Establishing the Informational Requirements for Modelling Open Domain Dialogue and Prototyping a Retrieval Open Domain Dialogue System

Trent Meier[0000-0003-3401-1384] and Elias Pimenidis[0000-0003-3593-8640]

Department of Computer Science & Creative Technologies,
University of the West of England, Bristol, UK
trent.meier@outlook.com, Elias.Pimenidis@uwe.ac.uk

Abstract. Open domain dialogue systems aim to coherently respond to users over long conversations through multiple conversational turns. Modelling open domain dialogue is challenging as both the syntactic and semantic features of language play a role in response formation. As similarity to human dialogue has been considered the goal of open domain dialogue systems, this paper takes the view that human linguistic reasoning research can be informative to the requirement engineering process of modelling open domain dialogue. Through a review of linguistic reasoning research and modern approaches in open domain dialogue systems, the authors present informational hypotheses impacting the modelling of open domain dialogue systems. Furthermore, this paper discusses the design and testing of an open domain dialogue system presenting response BLEU-1 scores of 35.41% based on the DailyDialogue Dataset.

Keywords. Natural Language Processing (NLP) · Open domain dialogue Modelling · DailyDialogue Dataset.

1 Introduction

Dialogue systems are a field of natural language processing (NLP) that seeks to produce conversational responses to user inputs. With dialogue systems finding application in diverse use cases ranging from chatbots to personal assistants and video game NPCs, advancements in deep learning techniques have led to significant improvements in system capabilities. Dialogue systems can largely be categorized between task-based dialogue systems and open domain dialogue systems. Task-based dialogue systems typically have fixed objectives while open domain dialogue systems attempt to conduct open conversation with users [25]. Open domain dialogue is challenging to NLP through the role of syntactics and semantics in forming conversational responses, compounded by the variety of topical domains found in open dialogue.

This paper presents a literature review of linguistic reasoning, comparing the findings with a review of related work in modern open domain dialogue techniques. The authors then propose informational hypotheses informed by the findings of the literature reviews which influence the requirements engineering for the paper's open domain dialogue system. Finally, this paper will discuss the dialogue system's design considerations, methodology and the results of testing.

2 Linguistic Reasoning

Linguistic reasoning describes the process of natural language understanding, reasoning, and response formation. While the cognitive processes of human linguistic reasoning are an open area of research, the authors compare literature regarding language and grammar with the findings of research discussing linguistic reasoning, cognitive sciences, and analogical reasoning. This review aims to identify the informational elements of language used in linguistic reasoning to inform the evaluation of open domain dialogue systems and to establish the informational requirements for language modelling in dialogue systems.

2.1 Language and Grammar

Literature widely acknowledges that language consists of elements that capture temporal, spatial, and objective information about the experienced environment [5]. Language encodes this information in structures that communicate these experiences to others or for self-reference. The shared nature of informational exchange has led Chomskain linguists to suggest that language has a common underlying structure leading to the suggestion that language models and formal grammar symbolize the cognitive processes that underlie them. It has been the focus of many language researchers to demonstrate the universality of grammar, with work including evaluating subject-verb sentence structures, reflexive bindings of pronouns, semantic dependency identification, and syntactic hierarchies [20]. The study of Universal Grammar (UG), while initially focusing on the English language, has been expanded to consider comparisons of other languages including multi-lingual translation. Findings from this cross-comparison have resulted in the criticism of UG with researchers suggesting that variations between languages invalidate the suggestion of UG [9]. While no definitive language structures have emerged as an answer to the suggestion of UG, consensus is growing that universality would occur at a higher abstraction of language than the commonly held grammar structures.

2.2 Linguistic Reasoning and Cognitive Science

Attempts to improve understanding of linguistic reasoning have used cognitive science research to identify key informational and structural attributes of human language processing [20]. The study of analogical reasoning includes disciplines ranging from linguistic reasoning to visuospatial and numeric reasoning including memory theories. Advancements in the field have resulted from analogical reasoning experimentation combined with eye tracking, brain imaging, and developmental studies [24]. It is beyond the aims of this review to discuss the physical processes of analogical reasoning, but the informational processing of natural language in linguistic reasoning is relevant to the aims of identifying the informational requirements for open domain dialogue modelling.

