Coherence disruptions in human-chatbot interaction: towards quantitative approach to conversation

Albert Maršík

Introduction

Recently, there has been a breakthrough in the way we interact with machines[1]. We can instruct a computer using natural language[2]. Besides making existing technology an extra step accessible, new ways to use technology appear. Automated interaction alone can solve previously unsolvable problems, such as notably accessing a knowledge base via semantic search[3]. Until recently a knowledge base would usually be accessed only via fulltext, meaning we would only be able to find information of which we knew a part of the formal encoding. Today, we can search for information simply by asking questions, all thanks to natural language computer interface.

The promise of much practical usage of the current wave of generative AI is ambitious and only brings its fruit slowly, perhaps slower, than was expected[4][5]. There is talk of a "plateau" in development of the technology powering the current cutting edge inventions[6]. And that is not the only issue

[1] D. Sharma et al. "Exploring The Evolution Of Chatgpt: From Origin To Revolutionary Influence". In: *Educational Administration: Theory and Practice* 30.5 (2024).

[2] G. G. Hendrix. "Natural-language interface". In: *American Journal of Computational Linguistics* 8.2 (1982), pp. 56–61.

[3] E. Makela. "Survey of semantic search research". In: *Proceedings of the seminar on knowledge management on the semantic web.* Department of Computer Science, University of Helsinki, Helsinki. 2005.

[4] R. Metz et al. "OpenAI, Google, and Anthropic Are Struggling to Build More Advanced AI". in: *Bloomberg News* (2024). Accessed: 2024-12-19. URL: `https://www.bloomberg.com/news/articles/2024-11-13/openai-google-and-anthropic-are-struggling-to-build-more-advanced-ai`.

[5] K. Hu and A. Tong. "OpenAI Rivals Seek New Path to Smarter AI as Current Methods Hit Limitations". In: *Reuters* (2024). URL: `https://www.reuters.com/technology/artificial-intelligence/openai-rivals-seek-new-path-smarter-ai-current-methods-hit-limitations-2024-11-11`.

[6] G. Ritter and W. Lu. "The AI Plateau Is Real — How We Jump To The Next Breakthrough". In: (Dec. 2024). URL: `https://www.emcap.com/thoughts/ai-s-curve-plateau-proprietary-business-data-breakthrough/`.

there is to the current wave of cutting edge AI. To name the most prominent ones:

- high electricity consumption[7]
- unpredictable and broad societal impact[8][9]

That being said, in context of conversation research, this development in technology promises to makes things possible that previously were not. With a partial control of what happens in the conversation and a decent certainty, that our system will simulate human-human conversation to the user, new kind of conversational data is in reach - logs of the human-bot interaction, that could be categorized based on

- which researcher controlled stimulus and
- which participant reaction to given stimulus

they contain.

In the 1960s the relatively recent emergence and adoption of telephone technology allowed for recording and transcribing authentic conversational data. This advancement took place thanks to developement in technology which is reminiscent of the current day situation. While human-bot conversational data is arguably less authentic than telephone conversation transcripts, experimental approach can be taken while the human element is present.

This papers intention is to provide a debate on a metaresearch question - is using generative AI a viable methodology for conversation research? This is done by attempting to develop that very methodology. Proceeding we operate in a frontier, our first steps should be establishing data backed baseline knowledge and assessing possible lines of research.

To understand what something is, it can be fruitful first understanding what that thing is not. One way to understand what makes the unraveling

[7] H. Ritchie. "What's the impact of artificial intelligence on energy demand?" In: *Sustainability by Numbers* (Nov. 2024). URL: `https://www.sustainabilitybynumbers.com/p/ai-energy-demand`.

[8] A Hagerty and I. Rubinov. "Global AI ethics: a review of the social impacts and ethical implications of artificial intelligence". In: *arXiv preprint arXiv:1907.07892* (2019).

[9] M. T. Baldassarre et al. "The social impact of generative ai: An analysis on chatgpt". In: *Proceedings of the 2023 ACM Conference on Information Technology for Social Good*. 2023, pp. 363–373.

text of a conversation a valid one - a coherent one - is obtaining conversational data containing coherence disruptions. This can be done using the discussed technology - it has the capability of conversing in a way that is found generally acceptable by humans and can drift away from the coherent interaction if appropriately instruced to do so.

The data this paper seeks to elicitate and analyse are actual human-chatbot exchanges containing moments which have the potential to be problematic for the human participant to process and follow up on. The line between what a coherent and an incoherent conversation is blurred. It is in no way a binary property of the text of the conversation. The goal is therefore to touch on the gradual divide between them.

While chatbots are evaluated for how natural and error free their way of conversing is, human-human conversation is rarely flawless:

- errors happen
- conversational coherence gets temporarily disrupted

In case of human-human communication, disruptions can however be cured easily.

It used to be (and remains so in legacy systems) that in human-bot communication, disruptions could derail a conversation completely, leaving the bot, who would only rely on surface level textual clues, in the dark.[10] This has become rare with generative AI. Even though it brings a set of its own problems like frequently lacking factuality or the difficulty to handle data responsibly, the cutting edge technology powered conversation systems are

- better capable of understanding and producing relevant answers
- instructed to return to their conversational point of departure

Human-bot communication is often single-purpose. Companies and institutions deploy voice applications to interact with customers and clients, so there is usually a goal to be achieved. The coherence of such conversation can then be described based on whether the goal has been achieved with success and ease. Another common scenario is an open-domain conversation, also known as chit-chat or smalltalk.

[10] M. F. McTear. *Conversational AI: Dialogue Systems, Conversational Agents, and Chatbots.* Springer International Publishing, 2020, pp. 43–70.

Some factors that influence coherence in conversational texts, whether in human-human or human-bot exchanges, have been extensively studied.

Namely:

Politeness Brown and Levinson's work on politeness strategies describes social alignment in smooth interactions[11]. Politeness strategies, such as using polite language, offering options, or softening potentially face-threatening comments, help to create a comfortable communicative environment. These strategies align with social norms, which people interpret as markers of respect, consideration, or even trust. A failure to employ these politeness strategies, or using them inconsistently, can disrupt conversational coherence. For example, blunt or overly direct responses may be perceived as abrupt or rude, diverting the conversation's flow or causing discomfort. In such cases, the breakdown of polite norms can lead participants to question intent, hindering effective and smooth communication.

Speech acts Following Austin and Searle's speech act theories, communication rely on expressing clear intentions and meanings that help build mutual understanding[12][13]. When speakers convey intentions explicitly through statements, questions, requests, or assertions, it signals to listeners the purpose and direction of the conversation. Effective communication strategies help maintain coherence by ensuring each contribution builds logically on the last. On the other hand, unclear intentions or ambiguous phrasing can create misunderstandings, disrupting the conversation's flow. Misalignment or mixed signals – such as using sarcasm without cues or making indirect requests without context— can leave listeners uncertain about how to respond, leading to off-track or irrelevant contributions and possibly creating need to address the communication to regain understanding.

Conversational Maxims Grice's conversational maxims are fundamental to coherent dialogue[14]. They suggest that participants should:

- provide truthful information (Quality)
- neither too much nor too little (Quantity)

---

[11] P. Brown and S.C. Levinson. *Politeness: Some Universals in Language Usage.* Politeness: Some Universals in Language Usage. Cambridge University Press, 1987.

[12] J. L. Austin. *How to do things with words.* Oxford University Press, 1962.

[13] J. R. Searle. *Speech Acts: An Essay in the Philosophy of Language.* Cambridge University Press, 1969.

[14] H. P. Grice. "Logic and Conversation". In: *Syntax and Semantics: Vol. 3: Speech Acts.* Ed. by Peter Cole and Jerry L. Morgan. Academic Press, 1975, pp. 41–58.

- remain on-topic (Relevance)
- communicate in an orderly, clear manner (Manner)

These maxims encourage effective exchange by setting a standard for contributions that are informative, truthful, relevant, and unambiguous. When violated, such as by giving excessive detail, omitting important context, or straying from the topic, coherence suffers. For instance, irrelevant tangents or over-detailed explanations may confuse the listener as to what is the main focal point of conversation in that moment. This misalignment can leave participants uncertain about the conversation's direction, ultimately diminishing coherence and the effectiveness of communication.

Sequence Structure The work of Schegloff and Sacks on sequence structure and turn-taking emphasizes that ordered interactions support predictability and continuity in dialogue[15]. Turn-taking conventions — where participants follow an implicit sequence of speaking and responding — help maintain the flow by structuring the conversation in a logical order. This sequence structure allows both parties to anticipate when to listen and when to speak, contributing to a well-paced, cohesive exchange. However, interruptions, abrupt changes in topic, or skipping expected responses can disrupt this sequence, introducing unpredictability that can confuse participants. These interruptions fragment coherence by shifting the conversation away from expected responses or structured flow, often leaving gaps in understanding or causing conversational breakdowns.

Disturbed coherence fill me in .. Bublitz Lenk () hearer knows best disturbed coherence - partial coherence topic

Message and Topic Interactional linguistics underscores that consistency in message and topic preserves continuity in conversation[16]. When speakers stick to a shared topic or make gradual, clear shifts, coherence is maintained because participants know what to expect. Frequent or abrupt topic shifts, however, or sending unclear or conflicting messages, can create disjointed exchanges. For instance, introducing a new topic without closure on the previous one can confuse listeners, leading to a scattered or fragmented interaction. By shifting focus unpredictably or offering unclear messages, co-

[15] E. A. Schegloff. *Sequence Organization in Interaction: A Primer in Conversation Analysis*. Cambridge University Press, 2007.

