Enhancing Natural Language Understanding through Cross-Modal Interaction: Meaning Recovery from Acoustically Noisy Speech

Özge Alaçam
Department of Informatics
University of Hamburg, Hamburg, Germany
alacam@informatik.uni-hamburg.de

Abstract

Cross-modality between vision and language is a key component for effective and efficient communication, and human language processing mechanism successfully integrates information from various modalities to extract the intended meaning. However, incomplete linguistic input, i.e. due to a noisy environment, is one of the challenges for a successful communication. In that case, incompleteness in one channel can be compensated by information from another one (if available). In this paper, by employing a visual-world paradigm experiment, we investigated the dynamics between syntactically possible gap fillers for incomplete German sentences and the visual arrangements and their effect on overall sentence interpretation.

1 Introduction

In recent years, a growing body of literature has investigated how and to what extent cross-modal interaction contributes to natural language understanding. Human language processing system integrates information from various modalities to extract the meaning of the linguistic input accurately, but the contribution of cross-modality to a successful communication goes beyond it. It facilitates early reference resolution while the sentence unfolds and allows disambiguation even without realizing that another (linguistic) interpretation would be possible, e.g. see (Altmann and Mirković, 2009; Knoeferle et al., 2005; Tanenhaus et al., 1995). Furthermore, it also prepares the grounds for the re-construction of the meaning from noisy/missing input. When the environment is noisy, or the communication partner suffers from a motor or cognitive impairment, text completion/prediction becomes a crucial element of a communication. Instead of waiting for or requesting spoken input, combining the uncertain information from the linguistic channel with information from the visual one increases the fluency and the effectiveness of the communication (Garay-Vitoria and Abascal, 2004).

In this study, by conducting an experiment with human-subjects, we address the problem of compensating the incompleteness of the verbal channel by additional information from visual modality. Investigating how humans reconstruct the meaning from a noisy data provides insights about how to incorporate human-like processing into communication systems. The psycholinguistic experiments help us to understand baseline preferences and the underlying mechanism of gap construction processes for meaning extraction. This capability for multi-modal integration can be a very specific yet crucial feature in resolving references and/or performing commands for i.e. a helper robot that aids people in their daily activities.

2 Meaning Recovery

The task of extracting meaning from a noisy input has been widely addressed by uni-modal approaches. In a uni-modal way, re-construction can be guided by e.g. morphological, syntactic, and semantic properties. In that case, a probability of a syntactic category in a certain context can be obtained from a language model (Asnani et al., 2015; Bickel et al., 2005). For example, using N-grams is a popular method for this task since they provide very robust predictions for local dependencies. However, their power is less when it comes to dealing with long-range dependencies. On the other hand, as several studies (Mirowski and Vlachos, 2015; Gubbins and Vlachos, 2013) show, a language model employing the syntactic dependencies of a sentence brings the relevant contexts

Proceedings of the 22nd Nordic Conference on Computational Linguistics (NoDaLiDa), pages 272–280
Turku, Finland, 30 September – 2 October, ©2019 Linköping University Electronic Press

closer. Using the Microsoft Research Sentence Completion Challenge (Zweig and Burges, 2012), Gubbins and Vlachos (2013) have showed that incorporating syntactic information leads to grammatically better options for a semantic text completion task. Semantic classification (e.g. ontologies) and clustering can also be used to derive predictions on the semantic level for meaning recovery. However, when it comes to the description of daily activities, contextual information coming from another modality would be more beneficial, since linguistic distributions alone could hardly provide enough clues to distinguish the action of *bringing a pan* from *bringing a mug*, which is a crucial difference for e.g. helper robots.

Cross-modal integration of two modalities can be addressed by various methods in a range from simply putting all features from both modalities together and then train a model to learn the associations, to more complex structures, e.g. relating uni-modal features from several modalities on a conceptual level by using common representations. Considering that the task of meaning extraction may benefit from not only low-level but also high-level knowledge representations, one meaningful method would be to utilize a triplet notation, consisting of $(argument, relation\_type, predicate)$ where $relation\_type$ is one of a predefined set of accepted relations, such as AGENT or THEME while $Predicate$ and $Argument$ are tokens of the input sentence. Within this framework, the reconstruction of content words can be formalized as recovering/predicting the predicates or arguments of a sentence. To put it simply, a sentence like *"the woman carries ...."* can be formulated into two triplets; $(woman_{\mathrm{i}}, \text{AGENT}, carry)$ and $(unknown_{\mathrm{i}}, \text{THEME}, carry)$. Here the task is to determine the unknown entity which has directly related to the $carry$ action and indirectly to the agent $woman$. In case the contextual information provided by the visual environment contains additional information (e.g. a scene that depicts a woman with a grocery bag), the missing part can be successfully filled.

