

Introduction

• Background

Given the speed and fluency with which speakers typically produce sentences, it is widely held that sentence planning and production occur incrementally^[1]. Instead of planning the whole sentence before starting to articulate it, speakers engage in planning and production simultaneously. Thus, the sentence is articulated piece-by-piece while it is being assembled, with pieces that are easiest to retrieve being assembled and produced first. This relative ease of retrieval of a sentence fragment/phrase - from hereon referred to as '**accessibility**' - is believed to be closely linked to the **weight of the phrase** i.e. the number of lexical items contained within it. A '**heavy**' phrase, containing more lexical items is harder to access than a '**light**' phrase with fewer lexical items to process and fewer alternate orderings from which to make a selection.^[2]

This **production-oriented account**^[3] has been widely used to justify the '**short-before-long preference**' exhibited by speakers of head-initial languages like English, who tend to produce short phrases before long ones in a sentence.^[4] This tendency has been attributed to the **greater accessibility** of short phrases compared to their when heavier(longer) counterparts. Many have pointed to this production accessibility paradigm to suggest that the 'short before long preference' might be a **universal pressure** across languages.^[5]

Recent work in head-final languages like Japanese and Korean, however, has brought conflicting speaker preferences to light. For instance, both corpus studies and online production experiments have shown that Japanese speakers have a '**long-before-short bias**' in the presence of a **heavy NP**.^[6] Specifically, Yashamita and Chang found that **Japanese speakers** tend to **shift long phrases**, such as an argument with a long modifier, **in front of shorter phrases** while producing sentences. These findings are **exactly opposite** to what one would predict under the production accessibility paradigm where shorter phrases are believed to be more accessible and hence more likely to be produced earlier on in the sentence.

In his 2004 paper, John Hawkins proposed a **comprehension-oriented account** to help reconcile the above-mentioned asymmetry.^[7] Specifically, he proposed a parsing principle called the **minimize domain (MiD) principle** which orders constituents so as to **minimize the linear sequence** of words which **must be processed** to construct **dependencies** among **constituents** within a mother phrase.^[8] In a head-initial language like English, this desire to **recognize immediate constituents as quickly as possible** manifests itself as a **short-before-long preference**. For instance, in the post-verbal VP domain given below, construction 1(b) allows the two constituents of the VP (*Mary* and *some friends*) to be recognized more quickly than construction 1(a) does.

1. (a) *I introduced [to Mary] [some friends that John had brought to the party.]*
1. (b) *I introduced [some friends that John had brought to the party] [to Mary.]*

Conversely, in pre-verbal domains of head-final languages, this preference for placing shorter constituents closer to the head predicts the opposite, **long-before-short preference**. Thus, the MiD principle makes **different predictions for languages with different headedness** and agrees with both 'English-like' and 'Japanese-like' results.

- **Hypothesis**

The goal of this experiment is to **replicate** the findings of **Yashamita and Chang** in another head-final language, namely **Hindi**. Specifically, I am looking for evidence of a '**long-before-short**' preference among native Hindi speakers.

Hindi, like Japanese, is a **predominantly head-final language** which allows considerable **flexibility** in word-order. This flexibility allows its speakers to use word-order to mark the prominence of different arguments. Under the assumptions of the MiD principle, I expect Hindi speakers to show a **similar preference for shifting long phrases before short ones** while producing sentences as Japanese speakers do.

My experiment design was slightly different from that of Yashamita and Chang's. I presented native Hindi speakers with sentences of the basic Subject-Object-Verb structure containing either a '**Long Subject**'(LS), a '**Long Object**'(LO) or '**All Short**'(AS) constituents. They were then asked to **reorder the constituents** (if needed) so as to **reproduce the sentences** in the form that felt **most natural** to them. Given the head-final nature of Hindi, I **expect participants to shift objects to the sentence-initial position more often** in the '**Long-Object**' condition than in the '**Long-Subject**' or the '**All-Short**' condition.

Methods

- **Participants**

Eight native Hindi speakers in the age range of 20-50 years were run in this experiment. Of these, five participants identified as **female** whereas three identified as **male**. All of the participants were **bilingual** speakers of **Hindi** and **Indian English**.

- **Materials**

Each participant was presented with a questionnaire which comprised of **20 sentences**. Of these 20 sentences, **5** belonged to the '**Long Subject**' condition, **5** to the '**Long Object**' condition, and the **remaining 10** were **fillers**. In the long condition, constituents (either Subject or Object) were modified by a relative clause, whereas in the '**All Short**' condition, they were left unaltered. I created two lists/versions of the questionnaire using the **latin square** method. Half of the participants were run on List1 of the experiment while the other half were run on List2. Below is a sample item set from the experiment

AS: *jasoos chor ko pakdega*
 detective thief-DO will catch
 “The detective will catch the thief.”

