On Korean Speakers' Knowledge of Unaccusativity in English

Eunjeong Oh (Sangmyung University, eoh@smu.ac.kr)
Sanghoun Song (Incheon National University, sanghoun@inu.ac.kr)

December 2, 2017

Empirical Data

- While much early work has not been particularly empirical, over the years the field has undergone a shift towards empiricism.
- Many studies now use
  1. observational data (often using methods from corpus linguistics) or
  2. experimental data of varying degrees of complexity (often using methods from psycholinguistics).

1. Introduction

2. Background
   - Unaccusative Hypothesis
   - Biased Overpassivization of Unaccusatives
   - Semantic Notions Associated with Split Intransitivity
   - Learnability Problem

3. Corpus Study

4. Experimental Study

5. Conclusion

Two Types of Intransitives

- Intransitives are classified into two classes: unaccusatives and unergatives.
  - Syntactically, the single argument of unaccusatives is base-generated in object position whereas the single argument of unergatives originates in subject position.
  - Semantically, while the former bears a Theme role, the latter bears an Agent role.
  - Despite such differences, the single argument of these two types of intransitives surfaces in subject position, thereby being identical on the surface.
- L2 acquisition of unaccusativity
  - The unaccusative-unergative distinction is presumably universal, but languages vary as to the syntactic and morphological reflexes of such a distinction.
  - Given the cross-linguistic variation, a learnability problem naturally arises for the L2 acquisition of unaccusativity.

Goals

- This talk addresses Korean speakers' knowledge of unaccusativity/unergativity in L2 English.
  1. whether Korean speakers are sensitive to the unaccusative/unergative distinction in English.
  2. whether they are able to distinguish unaccusatives from transitives.
- Overpassivization of unaccusatives
  - Overpassivization refers to a phenomenon defined as non-target-like passivization of intransitives by L2 learners.
  - Interestingly, ungrammatical passive unaccusatives (e.g., **An accident was happened*) are frequently produced and judged as acceptable by learners from various L1 backgrounds.
  - Thus, these errors are language universal rather than language specific.
  - By contrast, unergatives are rarely passivized.

Unaccusative Hypothesis

- Intransitive verbs are of two classes; unaccusatives vs. unergatives.
  - Burzio, 1986; Perlmutter, 1978
- Uniformity of Theta Assignment Hypothesis (UTAH, Baker 1988)
  - The single argument of unaccusatives, bearing a Theme role, is base generated in object position, whereas the single argument of unergatives, bearing an Agent role, originates in subject position.
  - unaccusative: [*e* [arrived John]]
  - unergative: [John [$_{VP}$ ran]]
- Nonetheless, unaccusatives and unergatives are identical on the surface: both feature a subject and an intransitive verb – viz. the saliency of the NP-V word order.

Overpassivization

- Overpassivization (see Yip, 1990) is a phenomenon defined as nontargetlike or nonnativelike passivization of intransitive verbs by L2 learners.
- Ungrammatical (and contextually inappropriate) passive unaccusatives (e.g., *An accident was happened*) are produced and judged as acceptable by learners from various L1 backgrounds.
  - The errors are language universal rather than language specific.

Overpassivization (cont'd)

- Unlike unaccusatives, unergatives are rarely passivized.
  - This disparity suggests a clear contrast in the ways unaccusatives and unergatives are used and perceived by learners.
- Such nontarget behaviors are particularly noticeable among upper-intermediate, advanced, and even higher proficiency learners.
  - U-shaped development in the acquisition of unaccusatives

Transitivization Hypothesis (Yip, 1990; 1995)

- Unaccusatives are underlyingly transitives.
- Learners somehow interpret unaccusatives as underlyingly transitive because only transitive verbs allow passivization in English.
  1. the acceptance of ungrammatical transitives (e.g., **We disappeared our heads.*)
  2. rejection of correct unaccusatives (e.g., *Our heads disappeared.*)
  3. acceptance of ungrammatical passive unaccusatives (e.g., **Our heads were disappeared.*).
- Evidence for the hypothesis
  - L1 errors from Bowerman (1983) and L2 errors from Rutherford (1987)

    (1) a. *Do you want to see us disappear our heads? (L1)
    b. *The shortage of fuels occurred the need for economical engine. (L2)
    c. *This construction will progress my country. (L2)
  - Treating unaccusatives as transitives by adding objects to them; hence, unaccusatives can be passivized.