Salama et al. (2018) address the problem of incomplete linguistic input referring to daily environment context by utilizing a context-integrating dependency parser. Their focus was to recover content words like nouns, adjectives and verbs given the contextual features (e.g. object properties, spatial relations among the objects or thematic roles). The results indicate that giving a strong influence to contextual information helps to fill a majority of gaps correctly.

While re-construction of content words is mostly about finding out either the argument or the predicate based on the relation between each other, re-construction of grammatical words is to determine the relation between argument and predicate. Furthermore, re-construction of grammatical words could be more challenging since they tend to occur with higher frequencies than the content words, yielding a very small type/token ratio (i.e. weaker a collocational relationship) that makes the reconstruction of them based on only linguistic information more difficult. Although this is beyond the scope of the current paper, it should be noted that a full-fledged cross-modal meaning recovery system is dependent on a success of visual relation extraction component as well. The state-of-art computer vision systems can be considered more effective to extract spatial relations among object and object properties compared to relations between the actors and their actions.

3 Situated Language Comprehension in a Noisy Setting

The noise in communication could be originated from various channels and sources. First of all, it can be a linguistic noise (e.g. spelling mistakes, complex attachments), or visual ambiguities (e.g. clutter of the environment, occlusions) or an acoustic noise.

The issues of how to comprehend noisy linguistic input and reconstruct the intended meaning have been addressed by both psycholinguistic and computational line of research (e.g. (Levy, 2011, 2008)).

According to a noisy-channel account, that mainly focus on linguistic noise, the sentence comprehension mechanism integrates all the information (at the syntactic, semantic and discourse level) from the existing words and use this linguistic evidence to predict the missing parts and infer the possible meaning (Gibson et al., 2013). Several studies have shown that in case of higher degrees of syntactic complexity, humans tend to choose an interpretation which is in line with the concurrent visual information or general world knowledge, even though this interpretation requires to accept grammatically unacceptable syntactic structures (Johnson and Charniak, 2004; Christianson et al., 2001; MacWhinney et al., 1984). Cunnings (2017)'s study on language learners also indicated when the perceiver processes (syntactically) noisy linguistic input, the other linguistic and non-linguistic constraints are prioritized compared to syntactic ones.

Based on noisy-channel framework, Levy (2008) proposes a probabilistic model of language understanding regarding situations where there are uncertainty about word-level representations. He addresses the problem in two different levels; a global inference that can be reached after processing the entire input, and incremental inference that is formed (usually) word-by-word the sentence unfolds. The main contribution of the proposed method is that it takes into account the prior and posterior probabilities calculated based on both linguistic and non-linguistic evidence, including e.g. the expectations about speaker's grammatical competence or about the environmental condition that can hinder the speech signals.

Gibson et al. (2013) describes language understanding as rational integration of noisy evidence and semantic expectations. In their study, they test their predictions by conducting reading experiments, in which mostly the prepositions in the sentences were altered (by deletion or insertion) keeping content-word same across conditions. For example, an ungrammatical sentence *"The mother gave the candle the daughter"* can be easily treated as plausible by inserting *to* before *"the daughter"*. The higher prior probability of the latter version of the sentence compared to that of the former one pulls the sentence meaning towards itself.

4 Negation Processing

One interesting question regarding the task of meaning recovery is how to recover a meaning communicated with a sentence that involves unclear negated statement.

Since negation is considered as a higher order abstract concept, it has its own uniqueness as a grammatical category. Identifying the scope and focus of negation is one of the challenging issues that gets particular attention from the NLP community (e.g. SEM 2012 shared task, Morante and Blanco (2012)). From a psycholinguistic perspective, the core discussion lies around whether both negated and actual situation of content is simulated or only the actual one. However, regardless of how this process happens, the literature agrees on that sentences containing negation are harder to interpret than affirmative sentences (Orenes et al., 2014; Khemlani et al., 2012; Kaup et al., 2006; Lüdtke and Kaup, 2006; Carpenter and Just, 1975; Clark and Chase, 1972).