[16] E. Couper-Kuhlen and M. Selting. *Interactional Linguistics: Studying Language in Social Interaction*. Cambridge University Press, 2017.

herence diminishes as participants lose track of the conversation's thread, resulting in exchanges that feel scattered or incomplete.

While all of the mentioned areas unveil much about the way conversation works, rarely do they concern themselves with the textual dimension of conversation. Most of the mentioned authors (with the notable exception of those operating within the interactional linguistics framework) would hardly be described as linguists, though their works significantly inform linguistics.

The lack of a true interpersonal dimension in human-chatbot communication allows to focus solely on the elements in conversational text, that make it cohesive and coherent or rather those that have the potential to prevent it from being that. The key concepts discussed in this paper are two closely related topics:

Coreference realized by anaphore and topic – what the text is about.

Chapter 1

Theoretical foundations

1.1 Textual dimension of conversation

The following concepts will be explored individually, in relation to one another and in relation to conversation: text, coherence, cohesion, coreference, anaphora, cataphora, endophora, exophora, topic, entity, and association. While the presented exploration draws on existing literature, it seeks to establish an independent and sustainable framework, rather than strictly adhering to established interpretations.

Text

Text, in its broadest sense, refers to any form of communication that conveys meaning through a combination of signs, symbols, or language[1][2]. These semiotic structures can take various forms, including written, spoken, visual, or even non-verbal modes of expression[3]. A text can be as simple as a single sentence or as complex as a novel, and it can exist across different mediums, from books and articles to advertisements and digital content. What defines a text is its ability to convey a coherent message or idea, often intended for interpretation by an audience or an adressee. Texts can serve a wide range of purposes, including storytelling, instruction, persuasion, or simply recording

[1] J. Hrbáček. *Nárys textové syntaxe spisovné češtiny.* Praha: Trizonia, 1994, p. 7.

[2] L. Hjelmslev. *O základech teorie jazyka.* Trans. by F. Čermák. 2. doplněné a upravené vydání. Praha: Academia, 2016.

[3] R. Barthes and S. Heath. *Image, Music, Text.* A fontana original. Fontana Press, 1977, p. 13.

information. Typically text is a structure that is linguistic, produced and percieved as intentional and coherent.

The text of a conversation is specific because it is multiproducer. Another example of a multiproducer text would be a sequence of commercial signs on a busy street. It is the spatial juxtaposition of the signs and temporal juxtaposition of utterances, that make them a text.

Another property of a conversation text is it is negotiated. This is given by its multiproducer and temporal nature. Other types of text which are also negotiated are relatively rare. There are occurences of debates which take place in written text, whether they are press columns or academic articles, which interact explicitly with each other, making them a negotiation. Such press discourse could however be considered a sequence of text units rather than a single temporarily juxtaposed text. This perspective could hardly be defended in regards to conversation, because its tight temporal coupling and cohesion, making conversation a unique phenomena.

Coherence

Coherence refers to the logical connections and consistent relationships that make a text easy to follow and possible to understand[4][5]. It is achieved when the ideas, sentences, and paragraphs within a text are linked together in a meaningful way, allowing the reader to grasp the author's message without confusion. Coherence often depends on the use of transitions, the logical flow of arguments, and the proper sequencing of information. It ensures that each part of the text contributes to the overall meaning, creating a unified whole[6]. Incoherent text can be difficult or impossible to understand, even if the individual sentences are grammatically correct[7]. It is a property of the whole text, but textual elements can be pointed out that contribute to or diminish the given texts coherence. Those elements are however not referred to as 'coherence elements'.

Coherence is a cognitive phenomenon[8] because it involves the mental processes of interpreting, organizing, and understanding information. When

[4] T. Givón. *Coherence.* John Benjamins Publishing Company, 2020, p. 83.
[5] Hrbáček, *Nárys textové syntaxe spisovné češtiny*, p. 9.
[6] Hrbáček, *Nárys textové syntaxe spisovné češtiny*, p. 28.
[7] Hrbáček, *Nárys textové syntaxe spisovné češtiny*, p. 30.
[8] R. M. Roberts and R. J. Kreuz. "Nonstandard discourse and its coherence". In: *Discourse Processes* 16.4 (1993), pp. 451–464.

reading a text, coherence arises not only from the structure and linguistic cues provided by the author but also from the reader's ability to make connections between ideas based on prior knowledge, expectations, and context. This cognitive interaction between the text and the reader's mind is what makes the content understandable.

In conversation, coherence becomes even more complex, as multiple participants are simultaneously contributing to and interpreting the flow of information. Each individual brings their own perspective and understanding to the interaction, which requires constant negotiation to maintain coherence. Misunderstandings, different backgrounds, and interruptions can disrupt the coherence of a conversation, making it a more dynamic and fragile process compared to written text.

- whether a written text is coherent depends mostly on the reader
- whether a conversation text is coherent depends on an ongoing negotiation

Coherence is a scalar property rather than a binary one. It is however tricky to measure. This paper seeks to explore one possible approach of declaring different levels of coherence disruptions and observing the acceptance rates in participants and corelation between them.

Cohesion

While coherence refers to the interpretative quality of a text, wherein the ideas form a logical and meaningful whole cohesion,focuses on the structural relations within a text, achieved through grammatical and lexical links. It should be seen as an umbrella term covering specific relations within the structure of the text, where cohesive elements can be directly pointed out. While coherent text does not necesarily need to be cohesive, cohesive elements often support it. A coherent text tends to be at least somewhat cohesive.

Halliday and Hasan[9] developed a detailed framework of cohesion, which includes endophoric references, relating parts of the text to each other, and exophoric references, which point outside the text[10]. Endophoric cohesion

[9] M. A. K. Halliday and R. Hasan. *Cohesion in English.* Longman, 1976.

[10] Halliday and Hasan, *Cohesion in English*, p. 31.

covers aspects like anaphoric references and cataphoric references[11][12]. Exophoric references, however, rely on shared context beyond the text itself, requiring readers to use prior knowledge. Their framework highlights how elements of cohesion contribute to textual unity and flow, even if coherence based on meaning is not fully achieved. Following concepts can be considered cohesive elements.

Cataphore and Exophore

In Halliday and Hasan's framework, cohesion in language is achieved through various devices that connect different parts of a text, forming a unified whole. They classify cohesive ties as references, substitutive forms, ellipsis, and connectors, with anaphoric references being one of the primary ways texts achieve cohesion[13]. When a text element cannot be mapped to a preceding referent, Halliday and Hasan suggest that cohesion is maintained through shared situational understanding, making the reference exophoric. Cataphoric references, though less common, involve elements that look forward in the text, showing intentionality by the author but contributing to cohesion primarily through the eventual resolution of the forward-pointing referent.

In conversation if a seemingly anaphoric text element is not succesfully mapped to a preceding textual coreferent the reference can still be understood, because shared context. Such element reaches out of the text with its reference, making it an exophoric one Cataphore is a related phenomena – a reference which points forward in the text. Such occurence is relatively rare in written text and even more so in conversation. In fact it is somehow pointless to account for cataphore in a multi-producer text. A cataphore denotes an authors intention to reveal the nature of a referent explicitly after first mentioning them. In conversation, where multiple contributors cocreate given text, and mutual understanding and agreement is the measure of how coherent the produced text is, later realisation of a vague reference does not

[11] E. Hajičová, J. Havelka, and P. Sgall. "Discourse Semantics and the Salience of Referents". In: *Journal of Slavic Linguistics* 11.1 (2003), pp. 127–140.

[12] S. Loaiciga, S. Dobnik, and D. Schlangen. "Anaphoric Phenomena in Situated dialog: A First Round of Annotations". In: *Proceedings of the Fifth Workshop on Computational Models of Reference, Anaphora and Coreference.* Association for Computational Linguistics, Oct. 2022, pp. 31–37. URL: `https://aclanthology.org/2022.crac-1.4/`.

[13] Halliday and Hasan, *Cohesion in English*, p. 68.

contribute to how coherent it is. Regardless, in case of a cataphore, only the referent is a cohesive element, not the cataphore, as it ties back to the previous text, creating bonds across large textual units.

Anaphore, Endophore and Coreference

A common cohesive text element is an anaphore[14]. It is a reference inside the text pointing back to a previously mentioned entity. Often it is realised via personal pronouns. Though there are other ways for anaphore to realise. In Czech, anaphoric references often rely on grammatical gender and number, making participial endings essential for identifying the referent. For instance, when a gramatically masculine entity is mentioned, later references might use a participle in the masculine form, such as šel ("he went"), connecting back to it without repeating the noun or using a demonstrative. Demonstratives, such as ten ("that") or tento ("this one"), also frequently serve anaphoric functions, guiding the reader to a previously mentioned subject. Temporal and locative adverbs, such as tam ("there") and tehdy ("then"), also contribute cohesion by indirectly referencing time and place details introduced earlier in the text. These anaphoric elements strengthen textual coherence by reducing redundancy and maintaining flow. The reader identifies coreferential links through these markers, following the cohesive threads without needing explicit repetitions.

An anaphoric element is by definition also endophoric. It points inside the text it appears in. By definition an anaphoric element has a referent, which occures earlier in the text. These two elements are then coreferent. As such they also share an identical exophoric reference – they point outside of the text.

