hypotheses and distinguish them from assertions. If the actual assertions the machine makes are sufficiently grounded in evidence and sensitive to changing circumstances, this exhibits truth-sensitivity even if the machine does not make any hypotheses. Also, if the machine is in a position to perform movements, its movements could correspond to its assertions, showing alignment between its utterances and behavior.

A particular problem for machine speech is the possibility of insincerity. It is this condition that Williams believes most clearly fails in the case of a machine. However, I think we can make sense of a kind of insincerity in automated speech, and therefore sincerity as well, by thinking about variations on ordinary NLG systems. For a system with a given purpose or function, the relation between its representational states and its statements is mediated by norms of function in such a way that we can speak meaningfully of insincere utterances. When the *function* of a machine is to report the truth, it is possible for its utterances to count as insincere in the event that it deviates from the norm of truthfulness for non-mechanical reasons. This is parasitic on ordinary norms of assertion, which has as a speech act the function of representing the truth. Suppose a coffee retailer pays the engineers of an automobile navigation system to design it so that it misstates the fastest route, guiding drivers past its coffee stores. In cases where we can specify a norm of assertion for the function of the machine ("stating the fastest route"), a kind of insincerity is possible if the statements are purposely unreliable as judged by that norm. When a designer is aware that the machine has this function, and that the corresponding norm of assertion applies, but intentionally programs it to not to meet the norm for some ulterior end, the machine creates insincere speech. Since this is a live possibility — existing systems are held to this norm — the relation between the internal representations and the outputs is never *purely* mechanical: reference to the design is necessary to explain why the machine is or is not truthful. Because this normative level of explanation applies, we can meaningfully judge computer speech as *sincere*, in a relevant sense, in cases where the system is truthful and where designing it to be untruthful never crossed the mind of the engineer.

With these points in mind, let us present a new set of conditions based on those of Williams. I propose that a speech actant is an entity that produces linguistically meaningful messages for which:

(1) the content and force of the message is causally due to the entity and conditioned by its generative inferential and linguistic activity;

(2) the message is delivered actively (it is *uttered*);

(3) the entity is usually sensitive to the evaluation-conditions for the utterance in the contexts of delivery (e.g., relevance);

(4) more specifically, if the entity presents something as true, it is (usually) responsive to relevant evidence, to its other logically related representational and behavioral states, and to the truth; and

(5) the message could in principle be insincere, in the sense that it deviates intentionally ('by design') from relevant norms of assertion.

For the speech act of assertion, sensitivity to evaluation-conditions will mean that the entity must by and large show itself to be attuned to the way the world is, normally by representing or attempting to represent it accurately.[4] Otherwise it will not (continue to) be interpreted as making assertions at all. Normally, we do our best to interpret apparently meaningful speech as meeting these conditions. If we cannot do so or if our interpretation proves consistently wrong in the long run, we cannot regard the entity as making assertions.

III. Applying the criteria to NLG technologies

In this section I will apply the criteria developed above to existing NLG technologies, arguing that some of them count as speech actants to at least a limited extent. I begin with the point that some message-producing technologies clearly fail to qualify as speech actants because they fail to contribute sufficiently to an original linguistic output. Some technologies are thus clearly excluded by the above criteria:

- *Simple instruments displaying or indicate a message based on a physical property.* A medicine cabinet thermometer gives a temperature reading by coupling the mechanism of glass tube and enclosed liquid with an output in a symbolic system. The numerals on the outside of the tube indicate such propositions as "The temperature of the item measured is 37 degrees." (We could even imagine a thermometer which spelled this out in very small letters.) Such an instrument is

[4] Insincere assertions, which show a kind of attunement to the world without attempting to represent it accurately, would be an exception.

highly truth-sensitive and points at a message, but does not contribute inferentially to the message or combine propositions synthetically.

- *Simple artifacts using language whose content is indexically or demonstratively determined by the context*. A mechanized sign that reads "You are our (*n*th) drive-thru customer," with a digital readout that increases the value of *n* by one with each passing vehicle, creates a modestly different, personalized, and informative semantic content for each vehicle. Although the semantic content differs for each car that drives through, the resulting message does not involve sufficiently complex interaction between different sources of information nor sufficiently complex propositional or linguistic manipulation to count as the author of the message.
- *Machine translators and summarizers.* Machine translation depends on large amounts of natural language data (parallel texts in two languages, the equivalent of facing-page prose) for converting text chunks from one language to another. Machine summarization depends on source texts from which whole sentences selected as highly salient by system algorithms are extracted and reformulated for coherence (Jurafsky & Martin 2009). Despite the complex linguistic manipulation of such systems, they make little or no authorial contribution to the message. Their function is to re-convey existing messages rather than to formulate new messages or utter them. They are media: they make no extra-linguistic observations and no propositional inferences going beyond their source texts.

