



Marking topic or marking case?

A comparative investigation of Heritage Japanese and Heritage Korean

Sixth Heritage Language Research Institute
June 18-22, 2012
UCLA

Oksana Laleko (SUNY New Paltz)
Maria Polinsky (Harvard)

General questions

- Which aspects of language structure can be difficult for heritage language speakers, and why?
- Can the range of difficulty be correlated with heritage language speakers' proficiency?
- What advantages, if any, do heritage speakers have over L2 learners?

By way of introduction

The acrobat likes his jokes and the clown does too.

- Reading 1: The acrobat likes his own jokes, and the clown likes his own jokes
- Reading 2: The acrobat likes the acrobat's jokes, and the clown likes the acrobat's jokes

Making it look formal and serious

The acrobat likes his jokes and the clown does too.

- Reading 1:

The acrobat ($\lambda x (x \text{ likes } x\text{'s jokes})$) & the clown ($\lambda x (x \text{ likes } x\text{'s jokes})$)

- Reading 2:

The acrobat ($\lambda x (x \text{ likes } a\text{'s jokes})$) & the clown ($\lambda x (x \text{ likes } a\text{'s jokes})$)
a = the acrobat

Which reading is easier?

- Reading 1: one needs to obtain the appropriate interpretation of the missing VP in the second clause
- Reading 2: one needs to go into prior discourse (the one introducing the acrobat) and establish coreference between discourse material and the material in the second clause

Which reading is easier?

- Reading I requires less information and less structure inclusion
- English and Dutch speakers prefer reading I and process it more quickly compared to reading 2 (Vasić 2006, Shapiro and Hestvik 1995, Shapiro et al. 2003, Frazier & Clifton 2000, Kornneef 2008).

Emerging generalizations

- Comprehension related to smaller domains is easier
- Comprehension related to syntactic computation is easier than comprehension including semantic and discourse domains

Why?

- Integrating various types of knowledge across domains is more difficult than computing structure

Our study

- Interaction of domains
- Phenomena mediated at different structural levels
 - NOM (internal to **basic clause structure**, TP) and TOP (external to basic clause structure, requiring more structure)
 - $[\text{CP}[\text{TP} \text{NP-NOM} \dots]]$
 - $[\text{CP} \text{NP-TOP} [\text{TP} \dots]]$



JAPANESE AND KOREAN: BASICS OF TOPICS AND NOMINATIVES

Where the two languages are similar

- Topic-prominent languages (Li & Thompson, 1976)
 - organize syntax around information structure
 - mark distinction between new and given (shared) information in addition to grammatical functions

(1)	Sakana-wa	tai-ga	oisii.	J
	fish-TOP	snapper-NOM	delicious	

‘Speaking of fish, red snapper is delicious’

(2)	Sayngsen-un	yene-ka	massissta.	K
	fish-TOP	salmon-NOM	delicious	

‘Speaking of fish, salmon is delicious.’

Functions of TOPIC (Kuno, 1976; Choi, 1999)

- Thematic generic

Kami -wa ki -kara tsuku-rare -masu J

Paper-TOP tree -from make-Passive-be

‘Paper is made from trees.’

- Thematic anaphoric

Watashi-wa sengetsu -hajimete chu-goku -o otozuremashi -ta J
I -TOP last month -first time China -ACC visit -PAST

‘I visited China for the first time last month.’ [China is introduced]

Ima chu-goku-wa ishiban sukina kuni -desu
now China -TOP first favorite country -is

‘Now, China is my favorite country.’ [China is expected to be familiar and is referred back to]

Functions of TOPIC (Kuno, 1976; Choi, 1999)

- Contrastive, in a matrix clause

Watashi-wa hudan **hougaku** -wa kikimasu ga J
I -TOP usual Japanese music-TOP listen but
yousaku -wa kiki -masen
western music-TOP listen –NEG
'I usually listen to **Japanese music** but I do not listen to **Western music**.'

- Contrastive, in an embedded clause

Taroo-wa [Hikari-wa kirei-da to] omou. J
Taroo-TOP Hikari-TOP beautiful-COMP think
'Taroo believes that Hikari is beautiful' (as opposed to someone else)
[we will return to the embedded contexts below]

Where the two languages are similar: I

- Case markers and Topic markers may be omitted in informal registers

Kicchin -ni oisii pai (ga) arimasuyo

Kitchen -in delicious pie (NOM) have

'There is a delicious pie in the kitchen.'

Kino-no -yoru tomodachi-to issyo -ni mi -ta eiga (wa) totemo yoka -tta.

Yesterday-GEN-night friends -and-together -with watch-PAST movie (TOP) very good -PAST

'The movie which I watched with my friends last night was very good.'

- But case marker drop is not allowed in embedded clauses

Kouta -wa [Mai *(ga) tsukut -ta]

Kouta –TOP Mai-NOM make –PAST

'Kouta ate a cake [which Mai made].'

keiki tabe -ta

cake eat -PAST



Where the two languages are similar: II

- When a topic marked DP occurs in an embedding, it **must** be interpreted contrastively
 - Particularly apparent in contexts which do not allow for topics at all, such as control structures

Where the two languages are similar: II

- When a topic marked DP occurs in an embedding, it **must** be interpreted contrastively
 - Cf. in English: regular topics OK in relative clauses but impossible in control clauses:
 - This is the man to whom, *liberty*, we will never grant
 - *I tried, *this book*, to read
 - *I persuaded Pat, *in this race*, to participate

Where the two languages are similar: II

- Contrastive topic in control clauses

Aika-wa okaasan-o [kuruma-wa uru-kotoo] settokushita **J**

Aika-TOP mother -ACC car-TOP sell-COMP persuaded

'Aika persuaded her mother to sell the CAR' (as opposed to a bicycle)

Chelswu-nun emeni-lul [cha-nun pha-tolok] seltukhayssta **K**

Chelswu-TOP mother-ACC car-TOP sell-COMP persuaded

'Chelswu persuaded mother to sell the CAR' (as opposed to the bicycle)

Where the two languages differ

- **Japanese**
- A single marker
 - TOP wa
 - NOM ga
- NOM must be interpreted as new:
- **Korean**
- Two sets of markers
 - TOP nun/un
 - NOM i/ka
- NOM can be given or new:

kore wa/*ga nan desu ka
this what is Q
'What is this?'

ikes (?)un /i mwues-i-eyyo
this what is Q
'What is this?'



