

Coding Scheme for Negative Utterances

Frank Förster¹

¹Adaptive Systems Research Group, University of Hertfordshire

July 31, 2018

Abstract

This document contains an abbreviated version of a coding scheme employed for the pragmatic 2-coder analysis of negation types and their felicity. It was used for the coding of negative utterances originating from human-robot dialogues gathered in the experiments described in [3, 2, 1]. Some theoretical parts as well as sections on future work have been removed for space reasons. The complete scheme is contained in [1]. The scheme was devised by the author who also acted as first coder. Additionally a second coder was employed, and those parts of the coding scheme handed to the latter as coding manual are marked as such.

1 Construction of the coding scheme

The purpose of the coding scheme described below is a description of how to determine the negation types for all negative utterances produced by the robot and the participants recorded during the experiments and, furthermore, which productions can be regarded as felicitous or adequate in the given situations.

Against the background of the limitations of speech act theory (SAT) with regard to the analysis of actual conversations (cf. [1]) it might not be surprising that these very limitations of SAT's single-utterance approach became apparent right from the start during the first round of coding. The first coding round was conducted by the author who also acted as 1st coder. The initial coding was conducted on the full set of negative utterances taken from the experiments and the resulting types are based on Pea's [4] taxonomy with eventual extensions wherever these were deemed necessary.

Despite the mentioned issues we still think that a mixed qualitative-quantitative approach is useful in order to determine what kind of negation types qua speech acts were produced in experiments of this type. Yet optimally, if the analyst has enough time to do so, a complete conversation analytic (CA) analysis would be preferable. The usefulness of this mixed

approach hinges on the concept of speech acts being extended in order to account of basic conversational phenomena such as adjacency in a way similar to Pea's approach. Without this kind of extension many utterances could not be coded due to them being 2nd part-pairs of adjacency pairs whose theoretical status in Searle's version of SAT is unknown. Thus many of the negation types listed below are, similar to Pea's types, 'types of use' and not 'pure' speech act types. They might be rather viewed conversation analytical and speech act theoretical hybrids due to the top-level inclusion of adjacency as defining property on one hand, and due to the refusal to accept Searle's short list of possible communicative functions on the other hand. It is unclear if Pea himself would have thought of his types in this way but the fact that his taxonomy uses adjacency as defining criterion on the top-level, and his refusal to have the elements of his scheme be classified as "types of early meaning" leads us to think that he indeed leaned towards a conversation analytical approach with a further leaning to Wittgenstein's "definition via use". In this context it is important to realize that at the time of Pea's publication CA was in its early days.

The underlying idea on how to decide on the felicity of a particular utterance (stage 1, see below) is to let a competent English speaker decide if an utterance was felicitous or not without taking into account the theoretical (and categorical) apparatus as proposed by Searle [5] but by relying instead on the judgmental process that any fluent speaker of a language employs when engaging in conversation herself.

The time stamps and durations in the table for each negative utterance were derived from the recorded audio and video, and the robot's ("Deechee") log-files. Deechee's utterances are in few cases hardly audible or not audible at all as the audio was recorded via the headset worn by the participants which suppresses external noise. Therefore, instead of relying solely on the audio recording, Deechee's utterances were extracted from the log files of the *languageing* module - the module that determines during any experimental session content and timing of Deechee's utterances (see *Materials & Methods* in [2, 3, 1]). Using this log data ensured that no robot utterance was missed.

Teaching episodes In terms of notation it might be helpful to introduce the concept of a *teaching episode*. Due to the way both experimental scenarios are set up, the interaction and conversation of any session naturally divides into a sequence of *teaching* episodes. Participants present and explain sequentially objects to the robot with small transition times in between presentations. A *teaching episode* starts with the participant picking up a particular object in order to present it to Deechee and ends when the object is either put back on the table or if the object is dropped by Deechee. Such a segmentation leads to some parts of the interaction and conversation being outside of any such episode. At the moment the data is not segmented into such episodes, but for future work we deem the notion of episode boundaries to be potentially advantageous. We hypothesize that the temporal location of episode boundaries have an impact upon how participants react to Deechee's utterances and bodily behaviour, which is encoded in the column *P* (*re-*)acts in accordance with *R*'s behaviour or speech.

2 Selection of sessions for 2nd coder

The percentage of negative utterances to be coded by the 2nd coder was set to 20%. Initially we were planning to select randomly 20% of the 100 experimental sessions. As the number of negative utterances per session and/or time unit vary greatly in between parties for both participants and robot, this method could lead to less than 20% of the total number of utterances being coded. This situation would occur if sufficient sessions from less-than-average communicative participants were randomly selected. For this reason we decided to select 20% of the negative utterances instead of selecting 20% of all sessions. The considerable variation of the number of negative utterances per session is true for both human and robot utterances.

Random selection of sessions for negative robot utterances The following procedure is applied in order to determine which sessions are selected for dual-coder coding of the robot utterances:

Let *TNU* be the total number of utterances across all sessions and participants. Split the (super)set of all sessions into two sets, each of which contains all (sub)sets of a particular condition, rejective condition vs. rejective-prohibitive condition. For each of these sets do:

1. Fill the *big_bag_of_sessions (BBS)* with all sessions of all participants in the selected condition
2. Select randomly a session from *BBS*
3. Determine the number of negative robot utterances for the selected session
4. Add this session to the *small_bag_of_sessions (SBS)* for dual-coding and remove it from *BBS*
5. If the total number of utterances in the *SBS* equals or exceeds 20% of *TNU* stop, otherwise go back to 1

Random selection of sessions for negative human utterances In order to select the 20% of the participants' negative utterances the same files were coded as the ones already selected by the procedure specified in the last paragraph. As there are more negative utterances produced by the participants than negative robot utterances, additional files had to be selected to reach the 20% margin: For each experiment all files across all sessions were numbered from 1 to 50. Subsequently a random generator was used to produce a new number upon any new run. The corresponding file was added to the set of files to be coded. This procedure was repeated until the 20% margin was reached.