In conversation, many aspects of which are subject to negotiation, also specific coreference relations can be questioned[15]. The reference realised by one communication participant may be unclear to the other resulting in a repair request coming from another particiapant. In conversation analysis, Sacks's concept of repair traditionally addresses misunderstandings related

[14] A. Nedoluzhko. "Rozšířená textová koreference a asociační anafora (koncepce anotace českých dat v pražském závislostním korpusu)". PhD thesis. Prague: Univerzita Karlova, Filozofická fakulta, Ústav českého jazyka a teorie komunikace, 2010.

[15] S. Loaiciga, Simon D., and D. Schlangen. "Reference and coreference in situated dialogue". In: *Proceedings of the Second Workshop on Advances in Language and Vision Research.* 2021, pp. 39–44.

to intentions and actions, loosely drawing on frameworks like Austin's and Searle's speech act theories. From this perspective, repairs often target interpretative gaps about what a speaker intends to do with their utterance. However, viewed from a broader, more abstract level, what is called repair triggers can extend beyond intentions alone, encompassing issues on the textual level as well. For instance, an nonassignable anaphora — a reference that lacks a clear antecedent — may lead to a repair request, thereby showing how textual ambiguities prompt interactional responses. This approach expands the causes of repair in conversation, integrating elements of reference and interactional misalignment, where a structural aspect of the language itself can become a repairable issue in the communicative exchange.

Topic

Topic is what a text is about. That makes topic very complicated to define. Among others, some issues with topic and annotating it in text are:

- A text can and typically does cover multiple topics
- Different framing will produce different topic annotations of text
- The span of a topic section can be impossible to delimit within text.
- Topic annotation is by its nature always more text, so even it can be annotated for topic. making topic annotations recursive. One cannot therefore achieve a definite topic description of a text.

Despite all these complications, topic cannot be ignored in conversation research as it is deeply intertwined with the aforementioned concepts. Topic progressions across text are realised via anaphore and association and tightly interact with coherence. An appropriate amount of time has to be spent on a given topic unit, enough information has to be said about a given topic in order for it to be possible to move on or add another one in the conversation. Closure has to be provided in order for a topic to be done. Transitioning from one topic to another has a potential to disrupt coherence, if the association between the topics is too distant. A divergence in topic has to be justified.

Association

Association is a textual realization of an isotopic relation[16]. By their exophoric properties, referents exist in a semantic web of relationships. Similarly to coherence, associative relationships are a cognitive phenomena. They come to exist when they are percieved. While association is a cohesive element it is difficult to formalize the way it can and has been done with anaphoric text relations. It is however a major factor in a coherence of text as in some cases a text can only rely on association in its coherence.

Entity

An entity is an exophoric referent, descriptions of objects, people, events etc[17]. Words or text elements which can be reffered to by an anaphore will be called entities. Since a phrase containing an anaphore typically adds more information about the referent the new information must be semantically compatible, in other words association has to be possible between the referent and the added information. Entity also has to do with topic. In text topic can be represented by a single or multiple entities. Coreferent words will be regarded as a single entity. It can serve to partially map a topic distance in the texts chronology.

1.2 Interactional dimension of conversation

In terms of introduced background, conversation is a text which is produced by multiple producers This complicates things:

- Conversation is an interactive process, distinct from static text, which is created collaboratively.
- Conversational content is continuously negotiated by participants, who continuously adapt one another.

Due to its temporal and cooperative nature, conversation allows for:

[16] T. Koblížek. *Interpretační sémantika. Úvod do textové teorie Francoise Rastiera.* Univerzita Karlova v Praze, Filozofická fakulta, 2015.

[17] D. Ayuso. "Discourse entities in Janus". In: *27th Annual Meeting of the Association for Computational Linguistics.* 1989, pp. 243–250.

- Overlaps in speech,
- Swift corrections of minor errors,
- Multiple layers of perspective, including:
  - Each participant's personal viewpoint,
  - Each participant's perception of others' viewpoints,
  - Each participant's understanding of the shared conversation as it's being co-created.

each of these perspectives can desynchronize resulting in misunderstandings. Humans however are excellent at correcting misunderstandings this is because under regular circumstances, people cooperate.

The Cooperation Principle, introduced by philosopher H.P. Grice, suggests that participants in a conversation typically work together to achieve effective communication. Grice proposed that, to ensure this cooperation, speakers follow four conversational maxims. In practice, people may not always follow these maxims but they do so in ways that still rely on shared expectations of cooperation. Even when misunderstandings arise, humans naturally engage in conversational repair, using their social intuition and mutual cooperativity to clarify intetion and realign perspectives.

Contemporary conversation research can be understood to draw from conversation analysis. Modern conversation research traces its roots to conversation analysis, a field pioneered by sociologists Harvey Sacks and Emanuel Schegloff in the 1960s. They sought to understand the structure and social rules of everyday interactions, focusing on the patterns and norms that govern turn-taking and response. Thanks to recordings of phonecalls, transcripts could be qualitatively analyzed. This research has lead to coining new terminology.

Adjacency pair

Adjacency pairs describe sequences of two related utterances by different speakers[18]. These pairs are characterized by their predictable and reciprocal nature, where the first part sets up the expectation for a specific type of

[18] H. Sacks. *Lectures on Conversation*. Ed. by G Jefferson. Oxford: Basil Blackwell, 1992, p. 188.

response. Common examples include greetings('Hi' $\rightarrow$ 'Hello'), questions and answers ('What time is it?' $\rightarrow$ '3 PM'), or offers and acceptances/declines ('Would you like some coffee?' $\rightarrow$ 'Yes, please' or 'No, thank you').

Sequence structure

Sequence structure refers to the organization of conversational turns into coherent patterns or sequences. It describes how interactions are shaped by predictable structures, such as adjacency pairs. These sequences provide order and meaning to conversations, guiding participants in understanding when and how to respond. Schegloff[19] emphasized that sequence structure is central to the social organization of talk, as it allows participants to manage and negotiate interaction effectively.

Topic shading

Topic shading, as discussed by Sacks[20], refers to the subtle way in which a conversation naturally shifts from one topic to another while maintaining coherence. Instead of abruptly changing the subject, speakers introduce a related idea or concept, gradually steering the discussion in a new direction. This process allows for smooth transitions in dialogue, helping participants maintain engagement and avoid confusion.

Dis/preferred answers

Preferred answers, according to Sacks[21], are responses in conversations that align with social norms and expectations, making interactions smoother and more cooperative. In conversation analysis, preferred answers typically follow the format or intent of the preceding question or statement . They contrast with "dispreferred" answers, which might include refusals or disagreements and often require additional explanation or mitigation to maintain social harmony.

---

[19] E. A. Schegloff. "On the organization of sequences as a source of 'coherence' in talk-in-interaction". In: *Conversational Organization and its Development.* Ed. by B Dorval. Ablex, 1990, pp. 51–77.

[20] E. A. Schegloff and H. Sacks. "Opening up Closings". In: *Semiotica* 8.4 (1973), pp. 289–327.

[21] Sacks, *Lectures on Conversation*, p. 410.

Conversational repair

Conversational repairs refer to how participants address and resolve problems in understanding, hearing, or speaking during interactions[22]. These issues, can occur at any point in a conversation. Repairs are classified into self-repair, where the speaker corrects their own error, and other-repair, where a different participant addresses the issue. They can further be classified into self-initiated repair and other-initiated repair.

As a descendant of conversation analysis interaction linguistics has emerged, building on its insights to examine language use in social contexts. It broadens the focus to study not only verbal exchanges but also multimodal cues like gestures, gaze, and intonation, analyzing how these elements contribute to meaning. Interaction linguistics aims to understand the dynamic aspects of conversations, such as how topics shift and how sequences of speech acts unfold, reflecting the fluid nature of human communication.

1.3 Disruptions in conversation coherence

While the question of what makes for a coherent text is too broad, the answer to what makes for a coherent conversation can be somewhat easier to answer. Because conversation participants negotiate understanding, it is up to them, when a conversation is and is not coherent to describe what a coherent conversation is, it is worth pursuing the moments, when the conversation stops flowing with ease. Such moments can be called coherence disruptions. A coherence disruption is a complex phenomena as

- it penetrates through some or all of mentioned perspectives on an ongoing conversation
- it can't be evaluated in a binary fashion

There are different degrees to which a conversation coherence can be disrupted:

- if a participant suddenly starts speaking in an a way that can hardly be considered interaction due to its irrelevance or

[22] E. A. Schegloff, G. Jefferson, and H. Sacks. "The Preference for Self-Correction in the Organization of Repair in Conversation". In: *Language* 53.2 (1977), pp. 361–382.

- if the utterance simply is not grammatical or understandable, while the conversation has been compromised and becomes incoherent, it has more to do with incoherent written text, because the incoherence is encapsulated on the level of a single utterance

Roberts[23] Discusses various types of incoherent text. He exemplifies so called giberish as incoherent text that is absent of structural relations. On the other hand he discusses experimental theater or literature as a type of text which is assumed to be coherent in the sense that there is an intention behind it but contains little to no structural relations. Lastly he mentions a so called "schizophrenic discourse" as a speech that is not assumed to be coherent even if it has structural relations to it. In any case Roberts definitively states that coherence is assumed and is therefore a receptive phenomena. The incoherence Roberts discusses is considerably different from when the source of incoherence stems from the structure of the conversational text or relationship between different utterances - this is when another participant assesses, they are simply speaking leading a different conversation perhaps with a differing intention or that they are conversing under differing set of circumstances which manifests formally in the linguistic fabric of the conversation - its text. All that despite everyone included being cooperative.