TOPIC VS. NOMINATIVE: PSYCHOLINGUISTIC EVIDENCE

Topic vs. Nominative

- Psycholinguistic evidence for TOP—NOM difference
 - LI Acquisition
 - Adult production
 - Heritage language speakers

Acquisition of *ga* and *wa* in Japanese

- Particles begin to appear early in child LI production (Okubo, 1967; Hayashi, 1982)
 - with high rates of omission and misuse
- Late full mastery (Ito & Tahora, 1985)
 - early LI acquirers overuse of the NOM and underuse TOP (Goto, 1988; Nakamura, 1993)
 - ...despite higher frequencies of TOP in adult speech
- NOM used instead of ACC and LOC (Fujitomo, 1977)

Acquisition of NOM vs. TOP in Japanese

- *wa* and *ga* emerge as early as 2;2 (Okubo, 1967; Miyahora & Miyahora, 1973)
 - high rate of ellipsis
- *wa* omission still high at 5 (Nakamura, 1993)
- Ito & Tahara (1985):
 - 4-5 year-olds: use of *ga* regardless of context
 - 6-12 year-olds: use of *wa* but not entirely consistent
 - complete adult-like differentiation by discourse function by 14

Particle omission in L1

- Particle ellipsis particularly frequent in speech addressed to young children
 - Miyazaki (1979):
 - adult to adult omissions: *ga* – 10.2%; *wa* – 25.5%
 - mother to 2-year-old omissions: *ga* – 30%; *wa* – 70%
- Age-related trend: with increasing age, the proportion of *wa* increases until it is used more often than *ga*

Particle omission in L1

- Both languages: NOM and ACC omission is sensitive to animacy (Lee et al. 2009; Kurumada and Jaeger 2012)
 - omission is more common with **inanimates**

Acquisition of NOM vs. TOP in Korean

- The acquisition of NOM precedes the acquisition of TOP by 2-4 months (however one study reports the opposite development, S.-Y.Kim, 1990)

Acquisition of cases in Korean

- Early use of NOM: *-ka* used correctly, but *-i* often replaced by *-i-ka*
- Acquisition of ACC is 6-8 months later than acquisition of NOM, and NOM is used instead of ACC up to 4;0
- Case omission more common with ACC than with NOM (mirrors the pattern in the input)

Acquisition of TOP marking in Korean

- Contrastive use of TOP is acquired about 6 months prior to the thematic use, possibly due to patterns in the input
- Earliest form: *-n*

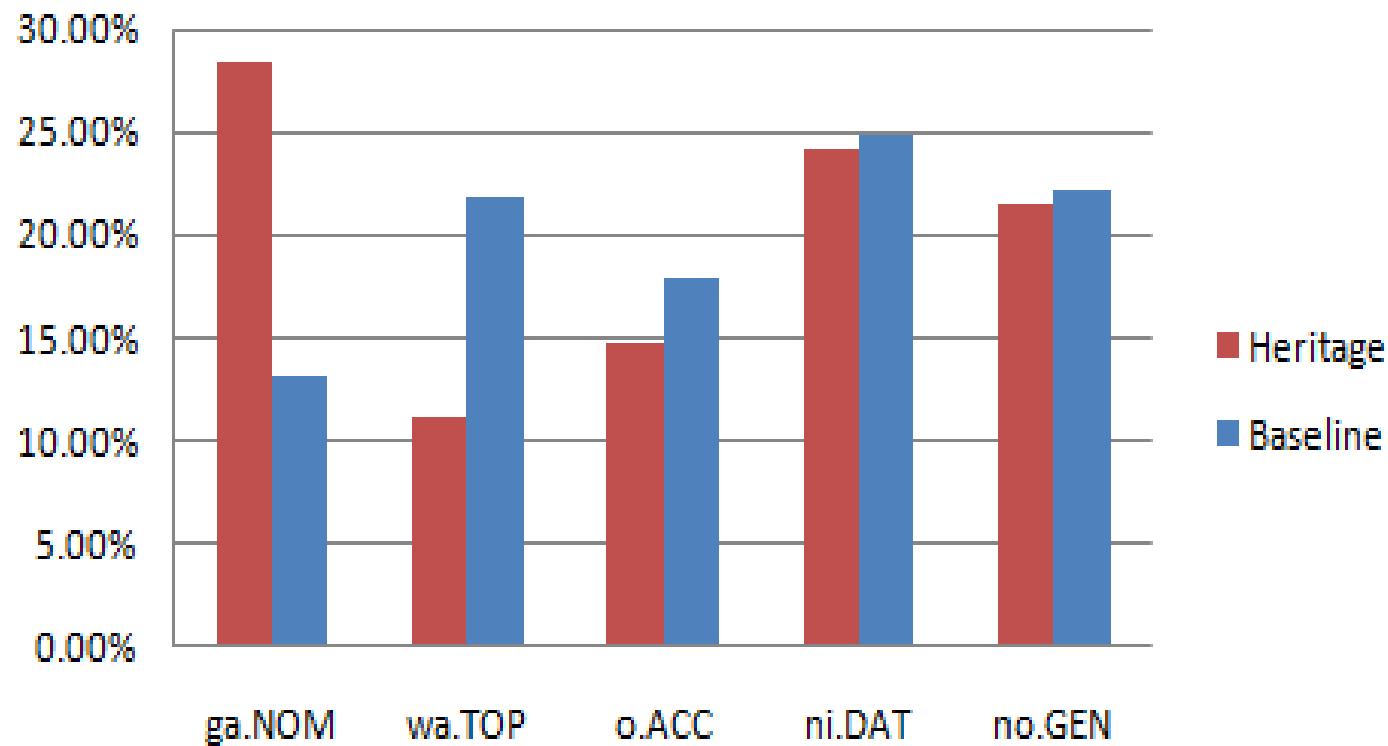
Laleko and Kawamura (2011)

- A pilot study of heritage Japanese (frog story)

Heritage	Participants	N	Current Age Mean (range)	Age of Arrival Mean (range)	Time in US (mean in years)	Mean Use of Japanese [%]	Score on Word Translation Task (out of 100)	Proficiency (self-ratings) <i>On a 10-point scale, how well do you...</i>				
								Speech Rate, Japanese (per minute)	Speech Rate, English (per minute)	understand spoken Japanese	speak Japanese	
Heritage	4	24.5 (21-31)	1.8 (0-5)	22.8	11	59.6	36.1	126.6	6.1 (3-8)	3.8 (1.5-6.5)	2.1 (0-5)	2.4 (0-7)
Baseline	4	21.5 (21-23)	21.5 (21-23)	>1	50	92.8	86.4	N/A	N/A	N/A	N/A	N/A

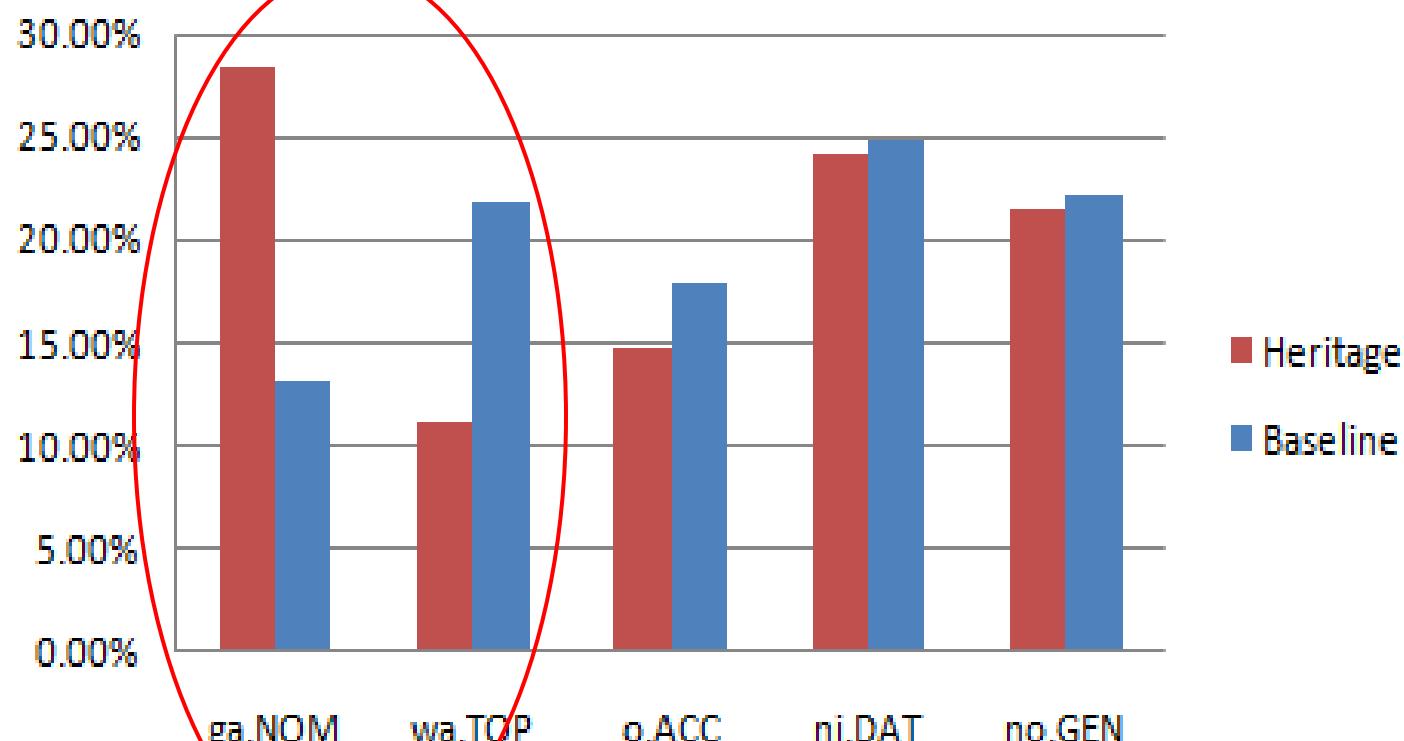
Laleko and Kawamura (2011)