It has been conclusively shown that a negative sentence is processed by first simulating the positive argument. For example, after reading a negative sentence *"The bird was not in the air"*, a response to image that depicts *a flying bird* was faster than to a image of *a bird at rest*, (Zwaan, 2012). In addition to an overall processing difficulties that negation entails, it has been also shown that it is only integrated into the sentence meaning at a later point (Lüdtke et al., 2008).

On the other hand, there are also some evidence that indicates that when negation is supported by right contextual support, the positive arguments is no longer need to be represented, yielding faster verification compared to no-context situations (Tian et al., 2016; Dale and Duran, 2011; Nieuwland and Kuperberg, 2008).

5 Experiment

This study focuses on humans' preferences for the reconstruction of unclear sentence parts (in German) by using visual-world paradigm. Moreover, the effect of contextual information on situated reference resolution and on gap re-construction has been also manipulated by restricting the affordances of the locative object in one of the visual arrangements.

Gibson et al. (2013) list four different criteria of language processing system that have an impact on meaning recovery; (i) how close the literal sentence is to the plausible alternative, (ii) what kind of change is involved (insertion or deletion), (iii) the expectations about the corruption (noise-rate), and (iv) the plausibility of implausible sentences based on context or speaker-based. Basically by keeping all the criteria described by Gibson et al. (2013) constant, we focus, in this experiment, on obtaining prior probabilities of three types of grammatical words; two common preposition of location (*on* and *next to*) and negation particle (*not*). All sentences are syntactically plausible regardless from which focus-of-interest gap filler is used and their semantic plausibility is dependent on the information coming from the visual-world. Here in this current paper, we focus more on how meaning recovery is affected by negation, instead of detailed discussion into negation processing. Thus we kept the focus of negation constant among conditions, and the scene has been designed to have low referential competition (i.e. there are two tables in the scene, making the decision a binary task instead of a multinomial one).

The task is simply, hearing a sentence that communicates the intended meaning and process it and extract the meaning as close as possible to the intended one (Shannon, 1948). The goals that need to determined are;

- the re-construction of the gap-word
- full-sentence interpretation ("which object needs to be moved, and where to")

5.1 Participants

20 students (native speakers of German) participated in the experiment (*Mean age* = 23.8, *SD* = 3.1). They were paid or given a course credit to participate. The entire experiment took approximately 45 minutes for each participant including the familiarization and instruction sessions.

5.2 Material

Linguistic Material. In their complete form without a gap, all sentences have the same structure except the negation/preposition part (NEG/PP) as given below. The sentences start with a verb in an imperative form preceding an object (*NP*) and a prepositional phrase that specifies the goal location (*PP*). Then the sentence continues with a disfluency (*umm*) and a repair/complement part consisting of a negation or one of the two preposition of location. Our focus-of-interest gap fillers are (*nicht (not), auf(on), neben (next to)*. These are chosen since they can fill the same position interchangeably.

- Stell den Becher auf den Tisch, umm [auf/nicht/neben] den blauen.
  put the mug on the table, umm [on/not/next to] the blue one.

The choice of filler-word given the visual information determines which object that the repair/complement part is attached to. In this setting, the repair/complement may have three different syntactic roles; referring back to the OBJECT which is the mug (with *not*), referring back to the ADVERBIAL which is the table (with both *on* and *not*) or providing new complementary ADVERBIAL which is an another mug (with *next to*). Due to filling different roles, all possible linguistic interpretations require different parsing results. In all cases, the object referred to in the repair/complement part shares either the property (e.g. blue) or the object class (e.g. mug) with the target object or location.

Pre-prossessing of the spoken material. The sentences were recorded by a male native speaker of German at a normal speech rate. Intonational differences between different linguistic entities have been found to have a significant effect on reference resolution (Coco and Keller, 2015; Snedeker and Trueswell, 2003). Therefore, we avoided unequal intonational breaks that may bias the interpretation. The breaks separating phrases were equalized.

A constant background noise (a sound recording from a restaurant) was added to an entire spoken sentence with the Audacity software [1]. In order to mask the target word completely, the volume of the NEG/PP part starting from the interjection (*umm*) was gradually decreased till the end of the gap-word. Concurrently, the volume of the background noise was increased during this segment.