Transitivization Hypothesis (cont'd)

- Yip (1990, 1995) claimed that there are inherent similarities between unaccusatives and agentless passives.
  - Both are intransitives on the surface.
  - They have patient-role subjects.
- Predictions from the hypothesis
  - Acceptance of ungrammatical transitives (e.g., **We disappeared our heads.*)
  - Rejection of correct unaccusatives (e.g., *Our heads disappeared.*)
  - Acceptance of ungrammatical passive unaccusatives (e.g., **Our heads were disappeared.*)

Postverbal NP Movement Hypothesis (Zobl, 1989)

- Passive unaccusatives are not produced by syntactic movement from transitives.
- Unaccusatives are subsumed under passives (Marantz, 1984).
- Evidence for the hypothesis

(2) a. The children frazzled her nerves.
b. ?Her nerves frazzled.

(3) a. The drought damaged the crops.
b. *The crops damaged.

- Because learners did not move objects in (2a) and (3a) to the subject slot, which resulted in sentences such as (2b) and (3b) not being found, Zobl ruled out the possibility of the transitivization of unaccusatives.

Postverbal NP Movement Hypothesis (cont'd)

- Unaccusatives have a single argument structure
  - [__ [V NP]] (i.e., [__ [sink the ship]])
- Learners acquire a lexical rule by which the postverbal NP is moved to subject position.
- Once learners acquire the passive rule, the lexical rule is subsumed under the passive rule. This is because the English passive rule is the core rule for marking the movement of the object into the subject slot, hence the overpassivization, with unaccusatives acquiring ungrammatical passive morphological markings.

Limitations

- Both failed to address great disparity in between-verb variation (Ju, 1997)
  - The car disappeared. (80% incorrectly rejected)
  - The accident happened. (20% incorrectly rejected)
  - If overpassivization errors are generated purely by syntactic movement, why such different rates?
  - There may be additional factors underlying overpassivization errors.

Limitations (cont'd)

- Both view overpassivization as a problem in mapping relations between lexical semantics and syntax.
- A satisfactory account of the overpassivization phenomenon will have to hinge on syntax.
- Nevertheless, the claim that the processing of English unaccusatives involves lexical and syntactic mapping relations only is too strong; it is needed to factor other dimensions into purely syntactic models.
  - Lexio-semantic underpinnings of unaccusativity
  - Discourse pragmatics (Ju, 2000)

Split Intransitivity

- Following Levin and Rappaport Hovav (1995), it is assumed that unaccusativity is syntactically represented but semantically determined (cf. Perlmutter, 1978).
- The distinction between the two classes is semantically predictable and syntactically encoded.
- Split intransitivity has been associated with two semantic properties: agentivity and telicity.

Split Intransitivity (cont'd)

- Unaccusativity has been associated with non-agentivity, whereas unergativity has been mainly associated with agentivity.
  - The class of unaccusatives includes predicates whose argument is assigned the theta-role of Theme (Perlmutter, 1978)
  - The single argument of unaccusatives undergoes a change of state or location and has no control over the action denoted by the verb.
  - The argument of unergatives is assigned the theta- role of Agent, and as such, the subject has control over the action.

Split Intransitivity (cont'd)

- One more semantic difference between the two classes of intransitives is related to their aspectual value.
- Unaccusativity is mainly associated with telicity, whereas unergativity is associated with atelicity.
  - Unaccusative verbs denote mainly telic events, i.e., events which have a natural end point.
  - Unergatives denote mainly atelic situations, i.e., which do not have a natural end point.

Split Intransitivity (cont'd)

- Unergatives
  - They denote volitional acts.
  - Their argument is the Agent of the event.
  - Their argument has control over the event.
  - They denote mainly atelic events.
- Unaccusatives
  - They denote mainly non-volitional acts.
  - Their argument is never the Agent.
  - Their argument does not have control over the event.
  - They denote mainly telic events.

Split Intransitivity (cont'd)

- Animacy plays a key role in choosing voice forms (Croft, 1995).
  - Animate subjects are preferred in active voice, whereas inanimate subjects are preferred in passive voice.
- Shin (2011) reports that overpassivization errors with the two verbs appeared and died were not found by Korean learners of English, whereas overpassivization errors with the verbs occurred and happened persisted.
  - The verbs *appeared* and *died* (e.g., *John died*) can take animate subjects, while the verbs *occurred* and *happened* cannot (e.g., *The accident occurred*).

Learnability Problem

- Despite syntactic and semantic differences, unaccusatives and unergatives are identical on the surface.
- The unaccusative/unergative distinction is presumably universal, but languages vary as to the syntactic and morphological reflexes of such a distinction.
- Although L2 learners are aware of the unaccusative-unergative distinction from their L1, they have to figure out which semantic notion is relevant for such a distinction from the L2 input, and how it is syntactically manifested.

