Nontrivial Lexical Convergence in a Geography-Themed Game

Amanda Bergqvist
Uppsala University
Uppsala, Sweden
amanda.bergqvist@nordiska.uu.se

Ramesh Manuvinakurike and Deepthi Karkada
Intel Corp
United States
first.last@intel.com

Maike Paetzel
Uppsala University
Uppsala, Sweden
maike.paetzel@it.uu.se

Abstract

The present study aims to examine the prevalent notion that people entrain to the vocabulary of a dialogue system. Although previous research shows that people will replace their choice of words with simple substitutes, studies using more challenging substitutions are sparse. In this paper, we investigate whether people adapt their speech to the vocabulary of a dialogue system when the system's suggested words are not direct synonyms. 32 participants played a geography-themed game with a remote-controlled agent and were primed by referencing strategies (rather than individual terms) introduced in follow-up questions. Our results suggest that context-appropriate substitutes support convergence and that the convergence has a lasting effect within a dialogue session if the system's wording is more consistent with the norms of the domain than the original wording of the speaker.

1 Introduction

The human habit of mirroring other's choices of words could potentially provide a neat shortcut in the challenging task of building dialogue systems capable of understanding human language. Simply put, the dialogue system could nudge speakers to use words that are in its vocabulary by itself using those words in its output speech. The adaptation, known as *lexical entrainment* (mutual alignment) or *lexical convergence* (one-way adaptation) (Brennan, 1996; Beňuš, 2014), does not only apply to human–human interaction, but extends to human–computer interaction (Gustafson et al., 1997), as well as human–robot interaction (Iio et al., 2009). While natural languages offer innumerable ways of expressing the same idea (Furnas et al., 1987), a strategically designed system vocabulary could thus narrow down the range of words used by a human when speaking with an artificial partner.

In previous work, however, lexical convergence to a dialogue system has mostly been assessed in simple tasks, and the words suggested by the computer were close synonyms to the ones that the participant originally used. For humans, it might not make that much of a difference if a ticket is booked by saying "I'd like to go to" or "I'd like to travel to" (Gustafson et al., 1997). Results from Parent and Eskenazi (2010)'s study on a bus information system suggest that words that are frequent in day-to-day speech get entrained more often than less frequent "unnatural or harder" words. So, what if the substitutes proposed by the computer require more thought from the human than their initial phrasing, or do not come naturally to them?

In this paper, we aim to examine to what extent people imitate a dialogue system when the substitutions it suggests are nontrivial. We conducted an experiment using a cooperative two-player game in which people are asked to describe the location of countries on the world map. We hypothesized that human speech converges when the substitution requires minimal effort (changing between using *next to* and *borders*), but that convergence to cognitively straining substitutions (changing between *egocentric* and *cardinal* descriptions) is suppressed.

2 The RDG-Map Domain

We tested the lexical convergence in the context of a dialogue-based collaborative two-player game between a human and an unembodied female agent called Nellie (Paetzel and Manuvinakurike, 2019). The goal of the game is to locate as many countries as possible on a world map within the game time of 10 minutes. The human plays the role of the *Director* who receives target countries (cf. Figure 1) that s(he) needs to verbally describe to the agent in the role of the *Matcher*. The targets were a predefined set of countries in a fixed or-