1.3.1 Sources of incoherence in conversation

Schegloff shows how incoherence arises when people interpret sequence structure differently, namely in terms of which turn is seen as an answer to which previously occuring turn. In his example, the participants misread each other's intentions, leading to confusion about how their turns fit together. They each project different expectations for how the conversation should unfold, which causes misaligned sequence structure interpretations. When this happens, they turn to brief metacommunication — comments about the conversation itself to try to clarify and re-align their understanding. Schegloff illustrates how these efforts to "repair" the misalignment are central to managing and resolving incoherent moments in conversation.

Coherence disruptions are also discussed in linguistic literature. Hrbáček's approach to coherence and cohesion in text distinguishes the two concepts, noting how they often interact but can also be independent. He highlights that while cohesion involves grammatical or lexical links that make sentences

[23] Roberts and Kreuz, "Nonstandard discourse and its coherence".

flow together, coherence relies on the logical and meaningful progression of ideas. This means that a text could be cohesive - using connectives, repetitions, and consistent lexical choices - yet lack coherence if the sequence of ideas doesn't make logical sense or follow a clear progression. Conversely, a text may be coherent in its narrative flow without relying heavily on cohesive devices. In Czech linguistics, the distinction between téma (theme) and réma (rheme), as used by Daneš, underlines the role of topic progression. Hrbáček illustrates this by discussing examples where a story progresses logically from one point to the next while being incoherent despite being clear about its topic structure due to never coming back to a previously mentioned topic.

Two kinds of phenomena are at hand when it comes to ways in which conversation coherence can be disrupted - topic shifts and nonassignable anaphore. While not unique to conversation both take on specific forms in it worth looking at.

Topic shifts

When conversations shift abruptly from one topic to another, it can create confusion for the conversation partner. They might find themselves trying to reconnect to the previous discussion or wondering how the new subject relates. This can lead to misunderstandings as the transition can feel jarring.

One interesting question is, how do we determine when a topic has run its course? What common traits do conversations share when a subject is truly exhausted? Perhaps observing transcripts could reveal repeating patterns in topic progression or sequence structure.

Moreover, what makes for a smooth transition between topics? Is it related to the cues participants give each other, or perhaps the context of the discussion? How do we navigate the flow of conversation and what indicates a natural shift versus a disruptive one?

Nonassignable anaphore

nonassignable anaphore is closely tied to topic progression. Currently established topic or topics help assigning anaphore and determining between an anaphore and an exophore. Even if an anaphoric device is not assignable, and the reference is presumably an exophoric one, The reason for employing this reference must be relevant to an established topic. In conversation meaning of demonstratives is to be negotiated. If an anaphores assignability causes confusion, chances are it is caused by one of the following

- there are no relevant assignment candidates

  this situation can be understood as a vague or unjustified exophore

- there are multiple equally relevant candidates

- candidate has occured in the conversation text too long ago

  can be understood as an abrupt return to previously established topic

1.3.2 What do people do about coherence disruptions?

In conversation, coherence disruptions often prompt participants to employ strategies to maintain understanding and flow. Schegloff suggests that people manage these disruptions through interactive repair or inference. Interactive repair often involves explicitly addressing misunderstandings or clarifying intentions, often by rephrasing or asking questions. Interactive repair refers to immediate, collaborative corrections within dialogue, where one speaker might correct the other or themselves to enhance clarity. Inference and pragmatic reasoning, the most seamless methods, allow participants to fill gaps based on context and social cues, helping conversations continue smoothly without explicit repair.

Dingemanse and Enfield[24] echoes this from a cognitive perspective, highlighting how inference and pragmatic reasoning are particularly effective. Participants rely on shared understanding and contextual knowledge to interpret ambiguous statements. Together, they use both explicit (metacommunication and repair) and implicit (inference and reasoning) methods work to restore coherence.

It needs to be noted however that both interactive repair and reasoning are deployed in a number of other contexts other than conversation coherence disruption. Inference takes place constantly[25]. Each of those moments could be hardly considered a coherence disruption. There is however always potential for it, particularily via unclear or nonassignable anaphore or abrupt unjustified topic shifts. Repair and metacommunication also takes place in a mutually informed and synchonized interaction. It is for example deployed

[24] M. Dingemanse and N. J. Enfield. "Interactive repair and the foundations of language". In: *Trends in Cognitive Sciences* 28.1 (2024), pp. 30–42.

[25] H. Garfinkel. "Studies in ethnomethodology". In: Routledge, 2023, pp. 209–261.

when it is revealed that the interaction participants intentions or opinions differ.

These uses of interaction management are however hardly possible to analyse on a textual level since they do not cooccur with coherence disruptions. What can be observed are – as mentioned above – troublesome anaforic references and topic progressions.

Chapter 2

Experimental framework

2.1 What are chatbots?

A chatbot is a dialog system powered application simulating conversation with a user. An attempt to make a machine conversate with a human user requires capturing the essence of human speech. What the essence speech is and what capturing it means are not defined and some definitions of what essence of speech is require specific definitions of what capturing it means. (Semantics and DialogueSchlangen, David2015) .. In case of the history of chatbots, the essence of speech is partially achieved via mimicking it. The intention was to have a user interact with a chatbot that would communicate so well that the user would be convinced this is a another human they are talking to. Whether just that has been achieved would be measured by a so called turing test proposed by Alan Turing in 1950[1]

Initial attempts at making a computer converse were rule-based[2] What that means is the content of the chatbot utterances would be predetermined and there would be a decision tree that would decide what to say next. In the early days as well as often times in modern day systems string matching would be used to analyse user input.

ELIZA[3] is regarded as a milestone what it was, it pretended to be a

[1] A. M. Turing. "Computing Machinery and Intelligence". In: *Mind* 59.236 (1950), pp. 433–460.

[2] Sacks, *Lectures on Conversation*, p. 43.

[3] J. Weizenbaum. "ELIZA a Computer Program for the Study of Natural Language Communication Between Man and Machine". In: *Commun. ACM* 9.1 (Jan. 1966), pp. 36–45.

therapist hiding behind general phrases doctor authority

as long as interaction frame is strictly defined and the robot has some level of authority in it the rule-based approach can work granted, it requires a lot of manual design and constant maintenance but it is used nowadays

Machine learning moved things forward in many ways Fuzzy matchin allowed for close matches without the neccesity to predefine the exact sequence of characters. IBM Watson uses classifiers build on examples[4] structure remains rule based reasoning about these rules is decided to a large degree by classifiers

The recent breakthrough pushed another thing in the mainstream it is now possible to generate near natural speech this gives the possibility to just let the conversation be taken over by one answer generator this way we lose tight control over what it does though For some use cases like open domain conversation or accessing knowledge base that is not an issue.

And so currently chabot interface text generators are very prevalent. In 2024 this technology is now closer to beating the Turing test than any other model or approach before it[5] by having 54% of participants thinking they are talking to a human. While Eliza convinced 22% of participants, actual humans only convince 67% of participants.

Turn taking in chatbot interactions

Even if the Turing test is passed, really fitting simulation of conversation can only be achieved if the low-level conversation mechanisms are simulated, like turn taking[6].

As established in previous chapter, turn taking is a crucial aspect of conversation. The way participants distribute who is to talk explains exhaustively the difference between a structure of the text of conversation and a single-producer text. The mechanism of turn taking differs between actual human conversation and an interaction between a chatbot and a user.

[4] D. Ferrucci et al. "Building Watson: An Overview of the DeepQA Project". In: *AI Magazine* 31.3 (July 2010), pp. 59–79. URL: `https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/2303`.

[5] C. R. Jones and B. K. Bergen. *People cannot distinguish GPT-4 from a human in a Turing test.* 2024. arXiv: `2405.08007 [cs.HC]`. URL: `https://arxiv.org/abs/2405.08007`.

[6] A. Raux and M. Eskenazi. "Optimizing the turn-taking behavior of task-oriented spoken dialog systems". In: *ACM Trans. Speech Lang. Process.* 9.1 (May 2012).

Interaction between chatbot and user typically take place in a strict fashion where both participants, human and virtual, have unlimited time to come up with the next answer. While the chatbot should be optimized to answer as fast as possible, the user has as much time as they need until fallback.

"Research in sociolinguistics, psycholinguistics, and conversational analysis has revealed that turn-taking is a mixed-initiative, locally coordinated process, in which a variety of verbal and nonverbal cues such as eye gaze, body pose, head movements, hand gestures, intonation, hesitations, and filled pauses play a very important role. We continuously produce and monitor each other for these signals and can coordinate seamlessly at the scale of hundreds of milliseconds across these different channels with multiple actors."[7]

People are capable of producing and picking up clues that indicate opportunities for turn taking flawlessly. There is a some way ahead for robots in this regard whether it is figuring out the correct time to start speech[8][9] or actually creating a system that will be able to produce such behavior[10][11]. This research field has the potential to push conversation technology closer to true conversation simulation.

2.2 Convform

An exploration has been carried out using a custom tool called Convform[12].

At its core Convform is a computer program which accepts a configuration, user input and context and determines next chatbot answer. Other than that it offers a collection of utilities to help design and run chatbots.

---

[7] S. Andrist et al. "Turn-Taking and Coordination in Human-Machine Interaction". In: *Ai Magazine* 37 (Dec. 2016), pp. 5–6.

[8] G. Skantze. "Turn-taking in Conversational Systems and Human-Robot Interaction: A Review". In: *Computer Speech & Language* 67 (May 2021), pp. 101–178.

