Title
Zipf's law of abbreviation and common ground: Past communicative success hampers the re-optimization of language

Permalink
https://escholarship.org/uc/item/9651z629

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

ISSN
1069-7977

Authors
Kuek, Jacob C
Ferdinand, Vanessa

Publication Date
2021

Peer reviewed

Zipf's law of abbreviation and common ground: Past communicative success hampers the re-optimization of language

Jacob C. Kuek (jacob.kuek@unimelb.edu.au)
University of Melbourne School of Psychological Sciences

Vanessa Ferdinand (vanessa.ferdinand@unimelb.edu.au)
University of Melbourne School of Psychological Sciences

Abstract

Zipf's Law of Abbreviation (ZLA) states that the more frequently a word is used, the shorter its length tends to be. This arises due to the optimal trade-off between competing pressures for accuracy and efficiency in communication, known as the Principle of Least Effort. Existing research has not focused on how individuals adapt their language use to remain optimal despite language change and whether social factors like common ground affect this. To investigate this, we replicated and extended the artificial language learning paradigm and communication game of Kanwal et al. (2017). We found participants were able to re-optimize their language use according to ZLA after a language change event, but this ability was hampered by common ground. This research identifies common ground as one potential cause for observed sub-optimalities in human languages and may have implications for understanding the dynamics of language change across communities where common ground varies.

Keywords: Zipf; Principle of Least Effort; optimality; language change; common ground

Introduction

Languages exhibit optimality in a remarkably large number of ways from their sound systems (Everett, Blasi, & Roberts, 2016; Everett, 2017), word length distributions (Zipf, 1935), and semantic systems (Kemp, Xu, & Regier, 2018), to the ways they categorize color (Zaslavsky, Kemp, Regier, & Tishby, 2018), kinship relationships (Kemp & Regier, 2012), and numeral systems (Xu, Liu, & Regier, 2020). Each one of these observations is presumably the outcome of an adaptive process which created a goodness of fit between linguistic structures and their environments of use. However, relatively little is known about the processes that give rise to and maintain optimality in language.

In this paper, we address one particular type of optimality in language, known as Zipf's Law of Abbreviation or ZLA (Zipf, 1935). ZLA identifies an inverse relationship between the frequency of words in human languages and their associated word length and posits that this relationship is due to the Principle of Least Effort (Zipf, 1949), where the competing communicative pressures of accuracy and efficiency (Kanwal, Smith, Culbertson, & Kirby, 2017; Kirby, Tamariz, Cornish, & Smith, 2015; Piantadosi, Tily, & Gibson, 2012; Tamariz & Kirby, 2016; Urbina & Vera, 2019) leads to the functional optimization of form-meaning mappings within a lexicon (Piantadosi, 2014; Piantadosi, Tily, & Gibson, 2011). When neither or only one of the competing pressures of accuracy and efficiency exist in a language, the distribution of the lexicon does not respect ZLA (Kanwal et al., 2017; Chaabouni, Kharitonov, Dupoux, & Baroni, 2019).

A key experimental study by Kanwal et al. (2017) provided direct evidence linking accuracy and efficiency pressures on individual language learners to ZLA optimality in the structure of their lexicon. Participants underwent an artificial language learning paradigm, mapping names (*zopekil* and *zopudon*) to two novel objects with one object being displayed more frequently. Importantly, both objects could also be referred to with the same clipped form, *zop*, making the referent ambiguous. They found that participants learned to refer to the more frequent object as *zop* and to the less frequent object with its respective long form, which is the optimal mapping for this task because it maximizes accuracy and minimizes the average length of the words being used. However, optimal mappings emerged only when participants were under *both* pressures for accuracy and efficiency. In other words, language users optimized form-meaning mappings in accordance with the Principle of Least Effort.

Nearly all human languages exhibit ZLA: it is a statistical language universal (Bentz & Ferrer Cancho, 2016; Koplenig, Meyer, Wolfer, & Mueller-Spitzer, 2017; Mahowald, Dautriche, Gibson, & Piantadosi, 2018; Pechenick, Danforth, & Dodds, 2017; Piantadosi et al., 2011) and remains stable over time (Pechenick et al., 2017). However languages do change over time (Sankoff, 2018) and the frequencies of referents change with their relevance to a speech community, making the words associated with important topics increasingly frequent themselves (Karjus, Blythe, Kirby, & Smith, 2020a, 2020b). ZLA's stability over time implies that languages have been able to re-optimize their lexicons as the frequencies of referents change. What factors facilitate or hinder the re-optimization of language in the face of change?

