Angus Addlesee

Heriot-Watt University
Edinburgh, UK

a.addlesee@hw.ac.uk
addlesee.co.uk

1 Research interests

Speech production is nuanced and unique to every individual, but today's Spoken Dialogue Systems (SDSs) are trained to use general speech patterns to successfully improve performance on various evaluation metrics. However, these patterns do not apply to certain user groups - often the very people that can benefit the most from SDSs. For example, people with dementia produce more disfluent speech than the general population (Boschi et al., 2017). The healthcare domain is now a popular setting for spoken dialogue and human-robot interaction research. This trend is similar when observing company behaviour. Charities promote industry voice assistants, the creators are getting HIPAA compliance, and their features sometimes target vulnerable user groups (Addlesee, 2023).

1.1 Data collection

Research on interactions between SDSs and people with dementia is stifled due to the severe lack of data (Addlesee et al., 2019). Collecting natural spoken dialogue data with vulnerable older adults is ethically challenging. Consent must be witnessed by the participant's carer, the collection location must be designed to be accessible, and collaboration with charities is often required to recruit participants (Addlesee and Albert, 2020). Bespoke tools are also required to collect data *securely* from vulnerable participants (Addlesee, 2022).

In order to tackle this challenge, we have collected two corpora of people with dementia interacting with SDSs. The first corpus, called DEICTIC, contains interactions captured between Amazon Alexa devices and family members in 10 family homes. One member in each family was diagnosed with dementia. This corpus is currently being filtered for personally identifiable information, so its exact size is unknown, but we expect to include over 300 interactions (including both multi-turn and multi-party interactions). Once complete, a sub-repository of TalkBank called DementiaBank[1] will be used to share data with other researchers studying communication in the dementia domain.

The second corpus, yet to be named, is currently being collected as part of the H2020 SPRING Project[2]. We noticed in DEICTIC that multi-party interactions take place at home, even though the system is only designed to have dyadic interactions. Hospital staff that work in a memory clinic also explained that patients typically attend their appointments with a companion. We designed a data collection framework to elicit a diverse range of multi-party conversations between patients, their companions, and a social robot called ARI (Addlesee et al., 2023). We have collected over 50 multi-party conversations with various versions of ARI (with a wizard-of-Oz setup, with a single user system, and with a multi-user system).

User:	EVA, Is Alex Rodriguez dating...
EVA:	Sorry, I didn't catch that. Dating who?
User:	Jennifer Lopez
EVA:	Yes, they are currently dating.

Table 1: Collaborative completion from understanding.

1.2 Mid-utterance interruption recovery

Voice assistants interrupt people when they pause mid-utterance, a frustrating interaction that requires the full repetition of the entire sentence again. This impacts all users, but particularly people with cognitive impairments (Boschi et al., 2017). We know, however, that natural spoken language unfolds over time. Our interlocutors process each token as it is uttered, maintaining a partial representation of what has been said (Marslen-Wilson, 1973; Madureira and Schlangen, 2020; Kahardipraja et al., 2021). That is, we understand the words that *were already said* if someone pauses mid-sentence. To avoid waiting indefinitely while a conversation partner is pausing, humans either prompt the turn-holder to collaboratively complete the question (Ginzburg and Sag, 2000; Fernández et al., 2007; Poesio and Rieser, 2010), as shown in Table 1, or suggest sentence completions themselves (referred to as cross-person compound contributions or gap-fillers (Purver et al., 2003; Howes et al., 2011, 2012)), shown in Table 2.

We implemented both approaches to answer people's incomplete questions and semantically parse their disrupted sentences. We constructed two novel cor-

[1] https://dementia.talkbank.org/

[2] https://spring-h2020.eu/

Proceedings of the 19th Annual Meeting of the Young Reseachers' Roundtable on Spoken Dialogue Systems, pages 9–11
September 11–12, 2023. ©2023 Association for Computational Linguistics

User:	EVA, when is the next solar...
EVA:	The next solar eclipse is on the 20th April 2023

Table 2: Prediction of question completion.

pora to measure a recovery pipeline's ability to complete these tasks. One corpus interrupts questions originally collected for Knowledge Base Question Answering (KBQA), where a semantic parser is used to convert questions into an executable meaning representation over some given knowledge. For example, a system may be asked to answer "What is the population of Portugal?" when given Wikipedia as a knowledge base. Both the questions and their semantic representations (in SPARQL, a knowledge graph query language) were interrupted, resulting in a corpus of 21,000 interrupted questions (see Tables 1 and 2) (Addlesee and Damonte, 2023a). The second corpus was generated by disrupting almost 80,000 sentences more generally, along with their abstract meaning representations (AMR) (Addlesee and Damonte, 2023b).

