# Current Challenges in Spoken Dialogue Systems and Why They Are Critical for Those Living with Dementia

**Angus Addlesee**
Heriot-Watt University
`ja204@hw.ac.uk`

**Arash Eshghi**
Heriot-Watt University
`a.eshghi@hw.ac.uk`

**Ioannis Konstas**
Heriot-Watt University
`i.konstas@hw.ac.uk`

## Abstract

Dialogue technologies such as Amazon's Alexa have the potential to transform the healthcare industry. However, current systems are not yet naturally interactive: they are often turn-based, have naive end-of-turn detection and completely ignore many types of verbal and visual feedback - such as backchannels, hesitation markers, filled pauses, gaze, brow furrows and disfluencies - that are crucial in guiding and managing the conversational process. This is especially important in the healthcare industry as target users of Spoken Dialogue Systems (SDSs) are likely to be frail, older, distracted or suffer from cognitive decline which impacts their ability to make effective use of current systems. In this paper, we outline some of the challenges that are in urgent need of further research, including Incremental Speech Recognition and a systematic study of the interactional patterns in conversation that are potentially diagnostic of dementia, and how these might inform research on and the design of the next generation of SDSs.

## 1 Introduction

Creating Spoken Dialogue Systems (SDSs) that are capable of natural, spontaneous conversation with humans is still full of many challenges (Eshghi et al., 2017; Porcheron et al., 2018; Coman et al., 2019). Some of these have begun to be addressed, and some have received too little attention.

In this paper, we outline some of these challenges, focusing on those that show up as in need of urgent consideration and research if we are to use interactive dialogue technology to ease growing pressures in healthcare and assisted living environments. We will then propose a number of avenues for further research.

### 1.1 Importance of Dialogue Technology in Healthcare

Carers are already stretched thin (Wright, 2015) and we have a rapidly ageing population in Scotland (Maclachlan, 2017) but this is not unique, in fact it is a global challenge (UN, 2018). This means that in the coming years, we will have a carer shortage and other problems, such as bed blocking in hospitals, will become increasingly problematic (Puttick, 2018).

Dementia is the leading cause of death in the UK but there is no treatment to prevent, cure or slow its progression (Alzheimer's Research UK, 2018). This is consequently a key area that we plan to focus on in this paper but the challenges we outline are not limited to dementia.

There are a huge number of IoT devices that could assist people with tasks that they find most difficult. Their embedded dialogue systems need to become more natural if they are effective however (Sakakibara et al., 2017; Helal and Bull, 2019) and they are usually accessible only through a disjoint range of apps. These app interfaces are becoming increasingly complex unfortunately as new devices and new features are released (Hargreaves et al., 2018). Humans most naturally communicate in conversation, so an SDS is very likely the best way to communicate with all of these devices (IBM, 2018). There are however a number of challenges still left to be addressed until such systems become usable by older adults or frail patients.

## 1.2 Challenges

**Incremental Processing:** Language processing in human conversation is inherently incremental, i.e. it proceeds word by word, or token by token rather than turn by turn (Ferreira et al., 2004; Purver et al., 2009; Howes et al., 2011) (among many others). This gives rise to many characteristic phenomena such as interruptions, backchannels, disfluencies, restarts, corrections, split utterances, and fragments: processing these correctly and effectively is crucial in building naturally interactive systems, and becomes critical in a healthcare context - see below. Yet, all commercial dialogue systems and most research systems are turn-based and ignore many of these phenomena. They are thus not user-friendly and seem unnatural, often frustrating the user.

The below dialogues contain examples of some of these problems from real conversations extracted from the British National Corpus (BNC).

**Example 1** (Howes and Eshghi, 2017a):
**A**: The doctor...
**B**: mhm
**A**: he examined me.

**Example 2** (Eshghi et al., 2015):
**C**: We went to see something called the Wedding Banquet.
**D**: Called the Wedding [Banquet]?
**C**:                      [Banquet].
**D**: Really?