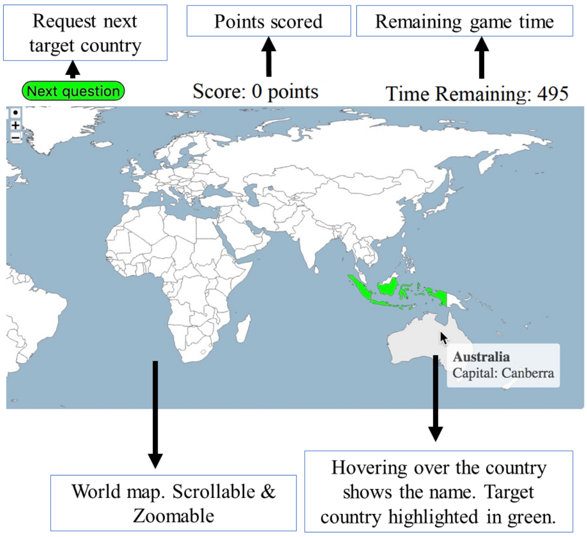

Request next target country
Points scored
Remaining game time
Next question
Score: 0 points
Time Remaining: 495
Australia
Capital: Canberra
World map. Scrollable & Zoomable
Hovering over the country shows the name. Target country highlighted in green.

Figure 1: Game interface, the director's view. The current target is Indonesia (in green). The director is hovering the mouse over Australia (gray) to see its name.

der and were selected to evoke spatial directions. While participants believed they were playing with an autonomous agent, Nellie was, in fact, remote-controlled by a researcher.

Each game was divided into three stages: In the *baseline stage* (targets 1^{st} to 4^{th}), the agent did not mention any directions and the operator registered the natural word choices of the interlocutor (*borders* or *next to* resp. *cardinal* or *egocentric*). In the *priming stage* (targets 5^{th} to 8^{th}), the agent made use of an opposing set of expressions in the form of follow-up questions. Below is a minimal example in which the agent tries to prime the speaker using cardinal directions:

EXAMPLE 1 - PRIMING:
HUMAN: Austria is directly above Italy.
AGENT: Is it to the west of Hungary?
HUMAN: West of Hungary, yes.
AGENT: Got it.

In the *post-priming stage*, the agent returned to its original speech pattern. This stage could later be used to understand whether participants continued to use the vocabulary suggested by the agent in the second stage, or whether they fell back to their original lexical choices. A longer dialogue excerpt is shown in Figure 2.

3 Substitute Words

Two main strategies can be used to make spatial references on a map: describing *general relations* between two countries and giving *directional descriptions*. General relations describe which countries border a certain country, but not which specific border they share. In this context, we identified *"A borders B"* and *"A is next to B"* to be simple substitutes that are interchangeable. Directional descriptions can be subdivided into *egocentric* (left, right, above, below) and *cardinal directions* (north, south, east, west). While bordering will always imply being "next to", the cardinal direction corresponding to, e.g., left, depends on the position in a global reference frame. Swapping between egocentric and cardinal directions is thus not a simple matter of one-to-one translation, but involves changing strategy and can be considered more challenging than changing from "borders" to "next to".

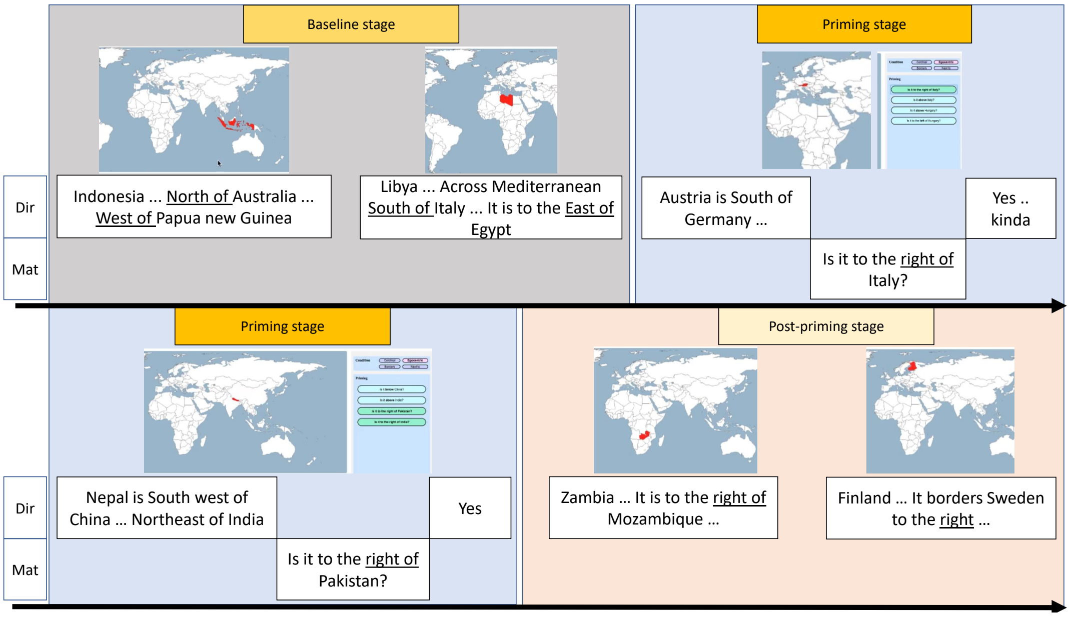
In contrast to most previous studies, we induced a swap of *referencing strategy* rather than a swap of *referencing terms*. In a study by Iio et al. (2009), participants adapt to the semantic framework that the system uses, not just individual terms, and Bell and Gustafson (2007) report a similar tendency in children playing a speech-enabled game. In our study, stimuli for *north* and *east* were thus expected to make players swap to *south* and *west* as well.

