Effects of Communicative Pressures on
Novice L2 Learners' Use of Optional Formal Devices

Yoav Binoun, Francesca Delogu, Clayton Greenberg,
Mindaugas Mozuraitis, and **Matthew W. Crocker**
Department of Computational Linguistics
Saarland University
66123 Saarbrücken, Germany
{ybinoun,delogu,claytong,mindauga,crocker}@coli.uni-saarland.de

Abstract

We conducted an Artificial Language Learning experiment to examine the production behavior of language learners in a dynamic communicative setting. Participants were exposed to a miniature language with two optional formal devices and were then asked to use the acquired language to transfer information in a cooperative game. The results showed that language learners optimize their use of the optional formal devices to transfer information efficiently and that they avoid the production of ambiguous information. These results could be used within the context of a language model such that the model can more accurately reflect the production behavior of human language learners.

1 Introduction

According to the Uniform Information Density hypothesis (Jaeger, 2010), language users optimize their production behavior to transfer information efficiently. More specifically, language users distribute information evenly across an utterance, avoiding peaks and troughs in information density (see Jaeger, 2010; Mahowald et al., 2013; Frank and Jaeger, 2008; Jaeger and Levy, 2006). Additionally, according to Grice's (1975) second Maxim of Quantity, language users avoid the use of redundant or ambiguous information in cooperative situations, although previous work suggests redundant utterances are sometimes preferred (see Arts, 2011; Engelhardt et al., 2006).

Previous work using the artificial grammar learning paradigm (AGL) has suggested that language

learners diverge from the statistical properties of the input language data to make the language more efficient (Fedzechkina et al., 2012). In that study, language learners optimized the use of optional case marking in sentences where animacy and constituent order (SOV vs. OSV) created ambiguity. We conducted a novel study, within the AGL paradigm, to explore whether this behavior extends to a dynamic communicative setting involving a cooperative game. We investigated whether, in this setting, language learners preserve the statistical properties of the input language data or whether they adjust to dynamic communicative pressures (conditions) that arise at production time. Three options were considered:

1. Language users prefer the most efficient structures for information transfer, regardless of the communicative setting and the learning process.

2. Language users are sensitive to the learning process and strictly follow (during production) the frequency of patterns to which they were initially exposed (during learning).

3. Language users consider the communicative setting and dynamically adjust their language production behavior according to changes in the communicative conditions, such as acoustic noise or ambiguities against the visual context.

To provide language users with controlled, yet variable structures, we presented participants with an artificial language with optional overt subjects (OS) and optional agreement affixes (AA) on the

verb. We examined the distribution of usage of these optional devices within the cooperative game.

2 Experiment

Our AGL experiment consisted of two parts. The first part, roughly 25 minutes long, was the learning part (*learning phase*); in this part, participants learned and were tested on an miniature artificial language. The learning phase was divided further into a noun exposure section and a verb exposure section. The second part, roughly 20 minutes long, was the game part (*game phase*); in this part, participants had to describe a target video to a confederate using the language they had learned, while a competitor video was also present. We recorded and transcribed utterances produced by participants during the game phase for the analysis.

The artificial language included two optional formal devices, namely optional overt subjects (OS) and optional agreement affixes (AA) on the verb (see Section 2.1.2 for examples). We manipulated three factors (one acoustic and two visual) during the interaction with the confederate throughout the game phase. The acoustic factor was a recording of coffee shop background noise in two levels of volume, high and low. The hypothesis was that with a higher level of acoustic noise, participants would include more of the optional formal devices in their utterances. The visual factors were determined by the potential overlap between the target and the competitor videos. More specifically, the two videos could have 1) same or different subject and 2) same or different verb. Thus, the experiment used a $2 \times 2 \times 2$ design crossing subject overlap, verb overlap and level of noise. We hypothesized that language learners would change their behavior online and prefer to include the optional formal devices of the input language in their utterances when the subject/verb overlap created ambiguity or when the acoustic noise level was high.

2.1 Method

2.1.1 Participants

Twenty nine Saarland University students (between ages 18-33) participated in the experiment and were monetarily compensated upon completion of participation. Since the optional formal devices

of the artificial language were borrowed from Hebrew, we ensured that all of the participants had no prior knowledge of Semitic languages. Rather, all participants were native speakers of German. Out of the twenty nine participants, three participants were removed from the data due to repeating errors in the artificial language production and two were removed from the data due to recording errors.