NLG technologies share some features with these other cases, but they also do more, combining observations from different sources and making situationally appropriate messages using generative language capacities.

But we must move carefully here. There is substantial difficulty in showing that both the “content” and the “force” from condition (1) in the criteria set out above are satisfied by non-human speech actants. The speech actant must not only say something, but also to *do something* with its words. It is useful to examine two relevant distinctions between content and force, both widely accepted by linguists and philosophers of language (although there have been disputes about their analysis and significance). The first is due to J.L. Austin, who distinguishes a *locutionary act*, by which one utters certain definite sounds in a certain construction in such a way that they have a referent (1962, 94ff.); an *illocutionary act*, by which one uses a locutionary act for one or another standard communicative purposes, e.g., making a statement, asking a question, threatening, etc.; and a *perlocutionary act*: an act of bringing about a change in the attitudes and behavioral dispositions of one’s audience by so speaking. The locutionary act is the content of the utterance in an attenuated sense: the proposition indicated by the utterance (e.g., that the prince is wearing a wig), abstracted away from being used in an assertion (‘The prince is wearing a wig’), a question (‘Is the prince wearing a wig?’), a threat (‘If the paparazzi insist on attending the dance then the prince will be wearing a wig’), etc. The illocutionary act and the perlocutionary act describe the force of the utterance or what it is used to *do*. *Asserting* that the prince is wearing a wig is one of the standard illocutionary acts I can perform with an utterance. Normally I make an assertion by using a declarative sentence, pertaining to a question that is relevant and unresolved in my conversation. This illocutionary act can have different effects on the hearer, and typically I intend to bring about certain of these effects. For example, my aim may be to get the believer to accept my claim. If I succeed, then in addition to the illocutionary act, I have also performed the perlocutionary act of *informing* or *persuading the hearer* that the prince is wearing a wig. Can NLG systems assert, inform and persuade?

The second distinction between content and force is due to H.P. Grice, who pointed out that the idea of meaning includes both *sentence meaning* and *speaker meaning* (Grice 1989). In many cases I can successfully communicate one thing by saying something else entirely, e.g., cases of *conversational implicature* in which the hearer is required to make an inference using norms for conversation (*conversational maxims*) as premises. Grice's classic example is a letter of recommendation for a student stating nothing more than that he comes to class regularly and uses grammatical English. By saying this, the recommender means that the candidate is unsuited to the position (speaker meaning), although this is not part of what is said (sentence meaning). The audience reasons that if the speaker's meaning were identical with the content of his letter, he would be violating norms of communication egregiously. Only a stupid or pigheaded letter-writer would think it appropriate to base a recommendation on the indicated qualities alone. Therefore the meaning must be something else. In the context, the most plausible supposition is that the letter-writer means to say that the candidate is very weak. This makes it clear that the literal (sentence) meaning and the intended (speaker meaning) come apart. However, even in cases where linguistic meaning appears identical to what is communicated and there is no conversational implicature, what is communicated is still the result of an interpretive act (see Barber 2010 for an argument to this effect, drawing on Davidson 1986).

I argue that existing NLG systems satisfy the conditions for being speech actants to a substantial degree. First I try to substantiate the claim that they make an authorial contribution to a message in the sense that they determine its locutionary force or sentence meaning. I then take up the issue of speaker meaning, illocutionary force, and perlocutionary force, arguing first, that NLG systems can easily be designed to make utterances that have these kinds of meaning or force; and second, that although they might not themselves be the "agent" behind this force, they can serve as proxy speakers on behalf of a legal or natural person.