Previous studies mainly primed by swapping specific terms. When provoking a swap of terms, there are two options: the correction can be either embedded or exposed (Jefferson, 1987).

EXAMPLE 2 - EMBEDDED CORRECTION:
HUMAN: Austria is directly above Italy.
AGENT: I have selected the country north of Italy, got it.

EXAMPLE 3 - EXPOSED CORRECTION:
HUMAN: Austria is directly above Italy.
AGENT: By above, do you mean north?

Priming for a swap of referencing strategy allows for a third option: embedding members of the substitute referencing strategy without touching on the specific word used by the person. In Example 1, the agent does not mention the cardinal equivalent to *above*. Instead, it hints at its preference for cardinal directions by simply using them in its requests for further information. This makes for a smoother flow, as the conversation is actually progressing with respect to the goal of the game. In this study, primes for a referencing strategy were thus embedded in clarification requests.



Indonesia ... North of Australia ... West of Papua new Guinea

Libya ... Across Mediterranean South of Italy ... It is to the East of Egypt

Austria is South of Germany ...

Is it to the right of Italy?

Yes .. kinda

Nepal is South west of China ... Northeast of India

Is it to the right of Pakistan?

Yes

Zambia ... It is to the right of Mozambique ...

Finland ... It borders Sweden to the right ...

Figure 2: Excerpts from a sample conversation between a human Director (Dir) and agent Matcher (Mat) playing the game. In the baseline stage, the director uses cardinal descriptions. The director then faces questions from the matcher using egocentric descriptions. In the post-priming stage, the user converges to egocentric descriptions.

4 Experimental Design and Procedure

We conducted a study with a between-subject design in which the experiment group was subject to the more challenging swap of words between the directional relations, while the control group was subject to the simple swap between the words marking general relations. Except for the stimuli words, the setup was identical between conditions. In alignment with previous work, participants in the control condition were predicted to pick up the agent's lexical choices, while those in the experiment group were predicted to converge less.

Participants first rated themselves in comparison to the average person in skills and pastimes involving navigation and travelling. In order to assess whether participants had a preference for egocentric or cardinal directions, they were also asked to fill in the revised Lawton's Wayfinding scale (Lawton and Kállai, 2002). Before entering the game, players were randomly assigned to the experiment or the control group. The groupings within the experiment and the control group were based on the preference as determined in the baseline phase of the game, and players were assigned the condition opposite to their preference.