Analogical reasoning comprises the processes of knowledge acquisition, storage, and memory reuse as part of a problem-solving paradigm. It has been used to explain human problem solving, creativity, and has even been suggested to underlie human cognition [5]. Analogical reasoning matches environmental stimuli with similar experiences through a mapping process that impacts the outcome of memory retrieval. Once retrieved, the memories of experience inform the outcomes of analogical reasoning through a process of further mapping and executive function [24].

Developmental studies of analogical reasoning show that children's objective focus leads to inferior analogical reasoning, while adult attenuation to relational information leads to more successful analogical reasoning [23]. Christie, Gao, and Ma [5] suggest that language plays a key role in conveying both objective information and relational information suggesting that these informational elements demonstrate that analogical reasoning rather than UG is common across linguistic differences. Temporal information conveyed through language forms another important element of linguistic reasoning and can be conveyed through language directly or as sequential information captured as experience through the consumption of language [2]. The preponderance of analogical reasoning research places emphasis on the importance of relational and temporal information in higher abstraction analogical mapping, memory retrieval, and reasoning.

2.3 Informational Attributes of Language

The following presents key informational components of language, each of which holds importance to linguistic reasoning outcomes.

Syntactic Information. The syntactic component of language describes linguistic elements including commonly experienced information (nouns) and named entities. Other syntactic information includes relational, temporal, and spatial information explicitly described by language [19].

Semantic Information. The semantic component of language describes implicit language meaning that captures relational, temporal, and spatial information conveyed through the consumption of languages. Christie, Gao, and Ma [5] emphasize the importance of relational information and temporal information in discourse processing, highlighting the importance of semantic information in modelling open domain dialogue systems.

Some authors note that even the linguistic elements widely accepted as syntactic are semantically influenced through the process of language learning [11]. This work acknowledges these findings but adopts the common definitions of syntax and semantics for the precision afforded to discussion by differentiating the explicit and implicit information in language.

3 Related Work in Open Domain Dialogue Modelling

Conversational dialogue systems have seen recent advances resulting from improvements in Deep Neural Network (DNN) architectures in modern approaches. An emphasis of research in open domain dialogue has focused on modelling the semantic information contained in multi-turn dialogue with the goal of more accurately forming natural language responses.

3.1 Dialogue Representation

DNN approaches to open domain dialogue require numeric language representations that encode the informational elements of language for processing. Embedding methods have emerged as the most successful methodology of representing dialogue in DNN techniques but vary in the embedding methodologies, informational capture, and intended application. This review of language encodings will evaluate the suitability of embedding methodologies for open domain dialogue systems informed by the findings of the review of linguistic reasoning.

One-hot encoding uses vector representations of a predefined set of vocabulary using the count of the occurrence of each word to represent the feature value in an embedding vector. Kim, Hong, and Cha [14] note that One-Hot encoding ignores the relational and temporal features of language, and as a result demonstrates lesser performance and increased memory requirements compared to alternative approaches. Due to the loss of relational and temporal language features, One-Hot encoded vector representations can be described as unsuitable for language encoding in open domain dialogue systems.

Skip-Gram encoding methods represent words as multi-dimensional vectors derived from the occurrence words adjacent to the word of the embedding. Skip-Gram language representation captures important informational elements of language but lacks direct representational capture of dialogues temporal or sequential information which forms an important element of linguistic reasoning.

Sequence-to-Sequence encoders describe the encoding element of encoder-decoder networks which form the basis for many modern approaches in open domain dialogue systems. Yang, Rong, and Xiong [25] state that Sequence-to-Sequence encoders have been demonstrated with Recurrent Neural Networks, Long-Term Short-Term Memory Networks, and Gated Recurrent Unit Networks, capturing improved semantics through high dimensional hidden state vectors. Through the capture of natural language semantics Sequence-to-Sequence based encoders demonstrate better suitability for open domain dialogue generation than other language encoding methodologies.

Zhou et al., [26] describe attention-based sequence embedding techniques as a key element in the advancement of dialogue representation in modern state-of-the-art language models. Attention-based Sequence-to-Sequence encoders use attention mechanisms to achieve embeddings by evaluating words in a sequence preserving the objective information, relational information, and temporal information at varying levels of abstraction [7]. This encoding scheme combined with transformer encoder-decoder networks has demonstrated state-of-the-art performance in open domain dialogue [1],

[22]. Despite these advances, deep semantic feature embeddings remain an open area of research due to the computational complexity of training such language models. More work is needed to integrate the deeper level relational and temporal informational encoding necessary for higher abstraction semantic reasoning. Other important areas of research correlate the incorporation of extra-lexical information, such as environmental information, with improved performance of dialogue systems [10].