[9] A. Gravano and J. Hirschberg. "Turn-taking cues in task-oriented dialogue". In: *Computer Speech & Language* 25.3 (2011), pp. 601–634.

[10] G. Jonsdottir and K. Thórisson. "A Distributed Architecture for Real-Time Dialogue and On-Task Learning of Efficient Co-Operative Turn-Taking". In: Oct. 2013, pp. 293–323.

[11] F. Gervits et al. "It's About Time: Turn-Entry Timing for Situated Human-Robot Dialogue". In: *Proceedings of the Special Interest Group on Discourse and Dialogue*. 2020. URL: `https://hrilab.tufts.edu/publications/Gervits2020Sigdial.pdf`.

[12] A. Maršík. *Convform*. `https://github.com/almarsk/convform`. 2024.

Participant facing chat interface

In order to handle the inputs, convform provides a chatting environment for the participants to interact with a chatbot. The convform environment differs from a usual chat log because it does not display the entire history the conversation. In attempt to simulate spoken conversation it only displays the last chatbot response. This way the participant has to rely on their memory in taking part in the conversation. Other than that the participant may enter their next response and send it. They are also instructed to end to conversation by a red button if the chatbot behaviour is "unnatural" (nepřirozené) After the conversation is over whether it has been ended by the user or the chatbot, there is a questionare which asks the participants to rate how "natural" the conversation was and mark and comment on utterances in the now fully displayed conversation.

Conversation design tool

Lets admin user create chatbots and define their behavior the behavior can be defined by string matching rules or prompts it is capable of working as a statemachine or a single state it provides a level of control over references within the design

Testing and debugging of various conversation contexts

while designing chatbots it is necessary to be able to simulate various situations to fine tune various possible scenarios that might occur in the conversation. To achieve this, there must be a way to encode required context to convform. The convform chatbots use a conversation status (CStatus) object to represent their current understanding of the conversation. It contains information about the history of the conversation which in conjunction with the configuration file and user input helps determine the next response. The configuration file is static CStatus changes automatically User input comes from the user. This conversation status can simulate any possible conversation context from the chatbots perspective For testing and debugging specific contexts, convform allows admin user to tweak the conversation status

Accesing the conversation data

Lastly convform naturally includes a convenient way to read user interactions and browse associated conversation status objects

2.3 Conversation design in theory

Designing the behavior of a dialog system is referred to as conversation design[13][14]. It is not the course of any one conversation that is being designed here but rather as many possible ways any conversation could go for a given use case. Conversation design as a profession is deeply connected with the rule-based approach that has been used in ELIZA. Maintaining all the possible utterances and rules under which they would be uttered in commercial dialog systems has proven to be a responsibility large enough to generate jobs.

A conversation designer operates between the business logic and use case of the dialog system the clients, customers or users interacting with the system and the developers maintaining the system.

2.3.1 Rule-based approach

In order to be able to design a rule-based dialog system, one needs to be able to encode the following:

- The possible utterances, that the dialog system can produce - "the mouth"
- Rules under which the next utterance will be chosen - "the ear"

If the conversation is supposed to be a state machnie e.g. it needs be able to use different sets of rules under different contexts in the conversation. This way a dialog system can be context aware to a degree. A conversation design of this sort can be displayed as a diagram. Then a way to maintain context of conversation is also necessary. This context needs to encode rules

[13] Š. Kološová. "Konverzační design: principy designu hlasového robota pro přirozenou komunikaci s lidmi". Diplomová práce. Praha: Univerzita Karlova, Filozofická fakulta, Ústav informačních studií - studia nových médií, 2022.

[14] McTear, *Conversational AI: Dialogue Systems, Conversational Agents, and Chatbots.*

to choose an immediate ruleset which helps determine the next utterance. This principle is a simplification of how people decide what they will say next in conversation.

Pros and cons

This approach to designing a dialog system has been the standard for decades. It offers a granular control over how a conversation should go. In case of the state machine variant it allows to guide the user through a relatively complex process. It however suffers from how unpredictable the user can be. It is up to the conversation designer to cover all the possible ways of answering which not only is hardly possible but also poses a neccesity to parcel the spectrum of possible answers which can generate conflict when a user input semantically spans across multiple determined categories. This issue is even stronger while using the string matching approach, because there the string literal can decide about the following dialog system answer as if meanings and their speech representations were a one-to-one map, which they are not. Even if a certain meaning is included in a ruleset, the system might not grasp the meaning and react in an incoherent way. With the state machine the distribution of various rules across various rulesets requires big effort. Extending the capabilites of a rule-based dialog system hardly scale and tend to have regressions. In case of dialog systems relying on user input by speech transcription the text input processed by the system is not guaranteed to represent what the user actually said. In conclusion rule-based approach to conversation design provides control over the dialog system behavior but tends to be inflexible.

2.3.2 Statistically driven approach

Some of the issues tied to rule-based systems are resolved using another approach. Especially in recent years the breakthroughs in the field of speech generation have been significant () allowing for letting the dialog system play a bigger role in what is being said next. In its simplest form, it is possible to just let the the answer be generated "end-to-end". The user input is sent to a model which generates an answer. It has been convincingly shown that this technology has the capability of reacting in a flexible way to much of what is being thrown at it (). While not perfect () this technology is capable of staying on topic (), mirroring () and other things that make for a coherent conversation.

Large language models

The main component that is responsible for this way of simulating conversation at this level of flexibility are so called large language models (). These models, powered by advanced neural networks, have revolutionized the field of natural language processing. Among the most influential architectures are transformers, which enable these models to handle vast amounts of text data and capture complex patterns of meaning, context, and grammar. They use their training data to generate the next most probable token.

and then it takes also its output and predicts the next probable output

llms for other things than conversation

The Generative Pre-trained Transformer family of models, exemplifies the capabilities of LLMs.

Researched in - exponential improvements over short period of time.

including ChatGPT, presented in ... ()

These systems are trained on diverse datasets containing billions of words, allowing them to generate coherent and contextually relevant responses across various topics. This flexibility has made them increasingly mainstream, being integrated into tools for writing, education, customer service, and more.

Nowadays there are multiple publicly available LLM services.

Unlike traditional rule-based systems, LLMs rely on deep learning techniques to process and predict language, enabling them to understand nuanced queries and provide human-like responses. This adaptability has set new standards for conversational AI, making it a valuable resource in numerous industries.

Prompt engineering techniques

The rapid advancement of LLM technology has outpaced research into optimal interaction strategies. Understanding how to engage effectively with these systems has been a developing area (), which illustrates both their power and their novel nature. The foundational idea is: an LLM might perform nearly any task if prompted correctly.

The quality of the task differs significantly task to taks in reality, but the anything is possible approach has proven to break new grounds when leveraging LLMs' capabilities.

Over time, researchers and practitioners have developed techniques for crafting effective prompts to optimize outputs. The simplest approach is known as "zero-shot" prompting, where a user poses a direct question or request without additional context. However, zero-shot prompting may not always yield the desired depth or accuracy. It is common for the model to "misunderstand" the assignment and generate tokens so that it will "confuse" itself and lead to generated answer in a completely irrelevat direction.

More sophisticated strategies include "few-shot" prompting, where examples are provided to guide the model's response style or focus. This way there is a reference for the structure of the answer and there is a protection to the answer leading somewhere it is not meant to. Since analogy is a task LLMs are doing really well in () framing the task as an analogy can help improve the output significantly.

Another very prevalent way that has proven to improve the performance of LLMs is a so called chain-of-thought (CoT) prompting. It encourages the model to articulate its reasoning step-by-step, enhancing logical accuracy. There are many ways to achieve this, but the primary one is a few-shot approach where a description of the logic is explicitly described. The model is then prompted to produce a similar chain of thought (hence the name) and end the answer with the sought after information.

This principle can be further improved by chaining several LLM calls and having one evaluate the previous one. Such strategy has been used for the GPT o1 which has proven to surpass other models in available metrics ().

Pros and cons

Using large language models as core component of dialog systems brings resolution to many issues rule-based systems introduce.

An LLM powered dialog system is flexible in understanding the user input. The user input is processed in a way much more sophisticated than the shallow string-matching approach. While the classifier approach is a lot more capable to understand, it is still forced to choose a predefined answer, whereas an LLM has the capability to taylor an answer for every input. It can do this in a way that would be very hard to come up with especially in advance with the help of a conversation designer, leveraging the fact that LLM is primarily a text generator and only functions as a component in a dialog system. It can be relatively well controlled as it can accept complex instructions as to how to behave and these instructions can be tuned in runtime.

Systems of this sort however introduce their own set of problems. The biggest issue are so called hallucinations (). Factuality is a challenge for LLMs overall. Being programs that are to output text no matter what, there have been instances of asking them questions, that do not have a correct answer (). It has been shown, that LLMs have an issue knowing that they do not know something. Recognizing that is the case requires an extra level of reasoning that is an object of research as of recently ()[dark matter of ai]

Since these models are their own agents, us humans also need them to be ethically aligned with us (). That can prove challenging since a lot of ethical problems exist, that do not have a simple answer. Alignment however means that an LLM must always side with humans. This is an ongoing research field which has been making some troublesome observations recently with the most intelligent models available ().

Even if all the programming and training is done to the most benefit of humans, information technology is succeptible to be broken by malicious action. In case of LLMs we talk of so called jail breaks (). LLMs being trained on vast amount of data, they hold knowledge that can be illegal or unethical to spread like steps to create explosives for example. The typical examples of jail breaks are ways to manipulate the LLM to give out this information which under regular circumstance it would not give.