One factor that could affect language re-optimization is *common ground*: the shared knowledge between communicative partners (Brown-Schmidt & Duff, 2016). The construction of common ground has been found to drive communicative success (Huggett, Peña, Sulik, & Spike, 2020), as communication can only occur when language users mutually accept that certain words are associated with certain meanings (Smith, 2014; Tamariz & Kirby, 2016). Common ground has also been linked to optimality in a maze game communication task by Castillo, Branigan, and Smith (2015), showing that participants with multiple partners converge on optimal strategies better than pairs do, who instead maintain the first strategy that works for them even if it is not optimal.

In this study, we investigate the effect of common ground

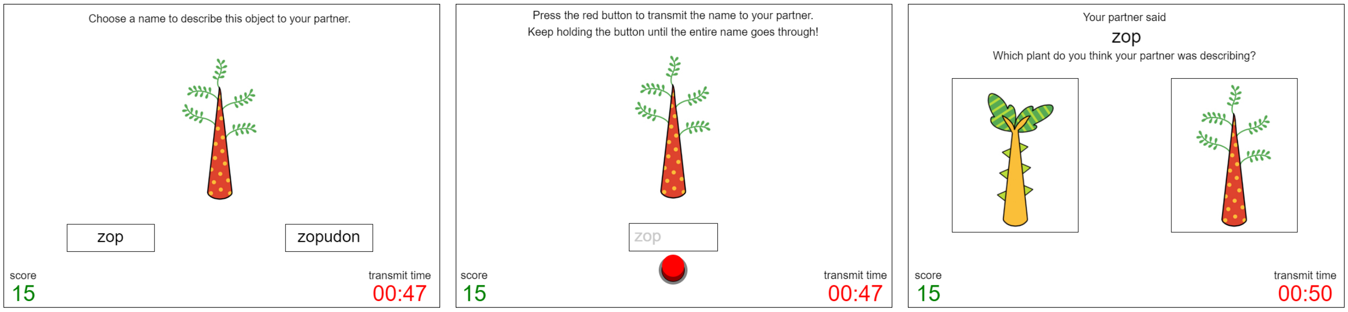

Figure 1: Screen shots from the communication game. Left: Director's view when choosing which word to send. Middle: Director's view when transmission button appears. Right: Matcher's view when guessing which object the director had.

on the re-optimization of language with respect to ZLA. We adopt the same language learning paradigm and communication game as developed by Kanwal et al. (2017) and add a novel second round to the task in which the frequencies of the two objects are reversed. *Are participants able to re-optimize their lexicon to the new object frequencies?* To manipulate common ground, participants either played a new partner in round two or the same partner again. *Do participants with common ground use less optimal lexicons after the frequency shift, than participants with no common ground?*

Experiment

Participants

Participants were 75 undergraduate students from the University of Melbourne (56 females, 17 males, and 2 non-binary people). Ages ranged from 18 to 36 (M_{age} = 20.07, SD_{age} = 3.52). There were 28 English-speaking monolinguals, with the remaining 47 participants speaking at least one other language in addition to English. Condition allocation was on an alternating basis in the order of who clicked our study link. Of those participants who completed the study, 38 were assigned to the no common ground condition and 37 were assigned to the common ground condition. If participants did not complete the study, their data was not saved.

Stimuli

Stimuli were the same as in Kanwal et al. (2017), consisting of two novel objects (see Figure 1), two novel 7-letter words (*zopekil*, *zopudon*), and one 3-letter short form (*zop*). Two drawings of a robot, which differed in fill color, represented the participant's partner(s) in the game (see Figure 2). All images are distinguishable under the three main forms of colour blindness according to Color Oracle (colororacle.org).

Procedure

Participants took the experiment online via Google App Engine. Each participant completed two rounds of the task. Each round was divided into two phases: a training phase and a communication game.

Training phase The training phase consisted of 32 trials. On each trial, one object was displayed with one word. The object appeared alone for the first 700ms and with the word for the next 2000ms. One object was displayed on 24 trials and the other was displayed on 8 trials. Each object was deterministically mapped to one of the long forms and the mapping was randomized across participants. Each object appeared with its long form on half of its trials and the short form on the other half. For example, object A would appear with *zopekil* 12 times and *zop* 12 times, and object B would appear with *zopudon* 4 times and *zop* 4 times. In round two, the frequencies were reversed (swapping the 12s and 4s above). A communication game followed each training phase.