3.2 Dialogue Response Formation

Modern open domain dialogue systems are generally classified by the manner conversational responses are generated as either retrieval dialogue systems or generative dialogue systems. Retrieval dialogue systems leverage an available corpus of dialogue using matching mechanisms to retrieve similar conversational responses to an input. Modern approaches have demonstrated neural network-based representation and matching methodologies to perform dialogue retrieval for conversational systems [18]. While retrieval based open domain dialogue systems have used a variety of language encoding methodologies, attention-based encoding methodologies such as BERT models have been demonstrated to produce state-of-the-art retrieval-based dialogue performance [12].

Generative dialogue systems differ from retrieval-based dialogue systems through the mechanism of response formation. Huang and Zhou [12] describe generative dialogue systems as forming responses by predicting each word of the response for each word of the input sequence. Sequence-to-sequence models have seen wide application in generative dialogue systems following an encoder-decoder network model. Other frameworks include Generative Adversarial Networks and Conditional Variational Autoencoders [13]. During the Convai2 challenge, Sequence-to-Sequence generative dialogue systems outperformed response-based approaches with the winning proposals all utilizing BERT based architectures [8].

The relative advantages of both dialogue system methodologies have led to hybridization of architectures that combine generative approaches with retrieval-based approaches. Facebook’s Blender open domain dialogue model which claims state-of-the-art performance in response formation, uses a hybridized poly-encoder retrieval methodology with a response ranking and blending module based on decoder networks [22]. Other hybrid solutions demonstrate improvements through the incorporation of increasingly complex semantic representations of language supporting the findings of the literature review of linguistic reasoning which identified semantic information as a key element of linguistic reasoning [6].

4 Hypotheses Development

The review of linguistic reasoning and related work in open domain dialogue systems identified the need for multi-turn semantic processing for language comprehension but drew no direct correlation between their findings. Drawing on the findings of this research, the authors propose the following informational hypotheses.

4.1 Preservation of Informational Atomicity

Training embedded representations of a dialogue corpus necessarily generalize corpus variations to an approximated fit of the distribution. Preservation of informational atomicity predicts that any informational generalization beyond its least subdivisible component holds the potential to obscure informational elements that may differentiate a conversational response. This hypothesis recognizes the necessity of generalization as subject to processing requirements, cost, and complexity, and notes that human linguistic reasoning similarly generalizes environmental experience. Despite this, the hypothesis suggests that lower informational atomicity in informational representation will result in improved open domain dialogue response formation.

4.2 Modelling Dialogue as Semantics, not Ontologies

Modelling language as semantics rather than ontology suggests that open domain language models should represent language as semantics, or discreet experiences of the environment and dialogue, rather than as universal types. This hypothesis is supported by the findings of the literature review that highlights the role of linguistic reasoning as a communication medium for experience rather than UG or ontological structures [10]. This hypothesis further suggests that natural language itself is the least atomic form of linguistic reasoning which is founded in the diverse experience of environmental stimuli and uses a process of mapping language to experiences rather than reasoning based solely on the information conveyed in language. This hypothesis predicts that systems that only model communicated language will be unable to match the distribution of user-generated dialogue. This hypothesis suggests that any environmental information that can be perceived by the users of natural language or that has an influence on response formation, has the potential to improve modelling of language for open domain dialogue.

4.3 The Role of Distant Semantic Features

The role of distant semantic features hypothesizes that distant or multi-sequence semantic features are impactful to response formation and linguistic reasoning outcomes. This hypothesis is supported by the growing evidence demonstrating the importance of distant semantic features to response formation both from linguistic reasoning and related work in open domain dialogue [1], [19]. While this hypothesis acknowledges that application environments often limit a deployed language models' ability to capture temporal semantics, it suggests that modelling of distant semantic features combined with preservation of the temporal continuum and preservation of informational atomicity will improve open domain dialogue response formation.