We used the current state-of-the-art systems on the corresponding original tasks, given the full original utterances, as performance upper bounds. Our best-performing systems performed remarkably well, identifying where the missing information is located in the utterance's semantic representation. In the KBQA domain, our best pipeline answered only 0.77% fewer questions than the SotA upper bound (Addlesee and Damonte, 2023a). When inspecting sentences more generally, our recovery pipeline lost only 1.6% graph similarity f-score (Smatch) compared to the AMR upper bound (Addlesee and Damonte, 2023b). We have therefore shown that interruption recovery pipelines could potentially be used to improve voice assistant accessibility, and general robustness to noisy environments like family homes, or public spaces (like hospital waiting rooms).

To confirm that our pipelines do improve accessibility in practice, a user study must take place. We have shown that our approach is feasible, but response generation would also be needed for an actual user study. We plan to use our interrupted corpora to elicit clarifications from humans. We can then evaluate whether today's LLMs can safely generate clarification requests to elicit the repair turn from the user.

1.3 Real-time semantic parsing

Our incremental semantic parsers in Section 1.2 work when given sentences interrupted at a single point before named entities (where mid-sentence pauses typically occur (Croisile et al., 1996; Seifart et al., 2018; Slegers et al., 2018)), but the next generation of SDSs need to process tokens in real-time (Addlesee and Eshghi, 2021). We have developed a fully incremental graph-based semantic parser by combining Dynamic Syntax (Kempson et al., 2001; Cann et al., 2005) with RDF (Lassila et al., 1998) – called DS-RDF (Addlesee and Eshghi, 2021). A prototype was built[3], but we have since extended the lexicon to be wider coverage. We are also working on an LLM-based approach. We plan to evaluate both of these approaches on our collected corpora. We expect to find that the LLM-based approach has a wider-coverage, but that DS-RDF does not hallucinate as frequently. This is particularly crucial when interacting with vulnerable users in a hospital setting (Addlesee, 2023).

2 Spoken dialogue system (SDS) research

The next generation of SDSs need to: (1) process language *incrementally*, token-by-token to be more responsive and enable handling of conversational phenomena; (2) *reason incrementally* allowing meaning to be established beyond what is said; and (3) be *transparent* and *controllable*, allowing designers as well as the system itself to easily establish reasons for particular behaviour and tailor to particular user groups, or domains. The boom of chatGPT (and co) is extremely exciting, but point 3 is a huge concern. Both startups and big tech companies are applying these new approaches to every domain they can, including healthcare. A disastrous news story seems inevitable when one of these systems provides a vulnerable user with a harmful response (e.g. a child, or person with a cognitive impairment). I think the controllability of these systems will be a huge focus for SDS researchers over the next few years.

3 Suggested topics for discussion

- Real-time time speech processing
- Multi-party dialogue
- Ethical Data Collection
- LLM controllability and grounding

Biographical sketch

Angus is currently studying his PhD in Artificial Intelligence at Heriot-Watt University. He has previously worked on machine learning and data science projects within The NHS, Scottish Government, and private clients in many sectors including finance. Angus is very passionate about 'AI for Good', hence his decision to move back into research from industry. He also enjoys bouldering and running.

[3] `https://youtu.be/nj-eaMDeEtc?t=903`

References

Angus Addlesee. 2022. Securely capturing people's interactions with voice assistants at home: A bespoke tool for ethical data collection. In *Proceedings of the 2022 NLP for Positive Impact Workshop at EMNLP*.

Angus Addlesee. 2023. Voice assistant accessibility. In *Proceedings of The 13th International Workshop on Spoken Dialogue Systems (IWSDS)*.

Angus Addlesee and Pierre Albert. 2020. Ethically collecting multi-modal spontaneous conversations with people that have cognitive impairments. In *LREC 2020 Workshop Language Resources and Evaluation Conference 11–16 May 2020*. page 15.