3 Coding process

Initially the author and first coder coded all entries for all participants and all sessions which lay the basis for constructing the coding scheme and for determining the initial set of negation types to be used by the 2nd coder. The coding process for the 2nd coder is separated into three stages in order to enable the two coders to work in parallel: While the processing of the 1st and 2nd stage by the 2nd coder was under way, the first coder sought to reduce the number of negation types for stage 3. Initially there were 24 negation types on part of the participants which were deemed too many. This number was subsequently reduced to 19 types.

Stage 1 Coding of robot utterances for felicity: The second coder codes the last row (*felicity*) of table 1. Based on her knowledge of English and by virtue of being a fluent speaker of the language she decides whether a particular negative utterance is felicitous/adequate, i.e. whether the utterance makes sense in the given situational and conversational context.

Stage 2 Coding of negative robot utterances for the remaining columns: The second coder codes the remaining columns listed in table 1. We expect all columns except for the columns '*P (re-)acts in accordance ..*' and '*negation_types*' to be unproblematic in terms of inter-coder variance. The set of candidates for each column is listed in the table and described in the subsequent section.

Stage 3 Coding of human negative utterances: The second coder fills in the columns '*negation word*' and '*negation type*' indicated in table 2 for each negative utterance listed there. These are all negative utterances produced by the respective participant in the stated session. The set of candidates for each column is listed in the table and described in the subsequent section.

Coding Table for Robot Utterances

Columns marked **bold** are given. They were derived from the log files of the module which is responsible for Deechee's speech. Columns for time stamps are omitted for space reasons. Column *negation_word* is omitted as it is not applicable for robot utterances.

spea-ker^a	utt.	body behaviour	ling. signaling that utterance was (mis-) understood	P (re-)acts in accordance with R's behaviour or speech	negation type	felicitous
R_utt	no go don't ...	positive rejective prohibited undecided neutral	signals underst. ^b signals misun-derst.(+word) ^c no signal	N/A: see text AB: acts on behaviour AS: acts on speech ASB: acts on both NoA: doesn't react	truth-func. denial neg. agreement mot. dep. denial mot. dep. exclam. neg. imperative persp. dep. denial rejection of offer self-prohibition mot. dep. assertion none	yes no n/a

^a fixed in advance by coder 1, based on speech log files and 1st round of video analysis

^b Example: Deechee: *No* P: *no, not a big fan?*

^c Example: Deechee: *Go* P: *no? ok*

Table 1: Coding Table for Negative Robot Utterances

Explanation of the single columns

speaker Specifies the speaker: robot (*R*) in case of *R_utt* or participant/human (*P*) in case of *H_utt*.

start_of_utt Time relative to the start of the video recording when the utterance started. Note that this time might be inaccurate by around 1 sec.

utterance (utt.) This field contains the negation word which the robot produced.

body behaviour This field contains a description of the robot's bodily behaviour at the time when it pronounced the given utterance. It is **only applicable to robot utterances**. The following five different types, also listed in table 1, are possible:

- **positive**: R is smiling and possibly reaching for an object
- **rejective**: R frowns and potentially looks away from object and participant
- **neutral**: the robot is neither smiling nor frowning and neither reaches nor avoids a presented object
- **undecided**: the robot starts to smile and starts to hold its hand out but flinches back again and stops smiling. This “approach and flinch back” move might be repeated in this situation several times
- **prohibited**: the robot is actively prohibited by the participant, that is, the participant restrains the reaching movement of the robot.

linguistic signaling .. This field contains an evaluation that indicates whether the participant signaled in some way that he or she understood the robot's utterance. It is **only applicable for robot utterances**. There are three possible values for this field:

- **signals understanding**: P somehow signals that she understood the word. Often this happens by repetition of the word with an assertive intonational contour (\rightarrow neg. type: *neg. agreement*) or an intonation contour of doubt (\rightarrow neg. type: *neg. question*). Sometimes utterances might be deemed signals of understanding that were not mere repetitions of the robot's word, especially if it was witnessed before that the participant in question understood this very utterance. A necessary requirement is that P says something in direct succession ($\Delta t \ll 1s$) to R's utterance, and that this utterance is deemed by the coder to be an affirmative signal that serves to indicate that the utterance was understood.

If the coder deems a signal to be such a “signal of understanding” (*uptake*) this does not imply that this very signal can not serve other functions as well. Latter functions are outside the scope of this analysis. For example a participant answering Deechee's *no* with a *no?* not only signals that she understood Deechee's *no*, but might also signal at

the same time, with the very same word, and by intonating it in a particular way that she is not sure or convinced that Deechee really means what it just said. The important point is that the coder thinks that one of possibly many functions of P's utterance is to signal understanding with regards to what Deechee just said.

Example:

P offers Deechee the heart and Deechee starts to frown.