5 Methodology

The paper's dialogue system implementation followed a methodology that used the proposed informational hypotheses to inform the selection of technologies and dialogue modelling architectures. The authors' hypotheses informed the requirements engineering of the system's implementation, but it was beyond the scope of the paper's work to control for all the hypotheses implications.

5.1 Dialogue Representation

Attention-based Sequence-to-Sequence encoding methodologies demonstrate state-of-the-art performance for generating language representations and can capture the objective and semantic features of language [1]. Another benefit of using attention-based encoders for language encoding is the availability of pre-trained and open-sourced language models for use in transfer learning. A disadvantage of using embedded language models, however, is the violation of the hypothesized preservation of informational atomicity which suggests that language encoding necessarily generalizes corpus variation. Another disadvantage of transfer learning is the lack of availability of models that incorporate other environmental information in addition to language. This makes pre-trained language models unsuitable for testing the hypothesized modelling of language as semantics rather than ontology. The drawbacks of Sequence-to-Sequence DNN approaches led the paper's implementation to develop two alternative dialogue representation and response formation architectures.

The first dialogue representation tested used directed graph-based language representation that stored dialogue as values in the nodes of a directed graph with conversational turns indicated using turn tokens. In this representation, a dialogue corpus was loaded into a graph database for corpus storage. A graph-based language representation was selected as it enables the preservation of atomic conversational features and implicitly preserves sequential semantics (Figure 1).

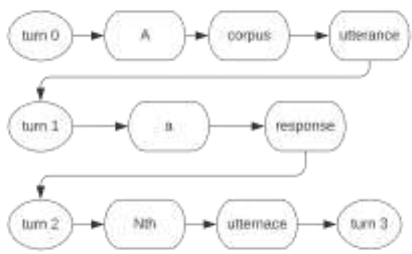

Fig. 1. Directed Conversation Graph

The second dialogue representation tested used feature vector-based language representation using the established approach demonstrated by Sequence-to-Sequence language encoding methodologies. This selection was justified by the findings of the review of related work suggesting that attention-based Sequence-to-Sequence encoding methodologies preserve semantic information better than alternatives. It was beyond the scope of the paper's work to train language encoding models, so two pre-trained dialogue encoding models were selected for evaluation through transfer learning. The first of these models, the Universal Sentence Encoder was trained on a variety of natural language including dialogue and is provided by Cer et al., [4]. The second, Sentence-BERT was trained for dialogue-based response generation and is provided by Reimers et al., [21]. In the vector-based representation, a dialogue corpus was encoded as feature vectors and stored in a vector database for corpus storage. A conceptual visualization of the language representation using feature vectors is given in Figure 2.

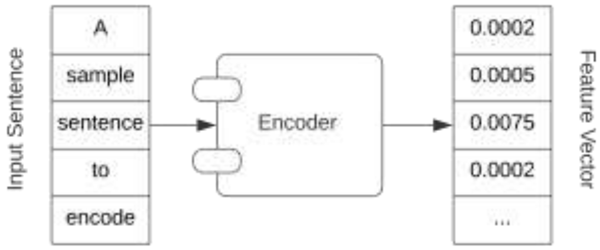

Fig. 2. DNN Encoding

5.2 Response Formation

Comparing architectures between generative and retrieval architectures demonstrates advantages and disadvantages between both approaches. Modern generative approaches demonstrate state-of-the-art performance compared to retrieval architectures but have trade-offs such as expensive computational requirements, reduced fluency, and decreased syntactic correctness compared to retrieval-based systems [12]. Lan et al., [15] note that hybrid architectures minimize some of these downsides but are comparatively more computationally expensive. Due to the prohibitive computational requirements of generative and hybrid architectures and paper's prototype selected a retrieval architecture for both the graph-based approach and the embedding-based approach.

Retrieval in the graph-based representations used graph transversal-based query matching combined with keyword feature extraction to match similar results from an input conversational graph returning a result set of similar corpus responses. In the graph-based approach, system responses were selected from the next conversational turn in the dialogue graph enabling corpus responses to be reused in the current conversational context.

The embedding-based approach used cosine similarity-based matching between embeddings of the current conversational input with the embeddings of the dialogue corpus in a vector database. System responses were selected by matching a key value stored in both the embedding and a paired natural language response stored in a

NoSQL database. As with the graph-based approach, corpus responses were reused in the current conversational context.