Communication game The communication game consisted of 64 trials in which the two partners alternated roles called *director* and *matcher*, such that each participant completed 32 trials in each role. On each trial, the director viewed an object along with the two words that had occurred with it during the training phase. The director then chose one of the words to send to the matcher using a transmission button. The button sent the word one letter per second, establishing a cost for sending the long form. Participants were instructed to minimize their cumulative transmission time, which was shown on screen. Then the matcher received the word and was told to select the object that the director was referring to. If the matcher guessed correctly, the director and matcher both received one point. Participants were instructed to maximize their score, which was also displayed on screen. The frequency of each object in the communication phase matched its frequency in the training phase.

Word learning algorithm Our procedure differed from Kanwal et al. (2017) in that we paired each human participant with a word learning algorithm, rather than a human partner. This choice allowed us to use a smaller participant pool and reduced the complexity of the data by ensuring that all participants played a partner who was using the same strategy. The algorithm stored the history of word-object mappings on each communication trial and selected words or objects on the basis of the most recent trial (see Figure 2 for the algorithm's decision tree). This design allowed the participant to easily drive the communication system to various word-object mappings. Participants were informed at the beginning of the

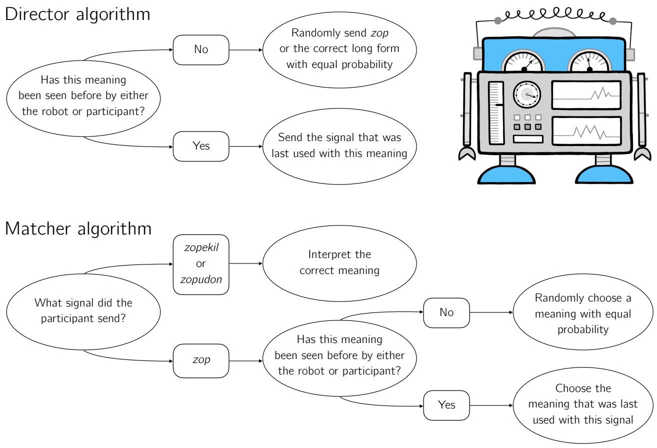

Figure 2: The director and matcher algorithms used by the robot partner in the communication game in the experiment.

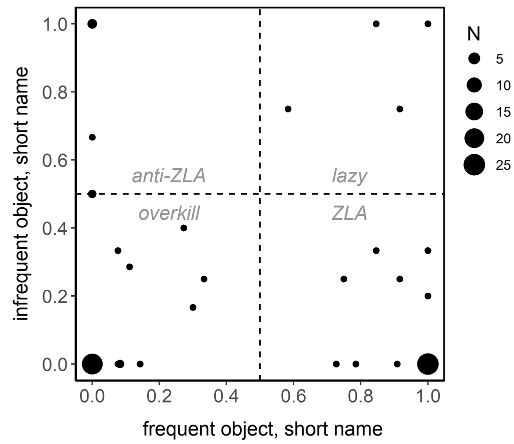

Figure 3: Proportion of trials in which the short name was used with the frequent object (x-axis) and the infrequent object (y-axis). Compare to Figure 3 in Kanwal et al. (2017).

experiment that their partner would be a word learning robot.

Condition Two conditions manipulated common ground. In the *common ground* condition, participants were told that they would be playing the same robot partner in both of the rounds and consequently played the same algorithm whose memory included all trials in rounds one and two. In the *no common ground* condition, participants were told they would be playing two different robots. In this case, the algorithm was re-initialized between rounds, causing it to lose all memory of the trials in round one.

Results

All analyses were conducted in R, with statistical significance being determined at the level of $p < .05$.

Experiment Replication

Round one of our experiment successfully replicated the *combined effects* condition in Kanwal et al. (2017) in a human-robot communication paradigm. Figure 3 shows participant behavior in terms of the four main strategy types: *ZLA*, using the short word for the frequent object and the long for the infrequent, *lazy*, using the short word for both objects, *overkill*, using the long word for both objects, and *anti-ZLA*, using the long word for the frequent object and the short for the infrequent. As in Kanwal et al. (2017), *ZLA* was the most common strategy, followed closely by the *overkill* strategy.

