

Eeke van der Wal

Dragon
NaturallySpeaking:
Being Listened
to and the
Subservience
of Speech

abstract Through an analysis of the relation between the speech recognition software Dragon NaturallySpeaking and myself, as user, I argue in this paper for an understanding of listening as an active determinant in the relation between listener and speaker, instead

of a conception that merely infers the act of receiving and obeying. I observe that although the software is marketed as a technology that would obey by listening to the user's commands, my experience with the software points to another direction.

As a computer operating subject, I am dependent on Dragon's recognition of speech. Drawing on Hegel's MASTER-SLAVE DIALECTIC, this paper argues that Dragon is an active participant in the relation between user and technology, rather than a mere tool. Following Karen Barad, I highlight the material-discursiveness of speech. Rather than focusing on meanings, Dragon attunes to — or listens for — the materiality of speech through its recognition of phonetic speech structures. As such, the article moves away from an anthropocentric understanding of listening.

Approximately two years ago, I incurred a repetitive stress injury due to excessive computer use, and it was then that I encountered Dragon. As a result of the injury, I am unable to type on the computer keyboard for a long period of time without pain, and to recover I have to unburden my arms, neck and shoulders by drastically reducing the amount of time spent (typing) behind a computer. In this digitalized and highly computerised 21st century, these circumstances are less than ideal. Dragon, however, offered me a solution.

Dragon NaturallySpeaking is a speech recognition software for the computer which allows the computer to be driven by the user's voice. On the website of *Nuance*, the company which sells this software, Dragon is presented as a tool for writing that enacts the same function as the computer keyboard: "Dragon lets users create and edit documents, send emails, and search the web through speech," making the work behind the computer "virtually hands-free." This software program is introduced as an invaluable piece of technology or tool for the writing and computer-using human subject — in particular for those who might otherwise be unable to use a computer due to (physical) disabilities such as arthritis, carpal tunnel syndrome, or repetitive strain injuries like my own. It is this intriguing marketing narrative through which Dragon becomes understood as the tool or technology that listens and adheres to the user's speech, as made apparent by quotes on the *Nuance*

website such as “control your computer by voice with speed and accuracy,” “let Dragon work for you,” and “say commands and your computer obeys.” Dragon is, thus, presented as a listening and obeying technology, therewith positing the speaking user in an undisputed position of control.

A conception of listening to speech in which listening merely infers the act of receiving and obeying emphasizes the importance of speaking at the expense of listening (Fiumara 31; Lipari, “Rhetoric’s other” 228). Departing from Hegel’s MASTER-SLAVE DIALECTIC, I instead set out to illustrate how Dragon’s ability to recognise speech through the act of listening disputes the conception of an independent subject who is able to ‘control’ the computer through speech. Instead, listening becomes the key concept and practice through which Dragon mediates speech, which is essential for the successful operationalisation of the computer. Moving away from the implicitly anthropocentric understanding of listening, whereby listening tends to be understood in relation to human consciousness and aurality, i.e. hearing through the ear (Gallagher et al. 619), I here adhere instead to an expanded conception in which listening attends to any and every kind of kinetic oscillation that occurs between humans, animals, technologies, materials and environment (Gallagher et al. 620). Thence, I here attune to the disparate and dynamic relation between listener and speaker, between Dragon and myself, analysing how

Dragon's ability to listen mediates speech, and by extension the operationalisation of the computer, therewith governing its own coming into existence as subservient software technology.

On Being Recognised

Ironically, any expectation of myself as an independent user in control was always already an impossibility, as it is only through Dragon that I can at present actualize my desire and potential to work with a computer. Nevertheless, this dubious presumption was reiterated in my own, somewhat naïve, expectations of engaging with the software. It was my implicit expectation that, after purchasing Dragon, the process of writing and working with a computer could be resumed without further ado — as the website promised: “your voice is ready for work.” The reality of working with Dragon, however, reveals that my voice was never ‘ready’ to work; instead, our relation is much more reciprocal, complex and frustrating. Firstly, Dragon is able to recognise pre-programmed commands that allow the user to operate the computer, which means that the user has to get acquainted with these commands first in order to be (en)able(d) to control the computer. *I, however, did not know beforehand that in English “()” are called ‘parentheses’ and I had to learn that Dragon responds to commands such as “open parenthesis,” “close parenthesis,” “open quote,” “comma,” “full stop,” “new line,” etc.* Secondly, Dragon attunes and responds to the speaker's voice and