5.3 Response Evaluation

Evaluating open dialogue systems presents unique challenges as many possible responses could be considered valid in a dialogue's context. While evaluation of open domain dialogue is an important area of ongoing research, current evaluation methodologies are divided between the retrieval-based and generation-based system architectures. Huang et al. [12], note that both BLEU and ROUGE, statistical metrics commonly used in dialogue evaluation, rely heavily on system responses lexical similarity with a target response and not with a human's judgement of coherence. Lan et al [15] suggest that the most reliable means of evaluating open domain dialogue is human annotation but adds that it is not reproducible and is further time-consuming.

While literature notes the limitations of using BLEU and ROUGE in scoring open domain dialogue, they remain popular accuracy measures for response formation. Campillos-llanos et al. [3] suggest that F-Measure scoring correlates more closely to human evaluation than BLUE or ROUGE and has the added benefit of correcting for false positives. Motivated by the widespread use of these measures, this paper's evaluation of the prototype open domain dialogue system used a combination of BLEU, ROUGE, and F1 to enable comparison and provide an objective measure of system accuracy.

6 Results

Proof-of-concept testing during the paper's implementation of the graph-based feature extraction and response formation architecture demonstrated unanticipated limitations. To verify the preliminary test results, a systematic review of query matching approaches was conducted. The result of this testing is presented below with reliability indicating the frequency of returned results-sets and semantic cohesion measured by subjective user evaluation (Table 1).

Table 1. Proof of Concept Testing Graph-Based Representation

Query Match Type	Reliability	Semantic Cohesion
Exact Node Values with Wildcard Pattern	Very Low	Low
Exact Node Values with Exact Path Pattern	Very Low	Low
Node Value Set with Wildcard Pattern	Medium	Very Low
Node Value Set with Exact Pattern	Low	Very Low
Limited Node Sets with Wildcard Pattern	Medium	Very Low

The methodical testing presented results indicating the limitation of the implementation. Further testing was conducted into alternative matching approaches such as Jaccardian Similarity, KNN similarity, ANN similarity, and Overlap Similarity. Testing of these matching methodologies produced similar results leading the authors to conclude that a graph-based representation would have to be more thoroughly explored in future work.

testing was undertaken to evaluate the embedding-based dialogue representation using the encoding models Universal Sentence Encoder [4] and Sentence-BERT [21]. Testing was conducted using varying conversational context over dialogue turns from the DailyDialogue corpus. The result of this testing is presented in Table 2.

Table 2. Performance Testing Embedding-Based Representation

Model	Turns	BLEU -1	ROUGE	F-measure (0.5)
Universal Sentence Encoder	1	34.812%	34.633%	33.565%
Universal Sentence Encoder	2	34.758%	34.499%	33.491%
Universal Sentence Encoder	4	22.355%	21.934%	20.928%
Sentence-BERT	1	33.188%	32.880%	31.930%
Sentence-BERT	2	35.409%	32.145%	31.347%
Sentence-BERT	4	19.821%	19.653%	19.574%

7 Discussion of Performance

The prototypes testing aimed to evaluate the suitability of the systems response formation of the course of open domain dialogue. While appropriate metrics for testing semantic cohesion remain an outstanding research question, BLEU, ROUGE, and F-Measure were identified as commonly implemented metrics and were used in testing the accuracy of the system's responses. To further evaluate the success of the system’s response formation, the performance test results are compared with the results of other open domain dialogue systems using the DailyDialogue dataset [16]. These results are presented in Table 3.

Table 3. Comparison of Performance of Vector-Based Representation

Author	Network Type	BLEU-1	ROUGE	F (0.5)
Li et al. [16]	Sequence-to-Sequence (Retrieval)	35.1%	-	-
	Attn-Sequence-to-Sequence (Retrieval)	33.5%	-	
	Hierarchical Encoder-Decoder (Generative)	**39.6**%	-	-
Luo et al. [17]	Sequence-to-Sequence (Generative)	12.43%	-	-
	Auto-Encoder-Matching (Generative)	13.55%	-	-
	Attn-Sequence-to-Sequence (Generative)	13.63%	-	-
	Auto-Encoder-Matching (Generative)	14.17%	-	-
This Paper	Universal-Sentence-Encoder (Retrieval)	34.81%	34.63%	33.57%
	Sentence-BERT (Retrieval)	**35.41**%	32.14%	31.35%

The testing demonstrated that the paper‘s implementation of open domain response formation compared competitively to the results of both paper‘s in BLEU scoring. Of the test results, the Sentence-BERT encoder implementation performed best with a BLEU score of 35.41% exceeding all but the Hierarchical Encoder-Decoder implemented by Li et al, [16]. The best accuracy was achieved using a conversational context of two turns in the systems‘s corpus encoding and matching process.