For purposes of comparability, we fit the same logistic regression model specified in Kanwal et al. (2017) (their Table 2) to our data from round one. The binary outcome variable was short form usage (rather than long form usage). Fixed effects were object frequency (frequent or infrequent), trial number (trials $1-32$ re-scaled to $0-1$ to help with model convergence), and their interaction. Random effects were by-participant slopes and intercepts for trial number and object frequency.

As in Kanwal et al. (2017), we find a significant interaction effect between object frequency and trial (reference value = infrequent; reference value = trial 1; $\beta = 2.026, SE = 0.828, z = 2.446, p = .014$), meaning participants were more likely to use the short form for the frequent object and were even more likely to do so as the trial number increased. We also replicate their non-significant effect of trial number ($\beta = 0.585, SE = 1.368, z = 0.427, p = .669$) meaning that participants were not more likely to use the short form on the basis of the trial increase alone (e.g., due to boredom with the task). The one difference we find is a significant effect of object frequency ($\beta = 5.593, SE = 1.688, z = 3.313, p < .001$) where theirs was marginally non-significant at $p = 0.079$. This means our participants also had a preference for using the short form for the frequent object, irrespective of trial number. Overall, this was a successful replication, demonstrating that human participants develop ZLA optimal strategies with A.I. communication partners, and validates our methodology.

Experiment Extension 1: Language Change

We found that participants were able to re-optimize their word usage to the new object frequencies in round two. To understand this re-optimization process, we compare participants' average word length in round two to what it would have been if they had continued to use their same mapping from round one. For example, take the following mapping used by a participant on round one: on the 8 trials with object B they used *zop* 2 times and *zopekil* 6 times, and on the 24 trials with object A they used *zop* 24 times and *zopudon* zero times. The average word length in round one is $\bar{S} = \sum_{m \in M} \bar{s}_m p(m)$ = 3.75 symbols, where $\bar{s}_m$ is the average word length for object m ($\bar{s}_A = 3$, $\bar{s}_B = 6$) and $p(m)$ is the probability of each object ($p(A) = 0.75$, $p(B) = 0.25$). To calculate this participant's average word length if they had continued to use this mapping in round two, we substitute the values for $p(m)$ in round two ($p(A) = 0.25$, $p(B) = 0.75$) and obtain 5.25 symbols. If we then find that the participant's actual round two average word length is less than 5.25 symbols, this suggests

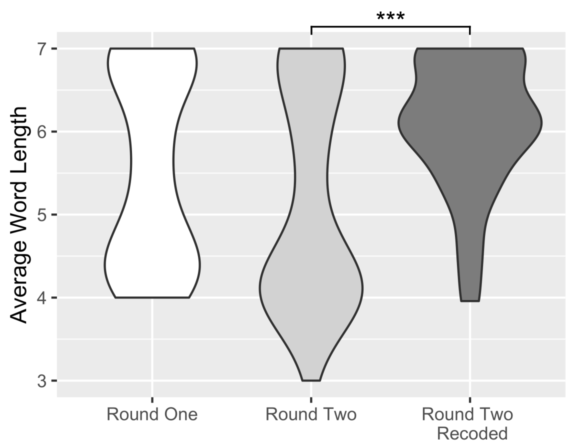

Figure 4: Participants re-optimize their word usage when frequencies change in round two. Average word length in round one (left), round two (middle), and in round two if participants had not re-optimized their word usage (right).

Fixed Effects	Beta	S.E.	z	p
Intercept	-2.308	0.613	-3.764	< .001
Frequency	0.638	0.714	0.894	.371
Trial	-0.185	0.832	-0.222	.824
Condition	-0.875	0.878	-0.996	.319
Round	0.374	0.469	0.799	.425
Frequency : trial	0.954	0.681	1.402	.161
Frequency : condition	1.539	1.018	1.513	.130
Trial : condition	-0.608	1.254	-0.485	.628
Frequency : round	1.095	0.537	2.038	.042
Trial : round	-1.249	0.844	-1.479	.139
Condition : round	0.477	0.694	0.687	.492
Frequency : trial : condition	2.561	1.099	2.330	.020
Frequency : trial : round	1.811	0.977	1.854	.064
Frequency : condition : round	-1.344	0.784	-1.714	.086
Trial : condition : round	1.323	1.264	1.047	.295
Frequency : trial : condition : round	-3.165	1.438	-2.200	.028

By-participant random effects	Variance	SD
Intercept	8.044	2.836
Object	11.723	3.424
Trial	11.234	3.352

Table 1: Model coefficients for the best-fit mixed-effects logistic regression (both rounds). Reference values: object frequency = infrequent; trial number = 1; condition = no common ground; round = 1. SE = Standard error. SD = Standard deviation. Number of participants = 75. Number of observations = 4800. Trial number each round (1 to 32) rescaled to be from 0 to 1.

they re-optimized their original mapping to the new object frequencies in round two. Figure 4 shows that participants were generally successful in achieving shorter word lengths in round two (middle bar) than they would have under the mapping they used in round one (right bar) and a paired samples t-test found that this difference is significant (mean of differences $= 0.893, t = 7.0029(74), p < .001$). This means that participants were successful in re-optimizing their language use after object frequencies changed.