Angus Addlesee and Marco Damonte. 2023a. Understanding and answering incomplete questions. In *Proceedings of the 5th Conference on Conversational User Interfaces*.

Angus Addlesee and Marco Damonte. 2023b. Understanding disrupted sentences using underspecified abstract meaning representation. In *Interspeech*.

Angus Addlesee and Arash Eshghi. 2021. Incremental graph-based semantics and reasoning for conversational ai. In *Proceedings of the Reasoning and Interaction Conference (ReInAct 2021)*. pages 1–7.

Angus Addlesee, Arash Eshghi, and Ioannis Konstas. 2019. Current challenges in spoken dialogue systems and why they are critical for those living with dementia. *Dialog for Good (DiGo* .

Angus Addlesee, Weronika Sieińska, Nancie Gunson, Daniel Hernández García, Christian Dondrup, and Oliver Lemon. 2023. Data collection for multi-party task-based dialogue in social robotics. In *The International Workshop on Spoken Dialogue Systems Technology, IWSDS 2023*.

Veronica Boschi, Eleonora Catricala, Monica Consonni, Cristiano Chesi, Andrea Moro, and Stefano F Cappa. 2017. Connected speech in neurodegenerative language disorders: a review. *Frontiers in psychology* 8:269.

Ronnie Cann, Ruth Kempson, and Lutz Marten. 2005. *The Dynamics of Language*. Elsevier, Oxford.

Bernard Croisile, Bernadette Ska, Marie-Josee Brabant, Annick Duchene, Yves Lepage, Gilbert Aimard, and Marc Trillet. 1996. Comparative study of oral and written picture description in patients with alzheimer's disease. *Brain and language* 53(1):1–19.

Raquel Fernández, Jonathan Ginzburg, and Shalom Lappin. 2007. Classifying non-sentential utterances in dialogue: A machine learning approach. *Computational Linguistics* 33(3):397–427.

Jonathan Ginzburg and Ivan Sag. 2000. *Interrogative investigations*. Stanford: CSLI publications.

Christine Howes, Ptarick GT Healey, Matthew Purver, and Arash Eshghi. 2012. Finishing each other's... responding to incomplete contributions in dialogue. In *Proceedings of the Annual Meeting of the Cognitive Science Society*. volume 34.

Christine Howes, Matthew Purver, Patrick GT Healey, Gregory J Mills, and Eleni Gregoromichelaki. 2011. On incrementality in dialogue: Evidence from compound contributions. *Dialogue & Discourse* 2(1):279–311.

Patrick Kahardipraja, Brielen Madureira, and David Schlangen. 2021. Towards incremental transformers: An empirical analysis of transformer models for incremental nlu. *arXiv preprint arXiv:2109.07364* .

Ruth Kempson, Wilfried Meyer-Viol, and Dov Gabbay. 2001. *Dynamic Syntax: The Flow of Language Understanding*. Wiley-Blackwell.

Ora Lassila, Ralph R. Swick, World Wide, and Web Consortium. 1998. Resource description framework (rdf) model and syntax specification .

Brielen Madureira and David Schlangen. 2020. Incremental processing in the age of non-incremental encoders: An empirical assessment of bidirectional models for incremental nlu. *arXiv preprint arXiv:2010.05330* .

William Marslen-Wilson. 1973. Linguistic structure and speech shadowing at very short latencies. *Nature* 244(5417):522–523.

Massimo Poesio and Hannes Rieser. 2010. Completions, coordination, and alignment in dialogue. *Dialogue & Discourse* 1(1).

Matthew Purver, Jonathan Ginzburg, and Patrick Healey. 2003. On the means for clarification in dialogue. In *Current and new directions in discourse and dialogue*, Springer, pages 235–255.

Frank Seifart, Jan Strunk, Swintha Danielsen, Iren Hartmann, Brigitte Pakendorf, Søren Wichmann, Alena Witzlack-Makarevich, Nivja H de Jong, and Balthasar Bickel. 2018. Nouns slow down speech across structurally and culturally diverse languages. *Proceedings of the National Academy of Sciences* 115(22):5720–5725.

Antoine Slegers, Renee-Pier Filiou, Maxime Montembeault, and Simona Maria Brambati. 2018. Connected speech features from picture description in alzheimer's disease: A systematic review. *Journal of Alzheimer's Disease* 65(2):519–542.