LS: *jasoos jiske zoron se paseene chhoot rahe the chor ko pakdega*
 Detective that profusely sweating was thief-DO will catch
 “The detective that was sweating profusely profusely will catch the thief.”

LO: *jasoos us chor ko pakdega jiske zoron se paseene chhoot rahe the*
 Detective that thief-DO will catch who was sweating profusely
 “The detective will catch the thief that was sweating profusely.”

- **Procedure**

Participant responses to the sentences were recorded via a questionnaire designed on **Google Forms**. Other than basic demographic information (such as age, gender identity and languages spoken), **no identifiable information** (such as name, email address) was recorded.

Each participant was presented with a sentence then asked to reproduce the sentence by **reordering its constituents** in a manner that felt **most natural** to them. In doing so, they were asked to preserve the content of the original sentence so as to avoid any (major) additions or deletions. To ensure that responses were as natural as possible, they were asked to only reorder a sentence if they felt that its current form wasn't the most natural-sounding one. Otherwise, they were asked to reproduce the sentence as is.

- **Analysis**

The dependent measure in this experiment was the proportion of sentences whose subject was shifted to the sentence-initial position. For each condition, **the proportion of sentences** for which the **object was shifted to the sentence initial position** was calculated. This measure reflects how the **weight** of a constituent **affects** its relative **placement** (ordering) within a sentence.

Results

Number of times Object is shifted to sentence-initial position by condition

CONDITION	COUNT (<i>c x n</i>)	NO. OBJECT SHIFT	% OBJECT SHIFT
AS	40	4	10%
LS	40	11	27.50%
LO	40	16	40%

We can see in the table above that the **proportion of sentences** whose **object was shifted** to the **sentence-initial position** was indeed **higher** for the **Long Object condition**. This observation is in conjunction with the hypothesised 'long-versus-short preference.' This preference, however, is not as strong as that observed by Yashmita and Chang and quite

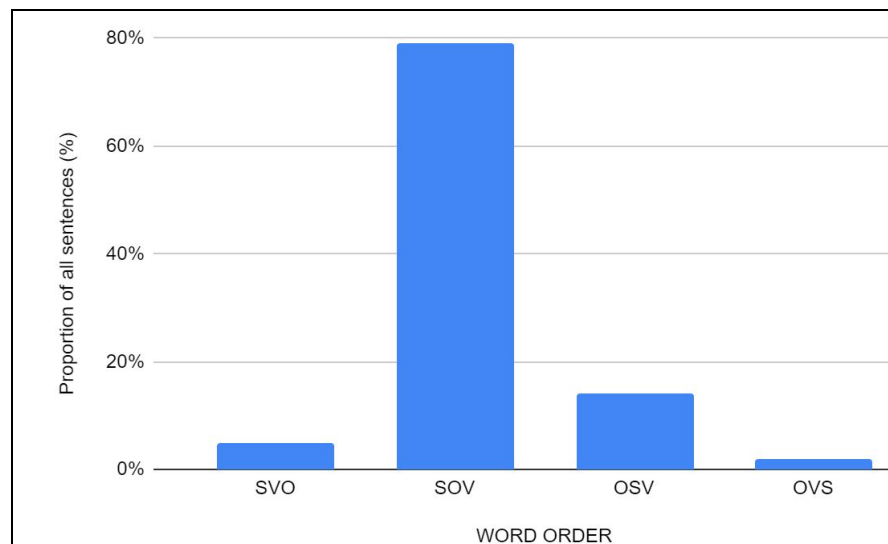
interestingly, reflects only a small proportion of the sentence structures constructed by the participants. The remaining are listed in the table below:

Breakdown of sentence structures constructed by participants

CONDITION	COUNTS
FRONTED CORELATIVE	9
EXTRAPOSE RELATIVE	26
SWITCHED THETA ROLES	1
NO CHANGE	53
SHIFTED OBJECT	31
TOTAL=	120

This variation in form (and consequent divergence from Japanese speaker tendencies) seems to be a consequence of the considerably flexible word-order at the disposal of Hindi speakers. The participants in this experiment alone produced sentences in four different word-orders, including the quite uncommon OVS construction.