Particles



Laleko and Kawamura (2011)

Particles



NOM vs. TOP

- *-wa* is mediated at the syntax-pragmatics interface
 - involves syntactic, semantic, and discourse knowledge (information structure);
 - gradient judgments
 - misuse results in infelicity
- *-ga* is governed by structural/syntactic constraints
 - categorical judgments
 - misuse results in ungrammaticality



CURRENT STUDY

What we already know

- Advanced HL speakers: syntax-pragmatics interface affected stronger than morphosyntax (Laleko, 2010; Polinsky, 2006)
- Low level HL speakers, including overhearers: strong deficits in morphosyntax

Questions

- (I) Are greater difficulties associated with the domain of morpho-syntax or discourse-pragmatics?
- Predictions:
 - If morpho-syntactic difficulties prevail, we expect less accuracy on NOM
 - If problems are with semantics and discourse-pragmatics, we expect less accuracy on TOP

Questions

- (2) Overt or null?
 - (OR: Tip of the iceberg or bottom of the iceberg?)
- Predictions:
 - If overt > null in heritage grammars (cf. a tendency towards redundancy of expression), we expect better accuracy on conditions involving overt markers compared to conditions involving omissions.

Study Design: Participants

Japanese	LI (N=13)	L2 (N=31)	HL (N=30)
Age	26.4	27.5	24.75
Age of arrival	N/A	N/A	4.5
Age of switch	N/A	N/A	5.3
% daily use	68	12.4	22.9

Study Design: Participants

Korean	L1 (N=14)	L2 (N=16)	HL (N=36)
age	21.7	25.8	24.5
Age of arrival	N/A	N/A	3.9
Age of switch	N/A	N/A	3.9
% daily use	40.8	23.5	32.9

Study Design

- Materials:
 - 56 sentences in each language
 - Conditions on the use of NOM and TOP:
 - appropriate use
 - misuse, including NOM instead of TOP and TOP instead of NOM
 - omission
 - Ratings elicited on Amazon Mechanical Turk
 - Demographic questionnaire

Mechanical Turk is a marketplace for work.

We give businesses and developers access to an on-demand, scalable workforce.

Workers select from thousands of tasks and work whenever it's convenient.

181,714 HITs available. [View them now.](#)

Make Money by working on HITs

HITs - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITs now.](#)

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work



or [learn more about being a Worker](#)

Get Results from Mechanical Turk Workers

Ask workers to complete HITs - *Human Intelligence Tasks* - and get results using Mechanical Turk. [Register Now](#)

As a Mechanical Turk Requester you:

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results



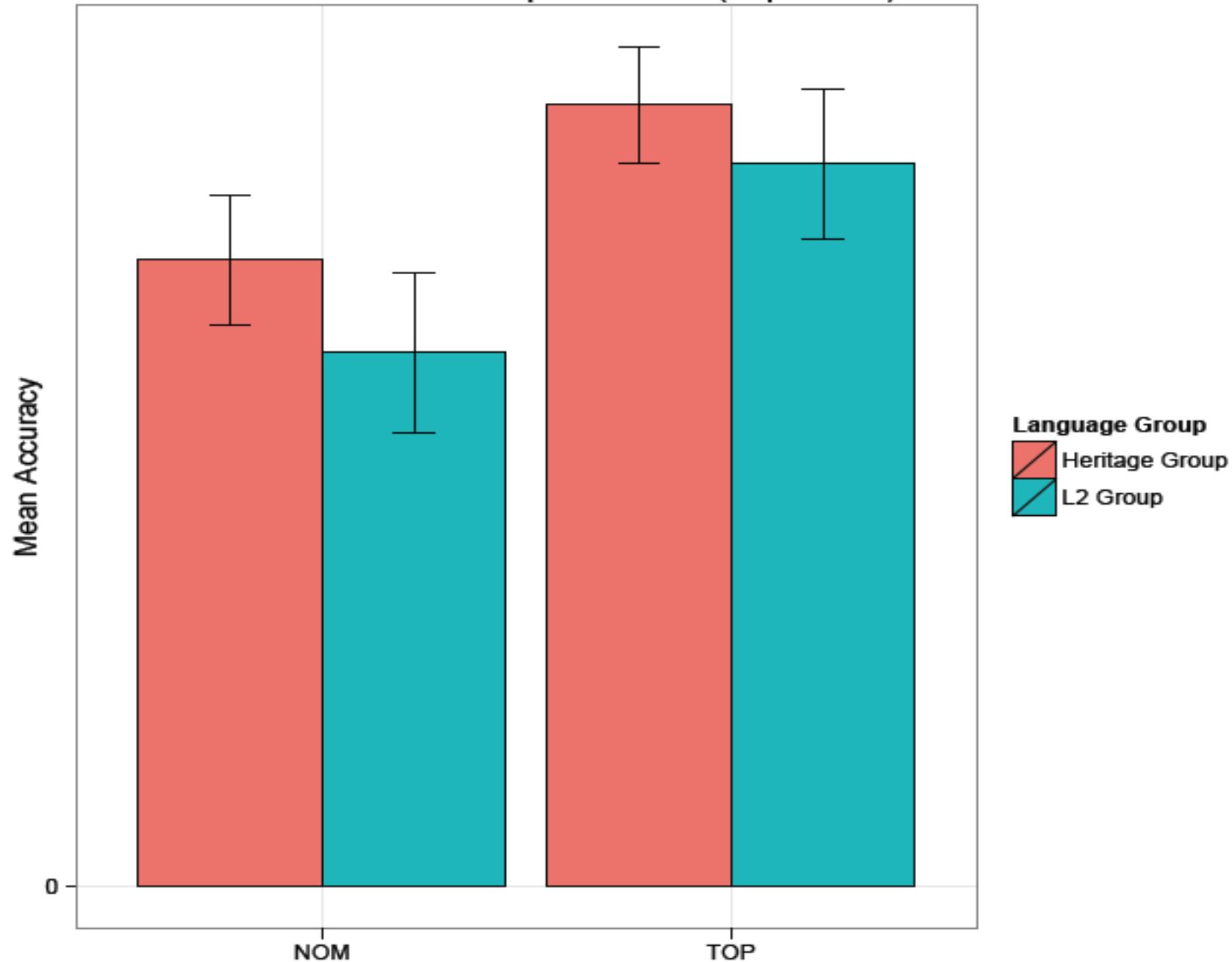


RESULTS

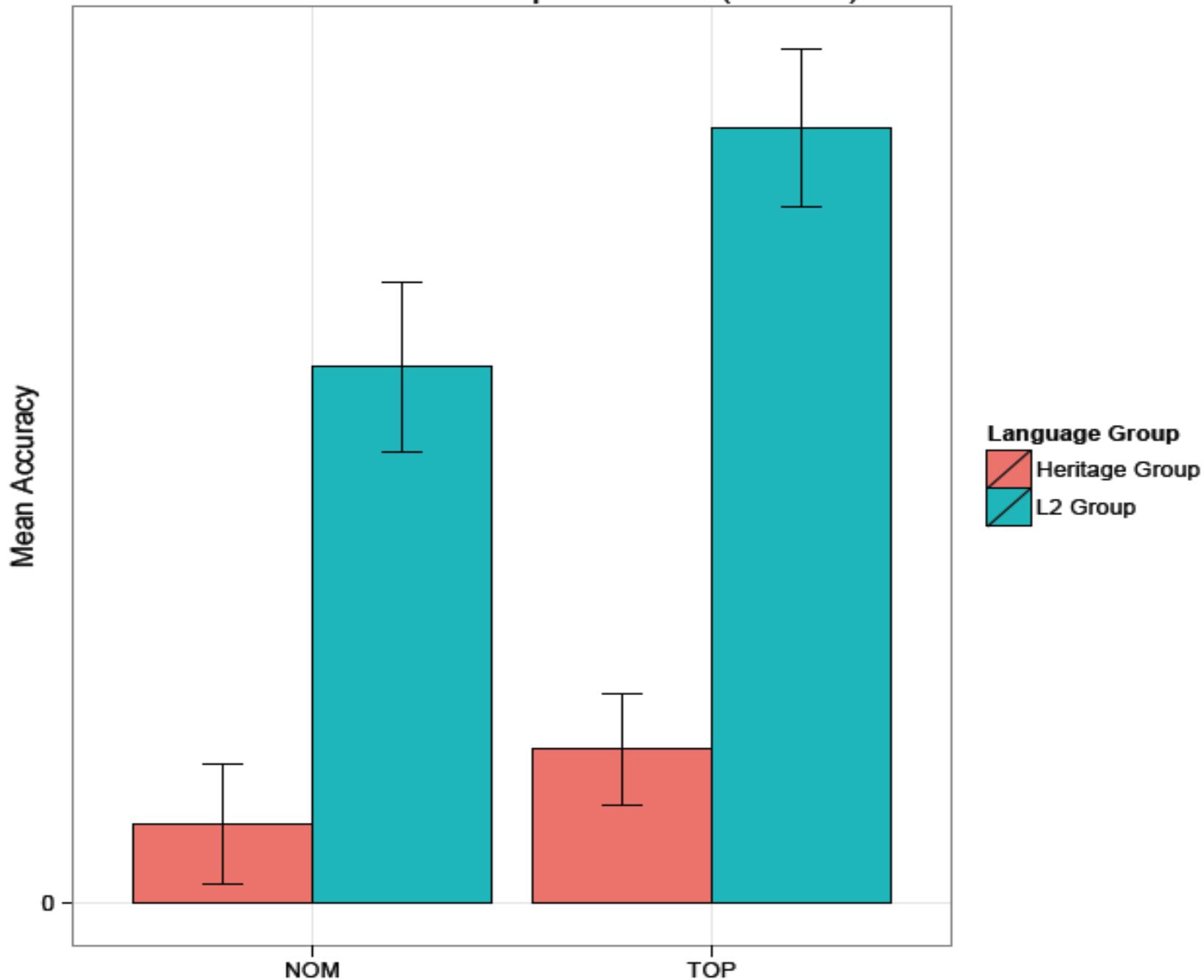
Results

- HL speakers are more accurate on conditions involving NOM than on conditions involving TOP
- The prediction is borne out for both languages
 - ... although with interesting differences:
 - Korean HL: closer to the baseline (compared to L2)
 - Japanese HL: further away from the baseline (compared to the Korean group); sometimes pattern with L2

Mean Accuracies in Top and Nom (Japanese)



Mean Accuracies in Top and Nom (Korean)



Japanese vs. Korean: I

	Korean HS, Percent YES	Japanese HS, Percent YES
Did your parents encourage you to speak the HL as much as possible in the house?	56	44
Did your parents correct you when you spoke the HL?	49	26
Did you study the HL as a foreign language at school or college?	39	20
Did you receive any other formal instruction in the HL (e.g., Sunday school)?	28	22
Would you like to improve your HL skills?	44	37

Japanese vs. Korean: I

	Korean HS, Percent YES	Japanese HS, Percent YES
Did your parents encourage you to speak the HL as much as possible in the house?	56	44
Did your parents correct you when you spoke the HL?	49	26
Did you study the HL as a foreign language at school or college?	39	20
Did you receive any other formal instruction in the HL (e.g., Sunday school)?	28	22
Would you like to improve your HL skills?	44	37

