The Role of Locality and Formulation Ease during Sentence Production

Mudafia Zafar, Samar Husain
Department of Humanities and Social Sciences, IIT Delhi
mudafia@hss.iitd.ac.in, samar@hss.iitd.ac.in

Understanding the factors determining word order during spontaneous speech informs us about the scope of planning and the nature of the language production architecture [1]. Keeping syntactically related words proximate to reduce memory load (Dependency Length Minimization/ **DLM** [2]) and formulation ease (**Easy-First** [3]) are two notable determinants of word order variation during sentence production. Interestingly, the two accounts assume a very different production mechanism. First, with regards to scope of planning, Easy-First entails incremental/sequential planning, while DLM calls for more non-incremental/structural planning. Second, while Easy-First is clearly a speaker-centric strategy (enabling production efficiency by faster lexical formulation), DLM could either be listener-centric (enabling parsing efficiency) or speaker-centric (enabling production efficiency by resource conservation [4]). These two accounts make similar predictions in SVO languages. In SOV languages, however, DLM predicts a long-before-short order while Easy-First predicts a short-before-long order (see 1a; 1b in OSV order).

We conducted two sentence production experiments[1] (Exp1 and Exp2) in Hindi, an SOV language, with the following **objectives**: (1) we tested whether (and to what extent) DLM governs word order choices during production (Exp1); (2) we investigated if DLM is listener-centric or speaker-centric by manipulating the presence and absence of a listener [5] (Exp1). Finally, (3) we examined if Easy-First also influenced word order (Exp1) and does this propensity increase when participants were induced to plan more incrementally (Exp2). Using the sentence recall task [6], participants saw the subject (S), object (O) and Verb (V) in different boxes on the screen and prepared a sentence using them. Number of OSV shifted responses was the dependent variable. Generalized linear mixed-effects models with a logit link function [7] were used for all analyses.

Exp1 crossed a between-subjects factor LISTENER (PRESENT, ABSENT) with a within-subject factor LENGTH of the argument (ALL-SHORT, S-LONG, O-LONG; see e.g., 1). Exp1 had 84 participants[2] (42 each for LISTENER-PRESENT and LISTENER-ABSENT). In the LISTENER-PRESENT condition, a confederate acted as the listener who did not have access to the speaker's screen and answered comprehension questions based on what was spoken by the speaker/participant. In the LISTENER-ABSENT condition, the experimenter left the room after practice. Results show that participants aligned with a DLM long-before-short preference (OSV in O-LONG >> ALL-SHORT + S-LONG; $p<0.001$). The interaction between LISTENER and O-LONG was not significant ($p=0.7$). Interestingly, participants also aligned with an Easy-First short-before-long preference (OSV in S-LONG >> ALL-SHORT; $p<0.001$). (See Table1 & Fig1 for more details).

The novel short-before-long effect in Exp1 for an SOV language highlights the incremental nature of the production system. In **Exp2** (n=42), we wanted to test if this preference would increase under time-pressure [8]. Following [8], we asked participants to finish speaking within a deadline. The time given was lesser than Exp1 but enough to complete speaking (determined from average speech durations in Exp1). Only the LENGTH of the argument (similar to Exp1) was manipulated. We were able to replicate the short-before-short and long-before-long effects found in Exp1 ($p<0.05$; 0.001 respectively). However, the short-before-long effect did not become stronger, i.e., the interaction between S-LONG and EXPERIMENT (EXP1 vs EXP2) was not significant ($p=0.6$). Contrary to predictions, we found a stronger effect for long-before-short (interaction between O-LONG and EXPERIMENT; $p<0.01$).

Together the two experiments show that (a) **BOTH** Easy-First and DLM together affect word order in an SOV language like Hindi, (b) DLM is **speaker-centric** and (c) that locality (i.e., efficiency in planning syntactic relations) **outweighs** formulation ease (i.e., efficiency in lexical planning) during sentence production.

Design/ Examples:

1(a) **O-LONG** *[$_O$ topi pehni hue chalaak mazdoor-ko] [$_S$ shikari-ne] [$_V$ dhakela]*
cap wear.PRF PTCP.OBL clever labourer-ACC hunter-ERG push.PRF
'The hunter pushed the clever labourer who was wearing a cap.'