With all this in mind, how much of a responsibility do we want to allow LLMS to have? Considering the direction our civilization is going in LLM powered technology is expected to be making decisions that will affect people on a daily basis. Hopefully that will not happen before the challenges of alignment and hallucinations are solved.

As far as low-stakes open-domain conversation simulation goes, LLM powered chatbots are relatively safe though there are cases of dangerous or tragic situations especially for vulnerable individuals ().

This is why for dialog systems that are supposed to achieve anything else on top of the conversation itself, if they are meant to be powered by LLMs, a regulating structure needs to be placed on top of the LLM.

2.4 Conversation design in practice

Conversation design in Convform attempts to combine elements of rule-based design with text generation. It allows creating a purely rule-based chatbots which analyse the user input based on string matching and say exactly what they are prescribed to. On the other hand it also allows to make the chatbot understand the user input by adding it in a prompt and answer using a generated response. Both these approaches can be combined in various ways. Other than that, convform also allows to predetermine the chatbot personality for the entire conversation. The building blocks of a convform chatbot are states and intents which are analogical to the previously mentioned "mouth" and "ear" of the chatbot.

State

A state is an object which carries several pieces of information bundled together. At its core it contains the utterance of the chatbot whether it is a hardcoded one or a prompt component which is to be called. A state however also contains information about which intens to listen to in the next user input, which states to add automatically to the next response and other navigation instructions like this one. Each convform status associated with a response can contain multiple states. This is to make convform generate more complex answers which can react in a flexible way. However it also comes with a challenge to order these responses correctly and make sure that they are not contradicting each other content-wise. Ultimately a master prompt with multiple components regulating itself via analysis prompts might be a smarter direction to go.

Intent

An intent is an object representing a category of a user response. It contains the information to determine whether user input fits in given category and the state or states to respond with next. Just like in state, the information about whether the user input corresponds with the intent can be encoded via string-matching patterns or a prompt. As mentioned intent is a problematic concept, because it forces an outside logic and categorization on user input, which might not be able to fit well in the framework declared by the current intent set. It is however also the necessary evil since it is the only way for a conversation designer to peek into what is going on in the conversation and to direct the dialog system in the correct way.

This way a convform chatbot can be created, that will be instructed to lead from one state to another make decisions based on intents while being able to use any combination of hardcoded responses and intent patterns and descriptions of responses or user inputs used in intents. Detailed description of how convform works can be found in the wiki of its github repository[15]

Coherence

With support of LLM powered responses convform can be used to simulate an open-domain conversation with a user and simultaneously using a combination of intents and prompting a convform chatbot can be created that will act incoherently under a predefined set of conditions allowing to create experimental stimuli. First however, regular conversation needs to be achieved using convform.

Conversation style

To simulate conversation, it is first necessary to simulate a persona. The persona can then have a simulated motive to conversate which can interest the user enough to engage in interaction with the dialog system. For rule-based systems, persona can be defined ahead of time and it can manifest itself via the specific writing of the hardcoded responses that the system is able to give. With generated responses, the persona of the chatbot has to be included in the prompt. The personality of LLMs and conversation technology more broadly is being discussed[16][17]. The general characteristics of a machine talking to a human are typically friendliness, helpfulness and submissivity. For conversation research with convform, the goal is to achieve just that. The chatbot persona needs to be friendly, polite and curious. It needs to be able to keep the conversation going but not change topic too often. It needs to be able to add a little bit of its own perspective.

The conversation style e.g. the amount of participation and initiativity in conversation is something people adapt in to their conversation counterpart. Since developing a system that would immitate this behavior requires additional effort and expands scope beyond the coherence research this paper

[15] **https://github.com/almarsk/convform/wiki**.

[16] A. Deshpande et al. *Toxicity in ChatGPT: Analyzing Persona-assigned Language Models*. 2023. arXiv: `2304.05335 [cs.CL]`. URL: `https://arxiv.org/abs/2304.05335`.

[17] G. Laban et al. "Robo-Identity: Exploring Artificial Identity and Emotion via Speech Interactions". In: *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 2022, pp. 1265–1268.

focuses on, this approach to conversation design has not been taken here. Instead two versions of conversation style have been developed and distributed evenly between participants.

The initial conversation style used in the experiments represents a curious and friendly chatbot who is instructed via prompt to ask lots of follow up questions. This tends to result in a conversation that moves forward in its topical structure in way deemed incoherent by Hrbáček[18]. It depends on the participants impression whether it would be perceived as curious and initiatve or shallow and dismissive.

A second version of conversation style has been introduced to get some insights on participants acceptance and the course of conversation itself. This one would interleave topical questions with remarks on the topic The intention behind this would be to slow down conversation tempo and give the participant the opportunity to bring their own initiative.

Prompting

special prompts used to achieve things beyond regular conversation all in czech including prompts czech versions of prompts will be in annex

Entity recognition

In order to track entities that could be referred to a few-shot prompt was deployed that would help keep track of which entities have been mentioned. Since GPT4o, the model used in the experiment, tended to consider too many things an entity, most examples are negative and do not capture an entity. It also contains some repetition as a result of fine-tuning the best wording.

.. (New or Old? Exploring How Pre-Trained Language Models Represent Discourse Entities)

Loose english translation of the prompt goes as follows:

What is an entity? a person, a thing; may not be animate;
It is always a noun and most nouns are an entity in the speech;
Look out! Pronouns such as "you", "he" and the like are never entities even if they represent persons.
Verbs are certainly by no means entities even if they represent a person.
nouns which are not entities in a speech are very generic or temporal;
a subject in a sentence to which reference may be made by a personal or reference pronoun;
the utterance usually has one entity, sometimes two, and rarely more. often the utterance lacks an entity altogether.
The participants in the conversation are never entities in any way.

example:
Alice likes speedball the best.
The entities in this sentence are ["Alice", "speedball"].

[18]Hrbáček, *Nárys textové syntaxe spisovné češtiny*, p. 30.

example:
We're meeting tomorrow at 2:00.
This discourse lacks an entity, so the output is [].
example:
What do you like?
This discourse lacks an entity, although there is a personal pronoun in it, the output is therefore [].
example:
How are you, Carl?
This speech lacks an entity, because the communicator is never an entity, so the output is [].
example:
I just hope it doesn't get too cold.
This speech lacks an entity, because the atmospheric phenomenon is never an entity, so the output is [].

Consider which words are entities in the next sentence:

$$userinput$$

Consider the meaning of each noun in a sentence and consider whether it could represent a specific entity,

and not just an abstract concept.

Consider all possible objects that could be mentioned in the speech

and consider whether their names could be considered entities.

Remember that verbs and pronouns are not entities, even if they represent persons.

Remember that the conversationalists are not entities.

anaphora in conversation .. (Annotating anaphoric phenomena in situated dialogue) but here we try to generate it instead

to be able to create conversation designs which contain various types of anaphore, convform first needs to be able to give a response that has an anaphoric reference to an entity from the previous conversation in it.

The GPT4o model used for this use case does not tend to generate sentences with anaphores in them. Instead it will rather mirror the entity phrase.

The anaphorization prompt therefore tasks the model to modify a generated response so that the mirrored entity is replaced with an anaphoric device. For this another few-shot prompt was used to modify a just generated response. Its loose english translation is as follows:

Find one main word in a sentence and swap it for a personal or reference pronoun.
Leave the other topic centers as they are.
Don't forget to omit words that may be part of the name phrase of the replaced word.
Don't forget that when a noun is replaced by a pronoun, you often need to change the verb sequence -
the verb will then often be at the end of the sentence; the order of the pronouns must be followed -
the replacement pronoun will come for the reversible and the personal pronoun. You also need to correctly recognize the more important word -

choose the one about which the question is.
To preserve the naturalness, you sometimes need to modify a sentence
to include a secondary sentence, especially if the replaced word is linked to a deverbative noun.

Example1:
sentence:
And what will your paper be about?
consideration:
the replaced word will be a seminar. The word "yours" belongs to the name phrase of the replaced word.
your answer:
And what will it be about?

Example2:
sentence:
how far from your house is your favorite park?
consideration:
the substituted word will be "park," is more important in a sentence than the word "house." the word "yours" belongs to the name phrase of the substituted word as well as the word "favorite."
your answer:
How far from your house is it?

Example3:
context:
I'll go to the cafe
sentence:
what is your drink in the cafe?
reasoning:
the substituted word will be "cafe" because it is in context. because of the natural word sequence, you will need to move the substitution pronoun.
your answer:
What is your drink there?

Example4:
sentence:
Do you have any tricks for ironing shirts quickly and efficiently?
reasoning:
the substituted word will be "shirt" and because the sentence is complex and you cannot easily move the verb, a secondary sentence will solve it.
your answer:
Do you have any tricks for ironing them quickly and efficiently?

Example5:
sentence:
What track do you most enjoy racing on?
your answer:
Which one do you most enjoy racing on?

This prompt has been tuned to catch as many tricky cases as possible. examples description ..

2.4.1 Stimuli

With these tools multiple chatbots were created that would generate conversation situations which serve the role of experimental stimuli. Participant reactions to these stimuli can be then compared. This way conversational experimental designs can be created. There are three types of stimuli created for the purpose of this paper. They are shallow anaphore, deep anaphore and nonassignable anaphore.