**Multi-modal Concurrent Feedback:** Concurrent feedback is very common within dialogue and these short utterances guide conversation (Charles, 1981; Bavelas and Gerwing, 2011). This has been known for a long time in the psycho-linguistic community but yet to be taken up by computational linguists. For example, hesitation utterances such as "umm", "err" and "hmm" are often picked up by current systems like any other token when in fact, these hesitations should be used to generate a more natural response based on whether the hesitation was used to indicate confusion, completion or used as a turn-holding device. Similarly, confirmation backchannels such as "yep", "uh-huh" and "mm-hmm" should be processed and integrated. This feedback often overlaps the SDSs turn, breaking the usual turn-by-turn nature that current systems expect. Additionally, hesitations and confirmation backchannels are not always audible as humans often use visual signals instead.

**Example 3**:
**A**: I went to see Ice Cube
**B**: ⟨screws face in confusion⟩
**A**: He's a rapper and acted as the police chief in 22 Jump Street.
**B**: ⟨nods and unscrews face⟩
**A**: I went to see him

Screwing of the face, brow furrows, looking up, nodding, smiling, eye-contact, etc... are all used by humans to subtly guide and support natural conversation and, though crucial in how a conversation unfolds, are lost completely by current systems.

**Interactional Patterns in Dementia:** Although some computational systems exist today for the detection of dementia from speech patterns (Luz et al., 2018; Zhu et al., 2018; Ammar and Ayed, 2018; Broderick et al., 2018), there is very little work on how dementia might affect *interactional patterns*. A systematic, empirical study of such patterns is essential for informing design of dialogue systems for this target group - see Sec. 3.2.

## 2 Current Work Towards Natural Conversation

Natural face to face conversation involves quick exchanges in which people frequently hesitate, hedge, restart, self-correct (Shriberg, 1996; Hough, 2015), interrupt each other (Healey et al., 2011), continue each other's sentences (Howes, 2012), backchannel (Heldner et al., 2013; Howes and Eshghi, 2017b), etc... with none of these phenomena respecting the boundaries of a sentence or turn. We will therefore have to investigate, extend and implement models of incrementality to ensure satisfactory system speed, naturalness and fluidity (Schlangen and Skantze, 2009; Skantze and Hjalmarsson, 2010a; Baumann and Schlangen, 2012; Eshghi et al., 2012, 2017). In this section, we give a very brief overview of current work towards capturing these phenomena and thus building more naturally interactive SDSs.

### 2.1 Incremental Dialogue Systems

Incrementality in dialogue puts several new constraints on how Dialogue Systems should be designed: Automatic Speech Recognition (ASR), Natural Language Understanding (NLU), Natural Language Generation (NLG), Dialogue Management (DM) and Text to Speech (TTS) need to process language word by word, or token by token,

where each token is fed immediately to downstream modules for processing rather than waiting for the end of turn. Furthermore, any token, word hypothesis from ASR, or piece of semantic analysis, etc... can be revoked and this has to percolate through the system. Thus, traditional turn-based pipelines for processing are immediately rendered inadequate.

(Schlangen and Skantze, 2011) provide an elegant abstract architecture in terms of Incremental Units (IUs) of processing which takes account of the constraints mentioned above. It has been shown that such an architecture is faster, more effective and perceived to be more natural (Skantze and Hjalmarsson, 2010b; Paetzel et al., 2015). This architecture has been implemented in the InProTK (Baumann and Schlangen, 2012) and Jindigo (Skantze, 2010) dialogue systems where conforming and suitably incremental ASR, NLU, NLG, DM and TTS modules can be plugged in and out.

**Incremental NLU:** Existing incremental NLU systems, like non-incremental systems before them, are either: (1) grammar-based and thus domain-general such as Dylan (Eshghi et al., 2011; Eshghi, 2015) which is based on the Dynamic Syntax Grammar formalism (Kempson et al., 2016) - Dylan maps linguistic inputs word by word to domain-general semantic representations; or (2) they map a sequence of words directly onto Dialogue Acts or Intent representations (DeVault et al., 2011; Rafla and Kennington, 2019). While the former approach is transferrable and principled, it is harder to develop and maintain because it is grammar based. The latter approach is highly domain-specific and thus not transferrable but instead enjoys the advantage of being easier and faster to create. This conundrum continues to this day, with most commercial systems preferring (2) for the reasons outlined.