intonation, which allows for Dragon to distinguish between utterances that are supposed to effectuate a command and utterances that are to be translated into text on screen. The user has to therefore learn when, for instance, the utterance ‘enter’ results in Dragon performing the function of ‘enter’ as the enter-key on a keyboard would, and when the same utterance effectuates the written word ‘enter’ on screen. As such, Dragon dictates the way in which I speak and compose my spoken sentences, as the spoken version of the italicized sentence above illustrates. This sentence was produced as a result of the utterance: “I *comma* however *comma* did not know beforehand that in English *open quote open parenthesis close parenthesis close quote* are called *open single quote* parentheses *close single quote* and I had to learn that Dragon responds to commands such as *open quote open parenthesis close quote comma open quote close parenthesis close quote comma open quote open quote close quote comma open quote comma close quote comma open quote full stop close quote comma open quote new line close quote comma* etcetera *choose option one full stop.*”⁽¹⁾

In the experience of working with the speech recognition software, this presumed unilateral relation between myself as the independent user (who speaks and is in control) and Dragon as my executive software who listens and obeys then becomes unsubstantiated, as Dragon imposes onto the user certain preconditions through which it can operate. The fallacy of a self-conscious human being who is in control has already been heavily

therefore the dominant and ostensibly independent SELF remains dependent on its relation with its subservient other (Kamal 461-464).

While Hegel analyses this complex relation of mutual dependence between two (human) self-consciousnesses, I infer here that a similar, complex relation of mutual dependence becomes discernible in the encounter between Dragon and myself. In order to focus on this particular relation, I adhere to a posthumanist perspective, which decentres the human, and thence the predominance of human agency — “the ‘man’ as alleged ‘measure of things’” (Braidotti 637). As Donna Haraway infers in her 1991 essay “A Cyborg Manifesto”:

It is not clear who makes and who is made in the relation between human and machine. It is not clear what is mind and what body in machines that resolve into coding practices. Insofar as we know ourselves in both formal discourse [...] and in daily practice [...] we find ourselves to be cyborgs, hybrids, mosaics, chimeras. Biological organisms have become communications devices like others. There is no fundamental, ontological separation in our formal knowledge of machine and organism, of technical and organic (177-178).

Building on this posthumanist perspective, Dragon is no longer simply a tool or technology to be used by its user, but instead becomes an active participant in this relation.

Dragon NaturallySpeaking is a speech recognition software; thence Dragon recognises speech. This software technology thus operates through the act of recognition. For Hegel, recognition is the subordinate, but vital, theme in his work on the phenomenology of SPIRIT. Although he does not carefully define this concept, it functions as the operative concept by which he works out the main thematic concept of spirit or GEIST (Williams 59). For Hegel, it is only in being recognised by an other that a self-consciousness can come into being (111). While I initially conceived of Dragon as a subordinate and obedient OTHER, it becomes apparent that in the reality of working with the software, Dragon's ability to recognise me (or, more specifically, my speech,) brings about, and thus determines, this relation in which I take up the role as computer-controlling subject. My speech can only actualise my desire and potential to operate a computer if it is recognised in such a manner that it effectuates a further response on the computer. Dragon's ability to recognise speech is thus a prerequisite if I want to realise my desire to operate a computer.

At the same time, however, the interaction with Dragon exposes and subsequently nullifies any fantasy of being an independent subject who is in control, firstly

as it dictates how I speak. When Dragon recognises my speech, it becomes recognisable again for myself in the form of text on screen or commands for the computer after translations. Whenever the utterances correspond with my intended message or command, I come to recognise them as mine. There are, however, frequent occurrences of mistranslation or misrecognition. As an example of this, my dictation “a modern proletarian industrial novel” resulted in the textual translation “a mortar proletarian industry novel.” It is these moments of misrecognition that expose that Dragon’s practice of recognising interferes with speech and mediates the way in which my speech becomes understood and translated.