While the hypothesised improvements resulting from the preservation of informational atomicity were not controlled for in the testing of the open domain dialogue system, the results indicated outcomes supporting the hypothesis. The higher-performing Sentence-BERT encoder represents its embeddings with larger feature vector lengths of 1024 compared to the Universal Sentence Encoder’s vector length of 512. While the finding does not control for encoder architecture or training data, the higher degree of informational atomicity enabled by the larger encoding provides limited supporting evidence for the hypothesis.

The hypothesised role of distant semantic features in improving response formation was tested in the system‘s accuracy testing. The authors of both encoding models suggest that near semantic features are preferred in the pre-trained encodings models training data [4], [21]. This suggestion was confirmed by the results of the systems testings. Despite this, a semantic depth of two conversational turns resulted in the most successful response formation using the Sentence-BERT encoder. The results indicate that distant semantic features play a role in response formation, but limitations of the pre-trained encoding scheme prevented testing of more distant feature extraction.

7.1 Future Work

The paper’s open domain dialogue system was able to demonstrate response formation using the syntactic and semantic features of language in an architecture enabling continuous learning through the preservation of experience. The hypotheses identified during the literature review played an important role in informing requirements analysis, design decisions, and implementation. Although the use of a transfer-learning model prevented controlled experimentation for the paper’s hypotheses, future work could more directly test each of the hypothesised informational elements of language to establish their veracity.

Future work could test the hypothesised role of informational atomicity by measuring differences in the statistical distributions or long-range dependencies between a language models responses and the distribution of a dataset. Results of such work could measure the role that informational atomicity plays in a models accuracy. Testing the hypothesis suggesting that language be modelled as semantics rather than ontology could take different approaches, but integrating environmental sensor data in modelling open domain dialogue could correlate environmental data to response outcomes. Such testing would require the availability of high-quality datasets incorporating environmental information.

Finally, future work could directly test the role of distant semantic features by varying the training regimes of open domain dialogue systems and comparing the semantic cohesion of the resulting responses. Analogical reasoning literature suggests that multi-sequence long-range semantic dependencies are integrated into linguistic reasoning through a mapping process from shorter sequences of inputs [24]. Modelling short-to-long sequence mapping could produce more efficient feature extraction and improved outcomes in open domain dialogue modelling.

8 Conclusion

The paper‘s view of linguistic reasoning provided informational hypotheses that correlated findings between the fields of Linguistics, Cognitive Sciences, and related work in Open Domain Dialogue modelling. While the informational hypotheses proposed by the paper were not directly tested in the paper‘s open domain dialogue system the hypotheses informed the requirement engineering and technology selection of the system‘s implementation. Following the implementation and testing the system responded with a BLEU-1 accuracy of 35.41% using the DailyDialogue Dataset using the vector embedding based language representation tested. Further work could build on the informational hypotheses proposed by the paper having the potential to impact open domain dialogue modelling and improving response formation.

9 References

1. Adiwardana, D. et al.: Towards a Human-like Open-Domain Chatbot. arXiv:2001.09977 [cs, stat]. (2020).
2. van den Broek, P. Helder, A.: Cognitive Processes in Discourse Comprehension: Passive Processes, Reader-Initiated Processes, and Evolving Mental Representations. Discourse Processes. 54, 5–6, 360–372 (2017). https://doi.org/10.1080/0163853X.2017.1306677.
3. Campillos-Llanos, L. et al.: Designing a virtual patient dialogue system based on terminology-rich resources: Challenges and evaluation. Natural Language Engineering. 26, 2, 183–220 (2020). https://doi.org/10.1017/S1351324919000329.
4. Cer, D. et al.: Universal Sentence Encoder. arXiv:1803.11175 [cs]. (2018).
5. Christie, S. et al.: Development of Analogical Reasoning: A Novel Perspective From Cross-Cultural Studies. Child Development Perspectives. 14, 3, 164–170 (2020). https://doi.org/10.1111/cdep.12380.
6. Cui, C. et al.: User Attention-Guided Multimodal Dialog Systems. In: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 445–454 Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3331184.3331226.
7. Devlin, J. et al.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805 [cs]. (2019).
8. Dinan, E. et al.: The Second Conversational Intelligence Challenge (ConvAI2). arXiv:1902.00098 [cs]. (2019).
9. Evans, N., Levinson, S.C.: With diversity in mind: Freeing the language sciences from Universal Grammar. Behavioural and Brain Sciences. 32, 5, 472–492 (2009). https://doi.org/10.1017/S0140525X09990525.

