

# Grounding a Social Robot's Understanding of Words with Associations in a Cognitive Architecture

Thomas Sievers<sup>1</sup><sup>a</sup>, Nele Russwinkel<sup>1</sup><sup>b</sup> and Ralf Möller<sup>2</sup><sup>c</sup>

<sup>1</sup>*Institute of Information Systems, University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany*

<sup>2</sup>*CHAI-Institut, Universität Hamburg, 20354 Hamburg, Germany*

**Keywords:** Human-Robot Interaction, Cognitive Architecture, Word Associations, ACT-R, ChatGPT.

**Abstract:** Social robots and humans need a common understanding of the current situation in order to interact and solve tasks together. They should know what the other one is talking about and refer to the same things. Word associations can help to find a common conceptual ground by enabling the robot to learn an association model of a human counterpart with regard to certain words and take them into account for its actions. This grounding of abstract words and ideas helps to constrain possible meanings. A model of a cognitive architecture connected to a social robot stores and processes chunks of memory from a language game between the robot and a human. The robot gives two words and keeps a third in mind. The human is asked to name a word associated with the two given words. In this way, an association model of the conceptual contexts of the human interaction partner is created. The dialog parts of the robot are generated with ChatGPT from OpenAI. An ACT-R model analyzes the data received from the robot, searches for suitable associations already in memory and, if applicable, provides feedback on these associations preferred by the human.


## 1 INTRODUCTION


Social robots that interact with humans and solve tasks together with a human partner must have some kind of model of the world, the situation, the task to be solved and the person with whom they are interacting. They must speak “*the same language*” and be mutually aware of what they are talking about. In addition, we need robots that can cope with complex and dynamic settings. Robots must be able to understand and take into account any human-related actions and behaviors as well as dynamic changes in the environment. They need a *mental model* of the situation.


Kambhampati introduced the term *human-aware AI* systems to refer to aspects of intelligence that enable successful collaboration between people (Sreedharan et al., 2024). This also includes modeling the mental states of humans in the loop. An algorithm-centric approach should transition to a human-centric perspective, which could improve human trust and thus the acceptance of social robots in human-robot interaction (HRI).

Flexible, grounded dialog systems are needed for autonomous social robots that enable an intelligible dialog with human partners. There are many linguistic ways to express concepts, and the ambiguities in language cause ambiguity and uncertainty for the dialog partner, as a single word can have different meanings. In addition, dialog participants may have different language skills. People generally have different experiences and associations with language and therefore use it in different ways. Language games can be a way to overcome these challenges as they represent a sequence of verbal interactions between two agents in a specific environment. Language games integrate unfamiliar words or expressions in a specific context, thereby constraining possible meanings (Steels, 2001). An example of different chains of association and understanding could be *team, grass : soccer* instead of *team, grass : tennis*.

Cognitive architectures, with their ability to use general concepts inspired by the human brain and create mental models based on human cognitive abilities, can be used to add a “human component” to robotic applications (Werk et al., 2024). A combination of robot sensing and data processing with such an architecture offers the possibility to deal with real-world information from the robot in cognitive models. Cre-

<sup>a</sup>  <https://orcid.org/0000-0002-8675-0122>

<sup>b</sup>  <https://orcid.org/0000-0003-2606-9690>

<sup>c</sup>  <https://orcid.org/0000-0002-1174-3323>

ating such models enables the robot to better understand the mindset of a human partner or to act in a way that is perceived as more natural by humans. Cognitive architectures refer both to a theory about the structure of the human mind and to a computational realization of such a theory. Their formalized models can be used to flexibly react to actions of the human collaboration partner and to develop situation understanding for adequate reactions. ACT-R (Adaptive Control of Thought - Rational) is a well-known and successfully used cognitive architecture (Anderson et al., 2004).

Using ACT-R's chunk system to store and process word associations from a language game could provide human-like associative behavior. Since the acquisition of initial word association data through a survey of people would be very effortful, the use of a large language model (LLM) from OpenAI's Generative Pretrained Transformer (GPT, commonly known as ChatGPT) (OpenAI, 2024) represents a practicable option for generating the initial associations. In the language game, one word is associated with two others. These chains of three words are stored in the ACT-R model via chunks. In this way, the cognitive model learns word associations and stores them in memory for later re-association of similar strings of words. The words linked by associations form an association model as a special kind of mental model for the conceptual interpretation of the language game player. Feedback from the model can influence the system prompt for ChatGPT to achieve convergence of word associations among players. If the players of the language game were a human and a social robot, the robot could develop and use an idea of the associative contexts of the human with this model. By storing predefined memory chunks as an a-priori knowledge in declarative memory, it is even possible to narrow down the topic area from which the words to be associated should originate. This leads to our hypothesis:

**Hypothesis.** Building an association model using ACT-R's chunk system enables a social robot to achieve human-like associative understanding.