Japanese vs. Korean: II

Japanese Demographic Information

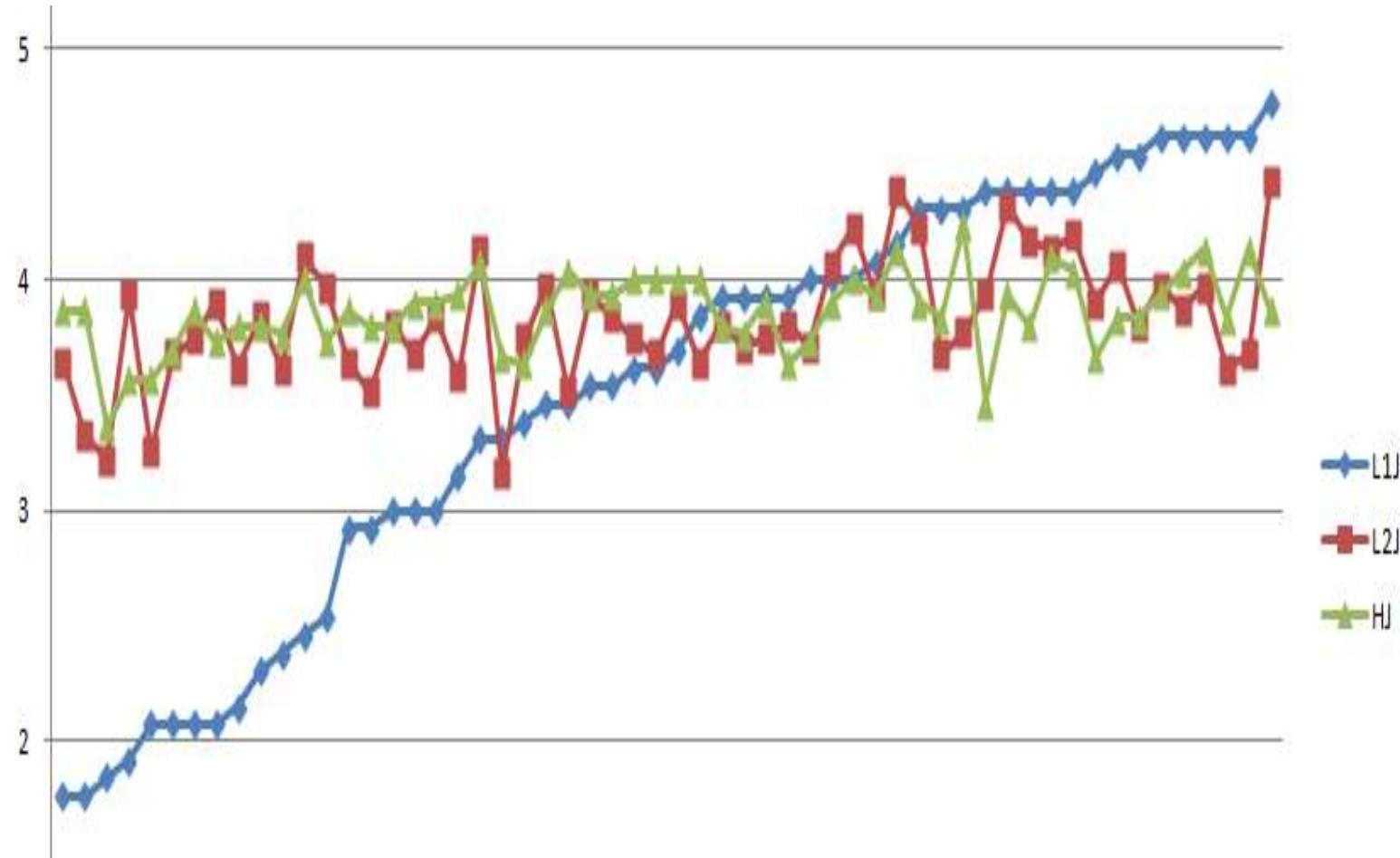
	L1 Group (N=13)		L2 Group (N=31)		Heritage Group (N=30)					
Background	mean	range	mean	range	mean	range				
Age	26.4	18 – 45	27.5	18 – 62	24.75	20 – 33				
Age of arrival to US	NA	NA	NA	NA	4.5	0 – 12				
Age of switch to English	NA	NA	NA	NA	5.3	0 – 15				
% daily use of Japanese	68	2 – 100	12.4	0 – 40	22.9	1 – 50				
1-5 Self-Ratings	%1	%2	%3	%4	%5	%1	%2	%3	%4	%5
Overall Japanese Ability	0	0	8	23	69	16	13	35	26	10
Overall English Ability	0	0	8	38	54	0	0	3	13	84
English Reading	8	15	8	31	38	0	6	3	3	87
English Speaking	8	15	23	23	31	0	0	6	19	74
English Listening	8	15	15	23	38	0	0	6	6	87
English Writing	15	15	8	31	31	0	3	3	6	88
Japanese Reading	0	8	8	23	62	6	23	39	19	13
Japanese Speaking	0	0	0	46	54	10	26	32	19	13
Japanese Listening	0	0	0	46	54	10	13	35	32	10
Japanese Writing	0	0	9	46	54	10	16	52	23	0

