Supplementary Materials A:

Codebook for identifying misunderstandings in online dialogue

Author note:

This codebook accompanies the article titled: *The Repeated Adjustment of Measurement Protocols (RAMP) method for developing high-validity text classifiers.* It is the final version of the codebook used in the RAMP method

Introduction

This document provides instructions for manually coding misunderstandings in online dialogue. Misunderstandings are defined as a break in dialogue participants mutual understanding. Mutual understanding is both cognitive and behavioral (Bavelas et al., 2017). A dyad *has* mutual understanding when they understand each other's perspective psychologically. A dyad *does* mutual understanding through dialogue, using linguistic and paralinguistic cues to indicate whether mutual understanding is being maintained.

The codebook was used to code for intersubjective misunderstandings, defined as a Self has an incorrect perspective of the Other's perspective. A subjective misunderstanding incorporates all intersubjective misunderstandings and is defined simply as an incorrectly held perspective. This distinction is derived from the Interpersonal Perception method (Laing et al., 1966), which distinguishes between *direct perspectives* (I think X), *meta-perspectives* (I think you think X) and *meta-meta-perspectives* (I think you think I think X).

Table 1 illustrates the difference between subjective and intersubjective misunderstandings. A subjective misunderstanding includes instances requiring only a direct perspective, "I don't understand X". An intersubjective misunderstanding is a subset where the object of misunderstanding contains an Other's direct or meta-perspectives. These can be *direct* where the Self misunderstands an Other's direct perspective (meta-perspective) or *felt* where the Self feels the Other has misunderstood the Self's perspective (meta-meta-perspective).

Table 1

Intersubjective and subjective misunderstanding

Misunderstanding type	Example	Direct perspective	Meta-Perspective	Meta-meta-perspective	Object of misunderstanding	N subjects
Subjective	"I don't understand the weather today"	"I don't understand"			"the weather today"	1
Intersubjective (direct)	"I don't understand what you just told me"	"I don't understand"	"you just told me"		"what you just told me"	2
Intersubjective (felt)	"I don't understand why you think I meant that"	"I don't understand"	"you think"	"I meant"	"why you think I meant that"	3

As with mutual understanding, intersubjective misunderstandings have a cognitive level, involving incorrect attributions of another's perspective, and a behavioral level, where these are expressed in language (see Table 1. Examples). This coding manual operationalizes the concept of misunderstanding through the construct of *reported* misunderstanding, defined here as sentences that provide explicit evidence of direct or felt intersubjective misunderstandings.

Methodology

The process of coding follows Content Analysis principles (see Krippendorff, 2018) which involves a deductive approach to classifying phenomena in text (Neuendorf, 2018). The coding manual was derived through an inductive process conducted by the authors in deliberation with four trained researcher assistants of MSc level. Through these deliberations, we sought to align our understandings of how reported misunderstandings appear in online dialogues and how they should be coded.

Datasets

This coding manual was designed and tested using the following three datasets:

- **Reddit**: random sample of 1,000 posts + comments from 30 subreddits gathered through API. Each subreddit has different rules and norms that participants abide by. There is a wide variation of topics and types of discussion on reddit. See example here: https://www.reddit.com/r/changemyview/comments/s4opx2/cmv_i_dont_see_how_a_person_can_be_a_christian/

- **Wikipedia talk pages**: pre-existing dataset (Danescu-Niculescu-Mizil et al., 2012)involves Wikipedia editors discussing articles and edits. Often these take place on an editors own talk page, meaning that people direct questions or comments to them. See example here: https://en.wikipedia.org/wiki/User_talk:U3964057[1]

- **Twitter**: pre-existing dataset (Thought Vector & Axelbrooke, 2017) involves dialogues between organisations & customers. See example here: https://twitter.com/SueHiles/status/1486337775746093062

Unit of analysis

We coded misunderstandings at the sentence level rather than dialogue turns. We therefore parsed a single dialogue turn into component sentences using the Python package NLTK (see: www.nltk.org/api/nltk.tokenize.html).