We do not here go into incremental NLG or DM but such systems do exist (Hough and Purver, 2012; Eshghi et al., 2017) and there is still plenty of room for further research in these areas.

**Turn Taking - End of Turn Prediction:** Human conversationalists are strikingly good at predicting when their interlocutors are about to finish speaking (Sacks et al., 1974; Schegloff, 2000). Various features of talk, such as intonation and morphosyntax enable this (de Ruiter et al., 2006). Re-

cently, computational models have been built for this task, so here we review some of these systems.

(Maier et al., 2017) used both acoustic and linguistic features to predict whether someone has finished their turn. Their long short-term memory network (LSTM) tagged 10ms windows as either speech, mid-turn-pause (MTP) or end-of-turn (EOT). Their system beat all baselines which were based on silence thresholds of different lengths to predict the EOT. This is a very valuable improvement for those with dementia as current systems are turn-based so, if the system cuts in, the user has to repeat their entire utterance. (Roddy et al., 2018) use acoustic features, linguistic features and voice activity to train an LSTM to predict the EOT. Their best performing model was obtained using voice activity, acoustic features and word-level linguistic features. The fact that using words outperformed part-of-speech (POS) tags is very significant as they are faster to process. This benefits real-time incremental prediction which is exactly what we humans do.

**Incremental Grounding Strategies:** In current commercial SDSs, grounding (Clark, 1996) is not incremental so users can only give feedback at the end of a system's turn. For a conversation to be natural, it must be grounded in an incremental manner. Incremental grounding is possible if overlaps are processed allowing reasoning over concurrent speech (Hough and Schlangen, 2016). If for example an SDS is embedded within a robot assistant in a care home, a user could ask it to "Bring me my scarf". In current systems the user would have to wait until the robot has brought them their scarf before they can say "no, the other one". If the system was grounded incrementally in a fluid manner however, the user could say "no, the other one" as the robot picks up the first scarf. To enable this, (Hough and Schlangen, 2016) provide a model of how the robot could know when it has sufficiently shown what it is doing to handle both repairs and confirmations through real-time context monitoring. The robot needs to know what the user is confirming and even more importantly, what is needing to be repaired.

## 3 Ongoing Work: Towards Natural SDS in Healthcare

In this section we motivate and outline our ongoing work towards building more natural, fluid SDSs in the general healthcare domain.

## 3.1 Evaluation and Improvement of Incremental ASR

For SDSs, none of the work summarised above would be of any use unless ASR and Gesture Recognition systems, (1) work on a token by toke basis; (2) output hypotheses *with minimal latency*; (3) produce as little 'jitter' as possible - i.e. the hypotheses are stable over time; and (4) capture all of the speech input *including* disfluency markers, hesitations, pauses and laughter for downstream modules such as NLU to work with. As (Baumann et al., 2017) outline, these requirements correspond to additional metrics in evaluating ASR systems and go much beyond the usual word error rate (WER) metric. (Baumann et al., 2017) found that Google outperformed the others on WER but filtered out disfluencies and did not provide word timings. Sphinx-4 and Kaldi both preserved material, provided detailed word timings and were both quicker than Google at deciding on a word once it has been uttered. Kaldi and Sphinx-4 performed similarly and can be retrained on in domain data which improves them.

While this evaluation work is very valuable, we can still go further: the open-domain systems tested are limited to Kaldi and Sphinx, but more importantly, the corpus used is highly domain-specific. It is in the end not entirely clear whether the systems' level of performance is due to the overall ML architecture of the systems themselves or simply due to out of domain training data; or indeed whether the performance gains achieved through re-training would generalise to more open-domain data. We therefore plan to extend this study to the much more open-domain Switchboard Corpus (Godfrey et al., 1992) whose more recent versions include both disfluency and dialogue act tags. We also plan to evaluate more of the open-source, trainable systems such as Wav2letter++ (Pratap et al., 2018) and Julius (Lee and Kawahara, 2009).

If current incremental ASR performs poorly, we will try re-training a current ASR using Multi-Task Learning (MTL) to achieve better generalisation across domains.