Participants were a convenience sample of 32 adult American native English speakers (age: $M = 34$, $SD = 9.38$; 45% female) who had never played the game before. All participants were recruited on Amazon Mechanical Turk and paid $3 upon completion of the experiment. In the experimental group (N = 17), 10 participants were exposed to cardinal directions and 7 to egocentric directions. In the control group (N = 15), 9 participants were exposed to "borders" and 6 to "next to". Participants rated themselves as being averagely experienced with reading maps and using a GPS, but less experienced than the average person in using a compass. On average, participants had a higher route strategy score ($M = 3.45$, $SD = 0.6$) than orientation strategy score ($M = 2.51$, $SD = 0.82$). Since egocentric directions are related to the route strategy, this shows that participants are overall more accustomed to using egocentric descriptions in their daily life.

Automatically generated speech-to-text transcripts of the dialogues were manually corrected. They were then parsed, and occurrences of keywords were automatically counted. In addition, transcripts were manually annotated for the usage of descriptive strategy and false descriptions were flagged by the annotators (Paetzel et al., 2020).

5 Results

In both the experimental and the control condition, participants' frequency of using the primed words increased during the priming stage of the experiment (cf. Table 1 and Figure 3). We performed a

		# words (frequency in %)				
Condition	**Stage**	**cardinal**	**egocentric**	**borders**	**next to**	**total # words**
cardinal	baseline	18 (15.25%)	108 (79.89%)	1 (0.63%)	4 (4.24%)	131
	priming	65 (52.81%)	73 (42.42%)	2 (1.05%)	7 (3.72%)	147
	post-priming	160 (65.4%)	116 (32.87%)	1 (0.26%)	5 (1.48%)	282
egocentric	baseline	34 (66.52%)	14 (18.4%)	7 (9.69%)	3 (5.39%)	58
	priming	41 (38.75%)	59 (53.47%)	8 (6.94%)	1 (0.84%)	109
	post-priming	98 (52.82%)	77 (40.11%)	8 (3.97%)	6 (3.1%)	189
borders	baseline	31 (37.87%)	50 (56.06%)	0 (0.0%)	5 (6.07%)	86
	priming	38 (38.31%)	29 (25.7%)	30 (30.9%)	4 (5.09%)	101
	post-priming	94 (37.98%)	79 (32.62%)	56 (26.26%)	9 (3.14%)	238
next to	baseline	29 (43.33%)	17 (35.6%)	3 (3.37%)	1 (1.04%)	50
	priming	27 (43.34%)	14 (21.34%)	5 (6.94%)	15 (28.38%)	61
	post-priming	80 (45.54%)	56 (33.84%)	26 (15.6%)	11 (5.03%)	173

Table 1: Instances of stimuli words (in absolute numbers and percentage) in player speech, grouped by condition and experiment stage. Cells representing the stimuli word(s) that a group was primed for are highlighted in green.

two-way ANOVA with the interaction stage (baseline, priming, post-priming) and the conditions (experiment: cardinal, egocentric; control: next to, borders) as independent variables.

The usage of the priming words “next to” and “borders” used for the control group was generally sparse. In the group primed for the word “borders”, the usage of the word increased significantly between the baseline and the priming stage, $p < .001$, and people continued using the word significantly more even in the post-priming stage, $p < .001$. In both the priming and the post-priming stage, the frequency of the word “borders” was significantly higher than in the same stage in all other three conditions. For the people being primed to use the words “next to”, we found a significant increase of the word usage during the priming phase, $p < .001$. However, the usage declines significantly after the priming stage, $p = .003$. During the priming stage, the usage of the word “next to” is significantly higher than during the priming stage in any other condition, while in the post-priming stage, it reaches the same level as in the other groups again.

In the experiment group, we found an increase of cardinal descriptions in the people primed to use the cardinal system. This increase is not significant between the baseline and the priming stage, $p = .15$, but becomes significant in the post-priming stage, $p = .009$. At the same time, the usage of egocentric descriptions in participants primed for the cardinal system declines between the baseline and the post-priming stage, $p = .012$. The group of people being primed to use the egocentric system slightly increase their usage of egocentric descriptions in the priming stage. This increase, however, is not significant, $p = .42$, and declines in the post-priming stage again.

