Netaya Lotze (Münster)

Künstliche Intelligenz im Dialog – Ein methodologisches Konzept zur Analyse von Mensch-Maschine-Interaktion

Der Mythos „Künstliche Intelligenz“ wird besonders von der sogenannten „trans-humanistischen“ Community im Silicon Valley propagiert, deren Vertreter wie der Physiker Ray Kurzweil davon ausgehen, dass wir in spätestens 30 Jahren mit KIs kommunizieren könnten, wie mit einem Menschen (Kurzweil 2005). Saudi Arabien hat 2017 bereits dem anthropomorphen Roboter mit Sprachinterface *Sophia* die Staatsbürgerschaft zugesprochen (Arab News 2017). Künstliche Intelligenzen wie Apples Assistenzsystem *Siri* oder Amazons *Alexa* halten derzeit Einzug in unseren Alltag. Chatbots und Social-Bots wie der Twitter-Bot *Tay* nehmen Einfluss auf öffentliche Diskurse und interaktives Spielzeug mit Dialogfunktion führt bereits unsere Jüngsten an die Interaktion mit dem artifiziellen Gegenüber heran. Hier entsteht eine völlig **neue Form der Dialogizität**, die wir aus linguistischer Perspektive noch kaum verstehen. Unabhängige Studien zur Mensch-Maschine-Interaktion stellen also ein großes Desiderat dar. Momentan wird in erster Linie die Benutzer/innen-Freundlichkeit, die *Usability*, der Systeme in den Entwicklungsabteilungen von Apple, Amazon oder IBM evaluiert, um die Dialog-Performance seitens der Bots zu verbessern. Dagegen gibt es wenige unabhängige Studien (Fischer 2006 (u.a.); Pitsch et al. 2017 (u.a.); Lotze 2016, 2018), die die Auswirkungen der Systeme auf das sprachliche Verhalten ihrer Nutzer/innen objektiv beschreiben und interpretieren können. Nur auf dieser Basis können Prognosen gestellt werden für eine Zukunft, in der wir in unserem Alltag häufiger mit Bots konfrontiert sein werden. Aus linguistischer Perspektive ist das Szenario der Interaktion zwischen Mensch und KI besonders interessant, weil hier Grenzfragen der Interaktionsforschung aufgeworfen werden: Worin bestehen überhaupt die Grundvoraussetzungen für Dialogizität? Inwiefern erfüllen Bots diese Grundvoraussetzungen (nicht)? Und wie reagieren User/innen, wenn sie sich mit dieser selten störungsfreien Illusion von Agentivität konfrontiert sehen.

1 Restriktionen von KIs in der Interaktion

Microsoft-CEO Satya Nadella postulierte 2016 anlässlich der Facebook-Zulassung für Social Bots diese Technologie als intuitiv nutzbare Schnittstelle der Zukunft: „Chatbots will fundamentally revolutionize how computing is experienced by everybody. In time, human language will be taught to all computers and become the new interface.“ Aus linguistischer Perspektive ist menschliche Kommunikation aber bedeutend komplexer! Ihre logisch-semantische Referenzialität und die genutzten Wissensressourcen lassen sich nicht ohne Weiteres in einen Bot implementieren. So gehen Dialogkohärenz und Kohäsion in der Mensch-Maschine-Interaktion nicht über einfache Adjazenz-Strukturen hinaus (meist in Form von Frage-Antwort-Paaren). Grund dafür sind die fehlenden Voraussetzungen für Agentivität seitens des Bots:

- ein **Bewusstsein**, das Aussagen versteht und denkt (im Sinne des Locke’schen „Self“ und Searles „Intentionality“; Locke 1975; Searle 1997),
- **Autonomie**, um Handlungspläne aufzustellen und **spontan** zu ändern (im Sinne Kants; Kant 1999),
- **Weltwissen** im Sinne Habermas’ (1993) als Wissensressource, um geteiltes Wissen mit dem Gegenüber aufzubauen (**Common Ground**; Stalnaker 2002).

Diese Aspekte sind unerlässlich, wenn sprach-ökonomisches und effizientes Dialogverhalten nach menschlichem Vorbild modelliert werden soll. Denn Menschen verbalisieren nur einen Bruchteil dessen, was mitgedacht wird. Zu inferieren, was nicht verbalisiert wird, stellt bis heute eine enorme Herausforderung für jeden Parser dar.

2 Methoden zur Erforschung der Interaktion mit KIs

Wie kann man sich also diesem äußerst komplexen Forschungsgegenstand methodisch annähern? Da unterschiedliche Ebenen der Interaktionssituation berücksichtigt werden müssen, empfehlen sich Methoden-Mix-Studien, die Ansätze aus Informationswissenschaften, Gesprächsforschung und Psycholinguistik konsistent zusammendenken:

Die interaktive Ebene: Die Analyse der Dialoge sollte sowohl **qualitativ-explorativ** im Sinne einer genauen Deskription dieser neuen Form der Dialogizität erfolgen als auch **quantitativ-kontrastiv**, um Vergleiche ziehen zu können.