R: *No*

P: *No? Why not? The last time you liked playing with the heart*

- **signals misunderstanding (+ word)**: P somehow signals that he or she misunderstood what the robot said. Upon R uttering a word, P says what she understood in direct succession to R's utterance. The intonational contours are typically identical to the ones mentioned in the last case: *assertive* or *doubtful*.

Example:

R: *Go*

H: *No? Alright ..*

Here P evidently mis-heard and took the *go* for a *no*. In this case please specify the word for which the 'real' word was mistaken for in brackets after the type.

- **no signal**: the participant does not signal understanding or misunderstanding.

P (re-)acts in accordance .. The idea behind this column is to see whether participants (re-)act on bodily behaviour/gestures and/or speech. When both gestures and speech are in accordance the decision is straight forward. A more interesting case is given when gestures and speech are incongruent and when the behaviour leaves ample room for interpretation (typically *neutral* and *undecided* behaviour). Below the term *teaching episode* is used with which the following is meant:

A *teaching episode* starts with the participant picking up a particular object in order to present it to Deechee. The *teaching episode* ends when the object is either put back on the table by the participant, or by Deechee, or if the object is dropped. There are five possible values:

- **N/A (not applicable)**: at the time when the negative is uttered participants cannot react to the robot at all or it is highly unlikely that they do so. The situations where this is case are the following:
 - R utters the negative word in between two *teaching episodes*, for example when P has just put down an object and is looking for the next object to present.

- P is constraining R’s arm with one hand, as asked for by the instructions, and holds the box with the other. In this situation P’s ‘freedom to act’ is diminished as she follows the experimental instructions.

- **AB (acts on behaviour):** the participant (P) reacts in accordance with the robot’s behaviour.

Example 1:

R exhibits positive behaviour and says *no*. P tries to put the box into the robots hand, effectively ignoring the *no*.

Example 2:

R exhibits rejective behaviour, P starts to put the box down, the robot says *no*.

- **AS (acts on speech):** P reacts in accordance with the robot’s speech.

Example 1:

R exhibits neutral behaviour and P is offering the box to it.

R: *No*

P puts the box down, possibly confirming the robots utterance with

P: *No? Alright*

Note that a linguistic confirmation is not necessary but is a clear indicator that P actually acts on what R said.

Example 2:

Identical to example 1 but with the robot exhibiting an undecided behaviour

- **ASB (acts on both behaviour and speech):** P reacts in accordance with both the robot’s behaviour and speech. This can only be the case if R’s behaviour and speech are congruent.

Example:

R exhibits rejective behaviour and shortly after says *don’t*. P puts the box down after hearing the *don’t*, but not before that.

- **NoA (no (re-)action):** The participant does not react to R’s behaviour nor to R’s speech.

Example:

R exhibits rejective behaviour and says *no*. P ignores both and continues

to offer and speak about the box.

negation type Based on a first round of analysis the following negation types for the robot's utterances were derived from the recorded interactions. These types can be split into two groups: adjacent types, and non-adjacent types. *Adjacent* means that the utterance is a linguistic reaction to what the conversation partner said - at this stage of coding the conversation partner is the participant. A response to an answer, for example, is adjacent, adjacent to the answer, whereas answers themselves count in our taxonomy as non-adjacent. A rejection of an offer is non-adjacent, if the offer was performed mainly non-linguistically.

Adjacent negation types

- **Truth-func. denial:** Truth-functional denials are negative responses to *truth-functional assertions* or *questions*. Truth-functional assertions are assertions about states of affairs in the world which can be evaluated objectively. With *objectively* we mean here, that the truth or falsity of the assertion does not depend on the motivational state of the speaker or hearer nor on their perspective.

Example:

A: *It's raining outside*
B: *No (I don't think so)*

The truth or falsity of "It's raining outside" is independent of A (or B) being happy, grumpy, or sad. It is also independent of the circumstance if A or B can actually see if it rains or not, that is, if any of the two is close to a window or if A is just coming from outside or not. This is what is meant with *independent of their perspective*. As A's "It's raining outside" is a truth-functional assertion, B's "No" counts here as *truth-functional denial*. Also note that the example is not a classic adjacency pair as "It's raining outside" is not a question. Also note, if A would have asked "Is it raining outside?", a "No" on part of B would qualify here as *truth-functional denial* as well.

- **Persp. dep. denial:** Perspective-dependent denials are negative responses to *perspectival (or perspective-dependent) assertions* or *questions*. The truth or falsity of a perspectival assertion depends on either the knowledge (epistemic), the ability, or the physical perspective of an agent.

Example 1 (epistemic):

A: *You know this one here, don't you?*
B: *No, never seen such a thing*

Example 2 (ability):

A: *A big sportsman like you surely can do 50 pushups or not?*

B: *No, not right after a match.*

Example 3 (physical perspective):

A: *Can you see that red bird there on Mr Burns fence?*

B: *No, I can't see over the hedges. You forget that you're quite a bit taller than I am.*

- **Mot. dep. denial:** Motivation-dependent denials are negative responses to *motivation-dependent questions* or *assertions*. The answers to *motivation-dependent questions* depend on the motivational state of the addressee. With motivational states things such as likes, and dislikes, or wants (and not-wants) are meant.

Note, that also those questions or assertions are considered motivation-dependent which assume or refer to motivational states without containing motivational or volitional verbs such as *like, want, fancy, feel like* etc. (implicitly motivational).