2.1.2 Materials

Artificial language stimuli During the learning phase, participants were exposed to 8 nouns: 4 subjects (man, woman, men and women) and 4 objects (apple, cheese, carrot and cake) in still images and text as well as to 2 verbs (eat and drop) in videos and text. All nouns were accompanied by the same determiner ("ha"). All sentences in the artificial language had SVO constituent order. Zero, one, or two optional devices could be present, therefore the translation for the sentence

"(The man) [eats]-<SG. MASC.> ⟨the apple⟩"

could be produced in the following four ways:

(OS)	Verb + (AA)	Object	Exposure
(ha dini)	[akal]-<ini>	⟨ha tapu⟩	25%
	[akal]-<ini>	⟨ha tapu⟩	25%
(ha dini)	[akal]	⟨ha tapu⟩	25%
	[akal]	⟨ha tapu⟩	25%

Table 1: Sentence type exposure during learning

The overt subjects in () and the agreement affixes on the verb in <>, could be dropped. During learning all four possibilities were equally probable, as shown in Table 1.

Visual stimuli The visual stimuli during the noun exposure part of the learning phase consisted of images of the nouns accompanied by written (and acoustic) descriptions in the artificial language. Each subject was presented one time, while objects were presented two times: one time with the object appearing alone in the screen (e.g. one apple) and one time with two images of the object on the screen (e.g. two apples). This was done in order to clarify that objects did not take a plural form (similar to "sheep" in English, for example). In total, 12 images were presented in the noun learning phase: 4

subjects, 4 objects (appearing alone on the screen) and 4 objects (appearing two times on the screen).

During the verb exposure part, video representations of simple transitive verbs between these nouns were played, also accompanied by their descriptions in text and audio form. Each verb was presented 32 times: 4 times per subject, across 4 different subjects and 2 objects. All images were created in Adobe Illustrator CS6, and the videos were created in Adobe Flash CS6 using these images.

The visual stimuli during the game phase consisted of videos showing the same representations of verbs performed by the same subjects and objects, but in different combinations than in the learning phase. For example, since in the learning phase the man was shown eating the cake and the carrot, in the game phase the man was only shown eating the cheese and the apple. Each target video was paired with a competitor video to create four different combinations:

same subject	same verb	diff. object
same subject	diff. verb	same object
diff. subject	same verb	same object
diff. subject	diff. verb	same object

Table 2: List of the visual communicative conditions.

Note that the game required some difference between the target and competitor videos, so it was necessary to have a distinction in the object for the same subject and same verb condition. An arrow indicated on every screen which video was the target. In total 64 screens were played during the game phase in 4 blocks. Each block was balanced for noise and visual communicative conditions.

Audio stimuli During the learning phase, audio and written descriptions in the artificial language accompanied the visual stimuli. Audio stimuli consisted of whole sentence recordings during the verb exposure part, and the nouns during the noun exposure. The audio stimuli were recorded by a male speaker of Hebrew, in a soundproof recording booth using Praat (Version 5.3).

During the game phase, acoustic noise was introduced in two levels, high and low. The noise was a 10 seconds long recording from a local coffee shop, with no intelligible speech in it. The noise at the low

level condition was set to 40 dB and at the high level condition was set to 70 dB.

Procedure The learning phase of the experiment was implemented in Microsoft PowerPoint 2013 and run on a laptop. During the noun exposure, participants were exposed to all the nouns from the artificial language vocabulary in picture, text and audio form. After the audio description ended, the text disappeared and participants had to repeat what they have heard and read, in order to facilitate learning. At the end of the noun exposure, a short noun testing part was played. Participants were presented with the same images and four text choices of nouns from the artificial language vocabulary. Participants had to choose the correct option. After choosing one of the possibilities, the correct choice was presented to the participants for personal feedback.

During the verb exposure part, participants watched videos showing the subjects performing actions denoted by simple transitive verbs ("eat" and "put down") on the objects in different combinations. Each video was consequently shown four times, each time accompanied by the description in a different sentence type. Participants were allowed to repeat the description during all screens and all except for 3 did so. Following the verb learning, a verb testing part was played. During this part, 34 test screens were played for the participants. On each screen, two videos were shown to the participant and only an audio description of one of them was played. After the description ended, participants had to indicate which of the videos was described. After making their choices, an arrow showed which option was the correct one providing feedback for the answers. At the end of the learning phase, a production test took place. Participants were shown 8 videos which they had to describe using the language they had learned. After production, all four possible sentences for the video were presented, and the experimenter indicated which option the participant had produced, thus hinting that all four options are equally usable in the language.