1(b) **S-LONG** *[$_O$ mazdoor-ko] [$_S$ topi pehni hue chalaak shikari-ne] [$_V$ dhakela]*
'The clever hunter who was wearing a cap pushed the labourer'

1(c) **ALL-SHORT** *[$_O$ mazdoor-ko] [$_S$ shikari-ne] [$_V$ dhakela]*
'The hunter pushed the labourer'

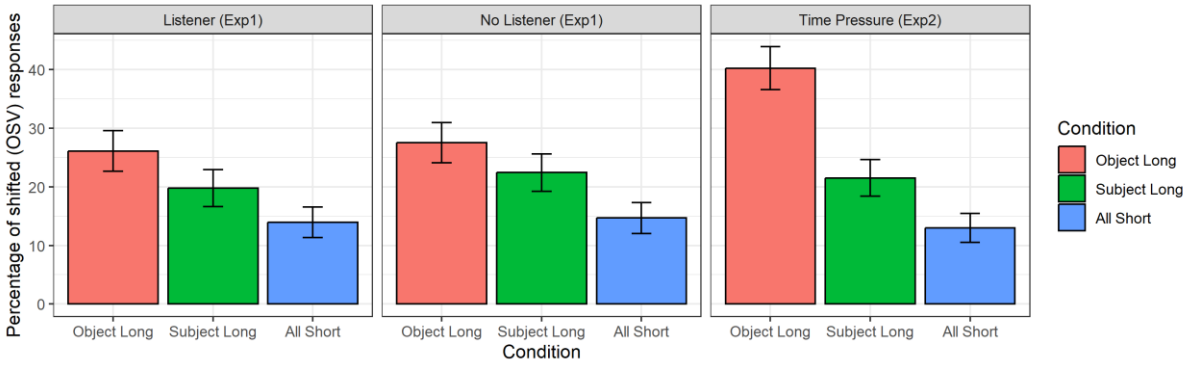
During the task S, O and V appeared in different boxes and participants were free to arrange them in any order during recall;

Notes (1) The predictions, design and other details of both experiments were preregistered on OSF. (2) sample size for both the experiments was determined after a power analysis using the *simr* [9] package in R.

Table 1 (Helmert contrast for LENGTH; Sum contrast for LISTENER and EXPERIMENT). Only the coefficients of interest have been reported.

Experiment	Co-efficient	Estimate (log-odds)	Standard Error	z-value	p-value
Exp 1	Intercept	-1.87	0.25	-7.4	1.35e-13***
	S-Long	1.44	0.39	**3.62**	0.0002***
	O-Long	1.05	0.19	**5.37**	7.60e-08***
	Listener * O-Long	0.1	0.33	0.32	0.74
Exp 2	Intercept	-1.34	0.22	-6.1	1.06e-09***
	S-Long	0.6	0.29	**2.1**	0.03*
	O-Long	1.26	0.29	**4.24**	2.21e-05***
Exp * Length	Intercept	-1.48	0.16	-9.19	< 2e-16 ***
	Experiment * S-Long	0.09	0.22	0.42	0.67
	Experiment * O-Long	0.74	0.28	**2.65**	0.007**

Fig 1



References:

[1] Ferreira, F., & Henderson, J. M. (1998). *Memory & Cognition*. [2] Gibson, E. (2000). *Image, language, brain*. [3] MacDonald, M. C. (2013). *Frontiers in Psychology.* [4] Futrell, R., Levy, R. P., & Gibson, E. (2020). *Language.* [5] Arnold, J.E., Losongco, A., Wasow, T., & Ginstrom, R. (2000). *Language* [6] Ferreira, V. S., & Dell, G. S. (2000) *Cognitive psychology* [7] Bates D, Mächler M, Bolker B, Walker S (2015). *Journal of Statistical Software* [8] Ferreira, F., & Swets, B. (2002). *JML.* [9] Green, Peter & MacLeod, Catriona. (2015). *Methods in Ecology and Evolution.*