Thence, Dragon performs a similar function to Hegel’s opposing or other self-consciousness, despite its inhuman and inorganic body, due to its ability to recognise. It is Dragon’s ability to recognise that simultaneously makes possible and negates the desire for control and independence that I, as the writing subject, aspire to. I come to realise that I can only actualise my potential as writing subject in *being recognised*, or more precisely, in being recognised *correctly* — an observation that I will address more extensively later. As shown above, however, recognition also entails mediation (Hegel 115), and therefore the act of recognising is in and of itself agential. It is this understanding that solicits further inquiry as to how Dragon’s ability to recognise speech is effectuated and how it, as a consequence, affects the process of operating a computer.

Listening through recognition

Let me note explicitly that Dragon's ability to recognise concerns the recognition of audible speech, meaning that Dragon is programmed to recognise sounds that make up the speech of a particular language in which it is operative.⁽²⁾ Spoken words enter the system through the computer microphone or the microphone on the headset. These auditory sounds that enter the microphone generate an analogue output signal that is indicative of the spoken words entering it. Subsequently, the analogue speech signal is converted into a sequence of digital values (i.e. commands or text on screen) that are representative of the microphone output signal (Baker et al., "Speech Recognition Systems").⁽³⁾ However, not all sound signals result in the execution of text or commands for the computer: Dragon also perceives sounds that it does not recognise, such as sneezes, sighs, laughter, grunts and background noises. Such sounds result in the appearance of a pop-up box saying "please say that again," therewith informing the user that the sounds it detects are not recognised.

aaahhhhhhhh

- (2) Dragon is a language-specific software. It can be purchased for and used in multiple languages, but when using Dragon, it can only function within the particular language selected. A user who has purchased Dragon for two (or more) languages can therefore either work within one language profile or within the other, but the languages cannot be used interchangeably within one profile.

it recognises as audible speech, which effectuates a further response on the computer. Dragon thus hears sounds but LISTENS to speech, which according to Lisbeth Lipari indicates the act of “hearing attentively; to give ear; to pay attention” (“Listening, Thinking, Being” 349), whereas hearing generally refers to the more passive form of receiving and perceiving sounds (Gallagher 622; Lipari, “Listening, Thinking, Being” 37). What remains indispensable in Dragon’s practice of listening is, however, that listening here connotes a type of recognition. That is to say that when Dragon listens to speech, Dragon listens to (or listens for) something that it already understands or knows as speech. As Heidegger infers in his 1927 book *Being and Time*: “Only he who already understands can listen” (208). Consequently, Dragon’s ability to listen requires a pre-existing knowledge or understanding of that which it listens to.

Listening to and Re-cognising Speech

As inferred above, Dragon is designed and programmed to know and recognise audible speech. Dragon operates by matching the acoustic description of the words that it is programmed to recognise against the audible signal that is generated by the utterance of the word (Baker et. al., “Speech recognition apparatus”). It is able to do so because it is programmed to recognise phonemes—the distinct units of a sound of a particular language such as *P*, *B*, *D* and *T*—and is trained to understand these

phonemes in relation to one another. Dragon is also programmed to understand “phonemes in context,” meaning that phonemes are understood in relation to the particular vowels, the duration of the utterance of these vowels, the number of syllables in the utterance, the emphasis placed on certain vowels, etc. (Bamberg and Gillick 163). Like Dragon, humans can also distinguish audible speech from other sounds, as they are able to perceive the systemic variation of fine phonetic detail of spoken utterances (Hawkins 374). Most human speakers and listeners are, however, not consciously aware of these phonetic sound structures. As Bamberg and Gillick point out, most humans are not aware of the fact that “the vowels in ‘will’ and ‘kick’ are identical according to dictionary pronunciations, [but] are as different in their spectral characteristics as the vowels in ‘not’ and ‘nut’, or that the vowel in ‘size’ has more than twice the duration of the same vowel in ‘seismograph’” (163).