In seeing what contribution existing NLG systems make to the content of their utterances, it is useful to consider what they actually do. Although there are many different applications of NLG, a familiar one is to provide driving, walking or other directions in natural language on the basis of a start- and endpoint and a database consisting of a map of navigable roads and paths (a Geographical Information Systems [GIS] dataset). Roughly speaking, given a start- and endpoint, the GIS is used to generate a fastest or shortest route consisting of points connected by distances. The NLG software transforms this ordered sequence of named points and distances into a set of verbal instructions for a driver or pedestrian. In the simplest version of sentence formation, the names of the points are simply inserted into sentence templates for each segment of the instructions. This is a rote task: "[I]t is straightforward to write a generator that produces impressive text by associating a sentence template (or some equivalent grammatical form) with each representational type and then using a grammar to realize the template into surface form" (Hovy 1990). A domain like route directions seems ideal for such a method. In the case of route directions, however, this results in a repetitive and overly-detailed set of instructions relying only on street names and distances to identify one's location, making the directions difficult for humans to remember and follow (Young 1999). The route directions can be improved in various ways: by adding other information, such as point-specific descriptions of salient geographical landmarks (Klippel & Winter 2005); through better discourse planning, structuring the information in parts ("chunking") to make it easier to remember (Klippel *et al.*, 2003, Dale *et al.*, 2005); and through personalization or tailoring to the user and situation, e.g., his degree of familiarity with the area, communication style, travel preferences, GPS location, and up-to-date information about road conditions or traffic.   As Hovy (*op cit.*) argues, in successful language use we adapt our message to the audience and situation. This can be accomplished by eliciting user input or incorporating data about user behavior into the inputs of the system (Richter *et al.*, 2008). Taking these factors into account, it is clear that the systems create complex, original, conversationally relevant contents, and goes some way toward showing that they meet criteria (1), (3) and (4) from the account of speech actants. Their linguistic productions are original, truth-sensitive, and relevance-sensitive.

I offer two main observations about such technologies. First, they involve narrow contexts of speech, in which the amount of linguistic generativity needed by the system is limited because the grammatical forms are routine and the vocabulary narrow. A system for producing route directions does not generally make observations about the scenery along the road or entertain its hearer with jokes about street names.[5] The system need not approach complete facility with natural language grammar or pragmatics in order to create original messages suited to the task at hand. In route direction technology, the language generation module consists of templates — predefined sentences with gaps that are filled in and structured based on the GIS database, global positioning system data and user input. In systems with a less routine type of linguistic output, probabilistic word sequence generators can be used with an accompanying filter that weeds out ungrammatical sentences (Macherey 2009, 24-25). A computer can draw on the human linguistic intelligence embodied in collections of grammatically well-formed texts available in databases or on the Internet. Using algorithms to analyze the most common word combinations in these texts it can select appropriate phrases and assemble them into a grammatical, naturally arranged discourse. These systems can produce a virtually infinite number of original messages.

Second, rationality or human intelligence is often built into the source data and the algorithms that use it. The GIS dataset, the "map" used by route directions software,

[5] However, there is an effort to use landmarks as points of orientation, as mentioned above (Klippel & Winter 2005). In addition, humor has been attempted in NLG using narrowly defined forms such as punning (Strapparava et al. 2011).

represents a domain that is already structured by millennia of human spatial engineering, e.g., centuriation and grid planning (see Castagnoli 1971). Beginning with what is already a structured network of navigable, named roads and paths, designers impose further rational structure by creating a coherent, digitally tractable representational model of planar space, based on satellites and existing maps, that is ready for linguistic use. The algorithm using this data to generate material for route directions assumes the perspective of a prioritized self with a location and finite movement within planar space according to that self's aims and priorities (speed, fuel efficiency). These building blocks of a route directions system are conceptually and etiologically inseparable from their manifestation in human language, because the relevant domain of our vocabulary and language practices for route directions co-developed with human land segmentation and the building of roads, as well as human aims in moving about a world with such a structure. Hence there is already linguistic rationality in the domain represented by the source data, in the structure of the source data itself, and in the algorithms that use the source data to provide inputs to NLG systems. Applied computational linguists draw on this rationality when they use such datasets and algorithms as source material for linguistic production. In this way even simple recombinations of words related to the GIS dataset produce utterances going beyond Descartes' push-button assertions in their generativity.

The large and highly structured dataset is partly what explains the independent authorship of NLG systems. It is doubtful that any individual human without a computer can provide me on short notice with driving directions from Amsterdam to Irkutsk, but Google Maps can provide them almost instantly, along with helpful remarks such as that the route partly consists of toll roads, and that after turning left on the way to Volgograd Street (in step 110) one has to drive through a roundabout. The content of these directions, complex and perhaps never considered by any human being until now, are causally due to the computer system, not directly traceable to any engineer or corporate entity. By contrast, it cannot be said of the content of the "utterance" of the thermometer or mechanized sign that it is due to the device. In the case of the thermometer, it is the intention of the device's designer that it always give a reading corresponding to the environmental temperature. Without even looking at the device, the designer can predict with great accuracy *exactly what it will say* so long as she already knows the "input." Similar remarks go for the mechanized sign. But this is not true of Google's route directions. Although mechanical, they are not a direct function of their input, but of a vast database, a complex computational algorithm, and a complex NLG system. The content of the directions is not the direct causal production of the engineer.