Example 1 ('straight-forward' motivational):

A: *Do you like dogs?*

B: *No, not really.*

Example 2 ('volitional'):

A: *Do you want to come along to the cinema tonight?*

B: *No, I don't feel like going to the movies today.*

Example 3 (implicit)

A: *Will you finally mow the lawn this afternoon?*

B: *No*

Example 4 (implicit)

A: *How about some ice cream?*

B: *No, I'm rather feeling like something savory at the moment*

Note that in the examples 3 and 4 there is no verb or adjective in A's questions that would qualify as *motivational*. Nonetheless the question involves the motivational state of the addressee by implicitly, that is without lexically referring to said states, asking if the addressee *feels like, is up to, or willing* to mow the lawn or if she wants some ice cream. Example 3 furthermore gives a hint that the listener must have responded negatively to the very same question in the past, but this is an issue out of our scope.

- **Neg. agreement:** Negative agreement is given if the participant produces a negative ut-

terance of some kind and R agrees with it by uttering a negative as well. Often but not always the negative is a repetition of the negative expression used by the participant. The participant's utterance can have an 'assertional' intonation contour or a question contour. In the latter case the question must already suggest or anticipate an answer as is the case in example 1. Without the question suggesting an answer there would be nothing with which B could agree with.

Example 1:

A: *So you don't like strawberries then?*

B: *No*

Example 2:

A: *No, evidently you're not very keen on strawberries.*

B: *No.*

Non-adjacent negation types

- **Rejection of offer:** Rejective utterances are very similar to *motivation-dependent denial*, the main difference being that the latter is adjacent to another utterance of the conversation partner. *Rejections* are always reactions to non-linguistic offers or proposals of some kind.

Example:

A is holding out an apple towards B, effectively offering it to B but not saying a word

B: *No, thanks! I'm not very much into fruit.*

It is important to emphasize that the timing of the robots utterances in relation to the human utterances is important to distinguish *rejection of an offer* from *motivation-dependent denial*. The crucial question is: Does the coder deem the utterance to be an answer or another kind of linguistic reaction to a recent utterance of the participant? Only if the coder thinks that the utterance is independent of the participant's utterance(s), it can be a case of *rejection*.

- **Self-prohibition:** Self-prohibition can only occur in the prohibition scenario. It consists of the repetition of a word which was previously used by the participant in a prohibitive way and/or while physically prohibiting the robot. Often participants counteract self-prohibition by saying things such as *No, you can have that*.

Example:

P is holding out a box to the robot

R: *Can't*

P: *Yes, you can hold it*

- **Neg. imperative:** *Negative Imperatives* are similar to *rejection of offers* but don't assume an offer on part of the conversation partner. For an utterance to count as a negative imperative it is necessary though, that the person that is addressed with the imperative is in the process of doing something or just about to do something that is not wanted by the speaker.

Example:

A: *And now we're going to put the chicken into the microwave.*

B: *No! Are you crazy?*

- **Mot. dep. assertion:** Motivation-dependent assertions are utterances other than *rejection of offers* or *neg. imperatives* that are linked to the motivational state to any of the conversation partners. They are in some way a residual class for non-adjacent motivational utterances that are too 'weak', intonationally or also by the mere context in which they are uttered, to count as *neg. imperatives*.

Example: (mot. dep. assertion in B's 1st utterance)

A: *I'm going to the cinema tonight! Can't wait to see the new Star Wars!*

B: *I'm still knackered from the weekend. So I guess, I'll give that one a miss.*

Negation types that can be both adjacent and non-adjacent

There is only one type of negation in our taxonomy that does not clearly fit into the adjacent - non-adjacent dichotomy:

- **Mot. dep. exclamation:** Motivation-dependent exclamations can stand in terms of adjacency in isolation. In this case they typically refer to a current event. But they can as well be adjacent and refer to an utterance of the conversation partner to signal disagreement. In the non-adjacent case they might be most similar to *mot. dep. assertions* but are typically less articulate and more spontaneous. As opposed to *mot. dep. assertions* they must refer to some current event of some kind which is disagreed with or negative in some other way. In the adjacent case they are most similar to *mot. dep. denials* but are not responses to questions or (linguistic) offers but rather express disagreement with an evaluation or assertion of the conversation partner which was not explicitly asked for.

Example 1 (non-adjacent):

A accidentally drops a glass of wine onto the white carpet.

A (or B): *Oh no! I'll never get that stain out.*

Example 2 (adjacent):

A and B are watching a football game together but side for opposite teams. Team X scores the first goal of the game 10 minutes before the end of the game and chances are that this will remain the only goal of the game.

A: *Finally! It was about time. That's it then, I guess.*

B: *No way! This game is far from over.*

Example 1 illustrates that the event which triggers the exclamation can be caused by the speaker itself or another agent. This type does not distinguish by whom the triggering event has been caused or if it was caused by any agent at all. Natural events such as a thunderbolt which are not caused by any agent could trigger such an exclamation.

Other types

- **None:** It can happen with those negation words which we qualified earlier as *pragmatic*, such as *go*, *down*, or *done*, that they are used and/or perceived as negatives, for example they may serve the function of rejection. This is not always the case though. So if one of these *pragmatic* negatives in a particular situation is none, their type should be qualified as such: *none*. This is only applicable to *pragmatic* negation words, not to regular lexical/ grammatical negatives such as *no*, *don't*, or *can't*.

Coding Table for Human Utterances

Columns marked **bold** are given. Each entry was extracted from the files coming out of the prosodic analysis. The time stamps were created by hand by the 1st coder when coding each entry. As opposed to the robot utterances the human utterances are tagged with a start and end time to roughly give an idea of where the automatic utterance boundary detection put the boundary. The utterance delimited through these boundaries does not necessarily coincide with what we intuitively think of as an utterance.