What becomes apparent then is that both Dragon and I can perceive and recognise speech. Yet, while we both perceive the same audible speech sounds, Dragon recognises speech through its acoustic phonetics, whereas I, with my untrained ear, cannot consciously identify its phonetic structures. This discrepancy can be understood by drawing on the work of Sarah Hawkins. In her article “Roles and Representations of Systemic Fine Phonetic Detail in Speech Understanding” she argues that, for humans, speech is intended to convey a message. Speech, therefore, cannot be reduced

to merely the audible. Instead, speech “perform[s] multiple roles, providing strictly linguistic information, as well as non-linguistic information and paralinguistic information” (373) — think, for instance, of intonation and pauses, but also gestures and the (social) context that might induce and support audible speech. For the human listener, speech thus becomes recognised and consciously understood through derivative and indexical meanings or messages that are attached to particular (speech) sounds. The *understanding of speech* is therefore, at least for the human listener, inextricably linked to language, semiotics and meanings (Hawkins 374), whereas Dragon understands speech through what might be considered the materiality of speech: the soundwaves that become perceptible as phonetic speech sounds.

The theoretical distinction and prioritization of language over matter has, however, been thoroughly debated and criticized by Karen Barad in her 2007 book *Meeting the Universe Halfway*. As she reasons:

it seems that at every turn lately every ‘thing’ — even materiality — is turned into a matter of language or some other form of cultural representation. [...] There is an important sense in which the only thing that does not seem to matter anymore is matter (“Posthumanist Performativity” 802).

But why is it that language and meaning have been deemed so much more trustworthy than matter? As Donna Haraway similarly wonders, do we not find ourselves in both discourse and in daily practice (178); in both imagination and material reality (150)? To overcome this bias of the lingual over matter, and vice versa, Barad proposes a profound ontological and epistemological shift, reasoning that ‘things’ come into being through entangled, material-discursive practices. Thence, matter and meaning are mutually articulated and inextricably fused together—they constitute one another (*Meeting the Universe Halfway* 4).

From such a material-discursive perspective, speech can then be understood as an entangled phenomenon which is no longer monopolized by its lingual meanings. Instead, speech consists of an entanglement of both lingual meanings, acoustic phonetics and numerous other constituents that I do not touch upon here. This explains why Dragon and I can both recognise speech, while understanding it differently. Whereas I, as human user, attune to and so foreground the speech through its semiotics and lingual meanings, Dragon listens to different constituents of speech. Thus, speech might come to matter differently for Dragon than for me, even though we perceive the same sounds and speak of the same phenomenon.

Listening that Speaks

Dragon's ability to listen to speech affects what speech does. Firstly, it confines speech to its aural spectrum, whereas humans might recognise speech through a vast array of sensory preceptors, including both the audible and visual (Lipari, "Rhetoric's other" 230). Secondly, speech becomes understood primarily as phonetic speech structures. It becomes apparent that my speech here no longer serves to "convey a message" in the lingual sense that Hawkins speaks of (374). Therefore, the derivative and indexical meanings of speech, which are traditionally imbued with agency, become the adjunctive rather than primary constituent of speech in the relation between me and Dragon.

This discrepancy between Dragon's understanding of speech and my own, occasionally results in misunderstandings and mistranslations of my speech. Distinct words, phrases or sentences might be confused by Dragon as they are phonetically similar (e.g. 'bird' instead of 'word', or 'attribute it' instead of 'attributed') or when the lingual meanings of words — which Dragon attempts to identify by analysing phonemes in context, so understanding the utterance in relation to its auditory context (i.e. duration and location of utterances of vowels, number of syllables, emphasis placed on certain vowels, etc.) — are misinterpreted (e.g. 'to' instead of 'too' or 'two').⁽⁴⁾ It is then notable that even though my speech or saying precedes the said, which is effectuated on

(620). Dragon listens to (or for) speech in order to effectuate a response on the computer. Although Dragon's ability to listen, then, does not effectuate a common understanding of speech, our common ground resides in the successful operationalisation of the computer.