Especially if a group converges towards the vocabulary of the dialogue system, it is relevant to examine whether communication suffers in other ways. If speakers comply with a computer by converging but commit errors because they are not accustomed to the proposed vocabulary, they may provide the computer with faulty information. However, in our system, we did not find a significant increase in the number of wrong descriptions given by participants in any condition. Similarly, we did not see an overall avoidance of giving directional descriptions in any of the conditions since the overall distribution between directional descriptions, size, or shape descriptions remained unchanged.

6 Discussion

As hypothesized, our results show that there was a statistically significant convergence of people’s vocabulary towards the vocabulary suggested by the agent in the control condition. This finding is in line with previous work and shows that *people are willing to adapt their vocabulary to an artificial agent even if substitute words are embedded in follow-up questions*, which is a weaker incentive for convergence compared to exposed corrections. Contrary to our expectations, however, we could also observe convergence in parts of the experimental group, specifically in the group exposed to cardinal directions. This finding is interesting as participants reported using egocentric directions more often in their daily lives, which would suggest they would be easier to adapt to than to the less common cardinal words.

A possible explanation for the higher convergence in the group naturally using egocentric descriptions lies in a core idea of lexical entrainment: conceptual pacts. According to Brennan and Clark

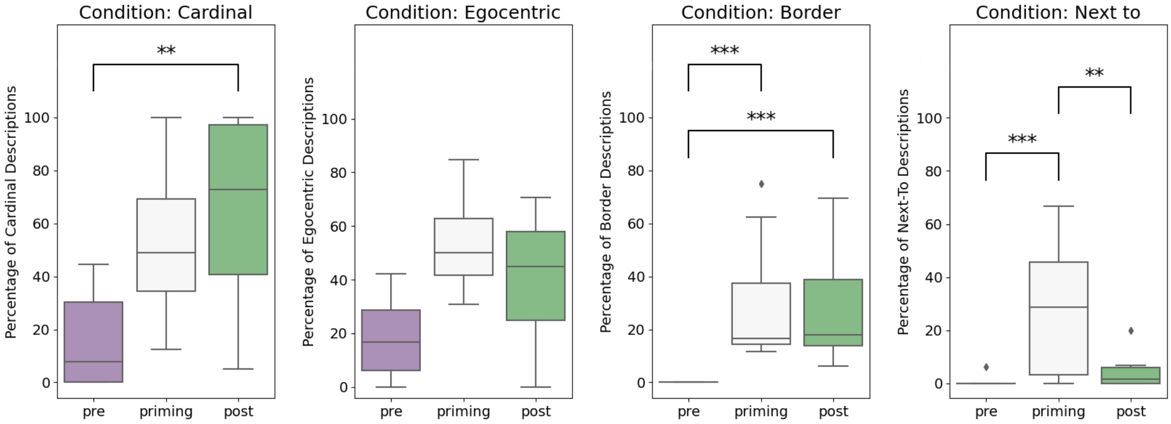


Figure 3: Visualization of the usage of the words participants were primed for in all four conditions. Significant differences are indicate with * ($p < .05$), ** ($p < .01$) and *** ($p < .001$)

(1996), entrainment is not merely a matter of repeating certain words, but rather a negotiation of a common reference system in a conversation. They suggest that in referring to an object with a certain word, the speaker is proposing a conceptualization of the object. In adopting the same word, the partner sends a message that they agree with the conceptualization. (S)he can also convey disagreement by rejecting the word and proposing a different one.