Note that there can be more than one negation word per utterance. If this is the case please specify both, separated through a semicolon in the column *negation word*. If you think that the two negation words belong to different negation types, also specify both negation types, separated through semicolon, in the column *negation type*. If you think that both negative words belong to the same negation type, it is ok to specify the type only once in this column.

speaker	start_of_utt	end_of_utt	salient word	negation word	negation type
H_utt	1:17:59	1:19:20	okay	no	truth-func. denial
	3:49:27	3:50:10	squares	no	neg. agreement
	1:01:49	1:03:40	don't	don't	neg. question
	5:17:39	5:20:30	hearts	no	rejection_of_request negating self-prohibition truth-func. question neg. persp. question neg. mot. question neg. persp. assertion mot. dep. assertion truth-func. negation prohibition disallowance neg. promise neg. tag question neg. intent interpret. quoted negation mot. dep. exclamation neg. imperative

Table 2: Coding Table for Negative Human Utterances; entries for columns in **bold** are given, entries for the remaining columns have to be entered by the coder

Explanation of the single columns

speaker Specifies the speaker: robot (*R*) in case of *R_utt* or participant/human (*P*) in case of *H_utt*.

start_of_utt Start time of the utterance, relative to the start of the video recording. Note that this time might be inaccurate by around 1 sec.

end_of_utt End time of the utterance, relative to the start of the video recording.

salient word This field contains the prosodically most salient word of the particular utterance as determined by the prosodic analysis. The most salient word can or cannot be the negation word. This is why this field may contain non-negative words.

negation word This field has to be filled with the negation word(s) in the utterance by the coder. It might be the same word as in the field “salient word” above or it might differ from it as not all negation words are salient.

negation type Based on a first round of analysis the negation types for participants’ utterances listed below were derived from the recorded interactions. These types can be split into two groups: adjacent types, and non-adjacent types. *Adjacent* means that the utterance is a linguistic reaction to what the conversation partner said. At this stage of coding the conversation partner is the robot. A response to an answer, for example, is adjacent, adjacent to the answer, whereas answers themselves count in our taxonomy as non-adjacent.

In all examples below the negative words which form part of the respective type are underlined like this.

Square brackets indicate an overlap of the robot’s with the participant’s speech: they pronounced the so marked words simultaneously. Square brackets containing a number indicate pauses, where the number corresponds to the duration of the pause in seconds.

Adjacent negation types

- **Truth-func. denial:** Truth-functional denial is a reaction of the participant to a truth-functional utterance. See also the explanation of the same type in the section on robot utterances.

Example 1:

P: *What’s this one?*
R: *Heart*

P: No, it's not a heart

Example 2:

P: *Heart!*

R: *Circle*

P: No, *bad*

Example 3:

P: *What about the circular one?*

R: *There*

P: No, *this one's a circle*

- **Neg. agreement:** See the explanation for the same type in the section on robot utterances plus the following extension: If a participant implicitly assumes a negative utterance, as in example 1 below, this also counts as negative agreement. The *either* in example 1 indicates that the participant believes that Deechee doesn't like the object. For all that matters P acts as if Deechee would have explicitly said *no* before.

Example 1:

P: *It's not my favorite either, I'll get rid of it.* (Deechee did not say *no* before P uttered this)

Example 2:

P: *You want to play*

R: *No*

P: No, *ok. Lets try a different one then.*

Example 3:

P: *You don't like the heart? No? It's turning away from me, you don't like the heart*

R: *No*

P: No! No, *ok.*

- **Neg. question:** Negative questions are very similar to *negative agreement*: they are adjacent negatives in which the negative word of the conversation partner is repeated. As opposed to *negative agreements* the intonation contour here is one of doubt or surprise. As opposed to *neg. perspective questions* and *neg. mot. questions* (see below) this question type is necessarily adjacent to the utterance of the conversation partner.

Example 1:

P: *Do you [like] squares*

R: *[No*

P: *No?*

Example 2:

P: *Got a circle here*

R: *circle*

P: *Well done, that's right*

R: *No*

P: *No?*

Example 3:

P: *What about the moon? The crescent [there?*

R: *[No*

P: *No?*

- **Rejection of request:** As the name indicates, a rejection of a request is given if Deechee asks the participant for something (or the participants interpret Deechee's utterance in this way) and P rejects linguistically to comply with Deechee's request. As opposed to *rejection_of_offers* (see section on the classification of robot utterances), *rejections of requests* are adjacent to a linguistic request of the conversation partner.

Example 1:

R: *Moon*

P: *No, you've had the moon already*

Example 2:

R: *Square*

P: *You no.. no I'm not gonna show you the squares any more*

Example 3:

R: *Moon*

P: *No, we've had the moon already*

- **Negation of self-prohibition:** Negations of self-prohibitions only occur in the prohibitive scenario. Deechee utters a prohibitive negative, that is, a negative which was previously used by the participant to prohibit Deechee from touching a box. Sometimes participants then interpret such an utterance, when produced by Deechee as a form of self-prohibition and counteract using this type of negation.

Example 1:

P presents and speaks about the heart box. R goes for it, but flinches back.

R: *No*

P: *No, you're allowed to touch this, it's ok*

Example 2:

P speaks about the hearts, R smiles and reaches for it and P hands the box to R

R: *No*

P: *No, you can hold it, I don't mind*

Example 3:

R is holding the moon box. P used *never* in previous sessions to explain Deechee which boxes were forbidden.