However, this by itself does not show that computers can *do* anything with the sentences they construct. For although the computer constructs the sentences composing the route directions, it is not clear that *it* performs any illocutionary or perlocutionary act with it, or can have speaker meaning. These may trace back to somebody or something else, or to nobody at all. I wish to make two points about this. First, I argue that NLG systems can easily be designed to carry out speech acts with a force going beyond the semantic meaning of the words they deploy, and can exploit speaker meaning. On their face, the utterances *have* force. But the crucial further question to be answered is *whose acts these are*: whether the force of the assertion and testimony traces to the system, or to the designer and/or deployer of the system, or to nobody at all. In our formulation of the conditions on being a speech actant, we held that both the force and the content must be original and due to the inferential and linguistic generativity of the system. Hence if the force traces not to the system, but to the designer, or nowhere, then it appears that the systems we have been discussing are not speech actants — since they do not perform speech acts in the relevant sense. In response to this further question I reply in the last section of the paper by arguing that the force does not have to trace to the system in order for them to engage in what I call *proxy speech*.

A computer system can easily be designed to employ the forms of speech associated with promising, threatening, assertion, questions, and so forth at appropriate moments in a (conversational) situation. Their utterances appear to have illocutionary force. They can equally be designed to exploit conversational implicature. One might imagine automatic text generated to inform a patient of a diagnosis after a test for cancer using the expression, “We are sorry to deliver bad news” without ever stating expressly that the recipient has cancer. The context is already given and the designer can easily anticipate it. The choice whether to use direct statement or implicature could easily be adjusted to the recipient based on prior dialogue, and the exact text used could vary with the context. This demonstrates that if NLG computers are speech actants at all, their outputs apparently can easily be given speaker meaning as well as linguistic meaning, force as well as content.

However, a potential objection to the idea that a computer’s utterances have illocutionary force or speaker meaning is that these concepts are typically analyzed in terms of a speaker’s intention to produce certain effects in an audience, through the audience’s recognition of those very intentions (Grice 1989, 99ff.; Strawson 1971). Since simple computer programs cannot have intentions, this appears to imply that they also cannot deploy speaker meaning. Two responses are in order here. First, it may be possible to modify the Gricean analysis of conversational implicature to allow for the case of computer speech. Grice himself extensively revised his initial analysis of speaker meaning to account for complex counterexamples, and although none of these revisions abandons the idea that speaker intentions are central, one should not prejudge whether and how we should reformulate the analysis without first looking at the phenomena.

Second, much of the interest of Grice’s analysis is found in the mental attitudes the speaker anticipates in the hearer, rather than the mental attitudes (i.e., the intentions) of the speaker himself. The audience is expected by the speaker to engage in a complex act of interpretation. It seems likely that we could describe this act of interpretation of a message by a hearer without making reference to the computer’s intentions, either by reference to the intentions of the designer, or by framing the analysis in terms of what a human speaker *would* intend by uttering such words in the context. The hearer can, if it is useful, temporarily adopt a self-consciously fictional “intentional stance” toward the computer’s utterances, along the lines of the process of verbal interpretation of experimental text described in detail by Dennett (1991, 74-78), in order to decode its meaning, without really ascribing consciousness or intentions to it. The fact that the computer does not actually have intentions is not a barrier to the hearer’s reconstruction of speaker meaning. Their utterances have apparent force, although this leaves it an open question whose force it is.

To sum up, then, existing NLG systems have some characteristic features related to linguistic agency: (a) they have narrow but highly productive grammatical generativity; (b) they can draw on a high level of background human rationality, including symbolic and linguistic rationality, embodied in their source data; and, (c) they can show sensitivity to situational and pragmatic features of language use, and their utterances can simulate speaker meaning. What these systems lack is a high level of overall situational awareness and general adaptive intelligence. They are tools designed for specific purposes and contexts, not intelligent agents with adaptive all-purpose observational, representational and intentional capacities. Coming back to our earlier criteria for being a speech actant, these features allow NLG systems to meet the first, causal, inferential and generative contribution condition for speech actants to a substantial degree.[6] Furthermore, they can deliver utterances actively (condition (2)). In addition, they have been successfully designed to be truth-sensitive, which is the relevant kind of evaluation sensitivity for assertion — allowing them to meet criteria (3) and (4) for being a speech actant. And finally, they can be sincere or insincere in the sense described in section II, meeting condition (5). In a minimal but significant (and useful) sense, then, they are capable of (co-)authoring artificial spoken and written messages. However, the powers of linguistic generation, situational awareness, and adaptive intelligence of existing applied NLG systems are very narrow and limited, and this restricts the extent to which they fully satisfy the first condition. Also, although their utterances have *prima facie* force, we are left with the question *whose force it is*, since we normally understand this in terms of agency and intentions. I address this in the last section.