We ultimately plan to work on incremental, multi-modal dialogue processing to guide an SDS's conversations in a fluid and more natural manner but this work on incremental ASR must be tackled first.

## 3.2 Interactional Patterns in Dementia Patients: A corpus study

Current dementia detection techniques are invasive, expensive, time-consuming and cause unnecessary stress for the patients so computational detection models are being developed with the aim to alleviate some of these problems (Luz et al., 2018; Zhu et al., 2018; Ammar and Ayed, 2018; Broderick et al., 2018). Language is known to be impacted by cognitive decline (Boschi et al., 2017) but unfortunately, suitable corpora to train and evaluate dementia detection models, and therefore our work, are rare and relatively difficult to access.

Most dementia detection models use the Pitt corpus on DementiaBank (Becker et al., 1994) as it contains audio with transcriptions of people with Alzheimer's Disease (AD) and healthy elderly controls. This corpus is elicited using a picture description task which ensures the vocabulary is controlled around a context but does not allow spontaneous conversational speech that we would expect an SDS to receive. (Luz et al., 2018) use a different corpus, the Carolina Conversations Collection (Pope and Davis, 2011), to develop their model. This corpus is relatively small however so they trained their model on only 21 interviews with people that have AD and 17 dialogues with control patients that did not have any neuropsychological conditions. The speech elicited is conversational though not spontaneous as they are interviews.

**Corpus Collection:** Fortunately, a new variant of the map task (Anderson et al., 1991) has been specifically developed recently to elicit spontaneous conversational speech from people with dementia (de la Fuente Garcia et al., 2019). This new work elicits spatial navigation dialogue, for example, which is a known cognitive marker of AD but is also required to guide assisted living robots around the home.

The creators of this task are working on a longitudinal collection with the aim to identify speech and dialogue features that can help *predict* cognitive decline leading to AD. We however, are working with the creators of this task and Alzheimer Scotland to collect a corpus from people with various types of dementia to improve computational *processing* of the above speech and dialogue features. We also plan to release this corpus on DementiaBank for use by other researchers working on socially responsible projects.

# References

Dementia Statistics Hub Alzheimer's Research UK. 2018. Deaths due to dementia. *Alzheimer's Research UK*.

Randa Ben Ammar and Yassine Ben Ayed. 2018. Speech processing for early alzheimer disease diagnosis: Machine learning based approach. In *2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA)*, pages 1–8. IEEE.

Anne H Anderson, Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, et al. 1991. The hcrc map task corpus. *Language and speech*, 34(4):351–366.

Timo Baumann, Casey Kennington, Julian Hough, and David Schlangen. 2017. Recognising conversational speech: What an incremental asr should do for a dialogue system and how to get there. In *Dialogues with Social Robots*, pages 421–432. Springer.

Timo Baumann and David Schlangen. 2012. The inprotk 2012 release. In *NAACL-HLT Workshop on Future Directions and Needs in the Spoken Dialog Community: Tools and Data*, pages 29–32. Association for Computational Linguistics.

Janet Beavin Bavelas and Jennifer Gerwing. 2011. The listener as addressee in face-to-face dialogue. *International Journal of Listening*, 25(3):178–198.

James T Becker, François Boiler, Oscar L Lopez, Judith Saxton, and Karen L McGonigle. 1994. The natural history of alzheimer's disease: description of study cohort and accuracy of diagnosis. *Archives of Neurology*, 51(6):585–594.

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.

Brianna Marlene Broderick, Si Long Tou, and Emily Mower Provost. 2018. Cogid: A speech recognition tool for early detection of alzheimer's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 14(7):P191–P192.

GOODWIN Charles. 1981. Conversational organization: Interaction between speakers and hearers.

Herbert H. Clark. 1996. *Using Language*. Cambridge University Press.

Andrei C Coman, Koichiro Yoshino, Yukitoshi Murase, Satoshi Nakamura, and Giuseppe Riccardi. 2019. An incremental turn-taking model for task-oriented dialog systems. *arXiv preprint arXiv:1905.11806*.

David DeVault, Kenji Sagae, and David Traum. 2011. Incremental interpretation and prediction of utterance meaning for interactive dialogue. *Dialogue and Discourse*, 2(1):143–170.