Shallow anaphore

A shallow anaphore is a kind of anaphore where the referent of the anaphoric device should be relatively easy to map as opposed to a deep anaphore. The referent will always occur in the preceding utterance of the participant. This type of stimuli is relatively simple to achieve in convform generating a response and using the anaphorization prompt on it afterwards.

Shallow anaphore is common in regular conversation () and should not pose a problem for participant to understand in a conversation with chatbot. It should therefore not have an impact on the user acceptance of the chatbot and should generally go unnoticed. It is regardless worth using as stimuli for a reference unproblematic case that still requires the same kind of processing as other more interesting stimuli.

Example:
Participant: I love coffee. Chatbot (not anaphorized): What kind of coffee do you like best? Chatbot (anaphorized): What kind of it do you like best?

Deep anaphore

A deep anaphore is a situation where the referent of the anaphoric device occurs several utterances ago. The depth is not measured by number of occurences but by number of new entities that occur since the referent which the anaphore refers to. Measuring depth of anaphore by the number of utterances does not capture the dynamic nature of topic progression in the text of the conversation. The number of utterances does not map on how many topics have been visited. While the number of new entities does not map exactly either, it is first of all a lot closer to the topic progression and second of all actually very close to what is being sought after here - how far in the conversation is an entity still acceptable or even available to speakers.

An entity is can represent a topic but can also be one of several entities to represent a topic or can cover several topics at once all depending on which way the covnersation goes.

As stated earlier, both topic and entity are difficult to define and their annotations tend to be recursive. A close-enough approach has been adopted in this paper. While runtime topic annotation by an LLM is not necessary for generating deep anaphore and has therefore not been attempted in this paper, entity tracking is made possible my entity recognition prompt.

This prompt runs in parallel with the next response generation and writes down its results in the conversation status. A chatbot that contains the deep anaphore stimuli chooses a participant mentioned entity relatively early on in the conversation (though not at the very beginning) and then tracks new mentioned entities. When there have been 4 new entities mentioned, the next response generation prompt will be modified. The modification lies in that the context of the conversation that has so far taken place will be cut so that the chatbot only has access to the conversation until the point of the mention of the entity. Given the response generation prompt the response will contain a question about the mentioned entity. Then the only thing that needs to be done is modifying the response via the anaphorization prompt.

The trick here the participant and the chatbot differ in their perspectives on what the conversation currently is. The chatbot refers to something that from the perspective of the participant has been mentioned a while ago.

This approach is relatively imprecise and relies on luck to a certain degree. Compared to the shallow anaphore however it is expected to be somewhat more problematic and perhaps cause the participant to request clarification.

Example ..

There has been one issue that has arisen while developing this stimuli that has proven to alter the character of the data in an unwanted way. Since the chatbot has no access to the conversation that happens between the occurence of the referred to entity and the participants present moment, chances are the chatbots question will be on a piece of information that has been mentioned in the meantime. Whenever that happens the degree of participant acceptance decreases significantly due to a topical incoherence rather than due to struggling to mapping a deep anaphore. This has been dealt with via providing the chatbot with the rest of the conversation in another component of the prompt with the instruction to avoid any of the topics mentioned there. LLMs are known to handle negative instruction with less success than positive ones () but this measure seems to have mitigated the problem as can be seen in the data attached in the annex.

nonassignable anaphore

The last type of stimuli used in this paper is called a nonassignable anaphore. It is a device that the participant will tend to interpret as an anaphoric device, typically a personal or demonstrative pronoun, but one such that the participant will not be able to map to any of the candidate entities in the previous conversation text. This stimuli is expected to lower the participant acceptance by the greatest amount.

To make a chatbot contain this stimuli entities are tracked to make sure there are candidates to be considered in case an anaphore occurs. Once there is a sufficient number of entities recognized in the conversation a hardcoded response is returned instead of an LLM generated one. The response contains a pronoun that to make sense of participant needs to interpret as an anaphore.

Since the response containing the stimuli is hardcoded, there is no guarantee that is actually is incoherent with the previous conversation and that there is no candidate to map the anaphore on to. Though odds are high enough every conversation that is supposed to contain this stimuli will have to be manually checked to confirm the required stimuli is present. This will be the case for all the conversations regardless because presence of stimuli is not guaranteed for shallow and deep anaphore either.

An approach not explored in this paper is achieving a nonassignable anaphore is also possible via generating a response using a prompt that instructs an LLM to come up with a question containing an unrelated entity avoiding all mentioned entities and anaphorize it before showing it to the participant. Although LLMs tend to perform worse with negative instructions () this could be achieved using a chain of prompts. The notion of nonassignable anaphore brings into scope the question of what makes an anaphora assignable. It is the semantic compatibility of the words around the anaphore that determine which of the candidates the anaphore is referring back to. The generation of the lexical surroundings of the anaphore needs to be handled carefully when coming up with an LLM based response.

2.4.2 Ending the conversation

While recognizing when the conversation is ending or especially when it should not end people rely on a set of clues similarly to knowing when to take turns speaking (). In open-domain conversation like the one a convform chatbot holds with experiment participants, the main challenge is to recognize when there is a topic at hand that interests the participant. Another discipline in the realm of ending the conversation is recognizing it is a good time to end the conversation due to the participants lack of interest. Conversation designs made for this paper do not take much of this into account. The main goal for a convform chatbot here is to present the participant with a stimuli. Once they manage that, if the participant is willing to continue the conversation continues for a hardcoded number of responses. This leads to participants sometimes noticing the conversation ending abruptly and mentioning they would like to continue in the questionnare or even at the very end of the conversation itself. This can potentially have effect on the score given by the participant and therefore brings noise into this parameter. In terms of quantization conversation is inherently noisy. Since each conversation has to be manually checked, an assessment of how to deal with this noise can be made while and after processing the data. A runtime topic annotation and other prompting techniques could potentially help make the convform chatbot converse in such a way that would be more aware of the general course of the conversation perhaps giving hints about ending the conversation ahead of time or reacting to and handling the participants hints of the same type.

2.5 Data

The data collected using a convform chatbot is a transcript of the conversation between a participant and the chatbot. Depending on the conversation design of the given chatbot the conversation may contain a record of the participant being exposed to a specific situation and their reaction to it. Other than that the collected data contains an information about whether the participant quit the conversation, the participants rating of the conversation and their comment on it.

Unfortunately it cannot be guaranteed the required stimuli actually occurs in the conversation. Though the probability is relatively high, the LLM technology responsible for most answers is nondeterministic and the participants tend to be unpredicatble. On many levels conversation can take an unintended direction which can spoil the stimuli. Whether thematic, textual or interactional, anything can go wrong. That is why as mentioned earlier, each conversation needs to be visited manually to confirm required stimuli is present. That does not present too much of a problem since a qualitative analysis of the data has the potential to bring lots of corrections to experiment design, relevant observations or inspiration for further steps. It would however pose a complication was one interested in scaling up this approach that would have to be dealt with.

2.5.1 Data analysis

Since the collected data is relatively complex multiple layers of analysis need to be performed on it. The steps of analysis will be following:

- Conversation design stimuli qualitative analysis
- Participant reaction qualitative analysis
- Quantitative analysis

In order to assess the datapoints for quantitative analysis each conversation needs to be visited to confirm or deny the presence of expected stimuli. This preliminary step of qualitative analysis provides feedback on whether the conversation design that is meant to facilitate the experimental design was succesful in doing so. Further analysis leading up to a quantitative assessment should only be done on conversations where convform succesfully exposed the participant to the stimuli. It has been observed in collected data that sometimes an unintended stimuli takes place instead of the intended one. Then simply changing the label of the conversation is appropriate. It is also necessary to filter the collected data for noncooperative participants who make it impossible for convform to perform the stimuli in the first place. The participant needs to take on a role of a friendly conversator if the experiment is to work. If they for example attempt to take advantage of the LLMs obedience and give it an unrelated task that involves any sort of text generation, they need to be pronounced noncooperative.

The qualitative analysis of the participant reaction to stimuli is mostly of explorative nature. The convform environment lets the user to input any text which even under the condition that they are cooperative can be unexpected and can derail the conversation. While these incidents are typically anecdotal and cannot be used to make generalizations they needs to be taken into account as a possible participant behavior. From the perspective of dialog system development such cases would be considered to be edge cases and dealing with them on the conversation design level is typically considered lower priority. However the conversation analysis perspective will be very interested in all the potential paths that a certain conversation situation can go in. This is ultimately why any weaknes to the conversation design does not matter too much as all conversation should be visited anyways in search for unique situations.

Ultimately the main reason for using an experiment environment solution like convform can be summed up in two points:

- Partial control over what happens in the conversation
- The ability to expose multiple people to a comparable conversational stimuli

The first point has been described in detail above. As far as the second point goes, this approach is a step towards a quantitative analysis of conversation. Though conversation is multifactored and various types of situation always come with a different set of circumastances if a quantitative analysis of a set of conversations containing a certain reoccuring pattern proves contrastive in some of the quantifiable parameters to a set of conversations containing a different reoccuring pattern case could be made this is caused by the observed patterns regardless of all the noise naturally present in a conversation text. In case of this paper, the parameter is mainly:

- participant acceptance score
- ratio of the participants quitting the conversation at the stimuli

Chapter 3

Data collection and analysis

3.1 Pilot experiment

3.2 Main experiment

Final thoughts

3.3 Further steps

Comparing different anaphore depths

Realtime topic annotation

Making the chatbot aware of the general course of the conversation

ending correctly

Relating tracked entities to one another

Prompting unassignable anaphora

Hybrid approach

An interesting next step that could be done on the contact surface of conversation design and LLM development is a hybrid system () that would allow a conversation designer to have the dialog system interact with a user in the required way including passing through a state-machine like process while keeping the flexibility of an LLM.