IV. Proxy speech and responsibility

In this section I provide a tentative answer to the question, "Whose speech is it?" regarding artificial speech. I argue that the answer can be given by dividing the authorial role between an artificial speech actant, which gives the speech act content and form, and a holder of ultimate responsibility, to which the force of the speech act traces back. Ultimate responsibility for artificial speech does not lie with machines, but either with persons or companies, or with nobody at all. It is absurd to hold a computer system accountable for a promise or assertion that it makes. Although I may be able to acquire knowledge from the assertions of a machine, at the same time the machine cannot vouch for its own reliability in the sense of taking responsibility for what it said. Therefore, if we are to avoid having nobody responsible for what an artificial speech actant says, we may need to assign ultimate

[6] Not all of the criteria are susceptible of satisfaction to a degree, but the criterion of originality of content is. There is no obvious threshold of originality beyond which we ascribe unambiguous authorship to a given entity. For example, in the adaptation of a story one has heard earlier for a new purpose or audience, there is no obvious point at which one becomes a co-author or sole author.

responsibility for it. If responsibility is assigned to the person or company deploying the artificial speech actant, then the force of the utterance traces back to a standard kind of agent.[7]

We can make better sense of this situation by pointing out a specific kind of linguistic agency of which NLG systems seem capable: *proxy speech*. Consider an analogy. Suppose Bart sends his eight-year-old daughter Lisa to the store with some money to buy a bag of whole wheat flour. He does not know exactly what it will cost, and he does not remember what brands or kinds of whole wheat flour there are. He therefore leaves the details to his daughter. Lisa needs to be a speech actant (and an agent in general) to some extent in order carry out Bart's request. She needs to be somewhat situationally adaptive and capable of using speech effectively in the narrow context in which she operates. But she need not have a general capacity for effective agency, or speech agency in order to represent Bart as his proxy for this type of transaction. She can focus on what is required in this limited context. Similarly, NLG systems do not need to have general situational awareness, adaptive intelligence and unlimited linguistic generativity in order to perform speech acts on behalf of some other agent. Proxy speakers need to be able to speak, but their autonomy in doing so can be limited compared with full-fledged speakers.

There are two points to be derived from this. First, it allows us to feel more comfortable with the conclusion that NLG systems can author speech. In the previous section I argued that NLG systems can easily emulate speaker meaning and carry out various apparent illocutionary and perlocutionary acts, but left it open whose acts they ultimately are — to whom the force of the speech traces. When artificial speech actants are embedded in a proxy speech relationship, this question is answered.

[7] For the moment we shall set aside doubts about whether the notion of agency can apply to artificial legal persons such as corporations.

Second, there is an important institutional and/or legal dimension to the proxy speech relationship. To some extent the idea of a proxy speaker is a natural and familiar one. Although Lisa acts on behalf of Bart, so long as Lisa did her job adequately Bart is ultimately responsible for her purchase. The question, “Why did you buy whole wheat flour?” is deflected to him. Similarly, in the case of proxy speech, where there is a clear relationship of proxy established in order for the speech actant to succeed, this proxy relationship indicates to whom justificatory questions are ultimately directed. When the proxy relationship is casual and little is at stake, there is no need for a formal framework to assign responsibility to the ultimate agent. But in serious matters there is a danger that the relationship is unclear, and even that the ultimate agent will use the proxy speaker to muddy the waters, claiming that its speech is independent of the intentions of the ultimate agent to such a degree that the ultimate agent is not responsible for what it said (or better, for what it did with its speech). For example, the company deploying a dialogue system that sells airplane tickets might claim, on a day when the system did not take recent price hikes into account, that the agreements it reached were not binding because they did reflect the company’s intentions. (They are arguing, in effect, that in this case nobody at all is responsible for the computer system’s speech.) Judges and legal scholars have begun to address the issue of whether contracts agreed to by electronic agents are legally enforceable, and under what conditions (Bellia 2001). Whatever is decided in different jurisdictions, these decisions will affect the way that actual hearers understand artificial speech, and how seriously they take it. Unless and until computers become fully-fledged agents, legal and institutional norms will partly settle the question on when, and on whose behalf, a computer speaks, and who (if anybody) is responsible when their speech is inadequate in some respect.