A. Eshghi, M. Purver, and Julian Hough. 2011. Dylan: Parser for dynamic syntax. Technical report, Queen Mary University of London.

Arash Eshghi. 2015. DS-TTR: An incremental, semantic, contextual parser for dialogue. In *Proceedings of Semdial 2015 (goDial), the 19th workshop on the semantics and pragmatics of dialogue*.

Arash Eshghi, Julian Hough, Matthew Purver, Ruth Kempson, and Eleni Gregoromichelaki. 2012. Conversational interactions: Capturing dialogue dynamics. In S. Larsson and L. Borin, editors, *From Quantification to Conversation: Festschrift for Robin Cooper on the occasion of his 65th birthday*, volume 19 of *Tributes*, pages 325–349. College Publications, London.

Arash Eshghi, Christine Howes, Eleni Gregoromichelaki, Julian Hough, and Matthew Purver. 2015. Feedback in conversation as incremental semantic update. In *Proceedings of the 11th International Conference on Computational Semantics*, pages 261–271.

Arash Eshghi, Igor Shalyminov, and Oliver Lemon. 2017. Bootstrapping incremental dialogue systems from minimal data: linguistic knowledge or machine learning? In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Fernanda Ferreira, Ellen F. Lau, and Karl G. D. Bailey. 2004. Disfluencies, language comprehension, and tree adjoining grammars. *Cognitive Science*, 28(5):721ñ749.

Sofia de la Fuente Garcia, Craig W Ritchie, and Saturnino Luz. 2019. Protocol for a conversation-based analysis study: Prevent-ed investigates dialogue features that may help predict dementia onset in later life. *BMJ open*, 9(3):e026254.

John J. Godfrey, Edward Holliman, and J. McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. In *Proceedings of IEEE ICASSP-92*, pages 517–520, San Francisco, CA.

Tom Hargreaves, Charlie Wilson, and Richard Hauxwell-Baldwin. 2018. Learning to live in a smart home. *Building Research & Information*, 46(1):127–139.

P. G. T. Healey, Arash Eshghi, Christine Howes, and Matthew Purver. 2011. Making a contribution: Processing clarification requests in dialogue. In *Proceedings of the 21st Annual Meeting of the Society for Text and Discourse*, Poitiers.

Sumi Helal and Christopher N Bull. 2019. From smart homes to smart-ready homes and communities. *Dementia and geriatric cognitive disorders*, 47(3):157–163.

Mattias Heldner, Anna Hjalmarsson, and Jens Edlund. 2013. Backchannel relevance spaces. In *Nordic Prosody: Proceedings of XIth Conference, Tartu 2012*, pages 137–146.

Julian Hough. 2015. *Modelling Incremental Self-Repair Processing in Dialogue*. Ph.D. thesis, Queen Mary University of London.

Julian Hough and Matthew Purver. 2012. Processing self-repairs in an incremental type-theoretic dialogue system. In *Proceedings of the 16th SemDial Workshop on the Semantics and Pragmatics of Dialogue (SeineDial)*, pages 136–144, Paris, France.

Julian Hough and David Schlangen. 2016. Investigating fluidity for human-robot interaction with real-time, real-world grounding strategies. In *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 288–298.

Christine Howes. 2012. *Coordination in dialogue: Using compound contributions to join a party*. Ph.D. thesis, Queen Mary University of London.

Christine Howes and Arash Eshghi. 2017a. Feedback relevance spaces: The organisation of increments in conversation. In *IWCS 201712th International Conference on Computational SemanticsShort papers*.

Christine Howes and Arash Eshghi. 2017b. Feedback relevance spaces: the organisation of increments in conversation. In *Proceedings of the 12th International Conference on Computational Semantics (IWCS)*.

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

IBM. 2018. How conversation (with context) will usher in the ai future. *IBM*.

Ruth Kempson, Ronnie Cann, Eleni Gregoromichelaki, and Stergios Chatzikiriakidis. 2016. Language as mechanisms for interaction. *Theoretical Linguistics*, 42(3-4):203–275.