In the geography game, the high convergence of the participants who started with egocentric directions might reflect an acceptance of not just the cardinal words, but of the concept of referring to positions in a map by cardinal directions. Even though participants were, on average, more used to egocentric words, they obliged with the agent because they recognized the norm that links maps to cardinal directions. The lesser convergence of participants who started with cardinal directions may convey their disapproval of using egocentric directions in the given context. In accepting or rejecting the terms proposed by the computer, participants are thus not simply trying to or failing at facilitating the conversation. They are taking a stand as to whether the words proposed by the dialogue system make sense or not in the present context. Similarly, bordering is more commonly used to describe relations between countries and the convergence to the word "borders" was thus more lasting in the remainder of the conversation compared to the phrase "next to". *Our findings suggest that people will replace their first choice of words if the alternative is more reasonable in a given context but will reject the alternative if they find it inferior to their initial choice.*

In our study, we did not measure whether people found the translation between the egocentric and the cardinal system to be more difficult than the swap between "borders" and "next to", which reduces the conclusions we can draw when it comes to limits of lexical convergence due to cognitive load. In the future, we plan to conduct a larger experiment in which we measure the participant's cognitive load explicitly. With a larger number of participants per group, we hope to be able to analyze further whether the differences in convergence in the experimental conditions are, in fact, the indicator of a significant trend.

7 Conclusion

The results of the present study provide further support for lexical convergence and the persuasiveness of lexical convergence in human–computer dialogue. They also indicate that convergence is related to the semantic appropriateness of the system vocabulary. More specifically, people are more likely to adopt substitute words that belong in the given context. In this particular study, the players of a geography-themed game rejected egocentric descriptions but adopted cardinal directions, likely since they were deemed better fitted for describing the location of a country. If high levels of lexical convergence are to be attained, we thus suggest that the vocabulary of a dialogue system needs to be harmonized with the domain at hand.

Acknowledgements

The authors thank K. Georgila and G. Castellano for their input and infrastructural support.

References

Linda Bell and Joakim Gustafson. 2007. Children's convergence in referring expressions to graphical objects in a speech-enabled computer game. pages 2209–2212.

Štefan Beňuš. 2014. Social aspects of entrainment in spoken interaction. *Cognitive Computation*, 6:802–813.

Susan E. Brennan. 1996. Lexical entrainment in spontaneous dialog. In *International Symposium on Spoken Dialog*, pages 41–44.

Susan E. Brennan and Herbert H. Clark. 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22:1482–1493.

G. W. Furnas, T. K. Landauer, L. M. Gomez, and S. T. Dumais. 1987. The vocabulary problem in human-system communication. *Commun. ACM*, 30(11):964–971.

J. Gustafson, A. Larsson, R. Carlson, and K. Hellman. 1997. How do system questions influence lexical choices in user answers. In *In Proc. Eurospeech '97*, pages 2275–2278.

Takamasa Iio, Masahiro Shiomi, Kazuhiko Shinozawa, Takahiro Miyashita, Takaaki Akimoto, and Norihiro Hagita. 2009. Lexical entrainment in human-robot interaction: Can robots entrain human vocabulary? In *IROS*, pages 3727–3734.

Gail Jefferson. 1987. On exposed and embedded correction in conversation. In *Talk and Social Organisation*, chapter 4, pages 86–100.

Carol Lawton and János Kállai. 2002. Gender differences in wayfinding strategies and anxiety about wayfinding: A cross-cultural comparison. *Sex Roles*, 47:389–401.

Maike Paetzel, Deepthi Karkada, and Ramesh Manuvinakurike. 2020. Rdg-map: A multimodal corpus of pedagogical human-agent spoken interactions. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 593–602, Marseille, France. European Language Resources Association.

Maike Paetzel and Ramesh Manuvinakurike. 2019. "Can you say more about the location?" The Development of a Pedagogical Reference Resolution Agent". *Dialog for Good - Workshop on Speech and Language Technology Serving Society (DiGo).*

Gabriel Parent and Maxine Eskenazi. 2010. Lexical entrainment of real users in the let's go spoken dialog system. pages 3018–3021.