Acknowledgements: For comments that helped improve the paper, I would like to thank audiences from Delft University of Technology, Eindhoven University of Technology, and the University of Twente, and especially an anonymous reviewer for this journal.

References

J.L. Austin, *How to Do Things with Words*, New York: Oxford, 1962.

Alex Barber, “Idiolects,” *The Stanford Encyclopedia of Philosophy* (Winter 2010 edition), E.N. Zalta, ed., URL=<http://plato.stanford.edu/archives/win2010/entries/idiolects>.

Anthony J. Bellia, Jr., “Contracting with Electronic Agents,” *Emory Law Journal* 50 (2001): 1047–1092.

Ferdinando Castagnoli, *Orthogonal Town Planning in Antiquity* (MIT Press, 1971).

Robert Dale, Sabine Geldof and Jean-Philippe Prost, Using Natural Language Generation in Automatic Route Description, *Journal of Research and Practice in Information Technology* 37, 1 (2005): 89–105.

Donald Davidson, “A Nice Derangement of Epitaphs,” in E. Lepore, ed., *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* (Cambridge, MA: Blackwell, 1986).

Daniel Dennett, *Consciousness Explained* (Boston: Little, Brown and Company, 1991)

René Descartes, *Discourse on the Method*, in *The Philosophical Works of Descartes, vol. I*, trans. John Cottingham, Robert Stoothoff, & Donald Murdoch, New York: Cambridge University Press, 1985.

H.P. Grice, *Studies in the Ways of Words* (Cambridge, MA: Harvard University Press, 1989).

Eduard H. Hovy, “Pragmatics and Natural Language Generation,” *Artificial Intelligence* 43 (1990): 153-197.

Daniel Jurafsky & James H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* 2nd Ed. (Prentice Hall, 2009).

Alexander Klippel & Stephan Winter, “Structural Salience of Landmarks for Route Directions,” in A.G. Cohn & D.M. Mark, eds., *COSIT 2005, LNCS 3693* (Springer-Verlag, 2005): 347–362.

A. Klippel, H. Tappe, C. Habel, “Pictorial Representations of Routes: Chunking Route Segments During Comprehension,” in C. Freksa et al., eds., *Spatial Cognition III, vol 2685* (Springer, 2003): 11–33.

Bruno Latour, “Morale et Technique: la Fin des Moyens,” *Réseaux* 100 (1999): 39–58.

Klaus Macherey, *Statistical Methods in Natural Language Understanding and Spoken Dialogue Systems* (Dissertation, RWTH Aachen University, 2009).

Mathias Winther Madsen, *The Limits of Machine Translation* (Masters Thesis, University of Copenhagen, 2009).

Marshall McLuhan, *Understanding Media: The Extensions of Man* (McGraw Hill, 1964).

Francois Portet *et al.*, “Automatic generation of Textual Summaries from Neonatal Intensive Care Data,” *Artificial Intelligence* 173, 7-8 (2009), 789–816.

Ehud Reiter & Anja Belz, “An Investigation into the Validity of Some Metrics for Automatically Evaluating Natural Language Generation Systems,” *Computational Linguistics* 35, 4 (2009): 529–558.

Kai-Florian Richter, Martin Tomko, & Stephan Winter, "A dialog-driven process of generating route directions," *Computers, Environment and Urban Systems* 32 (2008): 233–245.

Martin Salinga and Matthias Wuttig, "Phase-Change Memories on a Diet," *Science* 332 (2011), 543.

Carlo Strapparava, Oliviero Stock, & Rada Mihalcea, "Computational Humour," in P. Petta et al., eds., *Emotion-Oriented Systems* (Berlin: Springer-Verlag, 2011): 609-634.

Peter F. Strawson, "Intention and Convention in Speech Acts," in *Logico-Linguistic Papers* (London: Methuen, 1971).

Joseph Weizenbaum, "ELIZA—A Computer Program for the Study of Natural Language Communication between Man and Machine," *Communications of the ACM* 9,1 (1966): 36–45.

Bernard Williams, "Deciding to Believe," in *Problems of the Self* (Cambridge, 1973): 136–151.

R. Michael Young, "Using Grice's Maxim of Quantity to Select the Content of Plan Descriptions," *Artificial Intelligence* 115 (1999): 215–256.