Akinobu Lee and Tatsuya Kawahara. 2009. Recent development of open-source speech recognition engine julius. In *Proceedings: APSIPA ASC 2009: Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference*, pages 131–137. Asia-Pacific Signal and Information Processing Association, 2009 Annual .

Saturnino Luz, Sofia de la Fuente, and Pierre Albert. 2018. A method for analysis of patient speech in dialogue for dementia detection. *arXiv preprint arXiv:1811.09919*.

Kirsty Maclachlan. 2017. Projected population of scotland with uk comparisons. *National Records of Scotland*.

Angelika Maier, Julian Hough, and David Schlangen. 2017. Towards deep end-of-turn prediction for situated spoken dialogue systems. *Proceedings of INTERSPEECH 2017*.

Maike Paetzel, Ramesh Manuvinakurike, and David DeVault. 2015. "so, which one is it?" the effect of alternative incremental architectures in a high-performance game-playing agent. In *Proceedings of the SIGDIAL 2015 Conference*. Association for Computational Linguistics.

Charlene Pope and Boyd H Davis. 2011. Finding a balance: The carolinas conversation collection. *Corpus Linguistics and Linguistic Theory*, 7(1):143–161.

Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice interfaces in everyday life. In *Proceedings of CHI*.

Vineel Pratap, Awni Hannun, Qiantong Xu, Jeff Cai, Jacob Kahn, Gabriel Synnaeve, Vitaliy Liptchinsky, and Ronan Collobert. 2018. wav2letter++: The fastest open-source speech recognition system. *arXiv preprint arXiv:1812.07625*.

Matthew Purver, Christine Howes, Eleni Gregoromichelaki, and Patrick G. T. Healey. 2009. Split utterances in dialogue: A corpus study. In *Proceedings of the 10th Annual SIGDIAL Meeting on Discourse and Dialogue (SIGDIAL 2009 Conference)*, pages 262–271, London, UK. Association for Computational Linguistics.

Helen Puttick. 2018. Bed-blocking at worst level since 2016. *The Times*.

Andrew Rafla and Casey Redd Kennington. 2019. Incrementalizing rasa's open-source natural language understanding pipeline. In *arXiv:1907.05403*.

Matthew Roddy, Gabriel Skantze, and Naomi Harte. 2018. Investigating speech features for continuous turn-taking prediction using lstms. *arXiv preprint arXiv:1806.11461*.

J.P. de Ruiter, H. Mitterer, and N.J. Enfield. 2006. Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, 82(3):515–535.

Harvey Sacks, E.A. Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735.

Seiji Sakakibara, Sachio Saiki, Masahide Nakamura, and Kiyoshi Yasuda. 2017. Generating personalized dialogue towards daily counseling system for home dementia care. In *International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management*, pages 161–172. Springer.

E.A. Schegloff. 2000. Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, (29):1–63.

David Schlangen and Gabriel Skantze. 2009. A general, abstract model of incremental dialogue processing. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 710–718. Association for Computational Linguistics.

David Schlangen and Gabriel Skantze. 2011. A general, abstract model of incremental dialogue processing. *Dialogue and Discourse*, 2(1):83–111.

Elizabeth Shriberg. 1996. Disfluencies in switchboard. In *In Proceedings of the International Conference on Spoken Language Processing*, volume 96, pages 3–6. Citeseer.

Gabriel Skantze. 2010. Jindigo: a java-based framework for incremental dialogue systems. Technical report, KTH, Stockholm.

Gabriel Skantze and Anna Hjalmarsson. 2010a. Towards incremental speech generation in dialogue systems. In *Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 1–8. Association for Computational Linguistics.

Gabriel Skantze and Anna Hjalmarsson. 2010b. Towards incremental speech generation in dialogue systems. In *Proceedings of the SIGDIAL 2010 Conference*, pages 1–8, Tokyo, Japan. Association for Computational Linguistics.

UN. 2018. World population ageing highlights. *United Nations Department of Economic and Social Affairs*.

Chloe Wright. 2015. Facts about carers. *Carers UK*.

Zining Zhu, Jekaterina Novikova, and Frank Rudzicz. 2018. Detecting cognitive impairments by agreeing on interpretations of linguistic features. *arXiv preprint arXiv:1808.06570*.