Scenes. In order to accommodate the intended interpretation(s) and to eliminate others, the object properties and their spatial relations among each other have been systematically manipulated for each scene. Although, other many more different visual arrangements could be possible, for the sake of systematicity, we have narrowed our visual conditions down to five scene arrangements, see Figure 1. Scene-1 conveys all possible interpretations for all the focus-of-interest fillers. Scene-2A and Scene-2B allows only *on* and *not*. However, the availability of the location signaled by *on* is occupied by another object in Scene-2B. The last two visual arrangements allows only one interpretation; signaled by *not* in Scene-3A and by *next to* in Scene-3B. The number of objects in the scenes was limited to eight and one additional object has been used in Scene-2B. For each visual condition, six different visual scene were designed resulting 30 main-trial scene. The 2D visual scenes were

[1]http://www.audacityteam.org/ - retrieved on 21.11.2018

created with the SketchUp Make Software [2].

To prevent participants' associating the focus-of-interest gap fillers with a particular visual arrangement, additional slightly changed sentences and scenes were introduced as filler items.

5.3 Procedure

Using the visual-world paradigm, we presented participants visual scenes with accompanying spoken sentences. We employed a simple "look-and-listen" experiment following clicking tasks to get user's preferences. The experiment started with filling out the written consent and demographic data form. Afterwards, the instructions were given in written format, preceding the 3 familiarization trials.

The participants were instructed that multi-modal stimuli always contain some background noise, and at one point, one word will be impossible to hear. Then they are expected to choose the gap-filler word and click on the target object and location communicated in the sentence. It was told that target object is always located on the middle stand and needs to be moved to one of the white trays on the scene located on various places. In order to be able to separate the task of identifying object from identifying location, target objects and locations are presented in a specific layout.

The stimuli were displayed on an SR Eyelink 1000 Plus eye tracker integrated into a 17 monitor with a resolution of 1280 x 1024 pixels. We utilized a total of 53 visual displays with accompanying spoken utterances (3 familiarization, 30 test trials and 20 fillers). Each trial began with a a drift correction and the presentation of a simple fixation cross for 2 sec, located at the middle-bottom of the screen. Afterwards, a 5 sec of visual preview before the onset of the spoken sentence was given. The preview gives a comprehender time to encode the visual information in advance of the linguistic information being presented. So, visual attention is intended to be free of recognizing the objects of the visual context during language processing. Then, the spoken sentence was presented accompanying the visual stimulus. A trial ended 2 sec after the offset of the sentence. Participants were asked to examine the scene carefully and attend the information given in the audio. The order of stimuli was randomized for each participant.

After the sentence is completed, the scene disappears and the participants are asked to click their preference for the gap position among five options. They were also informed about that the gap could be accurately filled by more than one option. These options are "nicht (*not*)", "neben (*next to*)", "auf (*on*)", "mit (*with*)", and "den/das/die (*the*)[3]". Whereas the focus-of interest gap-fillers are syntactically acceptable for the gap position, two other grammatical words were provided among the options as distractors; *mit (with)* and *den/das/die (the)*. In German, the preposition *mit (with)* requires a dative object, therefore the gender of the following article should be different than the nominative or accusative forms of the article in the repair/complement part. Furthermore, *den/das/die (the)* can be understood in two ways, either as a repetition of the definite article or as a relative pronoun. In the former case, as a gap filler, it does not provide any additional information. On the other hand, the lack of relative clause verb makes the second interpretation unacceptable.

After the preference has been explicitly made, the scene appears again so that the participant can click on the scene to answer two questions respectively; *which object is the target?* and *where is the target location?*.

Although a time-course analysis of fixated items and locations when the sentence unfolds is very relevant in understanding the underlying mechanisms of language processing, in this paper, we narrow down our scope into participants' explicitly made choices after each multi-modal stimulus.