Word order of sentences constructed by participants



Discussion

The results demonstrate that of the sentences whose object *did* undergo a “shift” or “movement” to the sentence-initial position, those belonging to the ‘long object’ condition are more common than those belonging to the ‘long subject’ or ‘all short’ condition. That said, movement is not the dominant strategy adopted by the participants. In fact, the majority of the sentences were reproduced without making any change to the presented prompt. This suggests that the ‘long before short preference’ in Hindi is weaker in Hindi than in Japanese.

This is likely due to the relatively high flexibility in word-order afforded to Hindi speakers in sentence production. This flexibility manifests itself in a host of interesting constructions in a task like the one they performed in this experiment. These include the fronted relative construction like in

“jiske paseene choot rahe the jasoos us chor ko pakdega”
he who was sweating profusely that thief the detective will catch

and extraposed relative clause constructions like in

“jasoos us subject ko paldega jiske paseene choot rahe the”
detective that will catch he who was sweating profusely

These results suggest that while head-final languages do show a proclivity for a long-before-short preference in sentence production, the strength of this bias is likely a factor of the relative rigidity of and subsequent variability in the canonical word order of the language. It would be interesting to test this theory on a more reliable head-final language like Turkish which allows little variability in its canonical SOV word order. I would expect Turkish to show a stronger ‘long-before-short’ preference than even Japanese.

References

- [1] De Smedt, Koenraad. 1994. *Parallelism in incremental sentence generation*. In *Parallel Natural Language Processing*, ed. Geert Adriaens and Udo Hahn, 421–447. Norwood, NJ: Ablex.
- [2] Arnold, Jennifer E., Anthony Losongco, Thomas Wasow, and Ryan Ginstrom. 2000. *Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering*. *Language* 76 (1): 28.
- [3] Bock, J.Kathryn. 1986. *Meaning, sound, and syntax: Lexical priming in sentence production*. *Journal of Experimental Psychology: Learning, Memory* 12 (4): 575–586.
- [4] Arnold, Jennifer E., Anthony Losongco, Thomas Wasow, and Ryan Ginstrom. 2000. *Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering*. *Language* 76 (1): 28.
- [5] Gahl, Susanne, Yao Yao, and Keith Johnson. 2012. *Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech*. *Journal of Memory and Language* 66 (4): 789–806.
- [6] Yamashita, Hiroko, and Franklin Chang. 2001. *“Long before short” preference in the production of a head-final language*. *Cognition* 81 (2): B45–B55.
- [7] Hawkins, John J.A. 1994. *A performance theory of order and constituency*. Cambridge (England) and New York: Cambridge University Press.
- [8] Jaeger, T. Florian, and Elisabeth J. Norcliffe (2009). *'The Cross-linguistic Study of Sentence Production'*, *Language and Linguistics Compass* 3: 866–887

Appendix

Below is the full list of experimental stimuli used in this experiment. The left-hand column contains Hindi sentences transliterated into Roman script, and the right-hand column contains the corresponding English sense translations.

1. All Short (AS)
2. Long Subject (LS)
3. Long Object (LO)

1.1	jasoos chor ko pakdegaa
1.2	jasoos jiske zoron se paseene chhoot rahe the chor ko pakdega
1.3	jasoos chor ko jiske zoron se paseene choot rahe the pakdega

2.1	shraddhaloo jyotish se milne gayaa
2.2	shraddhaloo jise ajebo gareeb awaaze sunaae deti thi jyotish se milne gaya
2.3	shraddhaloo jyotish se jise ajebo gareeb awaze sunai deti thi milne gaya

3.1	gautam kutte ko paani pila rahaa tha
3.2	gautam jo kai ghanto se baahar baitha hai us kutte ko paani pila raha tha
3.3	gautam us kutte ko jo kai ghanto se baahar baitha hai paani pila raha tha

4.1	adhaapika us ladki ko dhoondh rahi hai
4.2	adhaapika jise bhagdad mein kaaafi chot aayi ius thi ladki ko dhoondh rahi hai
4.3	adhaapika us ladki ko dhoondh rahi hai jise bhagdad mein kaaafi chot aayi thi

5.1	natasha vah ghodi paalegi
5.2	natasha vah ghodi jo haal hi mein videsh se aayi thi paalegi
5.3	natasha jo haal hi mein videsh se aayo thi vah ghodi paalegi

6.1	aisha chappal chipka rahi hai
6.2	aisha vah chappal jise kal padosi ka kutta kaat gaya chipka rahi hai
6.3	aisha jise kal padosi ka kutta kaat gaya chappal chipka rahi hai