Experiment Extension 2: Common Ground

To investigate the effect of common ground on individuals' abilities to re-optimize language use, we constructed a logistic regression model where the binary outcome variable was short form usage, the fixed effects were object frequency (frequent or infrequent), trial number (rescaled $0-1$), round (one or two) and condition (*common ground* or *no common ground*), and random effects were by-participant slopes and intercepts for trial number and object frequency. What we are looking for is an effect of common ground in round two that links long forms to the frequent item. The best-fit model was the full model, involving a four-way interaction of all predictors, which explained significantly more variance than the next-most complex model with no four-way interaction ($\chi^2(1) = 4.85, p = .028$). The coefficients of the best-fit model can be seen in Table 1. Three effects were significant and are highlighted in the table.

The highest-order effect provides the most comprehensive interpretation of what is going on in the data: this is the four-way interaction between frequency, trial, condition, and round. Its negative coefficient, $\beta = -3.165$, means that participants were more likely to use the *long* word for the frequent object in round two if they were in the common ground condition, and this effect increased as the trials increased. This result supports our hypothesis that participants with common ground are less likely to re-optimize their language after a change in object frequencies than participants without common ground. It also shows that this effect develops over time, given the interaction with trial number.

The next effect, a three-way interaction with frequency, trial, and condition, means that when the frequent object was presented, regardless of round, participants in the common ground condition were more likely to use the short word as trials went on. This effect is outside of our hypothesis, as we expected common ground would only interact with round, and have no effect on ZLA optimality on its own. However this effect suggests that something about playing one partner, as opposed to two partners, causes participants to use language more optimally in general.

The last effect, a two-way interaction between frequency and round, simply means that participants were more likely to use the short form for the frequent object in round two. This means that experience with the task (i.e., playing the communication game a second time) helped participants find the optimal solution.

Finally, none of the fixed effects are significant on their own. This makes sense because we expected each of them to affect short form usage in conjunction with object frequency.

Figure 5 visualizes the main effect by plotting the optimality of participants' word usage per condition, per round. We constructed an *optimality score* that ranges from one (perfect ZLA behavior) to zero (anti-ZLA behavior) and places the lazy and overkill strategies at the midpoint, 0.5. This score is calculated as the mean probability mass of the di-

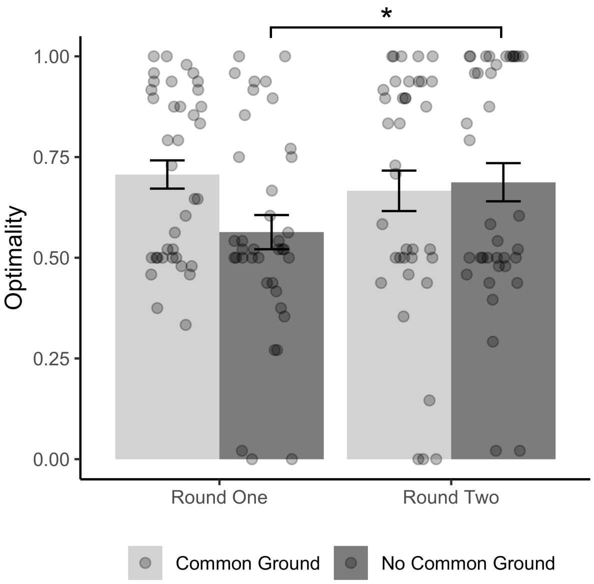

Figure 5: ZLA optimality as predicted by the experimental manipulation of common ground. Each point shows one participant's optimality score and bars show the mean per condition, per round.