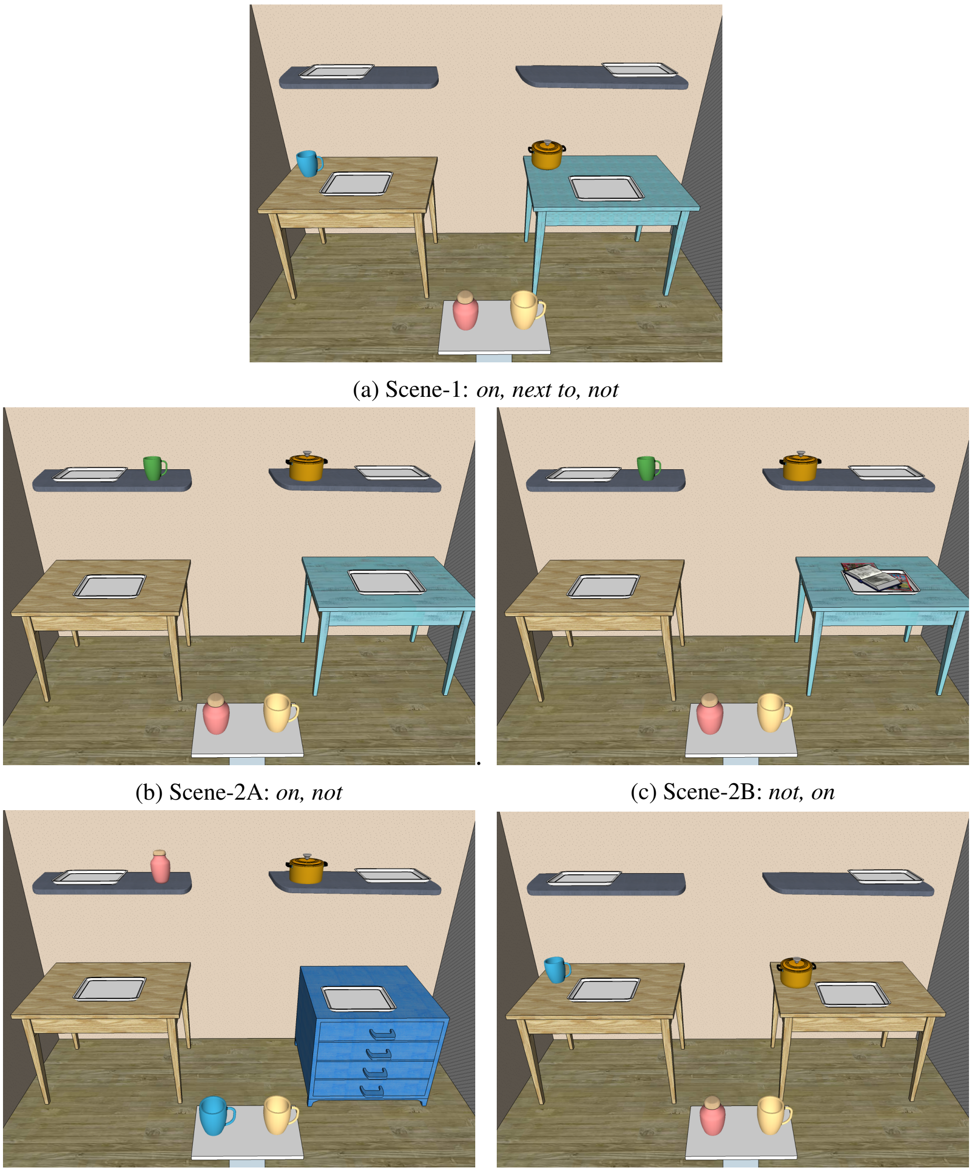
Our hypothesizes are listed as

- Syntactically all gap positions require one insertion to correctly accommodate the intended meaning, however unlike preposition of locations, negation operation is considered as a high-level (abstract) concept. Therefore, the sentences with *on* and *next to* should be more easier to disambiguate, therefore more preferred compared to ones with *not*.

- Conceptual information, i.e. target location's being not available as illustrated in Figure 1e may force to change the interpretation, accordingly the preference, from *on* to *not*.

[2]http://www.sketchup.com/ - retrieved on 06.05.2018

[3]The respective article/relative pronoun was shown among options in accordance with the grammatical gender of the noun it modifies



(a) Scene-1: *on, next to, not*

(b) Scene-2A: *on, not*

(c) Scene-2B: *not, on*

(d) Scene-3A: only *not*

(e) Scene-3B:only *next to*

Figure 1: Sample scenes that illustrate five different visual manipulations

6 Results

In this section, the gap-filler preferences and the global sentence interpretation by analyzing accurately chosen object given their preference have been reported. Figure 2 shows the distribution of the preferences for each visual condition. In total, participant preferences for 600 trials (20 participant * 30 scene) were taken into account.