In the following, we explain our ideas on this approach, provide an insight into the ongoing work and summarize initial findings.

## 2 RELATED WORK

Word associations are useful for generating creative combinations of related words (Gross, 2016). Semantic coherence, on the other hand, is seen as a metric related to point-by-point mutual information, or the ability to use the broader context of a story or sentence

and the semantic relationships between words to aid understanding and interpretation of spoken and written language (Mimno et al., 2011; Silverman, 2013).

Semantic coherence can be used to distinguish machine-generated text from human-generated text. Bao et al. proposed an end-to-end neural architecture that learns semantic coherence of text sequences for this purpose (Bao et al., 2019). Hüwel et al. used semantic interpretations of utterances as a basis for multi-modal dialog management in a speech understanding component for situated human-robot communication (Hüwel et al., 2006). In combination with gestures they generated the most likely semantic interpretation of the utterances and an interpretation with respect to semantic coherence.

Significant word associations that are unique to a document were used as a method for automatically composing poems using this document as inspiration (Gross, 2016). In an earlier approach Gross et al. achieved the semantic coherence of new poems by using semantically connected words (Gross et al., 2012).

The association of meanings to words in manipulation tasks on objects was investigated by Krunić et al. (Krunić et al., 2009). A word-to-meaning association could be learned even without consideration of a grammatical structure and led to the robot's own understanding of its actions. Rasheed et al. investigated the learning and understanding of abstract words with respect to the symbol grounding problem, focusing on the development of cognitive processes in robots (Rasheed et al., 2015). They elucidated concepts of grounded cognition with respect to the representation of abstract words using the robot's sensorimotor system.

Birlo et al. used a cognitive system based on ACT-R for their concept of robot self-awareness in an embodied robotic system (Birlo and Tapus, 2011). They focused on the representation of internal states of the robot, which processed external states from the real world and created its own interpretation of what it perceived. This robot self should be able to intervene in processes that control the robot.

Words can be understood as propositions and associated words can improve subsequent processes. Zhang et al. showed how an LLM could be controlled with a probabilistic propositional model (Zhang et al., 2022).

In this work, we use an LLM to generate word associations for a social robot, and a cognitive architecture connected to the robot to store and process human associations to these words.

### 3 METHODS

We thought of a language game in which a third word is to be associated with two related words. A robot names two words to its human teammate, to which it and the human should associate a third word. We used GPT-4o to create the conversational parts of the robot. The LLM was instructed via system prompts to understand the rules of the game and everything we needed it to output.

Our aim was for the robot to learn the conceptual interpretation of certain words of the human with whom it is playing this game. To do this, it stored the respective word associations that the human mentioned for the two given words in the memory of the cognitive model as a chain of these three words. If two of these three words occurred again in the course of the language game, the association model remembered the combination and sends the word that the human thought of in this context to the robot application.

For our studies we created a robot application for the humanoid social robot Pepper (Aldebaran, United Robotics Group and Softbank Robotics, 2024). In a dialog with humans, this robot application forwarded utterances of the human dialog partner as input to the OpenAI API, which returned ChatGPT’s answer as response. With each API call, the entire dialog was transferred to the GPT model. This allowed the model to constantly ‘remember’ what was previously said and refer to it as the dialog progresses. The text returned by the API was forwarded to the robot’s voice and tablet output. The OpenAI API provides various hyperparameters that can be used to control the model behavior during an API call. To obtain consistent responses and exclude any randomness as far as possible we set the value for *temperature* to 0.

Since ACT-R offers the technical possibility of integrating a cognitive model with bidirectional communication into a robot application and thus also into the control of an LLM (Sievers and Russwinkel, 2024), it should be possible to influence the system prompts and thus the interpretation of an LLM output via an association model stored in ACT-R’s declarative memory. To establish a remote connection from the robot application to ACT-R, the remote interface – the *dispatcher* – was used. The ACT-R core software is connected to this dispatcher to enable access to its commands. It accepts TCP/IP socket connections that allow clients to access these commands and provide their own commands for use. Since a real robot is not absolutely necessary for testing the system, we mainly used the robot emulator of the Android Studio development environment.