The sentences were presented to coders in context of the wider turn and dialogue, meaning they were provided contextual clues to the meaning of the phrases. Misunderstandings emerge through taking turns (e.g. Corti & Gillespie, 2016), meaning sentences were coded sequentially. This allowed for dialogues, turns, and sentences to be coded simultaneously and reconstructed at multiple resolutions of analysis. Misunderstandings are coded at the sentence level as a binary variable, where one indicates the presence of the variable:

Example coding

Misunderstanding example

[1] Note: the Wikipedia Talk Pages format is unusual because it is based on the same platform and management system as the articles. A new turn is indicated by an indent and ends in an editor signing off their name.

Turn id	Sentence	Misunderstanding
1	I get your point	0
2	I don't get yours	1

Coding misunderstanding

A sentence was coded as a misunderstanding when it involved a statement or a question expressing direct misunderstandings (see Table 2). These are statements that, unambiguously, involve the participants discussing a break in mutual understanding.

Table 2

Misunderstanding types

Sentence Type	Linguistic indicators	Misunderstanding types	
		Direct	Felt
Statement	Second or first person perspective as object	"I don't understand your point"	"You didn't understand my point"
	Generic pronoun as object	"I don't understand this"	"You don't understand this"
	Quote or paraphrase as object	"I don't understand 'halogen is best'"	"You don't understand 'halogen is best'"
Question	Second Person Pronoun in question	"What did you mean by X?	"Is that what you think I meant?"
	Generic pronoun in question	"What does this mean?"	"Is X what you think that meant?"
	Quote or paraphrase in question	"What does 'halogen is best' mean?"	"Is "X" what you meant by 'halogen is best'?"

When coders suspected a misunderstanding statement a sentence, they considered first whether:

a) the sentence's object contains a verb or verb phrase indicating misunderstanding (e.g. "don't understand", "misunderstand", "didn't mean", etc.)

b) the sentence's object contains a second or first person perspective (e.g. "I meant", "you said"), a *generic pronoun* (e.g. "this", "that") or a *quote or paraphrase* in the sentence's object

Coders were advised to consider the contextual meaning of a suspected statement of misunderstanding due to ambiguous terms. For instance, "I don't understand this" doesn't necessarily refer to something someone else has said. Some cases of disagreement were expressed by the Other in reply to a direct perspective by the Self through an observation of misunderstanding (e.g. Self: "Cats are evil.", Other: "that's a common misconception"). These accusations of misunderstanding were not coded as they indicated a disagreement rather than an instance of misunderstanding. These disagreements were considered as statements of subjective misunderstanding, meaning the object of confusion was a topic and not a person.

When a sentence is a question, misunderstandings were coded when the sentence's object contains a second or first person perspective (e.g. "I meant", "you said"), a generic pronoun (e.g. "this", "that") or a quote or paraphrase. A question is generally asking for more information and, therefore, indicative of an incomplete perspective. However, a question was only considered evidence of intersubjective misunderstanding when referring directly to an interactant's perspective. Questions were not considered a misunderstanding when requesting new information (e.g. "What do you think about the new film?"). A question was considered a request for new information if its content had not been previously mentioned in the dialogue. However, if the question related to the content of the previous dialogue, it was viewed as a request for new information unless it is directly addressing what someone else has said.

Some questions soliciting new information resembled clarification questions due to the use of pronouns. For instance, “can you help me?” or “Do you want me to change this for you?” contain the second person pronoun and, therefore, could be confused as a clarification request. The coders did not code for these examples as the speaker is asking for new information not previously provided by the Other, instead of rectifying a misunderstanding on information already shared.

In some questions, an adverb may indicate whether a question is clarifying or not (e.g. “can you *actually* help me or not?” uttered after already asking for help in a previous turn). Adverbs are not a definitive indicator of clarification and, as with their grammatical purpose, can be used to exaggerate or enunciate (e.g. “when exactly will this be fixed?” if uttered in the first turn). Coders were instructed to pay attention to whether adverbs were present and only coding the question as misunderstanding when confident in the coding.

Examples of misunderstanding

Direct misunderstanding: ***I misunderstood you, I don’t understand what you mean, I don’t understand, no idea what you’re saying, what did you mean, that doesn’t make sense, that makes no sense, we’re misunderstanding each other, can you clarify what you meant***

Felt misunderstanding: *I didn’t mean, you misunderstood me, I wasn’t saying, I wasn’t saying that, I never meant to, you missed my point, that’s not what I meant, you took this the wrong way, that’s not what I was talking about, I wasn’t talking about, my comment was meant to*

Frequently Asked Questions

These questions were used for guiding coders on how to approach edge cases.