Although I have focused primarily on Dragon's ability to recognise the speech of a particular language, the successful operationalisation depends on multiple constituents of speech. Dragon is, therefore, also (en)able(d) to attune to the voice of its user. Every new user is required to make a user profile, whereby s/he provides Dragon with different information, such as age and the country where the user is from. This information allows Dragon to better attune to its speaker's voice. Dragon is, for instance, able to specify the pitch and tone of people from a specific age group, based on all the data that Dragon has collected over the past twenty years ("Dragon NaturallySpeaking"). Dragon also has a number of built in language accents, which are algorithms that attempt to emulate those types of speech (Lenke). The user is asked to "[h]elp Dragon understand how you pronounce words" and can then choose between a limited selection of accents: 'standard', 'Australian accented English', 'Indian accented English' or 'Southeast Asian accented English'. Finally, the user has to do a mandatory tutorial during which Dragon gets acquainted with and attuned to the user's actual voice and speech patterns.

As such, Dragon is able to listen to the speech of its particular user. By drawing on and adding to its existing knowledge of speech, Dragon is (en)able(d) to better predict the acoustic realisation of spoken utterances and to actualise the user's intention that is vocalized through speech. Realising that listening creates relationships and shapes the content of the encounter also changes the user's relationship to and utilization of speech (Srader 95). In relation to Dragon, my speech no longer serves me to express myself or to be understood. Instead, my speech has to be recognised 'correctly' by Dragon in order to actualize my potential to operate a computer. As Tyler Reigeluth infers, a relation to one's self is always simultaneously a relation to others, by which the subject becomes object for another and inversely (247).

Thus, while I initially conceived of Dragon as my tool for writing, it is my speech that has become the object of Dragon's listening. I come to recognise myself as *being* listened to. In this relation, my speech is only effective and successful if it is recognised by Dragon in the way that I intend. Listening to speech is, however, essentially selective, as any listener of speech is determinative of how speech comes to matter. As set out above, Dragon's listening decentres the importance of semiotics and lingual meanings from the act of speaking, therewith effecting and mediating the way in which my speech is put into operation. The act of listening is, therefore, not merely an act

of simply receiving and obeying: listening *does*. It is this *doing* of listening that is then greatly determinative of the relation between listener and speaker — between Dragon and I — and which consequently prevents the listening party from being discarded as a mere obedient subordinate.

biography Eeke van der Wal holds a master's degree in Organisation, Change and Management (UU). Currently, she is completing a research Master's Cultural Analysis (UvA). Fascinated by the (e)mergence of material-discursive practices, her recent research revolves primarily around material encounters and the way in which they (re)configure, organise and resist (cultural and individual) understandings of self and other.