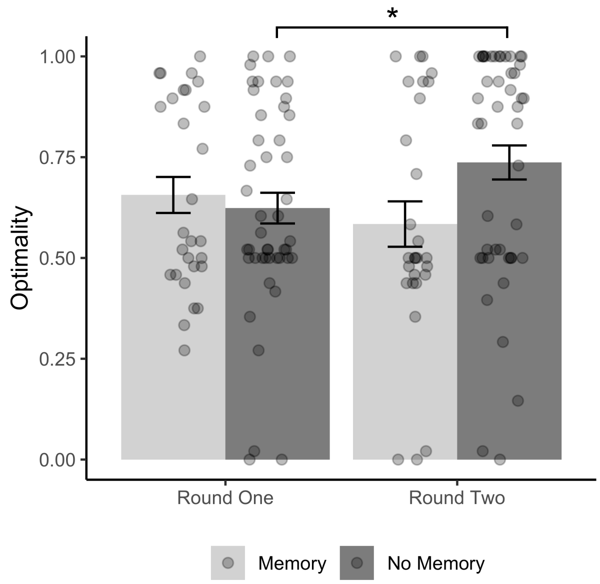

Figure 6: ZLA optimality as predicted by participants' answers to the exit question. *Yes* = memory = common ground. *No* = no memory = no common ground.

agonal values of $p(word|object)$, the proportion of trials that participants used the two word types (long and short) with each object type (frequent or infrequent). Some example $p(word|object)$ mappings and their associated scores and strategy names are:

	ZLA		*semi-ZLA*		*lazy*		*overkill*		*anti-ZLA*	
	F	I	F	I	F	I	F	I	F	I
short word	1	0	.75	.50	1	1	0	0	0	1
long word	0	1	.25	.50	0	0	1	1	1	0
ZLA score =	1		62.5		0.5		0.5		0	

In Figure 5, we see that participants in the *no common ground* condition produce more optimal systems in round two, whereas participants in the *common ground* condition show no change in optimality scores between rounds. However, we expected to see no difference in optimality scores between conditions in round one and were surprised to find that the scores in the *common ground* condition are higher.

Exit Question

In this experiment, we manipulated participants' ability to form common ground in a rather subtle way: by having them play the same partner vs a different partner in round two, where partner identity was indicated simply by a change in fill color to the image shown in Figure 2. Furthermore, the manipulation was fairly novel because it relies on human ability to form common ground with an A.I. partner. Therefore, we included an exit question to better understand how participants interpreted this manipulation. This question showed participants the image of the robot they played in round two and asked "Did you think this robot knew or remembered anything you did during Game 1?". We expected that participants in the *common ground* condition would answer *Yes* and participants in the *no common ground* condition would answer *No*, however the results showed considerable variation. In line with our prediction, more *Yes* responses were obtained in *common ground* than *no common ground*, but participants were more likely to answer *No* across the board:

condition	*Yes*	*No*
common ground	16	20
no common ground	12	26

The intention of our common ground manipulation was to vary whether or not participants believed that their partner remembered their interactions from round one or not. The results of this exit question suggest that 1) our manipulation did not have a strong uniform effect across participants or 2) that this single question at the end of the experiment captured little information about participants' behavior during the experiment. If point 1 above is true, participants' response to this question could serve as an alternative window into their subjective experience of common ground in this task. In response to reviewers' suggestions, we re-analyzed our data using participants' response to the exit question in the place of the common ground condition. We assume that participants who answered *Yes* were the ones who established common ground with their robot partner and participants who answered *No* did not. All other variables where kept the same, including the same reference values, and the reference value for the exit question was *No*.

The best-fit model mirrored that of the previous analysis, but without a significant effect of trial. The main effect was a three-way interaction between frequency, round, and exit question ($\beta = -3.271, SE = 0.436, z = -7.503, p < .001$). From this result, we come to the same interpretation as our previous analysis: participants who responded *Yes* (i.e., believed that the robot remembered round one) were less likely to use the short form for the frequent object in round two. However, this effect did not increase as trial increased. Figure 6 plots ZLA optimality re-grouped by exit question. The main effect patterns with Figure 5: participants without common ground produced significantly more optimal systems in round two than in round one. Overall, this result patterns more closely with our original expectations, where the difference between conditions only occurs in round two. In summary, participants' subjective common ground provides a clearer picture of the main result: when language changes, common ground hampers language re-optimization and a lack of common ground facilitates it.