R: *Never*

P: *No, no, good. This is for you*

Non-Adjacent Negation Types

- **Truth-func. question:** Truth-functional questions, as the name indicates, are questions that refer to or ask for some state of affairs being or not being the case. When used by our participants they often contained suggestions to possible answers. Open truth-functional questions, that is questions which do not already suggest an answer or a set of possible answers, albeit being the norm within the experiments are typically not listed in our table due to a lack of negative words within them.

Example 1:

P: *Is that a square? Yeah, no?*

Example 2:

P: *Is that a heart? No?*

- **Truth-func. negation:** Truth-functional negation is supposed to capture all kinds of truth-functional negation which are not *truth-functional denials*. Truth-func. negation is in this sense a residual class that captures all non-adjacent truth-functional utterances, be they negative assertions, suggestions, speculations, or guesses about state of affairs, which are in essence truth-functional. Also negative normative assertions such as the one in example 2 below count as a member of this class. Normative assertions are assertions about rules, laws or general practices in society such as “Thou must not kill” (law), “When driving a car one must stop in front of a red traffic lights” (rule), or referring to social practices in Italy, “When you greet somebody it is common to give the person two kisses, one on each cheek” (social practice).

Example 1:

P: *My heart beats. Have you got one? [1.5s] No robot heart maybe? Maybe not.*

Example 2:

In the context of explaining round traffic signs:

P: *They will tell you 30, which means you mustn't go any faster than 30.*

Example 3:

P: *Which one didn't we look at? We didn't look at the moon.*

- **Neg. persp. question:** Negative perspective questions, together with positive perspective questions, are the counterpart to *perspective dependent denial* on the side of the robot utterances. As is the case with the latter, they encompass questions, where the truth of the answer depends on either the knowledge (epistemic), the ability, the physical perspective of the agent, or any other state of affairs which can be only judged by the agent that is addressed. In example 2 below, for example, only the addressee can decide whether he or she is hungry or not. These questions either contain a lexical negative such as *no* or a grammatical one such as *don't*.

Example 1 (epistemic):

P: *Do you remember the moon? Don't you remember the moon?*

Example 2 (physical perspective):

P: *Can you see the squares? No? Ok*

Example 3 (“biological perspective”):

P is speaking about lollipops

P: *No? You're not feeling very hungry today?*

- **Neg. persp. assertion:** Negative perspective assertions, together with positive perspective assertions, can be found as counterpart to *perspective dependent denial* on the side of the robot utterances. All remarks on the dependencies of truth values mentioned under *neg. persp. questions* apply here as well. Furthermore perspectival assertions are captured here which are not about some perspectival aspect of the addressee but about such an aspect of the speaker (see example 1). Sometimes it's very hard to distinguish *neg. persp. assertions* from *neg. persp. questions* such as in example 2.

Example 1 (epistemic, regarding the speaker herself):

P: *Do you like the one with the squares? I can't remember*

Example 2 (epistemic, regarding the addressee)

P: You don't remember this one?

Example 3 (physical ability)

P tries to balance a box on Deechee's hand

P: No, I don't think you can hold it.

Example 4 (physical perspective)

P: Can you see it? You can't? Or you're looking ..

Example 5 (other ability)

P: Can you say moon? No

- **Neg. mot. question:** Negative motivational questions are questions that contain a lexical or grammatical negative. In the extreme case they consist of nothing else than this very negative, which has the intonational contour of a question. They may refer to the motivational state of the addressee directly (example 2). But they may also refer to stances or preferences (example 1) or intentional actions (example 3) that are indirectly linked to motivational states. In the direct case the question contains motivational or volitional verbs or constructions such as *want, like, feeling like, being keen on* etc. In the indirect case they do not contain such 'motivational markers' but clearly refer to the preferences of the addressee or her willingness to perform a certain action based on her current motivation. As it happens, example 3 contains the volitional word *want*, but a pragmatically equivalent question in the given context might be "Are you going to hold it? No?" which does not contain any such markers.

Example 1:

P: They're pretty. Don't you think hearts are pr.. . I think hearts are very pretty.

Example 2:

P: You wanna look at the circles again. Do you not like the heart?

Example 3:

P: Do you want to hold it? No?

- **Neg. intent interpret.:** Negative intent interpretations are assertions in which the participant interprets Deechee's intentional or motivational state utilizing lexical and/or grammatical negatives. Typically the semantics of these expressions is negative as well, i.e. the participants expresses that she thinks that Deechee does *not want* or *not like* either a particular object or does *not want* or *not like* to perform a particular action such as holding the box. Neg. intent interpret. are in some way a sub-type of *mot. dep. assertions*

(see below). Whereas *mot. dep. assertions* can refer to present, past, or future motivational states of speaker or addressee, *neg. intent interpret.* refer to the motivational states of the addressee only and only of his or her states right here and now. They are thought to have a special importance in early language acquisition in that toddlers might learn what we call here motivational words and their meaning by way of caretakers interpreting the toddler's emotional states or intents linguistically.

At times it can be hard to distinguish between *neg. mot. questions* and *neg. intent interpretations* as the main difference between the two types is the judgment as to whether the utterance is a proper question or not.

Example 1:

P: No you don't like circles do you like triangles (no transcription error here, the participant indeed merged two expected *you's* into one)

Example 2:

P: You don't want to hold the box. [1s] No

Example 3:

P: Do you like the circles box?