How much should I rely on context for coding a sentence?

How a sentence is coded should be decided, first and foremost, based on the grammatical construction of the sentence. Occasionally, however, the correct code for a sentence may be determined by its relation to the prior turns. There are three scenarios when it is likely that the context will influence the sentence's coding:

1. *When a question or statement in the first turn of a dialogue resembles a misunderstanding.*

   A question or statement in the first turn of a dialogue should only be coded for misunderstanding if it is in reaction to something said or done by another individual expected to reply. To elaborate, a conversation might begin with the question "*can you explain to me what you meant by X?"* or the statement *"I didn't understand when you said X"*. These sentences presume a previous dialogue the reader is not privy to and is addressing a single individual expected to reply. In these cases, the first turn is not a true first turn, but rather a second turn following the behavior or utterances elsewhere.

   If a statement or question in the first turn is not addressing a particular individual expected to reply – e.g. "I don't understand why any of you think X" – they should not be coded as misunderstanding. For instance, on Reddit, it is highly improbable for a first turn ever to be addressed as a single individual, meaning a statement or question will not be a

misunderstanding.

2. When a question is ambiguous in whether it is asking for new information.

   A question should not be coded as a misunderstanding if it is asking for new information. It is sometimes ambiguous whether a question is asking for new information or clarification on information previously given. For instance, the question “what does X mean?” if uttered without X being previously mentioned, is not a clarification question. However, if X has been mentioned previously, then the question is asking for clarification. In the case of ambiguous questions, the coder should rely on what has been said in the previous turns to determine whether new information is being elicited.

3. *When a question is rhetorical.*

   A rhetorical question should only be coded as misunderstanding when it is directly addressed at something somebody has said. For instance, the question “do you REALLY believe X?” if uttered in the first turn is unlikely to be a misunderstanding. However, if addressed directly to another participant in the dialogue who has previously spoken about X, then it should be coded for misunderstanding. For rhetorical questions with the second person pronoun “you”, the coders should ask themselves whether it is a general “you” meaning an ambiguous audience, or a specific “you” meaning another dialogue participant.

What should I do if I suspect a misunderstanding but am unsure?

Code conservatively, meaning only code if the above rules are met and the statement is an unambiguous statement of misunderstanding. Remember that *having* misunderstanding without linguistic evidence (e.g. we notice someone has misunderstood something but aren't aware of it) should not be coded.

References

Bavelas, J., Gerwing, J., & Healing, S. (2017). Doing mutual understanding. Calibrating with micro-sequences in face-to-face dialogue. *Journal of Pragmatics*, *121*, 91–112. https://doi.org/10.1016/j.pragma.2017.09.006

Corti, K., & Gillespie, A. (2016). Co-constructing intersubjectivity with artificial conversational agents: People are more likely to initiate repairs of misunderstandings with agents represented as human. *Computers in Human Behavior*, *58*, 431–442. https://doi.org/10.1016/j.chb.2015.12.039

Danescu-Niculescu-Mizil, C., Lee, L., Pang, B., & Kleinberg, J. (2012). Echoes of power: Language effects and power differences in social interaction. *Proceedings of the 21st International Conference on World Wide Web*, 699–708. https://doi.org/10.1145/2187836.2187931

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. SAGE Publications.

Laing, R. D., Phillipson, H., & Lee, A. R. (1966). *Interpersonal perception: A theory and a method of research* (pp. vii, 179). Springer.

Neuendorf, K. A. (2018). Content analysis and thematic analysis. In *Advanced Research Methods for Applied Psychology* (pp. 211–223). Routledge.

Schegloff, E. A. (1992). Repair after next turn: The last structurally provided defense of intersubjectivity in conversation. *American Journal of Sociology*, *97*(5), 1295–1345.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, *53*(2), 361–382. https://doi.org/10.2307/413107

Thought Vector, & Axelbrooke, S. (2017). *Customer Support on Twitter (v10)*. https://kaggle.com/thoughtvector/customer-support-on-twitter