General Discussion

In this study, we found that language users are able to re-optimize their language use after the frequencies of the objects they talk about changes. This work adds a dynamic component to the existing literature on Zipf's Law of Abbrevation (ZLA) and the competing pressures of accuracy and efficiency on linguistic structures (Kanwal et al., 2017; Chaabouni et al., 2019; Kirby et al., 2015; Piantadosi et al., 2012; Tamariz & Kirby, 2016; Urbina & Vera, 2019) and helps explain why ZLA manages to remain stable over time (Pechenick et al., 2017), despite fluctuations in topic frequencies and need probabilities over time. A real-world example of re-optimization to object frequency reversals occurs when the prototype in a set changes for socio-cultural reasons. For example *phone* has several long forms: *telephone*, *cellphone*, *payphone*, etc. As the home telephone gained popularity it took on its clipped form *phone*, but as cellular phones became the new assumed referent of *phone*, home phones began to require marked forms again for disambiguation.

Our second finding is that common ground affects the re-optimization process, causing partners who have a shared history of communicative success to use less optimal systems after topic frequencies change. This result provides one possible explanation for observed sub-optimalities in language, supporting the idea that communities' communication strategies may become "stuck" in local optima. This is in line with other work on language evolution showing that population turnover (i.e., playing new partners gradually over the course of a game) eliminates unnecessary complexities in communication systems (Castillo et al., 2015; Fay, Garrod, Roberts, & Swoboda, 2010; Tamariz, Cornish, Smith, Roberts, & Kirby, 2012; Granito, Tehrani, Kendal, & Scott-Phillips, 2019) and that smaller communities, with presumably higher pairwise common ground, are able to maintain more complex languages over time (Raviv, Meyer, & Lev-Ari, 2019; Lupyan & Dale, 2010). In other words, as more people become involved with communication, the more likely it is that an optimal solution will be found at a population level.

Although our results were obtained in a human-robot communication framing, we believe that they will extend to human-human interaction. First, we replicated the results of Kanwal et al. (2017), showing that human participants are able to develop ZLA optimal word usage with both a human partner and an A.I. partner. Both of these studies, however, are technically human-computer interaction paradigms, but we suspect the results will replicate in a face-to-face communication game between humans. The one aspect that may vary across such framings is the relative frequency of the various strategies. For example, our study obtained more *overkill* strategies than Kanwal et al. (2017) did, perhaps because human players perceived the robot to have lower linguistic competence than a human and therefore erred on the side of literally spelling everything out for them.

As for common ground, it is possible that our participants formed a different sort of common ground with their robot partner than they would have formed with a human partner. However, several studies show that humans employ human-human communication norms when communicating with robots. For example, Krämer, von der Pütten, and Eimler (2012) show that humans readily engage in human-normative communicative behavior with robots as soon as the context is sufficiently social and Powers et al. (2005) demonstrate that simple cues, such as the pitch of a robot's voice, cause participants to assume gender-normative common ground with robot communication partners. In this latter study, participants spent more time talking to male robots about topics that males are normatively assumed to have little knowledge about, demonstrating that assumed common ground directly affects the communicative effort humans are willing to put in with a robot communication partner.

In future research, we would like to extend this work to populations to understand how sub-optimalities are formed and maintained in speech communities. Real-world populations are structured into sub-populations that exhibit different topic frequencies (Sylwester & Purver, 2015) as well as asymmetries in common ground, where disempowered minorities display more knowledge about those in power than vice-a-versa (Graeber, 2012). As diverse populations negotiate shared linguistic conventions, *which* topic frequencies are they optimized to fit and why? Are certain sub-populations left with an information processing burden, or is a fair trade-off achieved?

Acknowledgments

We would like to thank Dr. Jasmeen Kanwal for sharing her data and thoughts, Jane Ferdinand for illustrating all of the stimuli used in this experiment, the helpful insights of three anonymous reviewers, and the members of the Computational Cognitive Science Lab at the University of Melbourne.

References

Bentz, C., & Ferrer Cancho, R. (2016). Zipf's law of abbreviation as a language universal. In *Proceedings of the leiden* *workshop on phylogenetic algorithms for linguistics.*

Brown-Schmidt, S., & Duff, M. C. (2016). Memory and common ground processes in language use. *Topics in Cognitive Science*, *8*(4), 722–736.

Castillo, L., Branigan, H., & Smith, K. (2015). Context influence vs efficiency in establishing conventions: Communities do it better. In *Proceedings of the 19th workshop on the semantics and pragmatics of dialogue.*

Chaabouni, R., Kharitonov, E., Dupoux, E., & Baroni, M. (2019). Anti-efficient encoding in emergent communication. In *Advances in information processing systems.*

Everett, C. (2017). Languages in drier climates use fewer vowels. *Frontiers in Psychology*, *8*, 1285.