Fig. 1 shows the setup for the bi-directional con-

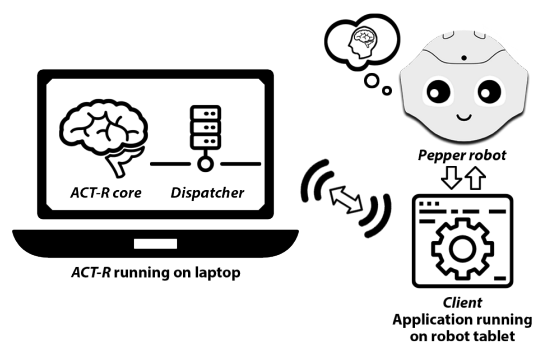


Figure 1: Connection between ACT-R / Dispatcher and the robot application.

nection between ACT-R and the client application on the robot. The dispatcher acts as a kind of server and is necessary to establish a remote connection between the robot application as a client and ACT-R.

#### 3.1 Prompting the LLM

We used prompts with Zero-shot prompting (Brown et al., 2020) for the system role to instruct GPT-4o to execute the tasks as a completion task. The system prompt for the LLM consisted of explanations on how to play the language game and instructions to output the words for which a word association was sought and the associated word, separated by commas in square brackets. The associated word should be in round brackets, for example [sun, beach, (summer)].

For testing and development purposes, we added a list of words to the system prompt from which the LLM could choose to form the triples for association in the language game. This list was dynamically expanded to include the words that the human associated. In this way, we had a manageable number of possible words for the association. This was helpful for testing and detecting effects of using the ACT-R model as opposed to not using it. At the beginning of the language game the list contained the words “summer, sun, cactus, beach, waves, sea, wind, water, heat, towel”. For the ACT-R model, some of these words were stored as chunk triples in the declarative memory as a-priori knowledge in the test phase in order to be able to perceive suitable associations more quickly. Prompting the LLM and storage of contents in the declarative memory of ACT-R was carried out in German.

The word associations generated by the LLM were transferred from the robot application to the ACT-R model to search for associated words in the declarative memory, as outlined in Figure 2.

In case of an association stored in memory to the words specified in the language game, the robot re-

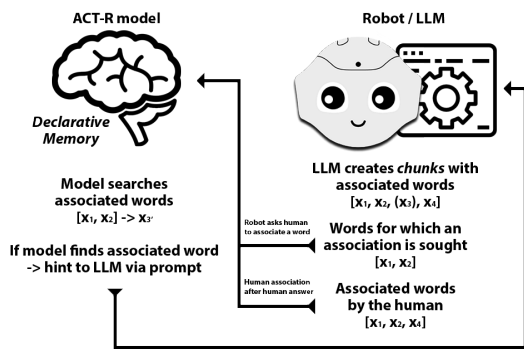


Figure 2: Transfer of word associations to ACT-R and possible feedback to the LLM.

ceived a hint about the word that had already been associated by the human. Such a hint would be added to the prompt as a group of three associated words, with the instruction that if two of the three words in this group of three are used, the third word must also be used in later rounds of the language game. In this way, information about human associations was passed on to the LLM.

### 3.2 Creating an Association Model with ACT-R

The ACT-R model we developed to guide attention to a representation of human word associations was quite simple for our initial tests. It used the chunk and production based concept of ACT-R with the goal buffer to exchange data with the robot application. Ideally, there should be no prefabricated chunks in declarative memory; all memory contents should be created as soon as they have been mentioned as associated words by the ChatGPT utterances. However, it turned out that the interaction time required to build up an association model with significant contents of associated words then became quite long. Therefore, to develop and test our system setup, we stored predefined memory chunks in the declarative memory as an a-priori knowledge and narrowed down the topic area from which the words for association should originate.

Perception, ideas and concepts of the real world must generally be reduced to abstract information for use in an ACT-R model. The abstraction is necessary because the ACT-R chunk system only allows a simple type of information storage in the form of strings. Our approach of word associations met this need very well, as we could store the words directly as strings in a chunk.

For each round of the language game, the model was given the first two words in the square brackets

– the words for which an association was sought – as a chunk. This chunk was stored in the declarative memory. In addition, this chunk was specified as the current goal in order to focus the model's attention on it. This started a search in ACT-R's memory for suitable word associations as shown in Figure 2. When the human partner had named their word association, the robot communicated its own association. The word named by the human was written as the fourth word in the series of associated words to complete the group of three associated words and stored as a memory chunk with the two initial terms in the declarative memory of ACT-R. Over time, this created an association model of the conceptual contexts of the human interaction partner. The order of the associated words within the chunk was irrelevant.