Gap Construction Preferences. The visual condition Scene-1 was designed to analyze user's general tendency among three focus-of interest gap fillers, since all are equally plausible w.r.t. the visual context. In this condition, *next to* was preferred more in 43.9% of the trials compared to *on* (30.8%) and *not* (20.6%). The other distractor options were preferred only in 3.7% of trials. The results of a Friedmans ANOVA indicated that preference rates for the three focus-of interest gap fillers significantly differs ($\chi^2(2) = 8.95,\ p < .001$). Wilcoxon tests were used to follow up this finding. It seems that this difference among groups is originated by the difference between *not* and *next to* ($z - score = -2.23,\ p < .0167$), *with a Bonferroni correction*).

The analysis on whether the participant could choose the object and the location in line with their explicitly made preference also demonstrated that all target objects are correctly identified. This result is highly expected, considering that the

PP/NEG part does not carry relevant information for the target identification in this visual setting. On the other hand, regarding the location, while 100% of the participants, who chose *next to*, correctly determine the target location, which is in line with their preference, this accuracy score is 90.9% for *on* and it drops drastically to 71.4% for *not*.

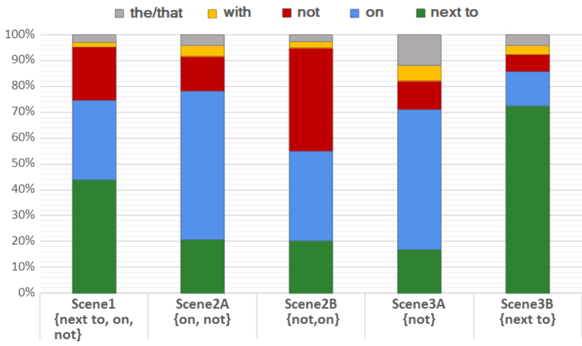


Figure 2: *Preference distributions regarding each visual condition*

The Effect of Contextual Cues. Whether the availability of a location signaled by one of the possible gap fillers has an effect on the preferences has been investigated by a mixed-design ANOVA comparing the number of preferred option across two visual arrangements; Scene-2A and Scene-2B. In these conditions, *on* and *not* are the two only semantically plausible gap fillers. The proportion results indicated that when the two locations are equally available (Scene-2A), participants prefer more *on* as a gap filler (57.5%), and the option *not* was chosen in only almost 13.3% of the trials. On the other hand, while the targeted location referred by the sentence with *on* repair is occupied (Scene 2B), then the participants' tendency to prefer *not* increases considerably by 21%. The preference of *next to* stays almost the same across the conditions.

The results of the ANOVA indicated no main effect of the visual condition ($p > .05$). However, the main effect of Preference was significant ($F(2,38) = 8.642,\ p = .001$). In general, *on* has been preferred more compared to *not* and *neben*; ($F(1,19) = 10.92,\ p = .004$) and ($F(1,19) = 11.61,\ p = .003$) respectively. Regarding our research question, the interaction effect between the visual condition and the preference is the relevant one, and it displays a significant interaction ($F(2,38) = 7.79,\ p = .001$). This indicates that the preference tendencies significantly differed in Scene-2A and Scene-2B. To break down this interaction, contrasts were performed comparing each level of focus-of-interest preferences across two scene types. These revealed significant interactions when comparing *on* and *not*, ($F(2,38) = 18.98,\ p < .001$). Looking at the interaction graph in Figure 3, this suggests that when the target location signaled by *on* is occupied, participants looks for alternatives and ending up with only other available interpretation in line with *not*. Moreover, the contrast between *not* and *next to* was significant as well, ($F(2,38) = 5.30,\ p < .05$).

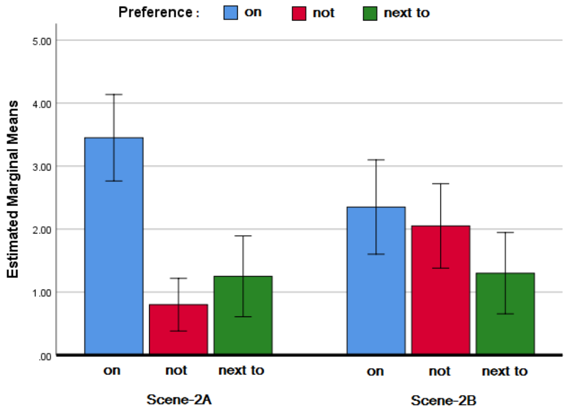


Figure 3: *Mean number of preferred focus-of-interest fillers across Scene-2A and Scene-2B*