During the game phase, participants were introduced to a confederate, supposedly a well-trained speaker of the artificial language. The game phase was implemented and run in E-prime (Psychology Software Tools, Inc.) on two desktop computers in

two opposite rooms, one computer for the participant and another for the confederate. The participants had to play a cooperative game with the confederate as follows: In each turn, the participant was shown two videos and had to describe one of them to the confederate, who in turn, selected the described video from the same set. The supposed goal of the game was for the confederate to correctly identify as many videos as possible. Thus, the participants were motivated to be understandable and efficient.

In total, 64 rounds of the game were played. Two short practice sessions were played before the game started. In the first practice sessions, the participant was playing the confederate's role, in order to understand the game from both sides. Four practice rounds were played and the confederate described the target video of each round using a different sentence type. The second practice session consisted of 8 additional rounds in which the participant could ask questions about the game.

2.2   Results

The raw counts of the occurrences of each sentence type by visual communication condition are presented in Table 3.

Sentence types production				
Condition	$-OS$ $-AA$	$+OS$ $-AA$	$-OS$ $+AA$	$+OS$ $+AA$
DSDV	73	106	18	132
DSSV	26	136	30	206
SSDV	115	86	23	153
SSSV	118	88	25	166

Table 3: Sentence type production. Condition: DSDV-Different Subject Different Verb, DSSV-Different Subject Same Verb, SSDV-Same Subject Different Verb, SSSV-Same Subject Same Verb.

The table suggests that visual communicative condition had an effect on use of the optional formal devices. Namely, participants diverged from the input language in the following ways:

1) Participants dropped the subjects more often when the competitor video showed the same subject as the target video. 2) participants preferred redundant utterances, mainly when the competitor video showed a different subject and the same verb (DSSV) as the target video. 3) Participants avoided

using the $-OS + AA$ sentence type, showing a possible bias towards the syntactic feature over the morphological one. Table 4 gives the raw counts of the occurrences of each sentence type by acoustic noise level.

Sentence types production				
Noise	$-OS$ $-AA$	$+OS$ $-AA$	$-OS$ $+AA$	$+OS$ $+AA$
H	73	106	18	132
L	26	136	30	206

Table 4: Sentence type production. Noise: H - High noise level, L - Low noise level.

The data was analyzed with linear mixed effects models constructed using the glmer() function of the "lme4" package in R (see Bates et al., 2015; R Core Team, 2015). We trained one model to predict use of OS, given in Table 5, and one model to predict use of AA, given in Table 6.

Fixed effects table $- OS$ model			
	Estimate	Std. Err.	P-Value
Intercept	2.36	0.66	
SO	-1.33	0.16	$< \mathbf{0.001}$
VO	0.63	0.15	$< \mathbf{0.001}$
AN	-0.05	0.16	0.75
SO:VO	-1.22	0.3	$< \mathbf{0.001}$
SO:AN	-0.11	0.31	0.72
VO:AN	-0.33	0.31	0.28
SO:VO:AN	0.8	0.61	0.19

Table 5: LME model. $OS \sim$ SO$*$VO$*$AN $+(1|$ participant$)$ $+(1|$ item$)$

Fixed effects table $- AA$ model			
	Estimate	Std. Err.	P-Value
Intercept	0.2	0.92	
SO	-0.46	0.16	$< \mathbf{0.01}$
VO	0.54	0.15	$< \mathbf{0.001}$
AN	-0.01	0.16	0.96
SO:VO	-0.87	0.3	$< \mathbf{0.01}$
SO:AN	-0.4	0.3	0.18
VO:AN	0.06	0.3	0.85
SO:VO:AN	-0.48	0.61	0.43

Table 6: LME model. $AA \sim$ SO$*$VO$*$AN $+(1|$ participant$)$ $+(1|$ item$)$

Each model included the effects of Subject Over-

lap (SO, same subject vs. different subject in the two videos), Verb Overlap (VO, same verb vs. different verb), Acoustic Noise (AN, high vs. low) and all possible interactions. We also included a by-item and a by-participant random intercept.