7.1	daakiya vah paarsal dhoondh lega
7.2	daakiya jo kaaafi dino se daakghar se laaptata hai paarsal dhoondh lega
7.3	daakiya vah paarsal jo kaaafi dino se daakghar se laapata hai dhoondh lega

8.1	thekedaar majdoor ko nikaal dega
8.2	thekedaar jise kai hafto se munaafa nahi hua majdoor ko nikaal dega
8.3	thekedaar majdoor ko jise kai hafto se munaafa nahi hua nikaal dega

9.1	raani vah bartan jod degi
9.2	raani vah bartan jo jadbaazi mein gir gayaa tha jod degi
9.3	raani jo jadbaazi mein gir gayi thi vah bartan jod degi

10.1	nurse mariz ko dawaa degi
10.2	nurse jisne subah se kuchh nahi khaaya hai mariz ko dawaa degi
10.3	nurse mareez ko jisne subah se kuchh nahi khaaya hai dawaa degi

Fillers

f.1	surbhi ne vah saadi jo uski maa ne kal dukaan mein dekhi thi khareed li
f.2	anil ne vah gullak jo uska bhai pehle hi khaali kar chuka tha phod daala
f.3	jyoti ne vah dukaan jiska maalik apni biwi se jhagdaa karta tha jaana chod diya
f.4	mantri us sadak ki jismein picchle hafte uski gaadi atak gayi thi maramat kara raha hai
f.5	aakanksha ne vah ladki jo bhookamp main baal baal bacchi god li
f.6	arjun vah gilahari jise ek cheel ghaayal kar chuki thi ghar le aaya
f.7	megha ne vah gaayak jise vah bacchan se sunte aa rahi hai dekha
f.8	karan ne vah biryaani jo bawarchi ne buri tarah se jalaa di thi mushkil se nigli.
f.9	kamal daftar mein vah shirt jise usne kal ragad ke dhoya tha pahane ga
f.10	zoya ne ve parde jo kaaleen ke saath jach rahe the pasand kiye

1. All Short (AS)
2. Long Subject (LS)
3. Long Object (LO)

1.1	the detective is chasing the thief
1.2	the detective who was sweating profusely is chasing the thief
1.3	the detective is chasing the thief who was sweating profusely

2.1	the pilgrim went to meet the astrologer
2.2	the pilgrim who would hear strange noises went to meet the astrologer
2.3	the pilgrim went to meet the astrologer who would hear strange noises

3.1	Gautam was giving the dog some water.
3.2	Gautam who has been sitting outside for hours was giving the dog some water.
3.3	Gautam was giving the dog who has been sitting outside for hours some water

4.1	the teacher is looking for that girl
4.2	the teacher who was seriously injured in the stampede is looking for that girl
4.3	the teacher is looking for that girl who was seriously injured in that stampede

5.1	natasha will raise the horse
5.2	natasha will raise the horse which only recently arrived from a foreign land
5.3	natasha who only recently arrived from a foreign land will raise that horse

6.1	aisha is fixing the shoe
6.2	aisha is fixing the shoe that the neighbors' dog bit yesterday
6.3	aisha who the neighbours' dog bit yesterday is fixing the shoe

7.1	the postman will find that parcel
7.2	the postman that has been missing from the postoffice for some days will find the parcel
7.3	the postman will find the parcel that has been missing from the postoffice for some day

8.1	the contractor will fire the labourer
8.2	the contractor who has not made profit in quite some weeks will fire the labourer
8.3	the contractor will fire the labourer who has not made profit in quite some weeks

9.1	Raani will repair the vase
9.2	Raani will repair the vase that fell while in a hurry
9.3	Raani who fell while in a hurry will repair the vase

10.1	the nurse will give the patient his medicine
10.2	the nurse who hasn't eaten all day will give the patient his medicine
10.3	the nurse will give the patient who hasn't eaten all day his medicine

Fillers

f.1	surbhi bought the saree that her mom had spotted at the store
f.2	anil broke the piggybank that his brother had already emptied out
f.3	jyoti stopped going to the store whose owner would fight with his wife
f.4	the minister is getting that road repaired where his car got stuck the other day
f.5	aakanksha adopted that girl who barely survived the earthquake
f.6	arjun brought home the squirrel that a hawk had injured
f.7	megha spotted the singer who she had been listening to ever since she was a kid
f.8	karan painfully swallowed the rice that the chef had badly burnt
f.9	kamal will wear to the office the shirt that he had so thoroughly cleaned yesterday
f.10	zoya selected the curtains that went well with the carpet.