Preferences under Restricted Conditions. The last comparison focuses on the cases, in which the visual arrangements and the properties of the objects only allow one interpretation. Scene-3A favors only the interpretation which is in line with the use of *not* as a gap filler. Yet, only in 53.3% of the trials correct option has been chosen as gap filler. Despite their conflict with the visual world, other gap fillers have been chosen in a considerable amount; *next to* (16.7%), *on* (10.8%), *with* (5.8%) and *the/that* (11.7%). On the other hand, Scene-3A syntactically allows only *next-to*. The results showed that in 72.5% of the trials, participants preferred *next to*. The comparison between the number of correct preferences across two visual condition revealed that on average, participants make more correct choices when they see Scene-3B (M = 4.35, SE = 1.72) compared to Scene-3A (M = 3.20, SE = .41), ($t(19) = -2.31,\ p < .05$).

7 Conclusion

In this study, by systematic manipulation of visual scene, we have obtained prior expectations regarding two locative prepositions and negation particle and we have also demonstrated how contextual cues pull an interpretation towards one side.

In order to accommodate different interpretations, five different visual arrangements have been utilized. Although our investigations into this area are still ongoing, the results could be a useful aid for developing models of situated natural language understanding that aims to account noisy data comprehension and meaning recovery. In this study, we particularly tried to put some spotlight into the special case "negation" as well.

The results indicate that when the visual world supports all interpretations, people have tendency to choose *next to* as a gap filler, that entails the repair/complement part referring to another object, which is not mentioned in the sentence before. Their second preference is to attach the repair part to the prepositional phrase (ADVERBIAL) by choosing *on* as gap filler. This selection also inherently assumes that the repair part is an affirmative statement. On the other hand, even in the cases where *not* is the only semantically plausible option as gap filler, the participants showed hesitance to choose it. This results are also in line with noisy-channel framework. A sentence with a gap is more harder to process compared to a complete sentence, since it requires at least two sub-tasks to be performed; predicting the gap-filler given the context and then confirming the inferred meaning. While the spatial relations like *next to* and *on* are easily graspable from an image, a negative statement requires additional operation to account for actual situation, that's why the listeners may prefer to override contextual expectations and stick to more easy-to-process one even it semantically, and sometimes syntactically creates a conflict (Ferreira, 2003). However, this preference (choosing *on* over *not*) still seems to be affected by the contextual cues like the availability of a target location.

It should nonetheless be acknowledged that the systematicity that we had to follow to single out all other effects becomes a limitation for generalization, thus further research is needed to better understand first the dynamics between the preference and the visual arrangements and second the dynamics between negation in detail and contextual cues. Moreover, none of the visual manipulations in this study was designed to address to explain the difference between choosing *next to* and *on*. Another set of experiments with reversed order; the first PP with a *next to* and the complement part with *on* would help us to gain some insights on this issue.

Acknowledgments

This research was funded by the German Research Foundation (DFG) in the project Crossmodal Learning, TRR-169.

References

Gerry TM Altmann and Jelena Mirković. 2009. Incrementality and prediction in human sentence processing. *Cognitive science*, 33(4):583–609.

Kavita Asnani, Douglas Vaz, Tanay PrabhuDesai, Surabhi Borgikar, Megha Bisht, Sharvari Bhosale, and Nikhil Balaji. 2015. Sentence completion using text prediction systems. In *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014*, pages 397–404. Springer.

Steffen Bickel, Peter Haider, and Tobias Scheffer. 2005. Learning to complete sentences. In *European Conference on Machine Learning*, pages 497–504. Springer.

Patricia A Carpenter and Marcel A Just. 1975. Sentence comprehension: a psycholinguistic processing model of verification. *Psychological review*, 82(1):45.

Kiel Christianson, Andrew Hollingworth, John F Halliwell, and Fernanda Ferreira. 2001. Thematic roles assigned along the garden path linger. *Cognitive psychology*, 42(4):368–407.

Herbert H Clark and William G Chase. 1972. On the process of comparing sentences against pictures. *Cognitive psychology*, 3(3):472–517.

Moreno I Coco and Frank Keller. 2015. The interaction of visual and linguistic saliency during syntactic ambiguity resolution. *The Quarterly Journal of Experimental Psychology*, 68(1):46–74.

Ian Cunnings. 2017. Interference in native and non-native sentence processing. *Bilingualism: Language and Cognition*, 20(4):712–721.

Rick Dale and Nicholas D Duran. 2011. The cognitive dynamics of negated sentence verification. *Cognitive science*, 35(5):983–996.