Both models revealed significant effects of Subject Overlap, Verb Overlap as well as an interaction between these two factors. Specifically, as predicted, participants used more often the OS or the AA to disambiguate the target video when the competitor video had a different subject performing the verb. Also, when the verb was the same in both videos, participants preferred to include the subject or the affix to better disambiguate the target, since the verb did not. The interactions between the Subject Overlap and Verb Overlap factors are shown in Figure 1 and in Figure 2. The graphs show that when the competitor video displayed the same subject, the formal devices did not help to disambiguate the target video. So, it is reasonable that in this case, the Verb Overlap factor did not have an effect on the production of the optional devices. On the other hand, when the competitor video displayed a different subject, the formal devices could help to disambiguate the target video. So, it is also reasonable that in this case, the Verb Overlap factor had a significant effect. In particular, participants produced more optional devices in the same verb condition, because the verb was not available for disambiguation.

2.3 Discussion and Conclusions

Three options of communicative behavior after recent exposure to the input language data were considered: 1) language learners favor efficient language use regardless of the learning process and the communicative setting, 2) the production behavior of language learners preserves the statistical properties of the input, 3) language learners are sensitive to dynamic communicative conditions and alter language use accordingly. The experimental data support the third option, since visual context affected production of optional formal devices. Acoustic noise, however, did not have an effect. It is possible that the acoustic noise levels were not different enough to provoke changes in behavior. Additionally, the data suggested that the use of the syntactic formal device (OS) was slightly preferred over the morphological one (AA). A possible explanation for

this is that since the affix attaches to the verb, the Verb Overlap factor was more salient. A possible systematic bias in favor of syntactic formal devices over morphological ones could be explored in future work.

The strong tendency of our participants to avoid global ambiguity (which occurred in the $-OS - AA$ condition) is fully consistent with the "make your contribution as informative as is required" part of the Gricean Maxim of Quantity. However, the most popular sentence type among our participants ($+OS + AA$) was redundant in nature, which does not strictly conform to the "do not make your contribution more informative than is required" part of the Gricean Maxim of Quantity.

Since the participants in this study optimized their usage of optional devices according to the presumed shared knowledge between the producer and comprehender, our experiment is quite consistent with models of language production that include Audience Design, such as the Uniform Information Density Hypothesis. Had we found an effect of acoustic noise, we could have made a stronger link to this hypothesis, but we remain hopeful that such information density-sensitive producer manipulations can be captured in future work.

The confirmed bias towards redundant structures, sensitive to assumptions about the knowledge of the comprehender, could be a useful behavior to exploit in both models of sentence processing and applied language models for technological applications. In particular, our results are informative about when language learners use these specific optional devices. Therefore, it would be reasonable for computers to leverage those expectations when processing human input, and to conform to the same expectations when producing linguistic output.