Japanese vs. Korean: II

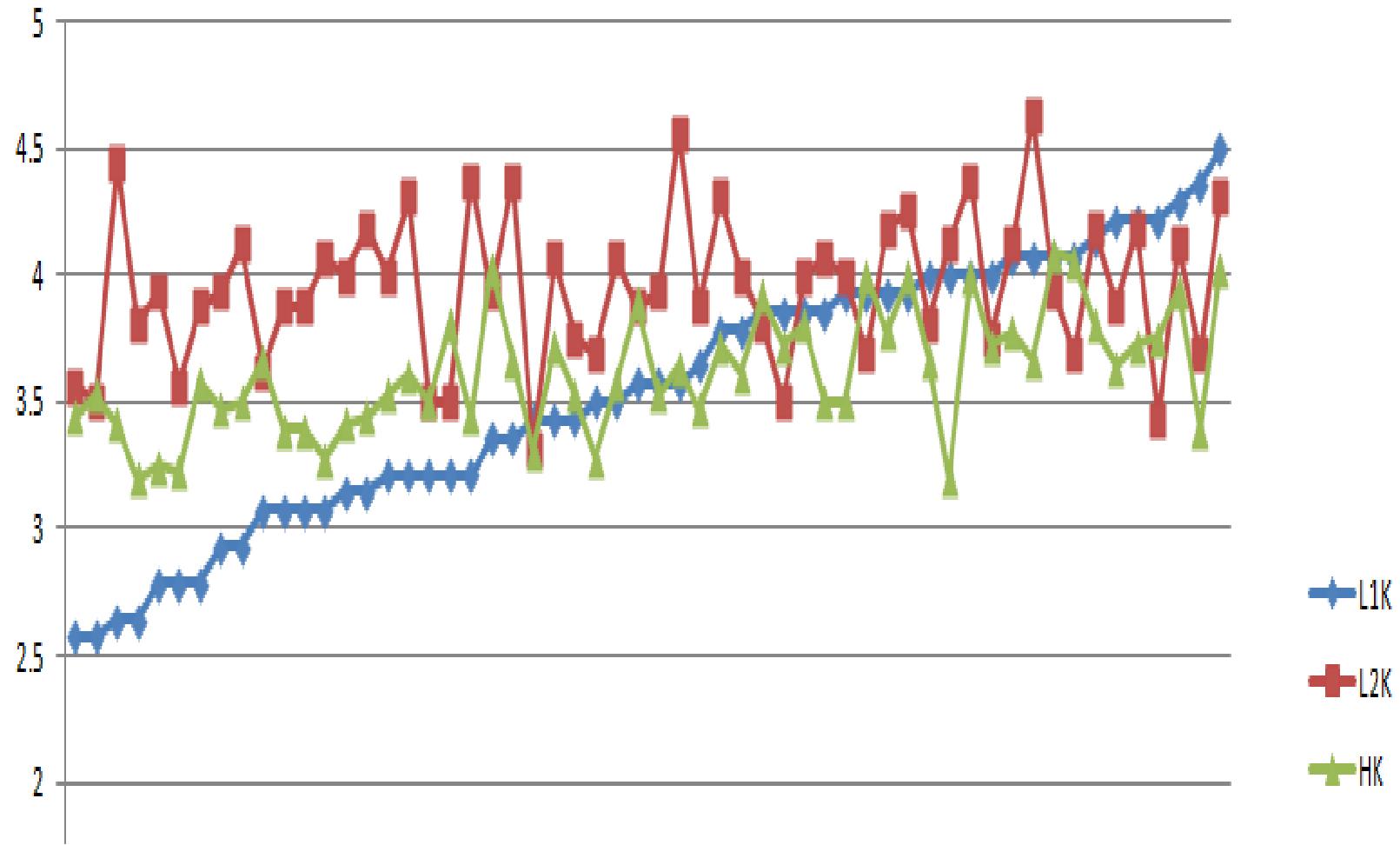
Korean Demographic Information

	L1 Group (N=14)		L2 Group (N=16)		Heritage Group (N=36)										
Background	<i>mean</i>	<i>range</i>	<i>mean</i>	<i>range</i>	<i>mean</i>	<i>range</i>									
—															
Age	21.7	17 – 25	25.8	20 – 36	24.5	18 – 38									
Age of arrival to US	NA	NA	NA	NA	3.9	0 – 19									
Age of switch to English	NA	NA	NA	NA	3.9	0 – 18									
% daily use of Korean	40.8	5 – 90	23.5	0 – 10	32.9	1 – 66									
1-5 Self-Ratings	%1	%2	%3	%4	%5	%1	%2	%3	%4	%5					
—															
Overall Korean Ability	7	14	21	14	43	6	12	38	25	19	3	11	6	11	69
Overall English Ability	0	7	7	21	64	0	6	0	25	69	8	3	3	22	64
English Reading	0	7	14	14	71	0	6	19	6	69	3	3	8	14	72
English Speaking	7	14	14	64	0	6	19	12	62	3	3	8	11	75	
English Listening	0	7	14	14	64	0	0	19	19	62	3	3	6	14	75
English Writing	7	0	14	7	71	0	0	25	6	69	3	3	8	11	75
Korean Reading	7	21	21	14	36	0	25	31	12	31	11	11	19	14	44
Korean Speaking	7	14	21	21	36	6	19	25	25	25	11	11	8	19	50
Korean Listening	7	14	21	21	36	0	19	31	25	25	8	8	14	19	50
Korean Writing	21	14	21	7	36	19	19	19	6	38	22	8	19	11	39

Japanese



Korean



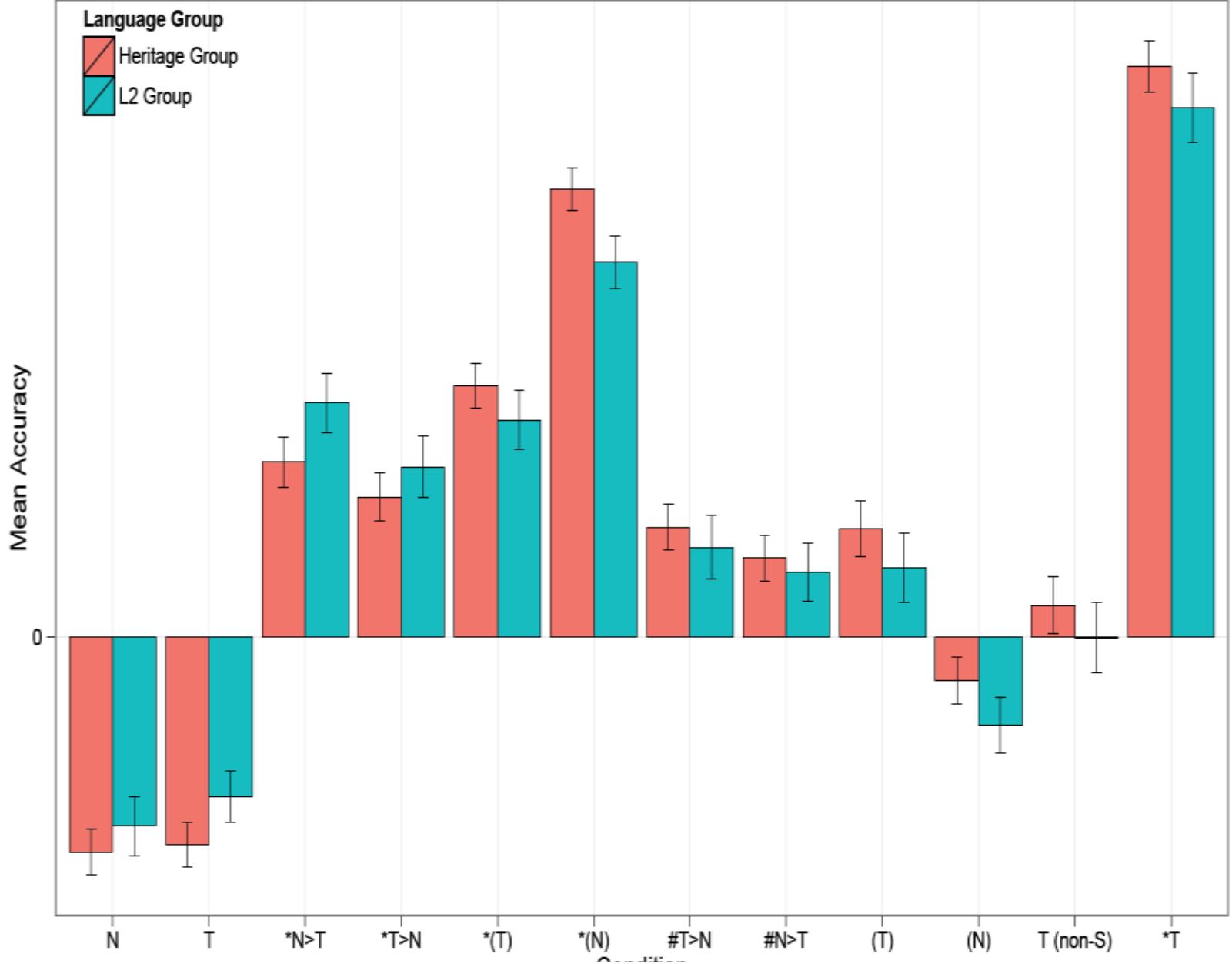
Japanese trouble spots

- Misuse of TOP (accepting the marker where it should not be used)
- Rules for omission
 - over-acceptance of incorrectly omitted markers
 - *omission of TOP ranked above *omission of NOM
 - acceptable omissions
 - (TOP) over-ranked and (NOM) under-ranked
- No effect of animacy on the use of markers (this effect is significant in LI group in condition 10, (NOM))