Fernanda Ferreira. 2003. The misinterpretation of noncanonical sentences. *Cognitive psychology*, 47(2):164–203.

Nestor Garay-Vitoria and Julio Abascal. 2004. A comparison of prediction techniques to enhance the communication rate. In *ERCIM Workshop on User Interfaces for All*, pages 400–417. Springer.

Edward Gibson, Leon Bergen, and Steven T Piantadosi. 2013. Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, 110(20):8051–8056.

Joseph Gubbins and Andreas Vlachos. 2013. Dependency language models for sentence completion. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1405–1410.

Mark Johnson and Eugene Charniak. 2004. A tag-based noisy channel model of speech repairs. In *Proceedings of the 42Nd Annual Meeting on Association for Computational Linguistics*, ACL '04, Stroudsburg, PA, USA. Association for Computational Linguistics.

Barbara Kaup, Jana Lüdtke, and Rolf A Zwaan. 2006. Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics*, 38(7):1033–1050.

Sangeet Khemlani, Isabel Orenes, and Philip N Johnson-Laird. 2012. Negation: A theory of its meaning, representation, and use. *Journal of Cognitive Psychology*, 24(5):541–559.

Pia Knoeferle, Matthew W Crocker, Christoph Scheepers, and Martin J Pickering. 2005. The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, 95(1):95–127.

Roger Levy. 2008. A noisy-channel model of rational human sentence comprehension under uncertain input. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '08, pages 234–243, Stroudsburg, PA, USA. Association for Computational Linguistics.

Roger Levy. 2011. Integrating surprisal and uncertain-input models in online sentence comprehension: formal techniques and empirical results. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 1055–1065. Association for Computational Linguistics.

Jana Lüdtke, Claudia K Friedrich, Monica De Filippis, and Barbara Kaup. 2008. Event-related potential correlates of negation in a sentence–picture verification paradigm. *Journal of cognitive neuroscience*, 20(8):1355–1370.

Jana Lüdtke and Barbara Kaup. 2006. Context effects when reading negative and affirmative sentences. In *Proceedings of the 28th annual conference of the cognitive science society*, volume 27, pages 1735–1740. Lawrence Erlbaum Associates Mahwah, NJ.

Brian MacWhinney, Elizabeth Bates, and Reinhold Kliegl. 1984. Cue validity and sentence interpretation in English, German, and Italian. *Journal of verbal learning and verbal behavior*, 23(2):127–150.

Piotr Mirowski and Andreas Vlachos. 2015. Dependency recurrent neural language models for sentence completion. *arXiv preprint arXiv:1507.01193*.

Roser Morante and Eduardo Blanco. 2012. * sem 2012 shared task: Resolving the scope and focus of negation. In ** SEM 2012: The First Joint Conference on Lexical and Computational Semantics–Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, volume 1, pages 265–274.

Mante S Nieuwland and Gina R Kuperberg. 2008. When the truth is not too hard to handle: An event-related potential study on the pragmatics of negation. *Psychological Science*, 19(12):1213–1218.

Isabel Orenes, David Beltrán, and Carlos Santamaría. 2014. How negation is understood: Evidence from the visual world paradigm. *Journal of Memory and Language*, 74:36–45.

Amr Rekaby Salama, Özge Alacam, and Wolfgang Menzel. 2018. Text completion using context-integrated dependency parsing. In *Proceedings of The Third Workshop on Representation Learning for NLP*, pages 41–49.

Claude Elwood Shannon. 1948. A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423.

Jesse Snedeker and John Trueswell. 2003. Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and language*, 48(1):103–130.

Michael K Tanenhaus, Michael J Spivey-Knowlton, Kathleen M Eberhard, and Julie C Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217):1632–1634.

Ye Tian, Heather Ferguson, and Richard Breheny. 2016. Processing negation without context–why and when we represent the positive argument. *Language, Cognition and Neuroscience*, 31(5):683–698.

Rolf A Zwaan. 2012. The experiential view of language comprehension: How is negation represented? *Higher level language processes in the brain: Inference and comprehension processes*, page 255.

Geoffrey Zweig and Chris JC Burges. 2012. A challenge set for advancing language modeling. In *Proceedings of the NAACL-HLT 2012 Workshop: Will We Ever Really Replace the N-gram Model? On the Future of Language Modeling for HLT*, pages 29–36. Association for Computational Linguistics.