R: circles

[1.5s]

P: No? Ok

- **Mot. dep. assertion:** Motivation-dependent assertions are assertions that refer directly (example 1) or indirectly to the motivational states in the present (examples 3), past (example 1) or future of the speaker, or the addressee (example 2), which are not *negative intent interpretations*. This type is in this regard another residual class.

Example 2 is a borderline case as it is questionable how tightly a personal judgment about a mishap (reading variant 1) is linked to motivational states. Another way of interpreting this utterance (reading variant 2, socio-linguistic) is the following: The purpose of the utterance is to soothe the potential fear of the conversation partner, a fear that might be directed towards the socially dominant teacher, in case of the teacher being angered by the child/robot. Teachers have, by virtue of their social status, the power to hand out punishments. If one accepts this reading variant, there are expected emotional states with both the speaker and the addressee: expected fear on the part of the student (S) based on his expectation of anger on the part of the teacher (T). T expects S to (potentially) display fear because T expects S to (potentially) expect T to become angry, because S dropped the box. Therefore T counteracts the expected fear by issuing the utterance in order to convey to S that T will not become angry because of what happened. As the term "motivational" here also captures emotional states, a link to motivation would be given under this reading.

Example 1:

P: *And I think the square you didn't like*

Example 2:

P: *Don't worry, not serious (when R drops the box)*

Example 3:

P: *Squares [1.5s] I don't like squares, I think they are boring.*

- **Prohibition:** (Linguistic) prohibition only occurs in the prohibition scenario. It encompasses occurrences of negation whose function is to keep Deechee from touching forbidden objects. Sometimes such an utterance taken in isolation does not indicate that its function is prohibitive, as for example in example 2, which looks rather like a *truth-func. negation*. But, in context, when looking at the video recording, it becomes clear that the utterance is used as prohibition. The prohibitive utterance can or cannot be accompanied by the participant physically restraining Deechee's arm movement.

Example 1:

P: *No, no, you're not allowed to touch* (no physical restraint on the part of the participant)

Example 2:

P: *No, you're not holding it, but you can look at them* (no physical restraint)

Example 3:

P: *No* (P pushes Deechee's arm away)

- **Disallowance:** Disallowance is similar to *prohibition* but, in contrast, captures those utterances that express general (negative) rules. In this sense disallowance utterances are more detached from the here and now of the interaction than *prohibitive* utterances. Whereas prohibitive utterances are always triggered by a current action on part of the robot, *disallowances* can or cannot accompany such an action. It can be tricky to clearly distinguish the two types from each other and possibly there is no clear-cut boundary. But there seems to be an important difference between *stating a (negative) rule* on one hand and *uttering a prohibition* with the purpose of stopping an agent's actions at that very moment on the other. The question, that the coder has to ask herself is: Is this utterance meant to act upon Deechee immediately or is it rather the expression of a (general) rule. We observed that both may happen more or less at the same time by uttering a prohibition followed by the statement of a negative rule.

Example 1:

P: *You can't have this one* (Deechee is neither being restrained when this is uttered nor shortly afterwards)

Example 2:

P: *You can't touch the moon* (Deechee is not even trying to touch the moon at the time of utterance)

Example 3:

P: *You're not allowed to touch the circles* (This utterance was uttered at times when P was restraining Deechee's arm as well as when not restraining it)

- **Neg. promise:** Negative promises are those negative utterances in which participants commit themselves not to do certain things (any more) *in the future*. This commitment is often triggered by a negative reaction of Deechee. "Promise" is actually a slightly too strong term as our category is supposed to capture all kinds of future commitments by the participants - also commitments whose force is weaker than that of a promise. In the examples 1 and 2 what is actually being said is the following: "We won't play with X any more". As the participants are in our setup the ones who decide what is played with, this utterance amounts to "In the future I won't pick up X any more".

Example 1:

P: *You don't like the circles, no? Ok, we won't play with the circles.*

Example 2:

P: *Alright, I'm not gonna force you.*

Example 3:

P: *Ok, we won't play with that one then.*

- **Neg. tag question:** Negative tag questions are negative grammatical constructions that are attached to the end of the utterance. They consist of the negated auxiliary verb of the main clause, if there is one, plus a personal pronoun (see example 2: *can* [main clause] → *can't you* [tag question]). As can be seen in the examples 1 and 3 the main clause does not always contain the non-negated form of the negated auxiliary verb in the tag question, but putting it there wouldn't make a semantic difference to the utterance (ex. 1: Oh you do like that, don't you, ex. 2: But you did like the circles box, didn't you). Tag questions, negated or non-negated, are not proper questions but are attached to assertions. The negation is purely grammatical and, as far as we know, does not serve any

of the functions that the other negatives in our taxonomy serve. Yet as they are distinct grammatical constructs, they are very easy to spot.

Example 1:

P: *Oh you like that, don't you?*

Example 2:

P: *You can say square for me, can't you?*

Example 3:

P: *But you liked the circles box, didn't you?*

Negation types that can be both adjacent and non-adjacent

- **Quoted negation:** In the case of quoted negation, the negative part of the utterance belongs to a part of reported speech, which, if written down, could be quoted or would constitute indirect speech. The speech reported can either be the participant's or Deechee's.

Example 1:

P: *I said 'no'. Not this* (uttered in a prohibitive situation)

Example 2:

P: *No, you don't like it. You said you didn't like the squares.*

Example 3:

P: *What you're saying Deechee? No? Ok*

- **Mot. dep. exclamation:** See explanation of the same type in the section on the coding of robot utterances.