1. **Qualitativ-explorative Methoden der linguistischen Gesprächsforschung** (u.a. Sacks/Schegloff/Jefferson 1992): In offenen Datensessions werden Mensch-Maschine-Dialoge gesichtet und wiederkehrende Muster extrahiert. So können relevante Forschungsparameter für weitere Untersuchungen isoliert werden: *Störungen und Reparaturen*, *sprachliche Höflichkeit* und „*Flaming*“ (Beschimpfungen), „*Computer-Talk*“ (Simplifizierung und Explizitheit), *Dialog-Kohärenz und Kohäsion*.
2. **Quantitativ-vergleichende Methode der Korpuslinguistik**: Annotation und Auszählung von Untersuchungsparametern sowie die statistische Aufbereitung von Dialogdaten kommen in der Mensch-Maschine-Interaktions-Forschung zum Einsatz, um Vergleiche ziehen zu können: a) zur Mensch-Mensch-Kommunikation und b) zwischen unterschiedlichen Anwendungskontexten (unterschiedliche Chatbots, unterschiedliche Interaktionssituationen). Hier haben **theorie-geleitete, hypothesen-prüfende Verfahren** ihren Platz, aber immer erst im zweiten Schritt auf Grundlage der Explorationsphase. Die Untersuchungsparameter müssen klar quantifizierbar sein, wie z.B.: *Turn-Längen* (in WF), *Adjazenzellipsen* (pro Turn/pro Dialog), *Konstruktionsübernahmen* (pro Turn/pro Dialog), *Emoticons/Emojis* (pro Turn/pro Dialog).

Die kognitive Ebene: Den herausgearbeiteten sprachlichen Phänomenen liegen kognitive Konzepte zugrunde (z.B. *Konstruktionsübernahmen – interaktives Alignment*, *Kohäsionsmarker – logische Kohärenz*), die in **psycholinguistischen User/innen-Experimenten** überprüft werden können. Hier wäre ein mögliches Szenario z.B., dass User/innen gemeinsam mit einem System kollaborative Aufgaben lösen. Der Vorteil von Dialogsystemen in experimentellen Settings ist, dass sie hochgradig kontrollierbar sind. Sofern man testen möchte, wie sich Nutzer/innen gegenüber einer noch nicht vorhandenen Technologie verhalten würden (z.B. einem störungsfreien Assistenzsystem), eignen sich sogenannte „Wizard-of-Oz“-Experimente, in denen ein/e Verbündete/r der Versuchsleitung sich als Bot ausgibt.

Die technologische Ebene: Bei der Interpretation der Sprachdaten müssen die technologischen Rahmenbedingungen immer mitgedacht werden. Unterschiedliche System-Architekturen (z.B. plan-basierte Systeme vs. statistische Big-Data-Ansätze) und Dialog-Designs (z.B. einfacher Parsing-Prompting-Mechanismus auf der Basis von Keywords oder hinterlegte Dialog-Skripte) begünstigen nämlich unterschiedliche sprachliche User/innen-Strategien. Zur Interpretation dieser Variable ist ein in Informatik und Informationswissenschaften fundierter Zugang unerlässlich. Darauf aufbauend können im letzten Schritt auf der Basis der Ergebnisse der Grundlagenforschung neue Systeme oder Dialog-Designs entwickelt

- Locke, John (1975): *An essay concerning human understanding*. Edited with an introduction, critical apparatus and glossary by Peter H. Nidditch. Oxford.
- Lotze, Netaya (2016): *Chatbots – eine linguistische Analyse*. (= Sprache – Medien – Innovationen 9). Frankfurt a.M.
- Lotze, Netaya (2018): *Zur sprachlichen Interaktion mit Chatbots – eine linguistische Perspektive*. In: Hug, Theo/Pallaver, Günther (Hg.): *Talk with the bots – Gesprächsroboter und Social Bots im Diskurs*. Innsbruck, S. 29–50.
- Pitsch, Karola/Gehle, Raphaela/Dankert, Timo/Wrede, Sebastian (2017): *Interactional dynamics in user groups: Answering a robot's question in adult-child constellations*. In: Wrede, Britta/Nagai, Yukie/Komatsu, Takanori/Hanheide, Marc/Natale, Lorenzo (eds.): *HAI '17 – Proceedings of the 5th International Conference on Human Agent Interaction*, Bielefeld, Germany, October 17–20, 2017. Bielefeld, S. 393–397.
- Sacks, Harvey/Schegloff, Emanuel A./Jefferson, Gail (1992): *Lectures on conversation*. 2 Volumes. Oxford.
- Searle, John R. (1980): *Minds, brains and programmes*. In: *Behavioural and Brain Sciences* 3, S. 417–457.
- Searle, John R. (1997): *Intentionalität – eine Abhandlung zur Philosophie des Geistes*. Frankfurt a.M.
- Stalnaker, Robert C. (2002): *Common ground*. In: *Linguistics and Philosophy* 25, S. 701–721.
- Zoeppritz, Magdalena (1985): *Computer talk?* (= Technical Reports/IBM Germany Heidelberg Scientific Center TN 85.05). Heidelberg.