10. Hammer, R. et al.: Individual differences in analogical reasoning revealed by multivariate task-based functional brain imaging. NeuroImage. 184, 993–1004 (2019). https://doi.org/10.1016/j.neuroimage.2018.09.011.
11. Horvath, S. et al.: Acquisition of Verb Meaning from Syntactic Distribution in Preschoolers with Autism Spectrum Disorder. Language, Speech & Hearing Services in Schools (Online). 49, 3S, 668–680 (2018). http://dx.doi.org/10.1044/2018_LSHSS-STLT1-17-0126.
12. Huang, M. et al.: Challenges in Building Intelligent Open-domain Dialog Systems. ACM Trans. Inf. Syst. 38, 3, 21:1-21:32 (2020). https://doi.org/10.1145/3383123.
13. Ke, P. et al.: Generating Informative Responses with Controlled Sentence Function. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 1499–1508 Association for Computational Linguistics, Melbourne, Australia (2018). https://doi.org/10.18653/v1/P18-1139.
14. Kim, H.J. et al.: seq2vec: Analyzing sequential data using multi-rank embedding vectors. Electronic Commerce Research and Applications. 43, 101003 (2020). https://doi.org/10.1016/j.elerap.2020.101003.
15. Lan, T. et al.: Self-attention Comparison Module for Boosting Performance on Retrieval-based Open-Domain Dialog Systems. arXiv:2012.11357 [cs]. (2020).
16. Li, Y. et al.: DailyDialog: A Manually Labelled Multi-turn Dialogue Dataset. arXiv:1710.03957 [cs]. (2017).
17. Luo, L. et al.: An Auto-Encoder Matching Model for Learning Utterance-Level Semantic Dependency in Dialogue Generation. arXiv:1808.08795 [cs]. (2018).
18. Mehndiratta, A., Asawa, K.: Non-goal oriented dialogue agents: state of the art, dataset, and evaluation. Artif Intell Rev. 54, 1, 329–357 (2021). https://doi.org/10.1007/s10462-020-09848-z.
19. Nahatame, S.: Revisiting Second Language Readers' Memory for Narrative Texts: The Role of Causal and Semantic Text Relations. Reading Psychology. 41, 8, 753–777 (2020). https://doi.org/10.1080/02702711.2020.1768986.
20. Progovac, L. et al.: Diversity of Grammars and Their Diverging Evolutionary and Processing Paths: Evidence From Functional MRI Study of Serbian. Frontiers in Psychology. 9, (2018). https://doi.org/10.3389/fpsyg.2018.00278.
21. Reimers, N., Gurevych, I.: Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. arXiv:1908.10084 [cs]. (2019).
22. Roller, S. et al.: Recipes for building an open-domain chatbot. arXiv:2004.13637 [cs]. (2020).
23. Simms, N.K. et al.: Working memory predicts children's analogical reasoning. Journal of Experimental Child Psychology. 166, 160–177 (2018). https://doi.org/10.1016/j.jecp.2017.08.005.
24. Westphal, A.J. et al.: Shared and distinct contributions of rostrolateral prefrontal cortex to analogical reasoning and episodic memory retrieval. Human Brain Mapping. 37, 3, 896–912 (2016). https://doi.org/10.1002/hbm.23074.
25. Yang, H. et al.: Open-Domain Dialogue Generation: Presence, Limitation and Future Directions. In: Proceedings of the 2019 7th International Conference on Information Technology: IoT and Smart City. pp. 5–12 Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3377170.3377248.
26. Zhou, M. et al.: Progress in Neural NLP: Modeling, Learning, and Reasoning. Engineering. 6, 3, 275–290 (2020). https://doi.org/10.1016/j.eng.2019.12.014.