Works Cited

- Baker, James K. et al. "Speech recognition apparatus and method." United States Patent 4,783,803A, 8 November 1988. *Google Patents*, patentimages.storage.googleapis.com/15/dc/98/f122f7d3f2cb50/US4783803.pdf. Accessed 7 Nov. 2018.
- Baker, James K. et al. "Speech recognition systems for languages with compound words." United States Patent 5,754,972A, 19 May 1998. *Google Patents*, patentimages.storage.googleapis.com/48/5b/57/ec08a444819911/US5754972.pdf. Accessed 7 Nov. 2018.
- Bamberg, Paul and Laurence Gillick. "Phoneme-in-Context Modelling for Dragon's Continuous Speech Recognizer." *Proceedings of the workshop on Speech and Natural Language*, 1990, pp. 163-169.
- Barad, Karen. *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Duke UP, 2007.
- Barad, Karen. "Posthumanist Performativity: Towards an Understanding of How Matter Comes to Matter." *Signs: Journal of Women in Culture and Society*, vol. 28, no. 3, 2003, pp. 801-831.
- Bodie, Graham D. and Nathan Crick. "Listening, Hearing, Sensing: Three Modes of Being and the Phenomenology of Charles Sanders Peirce." *Communication Theory*, vol. 24, 2014, pp. 105-123.
- Braidotti, Rosi. "Posthuman Feminist Theory." *The Oxford Handbook of Feminist Theory*, edited by Lisa Disch & Mary Hawkesworth, Oxford UP, 2016, pp. 673-698.
- "cognize | cognise, v." *Oxford English Dictionary*, 2nd ed. 1989, OED Online. www.oed.com/oed2/00043407. Accessed 6 Nov. 2018.
- "Dragon NaturallySpeaking Tutorial 1: Getting Started." YouTube, uploaded by Acces2it, 19 august 2011. www.youtube.com/watch?v=rGy5SHYV38E&t=133s. Accessed 7 Nov. 2018.
- "Dragon Speech Recognition Software." *Nuance*, 2018, www.nuance.com/dragon.html.
- Fiumara, Gemma Corradi. *The Other Side of Language: A Philosophy of Listening*. Translated by Charles Lambert. Routledge, 1990.
- Gallagher, Michael, et al. "Listening geographies: Landscape, affect and geotechnologies." *Progress in Human Geography*, vol. 4, no. 5, 2017, pp. 618-637.
- Haraway, Donna. A Cyborg Manifesto: Science, Technology and Socialist-Feminism in the Late Twentieth Century." *Simians, Cyborgs and Women: The Reinvention of Nature*. Routledge, 1991, pp. 149-181.

- Ghosh, Amitav. *The Great Derangement: Climate Change and the Unthinkable*. U of Chicago P, 2016.
- Hawkins, Sarah. "Roles and representations of systemic fine phonetic detail in speech understanding." *Journal of Phonetics*, vol. 31, no 3-4, 2003, pp. 373-405.
- Heidegger, Martin. *Being and Time*. 1927. Translated by John Macquarrie and Edward Robinson, Blackwell, 1962.
- Hegel, Georg W. F. *The Phenomenology of Spirit*. 1807. Translated by Arnold V. Miller. Oxford UP, 1979.
- Kamal, Muhammad. "Master-Slave Relationship in Hegel's Philosophy." *Indian Philosophical Quarterly*, vol XXV, no.4, 1998, pp. 455-466.
- King, Anthony. *The Structure of Social Theory*. Routledge, 2004.
- Lenke, Nils. "Dragon, do you speak my dialect?" *What's Next*, Nuance Communications. 26 January 2017, whatsnext.nuance.com/in-the-labs/speech-recognition-understands-language-dialects/. Accessed 6 Nov. 2018.
- Lipari, Lisbeth. "Listening, Thinking, Being." *Communication Theory*, vol. 20, 2010, pp. 348-362.
- Lipari, Lisbeth. "Listening Otherwise: The Voice of Ethics." *International Journal of Listening*, vol. 23, no.1, 2009, pp. 44-59.
- "re-, prefix" *Oxford English Dictionary*, 3rd ed. December 2008, OED Online. www.oed.com.proxy.uba.uva.nl:2048/view/Entry/158795?rskey=ZDCPdd&result=10#eid. Accessed 6 Nov. 2018.
- Lipari, Lisbeth. "Rhetoric's Other: Levinas, Listening and the Ethical Response." *Philosophy & Rhetoric*, vol. 45, no. 3, 2012, pp. 227-245.
- Srader, Doyle W. "Performative Listening." *International Journal of Listening*, vol. 29, no. 2, 2015, pp. 95-102.
- Stuart, Caleb. "The Object of Performance: Aural Performativity in Contemporary Laptop Music." *Contemporary Music Review*, vol. 22, no. 4, 2003, pp. 69-65.
- Reigeluth, Tyler. "Why data is not enough: Digital traces as control of self and self-control." *Surveillance & Society*, vol. 12, no. 2, 2014, pp. 243-254.
- Williams, Robert R. "The Concept of Recognition in Hegel's Phenomenology of Spirit." *Hegel's Phenomenology of Spirit*, edited by Alfred Denker and Michael Vater, Humanity Books, 2003, pp. 59-92.