Everett, C., Blasi, D. E., & Roberts, S. G. (2016). Language evolution and climate: the case of desiccation and tone. *Journal of Language Evolution*, *1*(1), 33–46.

Fay, N., Garrod, S., Roberts, L., & Swoboda, N. (2010). The interactive evolution of human communication systems. *Cognitive Science*, *34*(3), 351–386.

Graeber, D. (2012). Dead zones of the imagination: On violence, bureaucracy, and interpretive labor. *HAU: Journal of Ethnographic Theory*, *2*(2), 105–128.

Granito, C., Tehrani, J., Kendal, J., & Scott-Phillips, T. (2019). Style of pictorial representation is shaped by intergroup contact. *Evolutionary Human Sciences*, *1*.

Huggett, E., Peña, P., Sulik, J., & Spike, M. (2020). The effects of interaction on establishing common ground in communication. In *The evolution of language: Proceedings of the 13th international conference.*

Kanwal, J., Smith, K., Culbertson, J., & Kirby, S. (2017). Zipf's law of abbreviation and the principle of least effort: Language users optimise a miniature lexicon for efficient communication. *Cognition*, *165*, 45–52.

Karjus, A., Blythe, R. A., Kirby, S., & Smith, K. (2020a). Communicative need modulates competition in language change. *arXiv preprint arXiv:2006.09277*.

Karjus, A., Blythe, R. A., Kirby, S., & Smith, K. (2020b). Quantifying the dynamics of topical fluctuations in language. *Language Dynamics and Change*, *1*, 1–40.

Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, *336*(6084), 1049–1054.

Kemp, C., Xu, Y., & Regier, T. (2018). Semantic typology and efficient communication. *Annual Review of Linguistics*, *4*, 109–128.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.

Koplenig, A., Meyer, P., Wolfer, S., & Mueller-Spitzer, C. (2017). The statistical trade-off between word order and word structure–large-scale evidence for the principle of least effort. *PloS One*, *12*(3), e0173614.

Krämer, N. C., von der Pütten, A., & Eimler, S. (2012). Human-agent and human-robot interaction theory. In *Human-computer interaction: The agency perspective.* (pp. 215–240). Springer.

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PloS One*, *5*(1), e8559.

Mahowald, K., Dautriche, I., Gibson, E., & Piantadosi, S. T. (2018). Word forms are structured for efficient use. *Cognitive Science*, *42*(8), 3116–3134.

Pechenick, E. A., Danforth, C. M., & Dodds, P. S. (2017). Is language evolution grinding to a halt? *Journal of Computational Science*, *21*, 24–37.

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review*, *21*(5), 1112–1130.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*(9), 3526–3529.

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*(3), 280–291.

Powers, A., Kramer, A., Lim, S., Kuo, J., Lee, S., & Kiesler, S. (2005). Common ground in dialogue with a gendered humanoid robot. *Proceedings of RO-MAN 2005*.

Raviv, L., Meyer, A., & Lev-Ari, S. (2019). Larger communities create more systematic languages. *Proceedings of the Royal Society B*, *286*(1907), 20191262.

Sankoff, G. (2018). Language change across the lifespan. *Annual Review of Linguistics*, *4*, 297–316.

Smith. (2014). Models of language evolution and change. *Cognitive Science*, *5*(3), 281–293.

Sylwester, K., & Purver, M. (2015). Twitter language use reflects psychological differences between democrats and republicans. *PloS One*, *10*(9), e0137422.

Tamariz, M., Cornish, H., Smith, K., Roberts, S., & Kirby, S. (2012). The effects of generation turnover and interlocutor negotiation on linguistic structure. In *The evolution of language.* (pp. 555–556).

Tamariz, M., & Kirby, S. (2016). The cultural evolution of language. *Current Opinion in Psychology*, *8*, 37–43.

Urbina, F., & Vera, J. (2019). A decentralized route to the origins of scaling in human language. *Journal of Statistical Mechanics: Theory and Experiment*, *2019*(9), 093401.

Xu, Y., Liu, E., & Regier, T. (2020). Numeral systems across languages support efficient communication: From approximate numerosity to recursion. *Open Mind*, *4*, 57–70.

Zaslavsky, N., Kemp, C., Regier, T., & Tishby, N. (2018). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, *115*(31), 7937–7942.

Zipf, G. K. (1935). *The psycho-biology of language*. Houghton Mifflin.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Addison-Wesley.