Acknowledgments
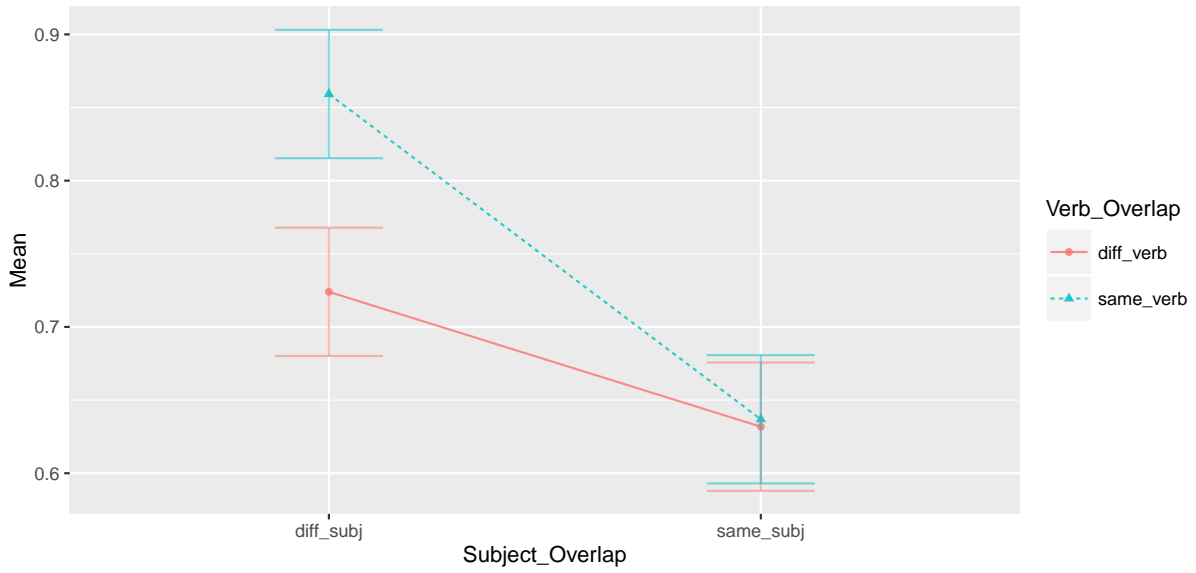
Figure 1: SO:VO interaction plot for the $OS \sim \text{SO}*\text{VO}*\text{AN} + (1|\text{participant}) + (1|\text{item})$ model
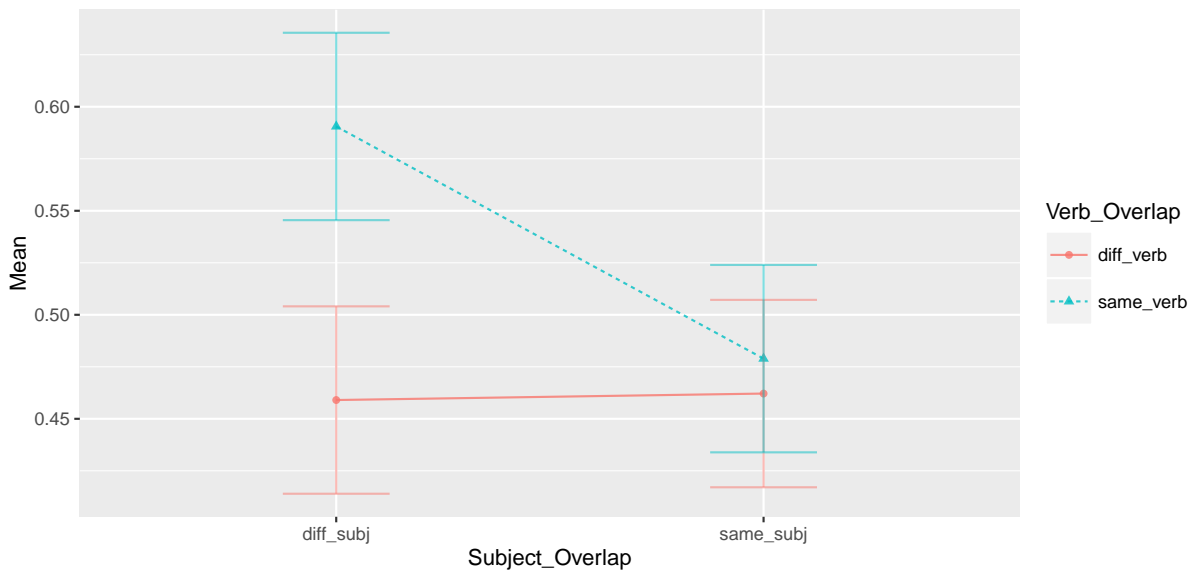
Figure 2: SO:VO interaction plot for the $AA \sim \text{SO}*\text{VO}*\text{AN} + (1|\text{participant}) + (1|\text{item})$ model

References

Arts, A. (2011). Overspecification facilitates object identification. *Journal of Pragmatics*, pages 361–374.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Engelhardt, P. E., Bailey, K. G., and Ferreira, F. (2006). Do speakers and listeners observe the gricean maxim of quantity? *Journal of Memory and Language*, 54(4):554–573.

Fedzechkina, M., Jaeger, T. F., and Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences*, 109(44):17897–17902.

Frank, A. and Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In *Proceedings of the 30th annual meeting of the cognitive science society*, pages 933–938. Cognitive Science Society Washington, DC.

Grice, H. P. (1975). *Logic and conversation*. Academic Press, University of California, Berkeley.

Jaeger, T. and Levy, R. P. (2006). Speakers optimize information density through syntactic reduction. In *Advances in neural information processing systems*, pages 849–856.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive psychology*, 61(1):23–62.

Mahowald, K., Fedorenko, E., Piantadosi, S. T., and Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, 126(2):313–318.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.