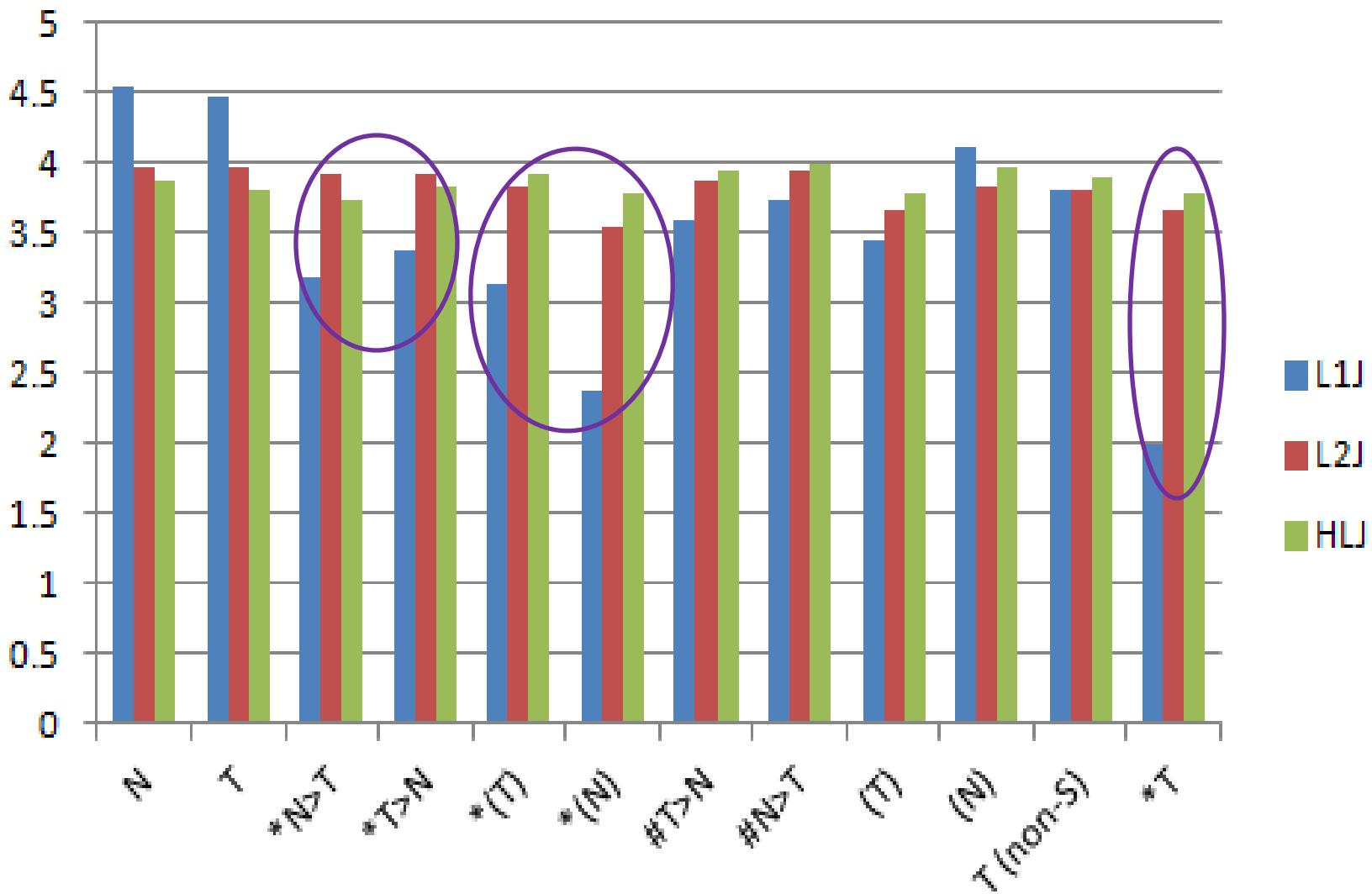
Japanese trouble spots

- With respect to the trouble spots, HLSs and L2 speakers are very much alike

Mean Accuracies by Condition (Japanese)



Japanese (means)



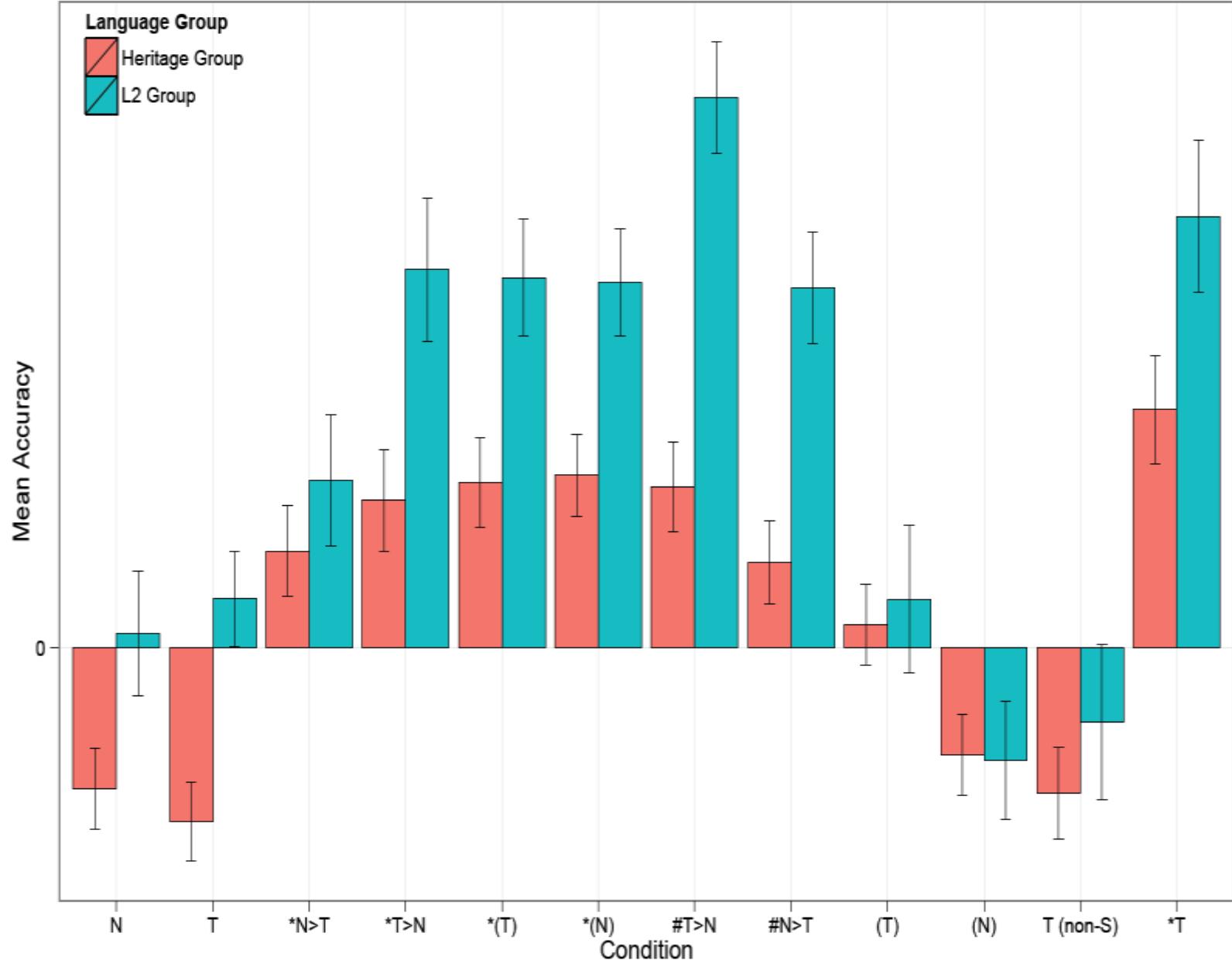
Korean trouble spots

- Less accuracy on TOP
- Particle omissions (accepting the omission where it should not occur)
- No effect of animacy on the use/omission of markers (esp. in Condition 10, (NOM))
- Most trouble spots in the L2 group