## 4 RESULTS

This is work in progress, but we already have some results that support our hypothesis and the underlying principle. The system prompt of the LLM was supplemented as desired with a three-word group from the declarative memory of the ACT-R model in the case of a suitable association, which ChatGPT should use completely instead of using two words from it with another third. Thus, over time, a list of three-word groups built up in the system prompt, which had to be taken into account by the LLM.

However, one difficulty lies in the precise formulation of the instructions in the system prompt so that the word associations suggested by the LLM always give priority to the three-word groups introduced via the cognitive model. Further optimizations and tests are necessary here.

## 5 CONCLUSION

A reference to words for which a human association is already known, for use, for example, in the prompt optimization of an LLM, could certainly also be implemented programmatically in a different way than with the help of a cognitive model. However, we were primarily interested in demonstrating the fundamental possibility of achieving this using a cognitive architecture such as ACT-R. Perhaps this paves the way for further applications for human-robot interaction that can use the special capabilities of cognitive models to simulate behavior similar to human cognitive abilities.

Initial tests suggest that the principle on which our hypothesis is based works and that it is possible to in-

fluence the LLM output in terms of matching associations between the game partners. However, further experiments are needed to optimize the ACT-R model and the generated system prompts for the LLM, as well as an evaluation of the quality of the association models in studies with different people.

In future work, language abstractions could also be used to describe, for example, situations or objects that a robot has to deal with. All kinds of real-world data that the robot can collect via its sensors could be used as a basis for this. Word associations could be a good way to find a common ground between robots and humans and a grounded cognition for the social robot.

## REFERENCES

- Aldebaran, United Robotics Group and Softbank Robotics (2024). Pepper. Technical report.
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., and Qin, Y. (2004). An integrated theory of the mind. In *Psychological review*, volume 111, pages 1036–1060.
- Bao, M., Li, J., Zhang, J., Peng, H., and Liu, X. (2019). Learning semantic coherence for machine generated spam text detection. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.
- Birlo, M. and Tapus, A. (2011). The crucial role of robot self-awareness in hri. In *Proceedings of the 6th International Conference on Human Robot Interaction*, pages 115–116.
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., and Amodei, D. (2020). Language models are few-shot learners.
- Gross, O. (2016). *Word Associations as a Language Model for Generative and Creative Tasks*. PhD thesis, Finland.
- Gross, O., Toivonen, H., Toivanen, J. M., and Valitutti, A. (2012). Lexical creativity from word associations. In *2012 Seventh International Conference on Knowledge, Information and Creativity Support Systems*, pages 35–42.
- Huwel, S., Wrede, B., and Sagerer, G. (2006). Robust speech understanding for multi-modal human-robot communication. In *ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*, pages 45–50.
- Krunic, V., Salvi, G., Bernardino, A., Montesano, L., and Santos-Victor, J. (2009). Affordance based word-to-meaning association. In *2009 IEEE International Conference on Robotics and Automation*, pages 4138–4143.
- Mimno, D., Wallach, H. M., Talley, E., Leenders, M., and McCallum, A. (2011). Optimizing semantic coherence in topic models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '11*, page 262–272, USA. Association for Computational Linguistics.
- OpenAI (2024). Transforming work and creativity with ai. Technical report.
- Rasheed, N., Amin, S., Hafeez, A., Raheem, A., and Shakoor, R. (2015). Acquisition of abstract words for cognitive robots [isi index x category]. *Jurnal Teknologi*, 72.
- Sievers, T. and Russwinkel, N. (2024). How to use a cognitive architecture for a dynamic person model with a social robot in human collaboration. In *CEUR Workshop Proceedings*.
- Silverman, L. B. (2013). *Verbal Semantic Coherence*. Springer New York, New York, NY.
- Sreedharan, S., Kulkarni, A., and Kambhampati, S. (2024). Explainable human-ai interaction: A planning perspective.
- Steels, L. (2001). Language games for autonomous robots. *IEEE Intelligent Systems*, 16(5):16–22.
- Werk, A., Scholz, S., Sievers, T., and Russwinkel, N. (2024). How to provide a dynamic cognitive person model of a human collaboration partner to a pepper robot. In *Society for Mathematical Psychology*.
- Zhang, H., Li, L., Meng, T., Chang, K.-W., and Broeck, G. (2022). On the paradox of learning to reason from data.