For such system one would need to decide the relationship between generated and hardcoded system utterances. Integrating a hardcoded utterance with a potential generated surrounding could be impossible, so the final layer of producing an answer would always be a prompt.

Prompt consists of multiple components as calling an LLM usually includes giving it a general instruction called the system prompt and whichever dynamically added information that is necessary to make the model output appropriate information like custom context and mainly the end user input. A hybrid approach to this would mean to allow a conversation designer to modify various components in the prompt based on context. The only way to know anything about what is going on in the conversation at design time however would be analyzing the user input and the evolving interaction via analysis prompts.

Bibliography

Andrist, S. et al. "Turn-Taking and Coordination in Human-Machine Interaction". In: *Ai Magazine* 37 (Dec. 2016), pp. 5–6.

Austin, J. L. *How to do things with words*. Oxford University Press, 1962.

Ayuso, D. "Discourse entities in Janus". In: *27th Annual Meeting of the Association for Computational Linguistics*. 1989, pp. 243–250.

Baldassarre, M. T. et al. "The social impact of generative ai: An analysis on chatgpt". In: *Proceedings of the 2023 ACM Conference on Information Technology for Social Good*. 2023, pp. 363–373.

Barthes, R. and S. Heath. *Image, Music, Text*. A fontana original. Fontana Press, 1977.

Brown, P. and S.C. Levinson. *Politeness: Some Universals in Language Usage*. Politeness: Some Universals in Language Usage. Cambridge University Press, 1987.

Couper-Kuhlen, E. and M. Selting. *Interactional Linguistics: Studying Language in Social Interaction*. Cambridge University Press, 2017.

Deshpande, A. et al. *Toxicity in ChatGPT: Analyzing Persona-assigned Language Models*. 2023. arXiv: `2304.05335 [cs.CL]`. URL: `https://arxiv.org/abs/2304.05335`.

Dingemanse, M. and N. J. Enfield. “Interactive repair and the foundations of language”. In: *Trends in Cognitive Sciences* 28.1 (2024), pp. 30–42.

Ferrucci, D. et al. “Building Watson: An Overview of the DeepQA Project”. In: *AI Magazine* 31.3 (July 2010), pp. 59–79. URL: https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/2303.

Garfinkel, H. “Studies in ethnomethodology”. In: Routledge, 2023, pp. 209–261.

Gervits, F. et al. “It’s About Time: Turn-Entry Timing for Situated Human-Robot Dialogue”. In: *Proceedings of the Special Interest Group on Discourse and Dialogue.* 2020. URL: https://hrilab.tufts.edu/publications/Gervits2020Sigdial.pdf.

Givón, T. *Coherence.* John Benjamins Publishing Company, 2020.

Gravano, A. and J. Hirschberg. “Turn-taking cues in task-oriented dialogue”. In: *Computer Speech & Language* 25.3 (2011), pp. 601–634.

Grice, H. P. “Logic and Conversation”. In: *Syntax and Semantics: Vol. 3: Speech Acts.* Ed. by Peter Cole and Jerry L. Morgan. Academic Press, 1975, pp. 41–58.

Hagerty, A and I. Rubinov. “Global AI ethics: a review of the social impacts and ethical implications of artificial intelligence”. In: *arXiv preprint arXiv:1907.07892* (2019).

Hajičová, E., J. Havelka, and P. Sgall. “Discourse Semantics and the Salience of Referents”. In: *Journal of Slavic Linguistics* 11.1 (2003), pp. 127–140.

Halliday, M. A. K. and R. Hasan. *Cohesion in English.* Longman, 1976.

Hendrix, G. G. “Natural-language interface”. In: *American Journal of Computational Linguistics* 8.2 (1982), pp. 56–61.

Hjelmslev, L. *O základech teorie jazyka*. Trans. by F. Čermák. 2. doplněné a upravené vydání. Praha: Academia, 2016.

Hrbáček, J. *Nárys textové syntaxe spisovné češtiny*. Praha: Trizonia, 1994.

Hu, K. and A. Tong. "OpenAI Rivals Seek New Path to Smarter AI as Current Methods Hit Limitations". In: *Reuters* (2024). URL: `https://www.reuters.com/technology/artificial-intelligence/openai-rivals-seek-new-path-smarter-ai-current-methods-hit-limitations-2024-11-11`.

Jones, C. R. and B. K. Bergen. *People cannot distinguish GPT-4 from a human in a Turing test*. 2024. arXiv: `2405.08007 [cs.HC]`. URL: `https://arxiv.org/abs/2405.08007`.

Jonsdottir, G. and K. Thórisson. "A Distributed Architecture for Real-Time Dialogue and On-Task Learning of Efficient Co-Operative Turn-Taking". In: Oct. 2013, pp. 293–323.

Koblížek, T. *Interpretační sémantika. Úvod do textové teorie Francoise Rastiera*. Univerzita Karlova v Praze, Filozofická fakulta, 2015.

Kološová, Š. "Konverzační design: principy designu hlasového robota pro přirozenou komunikaci s lidmi". Diplomová práce. Praha: Univerzita Karlova, Filozofická fakulta, Ústav informačních studií - studia nových médií, 2022.

Laban, G. et al. "Robo-Identity: Exploring Artificial Identity and Emotion via Speech Interactions". In: *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 2022, pp. 1265–1268.

Loaiciga, S., Simon D., and D. Schlangen. "Reference and coreference in situated dialogue". In: *Proceedings of the Second Workshop on Advances in Language and Vision Research*. 2021, pp. 39–44.

Loaiciga, S., S. Dobnik, and D. Schlangen. “Anaphoric Phenomena in Situated dialog: A First Round of Annotations”. In: *Proceedings of the Fifth Workshop on Computational Models of Reference, Anaphora and Coreference*. Association for Computational Linguistics, Oct. 2022, pp. 31–37. URL: `https://aclanthology.org/2022.crac-1.4/`.

Makela, E. “Survey of semantic search research”. In: *Proceedings of the seminar on knowledge management on the semantic web*. Department of Computer Science, University of Helsinki, Helsinki. 2005.

Maršík, A. *Convform*. `https://github.com/almarsk/convform`. 2024.

McTear, M. F. *Conversational AI: Dialogue Systems, Conversational Agents, and Chatbots*. Springer International Publishing, 2020, pp. 43–70.

Metz, R. et al. “OpenAI, Google, and Anthropic Are Struggling to Build More Advanced AI”. In: *Bloomberg News* (2024). Accessed: 2024-12-19. URL: `https://www.bloomberg.com/news/articles/2024-11-13/openai-google-and-anthropic-are-struggling-to-build-more-advanced-ai`.

Nedoluzhko, A. “Rozšířená textová koreference a asociační anafora (koncepce anotace českých dat v pražském závislostním korpusu)”. PhD thesis. Prague: Univerzita Karlova, Filozofická fakulta, Ústav českého jazyka a teorie komunikace, 2010.

Raux, A. and M. Eskenazi. “Optimizing the turn-taking behavior of task-oriented spoken dialog systems”. In: *ACM Trans. Speech Lang. Process.* 9.1 (May 2012).

Ritchie, H. "What's the impact of artificial intelligence on energy demand?" In: *Sustainability by Numbers* (Nov. 2024). URL: https://www.sustainabilitybynumbers.com/p/ai-energy-demand.

Ritter, G. and W. Lu. "The AI Plateau Is Real — How We Jump To The Next Breakthrough". In: (Dec. 2024). URL: https://www.emcap.com/thoughts/ai-s-curve-plateau-proprietary-business-data-breakthrough/.

Roberts, R. M. and R. J. Kreuz. "Nonstandard discourse and its coherence". In: *Discourse Processes* 16.4 (1993), pp. 451–464.

Sacks, H. *Lectures on Conversation*. Ed. by G Jefferson. Oxford: Basil Blackwell, 1992.

Schegloff, E. A. "On the organization of sequences as a source of 'coherence' in talk-in-interaction". In: *Conversational Organization and its Development*. Ed. by B Dorval. Ablex, 1990, pp. 51–77.

— *Sequence Organization in Interaction: A Primer in Conversation Analysis*. Cambridge University Press, 2007.

Schegloff, E. A., G. Jefferson, and H. Sacks. "The Preference for Self-Correction in the Organization of Repair in Conversation". In: *Language* 53.2 (1977), pp. 361–382.

Schegloff, E. A. and H. Sacks. "Opening up Closings". In: *Semiotica* 8.4 (1973), pp. 289–327.

Searle, J. R. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, 1969.

Sharma, D. et al. "Exploring The Evolution Of Chatgpt: From Origin To Revolutionary Influence". In: *Educational Administration: Theory and Practice* 30.5 (2024).

Skantze, G. "Turn-taking in Conversational Systems and Human-Robot Interaction: A Review". In: *Computer Speech & Language* 67 (May 2021), pp. 101–178.

Turing, A. M. "Computing Machinery and Intelligence". In: *Mind* 59.236 (1950), pp. 433–460.

Weizenbaum, J. "ELIZA a Computer Program for the Study of Natural Language Communication Between Man and Machine". In: *Commun. ACM* 9.1 (Jan. 1966), pp. 36–45.