Example 1:

P: *Clever boy, I didn't even need to say the name!*

Example 2:

P: *Oh we go back to the crescent moon then I think you quite .. oh dear, oh no!*

- **Neg. imperative:** Negative imperatives are a residual class for all those imperatives which are neither *prohibitions* nor *disallowances*. The two latter types cover all imperative negatives which are linked to the prohibition task set out in the experimental

instructions, that is, that Deechee must never touch the forbidden objects. Example 2 below was not conceived of as a form of *disallowance* as it was judged to be more general than utterances that are tightly linked to the prohibition task. In the case of the example this judgment is supported by the fact that the participant used different and more specific prohibitive utterances prior to this negative, which referred specifically to the particular situation at hand (the prohibition task).

Example 1:

P: *Oh you're holding that very nicely. Don't throw it away*

Example 2:

P: *You can't have it!*

(uttered while P is

P: *It's no good, it's no good putting a face like that* restraining Deechee)

Example 3:

P: *No, don't say 'done'*

END OF SECTION FOR 2ND CODER

Fused Negation Types

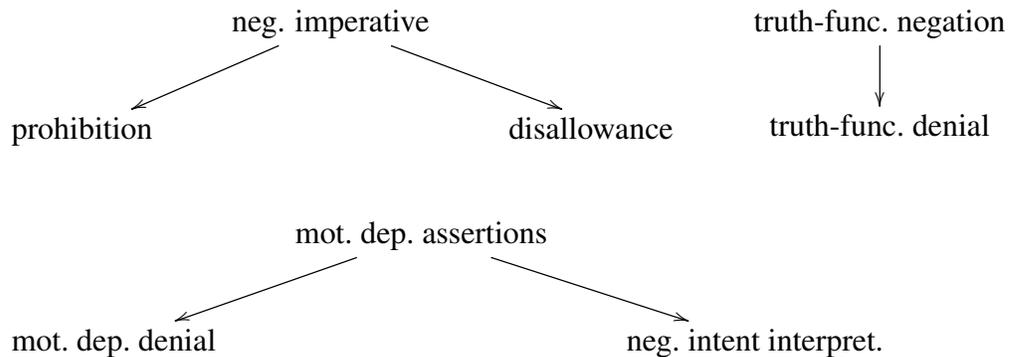
The following types were fused in order to reduce the total number of negation types observed amongst the human utterances.

$$\left. \begin{array}{l} \text{Neg. epistemic assert.} \\ \text{Neg. persp. assert.} \end{array} \right\} \text{Neg. persp. assert.}$$
$$\left. \begin{array}{l} \text{Neg. epistemic question} \\ \text{Neg. persp. question} \end{array} \right\} \text{Neg. persp. question}$$

Apart from an overall reduction of the number of negation types the indicated types were fused into the indicated types because the resulting type subsequently matches the type *persp. dep. denial* on the robot's side as first pair parts.

Super-/Sub-Types

The following types stand in a super-/sub-type relationship to each other, that is, the ones listed at the end of an arrow are more specialized than their super-type. All utterances that are instances of the sub-type are therefore also instances of the super-type.



For the sake of simplicity of the coding scheme one might argue that it would be advantageous to eliminate these sub-types. Yet in order to keep the human negation types as synchronized with the robot negation types as possible, we decided to maintain *truth-func. denial*. Moreover this is also one of the negation types listed in Pea's taxonomy. Another argument for maintaining this type is the circumstance that it indicates a genuine interaction between participant and robot, whereas the super-type *truth-func. negation* is more detached from the actual interaction both in terms of adjacency and in terms of the topic. The same argument can be made for the relationship between *mot. dep. assertions*, *mot. dep. denial*, and *neg. intent interpretations*.

Furthermore *neg. intent interpretations* are linked to one of our hypotheses and are thought to play a central role in the acquisitions of emotional/affective words.

With regards to the three types *neg. imperative*, *prohibition*, and *disallowance*, we decided not to eliminate the two subtypes, as *prohibition* is tightly linked with the prohibitive task and shows that actual linguistic prohibition took place, which is not necessarily the case given a *neg. imperative*.

Acknowledgments

Large parts of the work that this document forms part of was conducted within the EU Integrated Project ITALK (“Integration and Transfer of Action and Language in Robots”) funded by the European Commission under contract number FP-7-214668.

References

- [1] F. Förster. *Robots that Say ‘No’: Acquisition of Linguistic Behaviour in Interaction Games with Humans*. PhD thesis, University of Hertfordshire, September 2013.
- [2] F. Förster, J. Saunders, and C. L. Nehaniv. Robots that Say ‘No’. Affective Symbol Grounding and the Case of Intent Interpretations. *IEEE Transactions on Cognitive and Developmental Systems*, 2017. ISSN 2379-8920. doi: 10.1109/TCDS.2017.2752366.
- [3] F. Förster, J. Saunders, and C. L. Nehaniv. Robots Learning to Say ‘No’: Prohibition and Rejective Mechanisms in Acquisition of Linguistic Negation. *ACM Transactions on Human-Robot Interaction*, submitted.
- [4] R. Pea. The Development Of Negation In Early Child Language. In D. R. Olson, editor, *The Social Foundations of Language & Thought: Essays in Honor of Jerome S. Bruner*, pages 156–186. W. W. Norton, New York, 1980.
- [5] J. R. Searle. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, Cambridge, 1969.