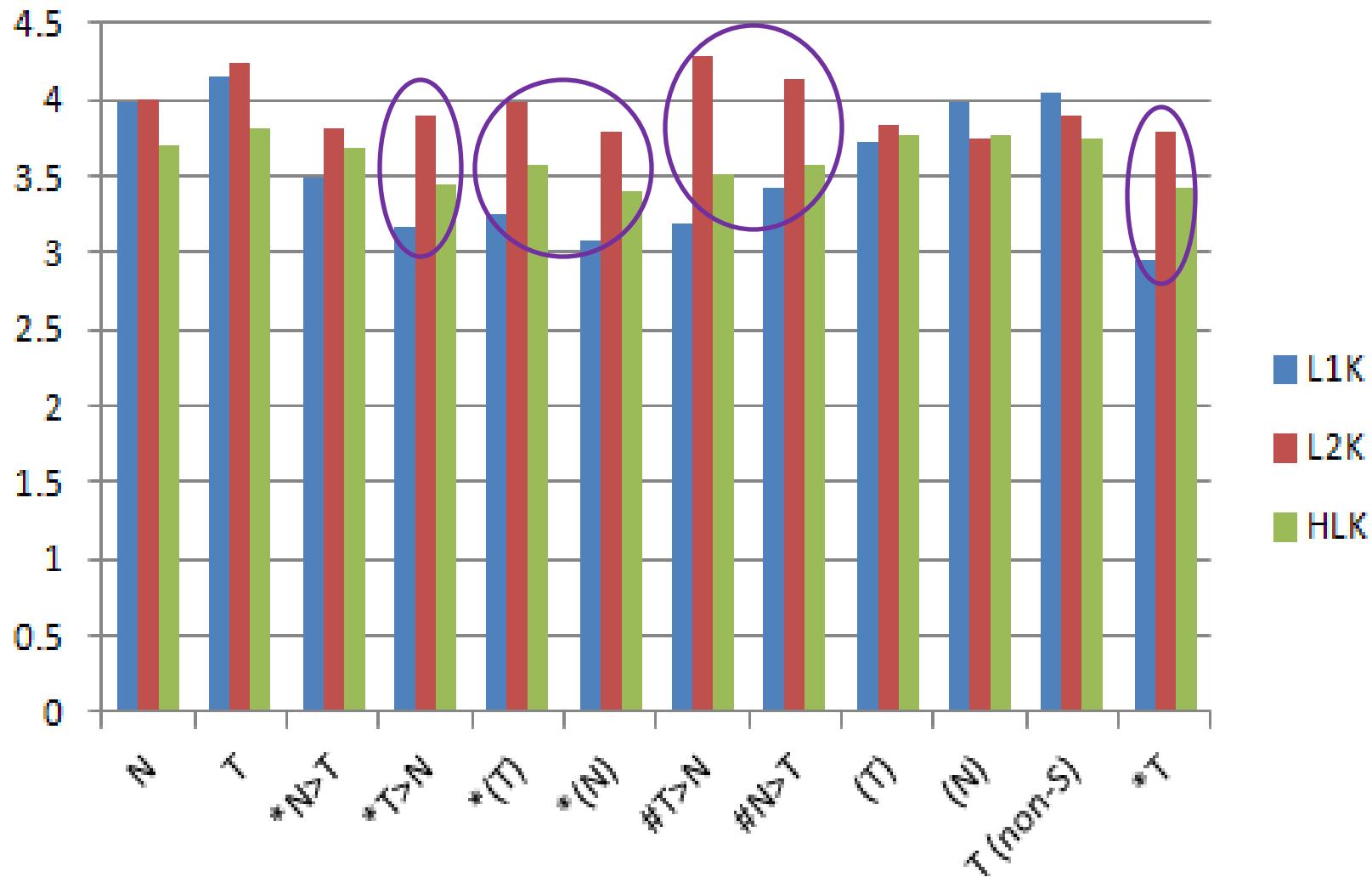
Korean trouble spots

- As compared to the Japanese group, HSs are closer to the baseline and are significantly different from L2 speakers
- However, HSs still differ from LI, especially in the TOP use

Mean Accuracies by Condition (Korean)



Korean (means)



Korean

- The greatest difference between HL and L2 is observed in conditions involving infelicity
 - HL significantly closer to baseline controls in recognizing pragmatically odd choices of particles
- Conditions for particle omission reveal an asymmetry between NOM and TOP in HL
 - knowledge of where NOM can and cannot be omitted
 - acceptable vs. unacceptable, $p=0.0000504$ (***)
 - no apparent knowledge of where TOP can and cannot be omitted
 - acceptable vs. unacceptable, $p= 0.69408$ (not sig.)

Heritage Japanese and Heritage Korean

• **WHY THE DIFFERENCE?**

Japanese and Korean in the US: two immigration scenarios

Languages spoken at home	1980	1990	2000	2007	Increase %
Japanese	336,318	427,657	477,997	458,717	36.4%
Korean	266,280	626,478	894,063	1,062,337	299%

% speak English	“very well”	“well”	“not well”	“not at all”
Japanese	53.8	29.1	15.7	1.4
Korean	41.8	29.3	24.2	4.7

- Source: American Community Survey 2007 (US Census)

Japanese and Korean in the US: Timeline differences

- Height of Japanese immigration to the US:
1868-1907
 - Longer assimilation
 - Repressive policies in WWII
 - Less compact settlements
- Height of Korean immigration to the US:
1965 and on (after the Hart-Celler Act)
 - More recent
 - Compact communities

Japanese and Korean in the US: Two immigration scenarios

- Community maintenance vs. assimilation:
 - Korean HLSs exhibit the highest rates of participation of HSs in a community or church school (72.3%) and involvement in community events (50.4%); HLs prioritize family/ cultural/ personal reasons--community actively involved in promoting the heritage language (Wiley, 2005; Cho, 2000)
 - Japanese HLSs prioritize professional goals, which outrank personal reasons, overall greater degree of assimilation into the American society
(based on the National Heritage Language Community Survey NHLCS)

Japanese and Korean in the US: Comparing populations

- “broad” and “narrow” conceptions of heritage language (Polinsky & Kagan, 2007)
 - **HL** = language with a cultural significance (cf. Fishman, 2001); may include some ‘overhearers’ (Au et al., 2002)
 - **L** = language spoken at home
- Japanese: **HL**, Korean: **L**

Japanese and Korean in the US: Comparing populations

- Japanese: **EOL**, Korean: **HL**
- As predicted, **HL** is more L2-like
- Question for the future: will the current positive attitude toward HLs help the maintenance of Korean as **HL** in the foreseeable future?



TAKING STOCK

Our initial questions

- (I) Are greater difficulties associated with the domain of morpho-syntax or discourse-pragmatics?
- Results:
 - less accuracy on TOP, hence more problems with semantics and discourse-pragmatics

Our initial questions

- (2) Overt or null?
- Results:
 - better accuracy on conditions involving overt markers compared to conditions involving omissions

What exactly matters?

- Integrating various types of knowledge across domains is more difficult than computing within one component of grammar
- Why? Unresolved question:
 - Interface
 - Structure-building and degree of embedding
 - ???
- This is testable: topics in embeddings (forthcoming experiment – stay tuned for the next Institute!)

What we have learned

- Language exposure at home does not guarantee advantage over L2 learners
- The umbrella term “heritage speaker” is very broad and may refer to different populations with different linguistic skills

What we have learned

- Cross-linguistic comparative studies of HLA should take into account possible variation in acquisition/maintenance scenarios for different communities
 - e.g., definitions of proficiency for testing particular theoretical claims

What we have learned

- Limitations of methodology: Mechanical Turk is not the best tool yet for studying “exotic” languages
 - MTurk has been widely used for English (Gibson et al., 2011)
 - Little control over subject populations, room for scam and fraud
 - Lack of familiarity with MT in Korea or Japan
- Researcher beware!
 - Prescreening, lang. competence; limiting IP addresses to specific countries; reviewing for red flags

Acknowledgments

- Sun-Hee Bae
- Shin Fukuda
- Sandy Kim
- Hiroki Nomoto
- Steve SanPietro
- Aika Taguchi
- National Heritage Language Resource Center
- Center for Advanced Study of Language
- National Science Foundation

THANK YOU