Collostructional Analysis

- a cognitive-linguistic toolbox by tweaking a corpus-linguistic method
- the use of statistical association measures to study collocations
  - the co-occurrence of words
  - the co-occurrence of grammatical patterns
  - the co-occurrence of constructions
- distributional hypothesis: the frequencies with which linguistic elements of interest co-occur with other linguistic/contextual elements

2×2 Contingency Table

	***e* is present**	***e* is absent**	**Totals**
***ty* is present**	a	b	$a+b$
***ty* is absent**	c	d	$c+d$
Totals	$a+c$	$b+d$	$a+b+c+d=N$

Fisher-Yates Exact

- Many association measures have been used in corpus-linguistic studies: MI, t, z, and Fisher-Yates Exacts (FYE).
- The negative log10 of the p-values of the FYE has been widely and reliably used.
- FYE
  - is an exact test (rather than asymptotic) which makes no distributional assumptions;
  - can therefore handle small and skewed frequencies better than, say, MI or chi-squared;
  - as a significance test, can distinguish between identical effect sizes by weighing those that are based on more data more heavily;
  - is not a linear function of the observed frequencies.

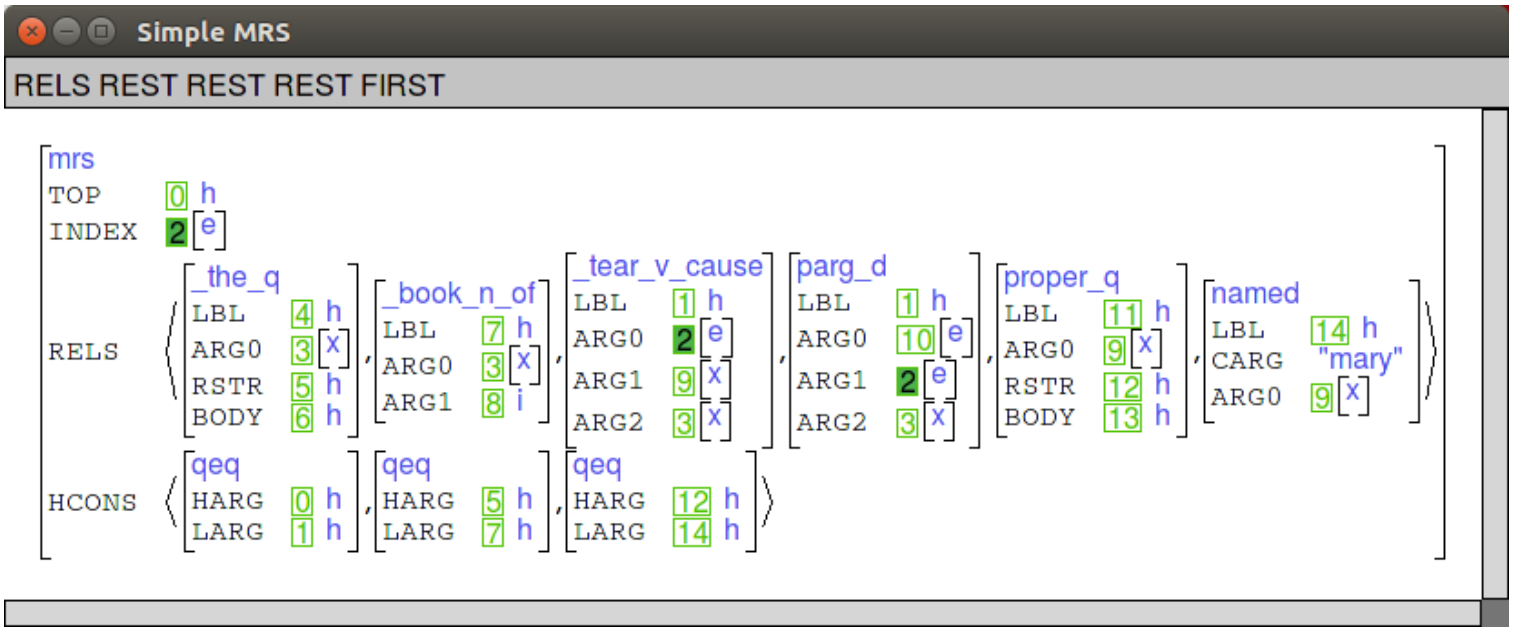
Corpora

- **EFL**: two development corpora
  - Gachon Learner Corpus
    - `http://thegachonlearnercorpus.blogspot.kr`
    - 2,507,899 words
  - Yonsei English Learner Corpus
    - Rhee and Jung (2014)
    - 1,082,295 words
- **COCA**: a reference corpus
  - Corpus Of Contemporary American-English (Davies, 2010)
  - randomly chosen 12 sections out of 125 sections
  - 36,643,094 words

Annotation

automatically construct, and then manually edit

- automatic pre-processing: ERG(MRS)+ACE
  - six CPUs, four days



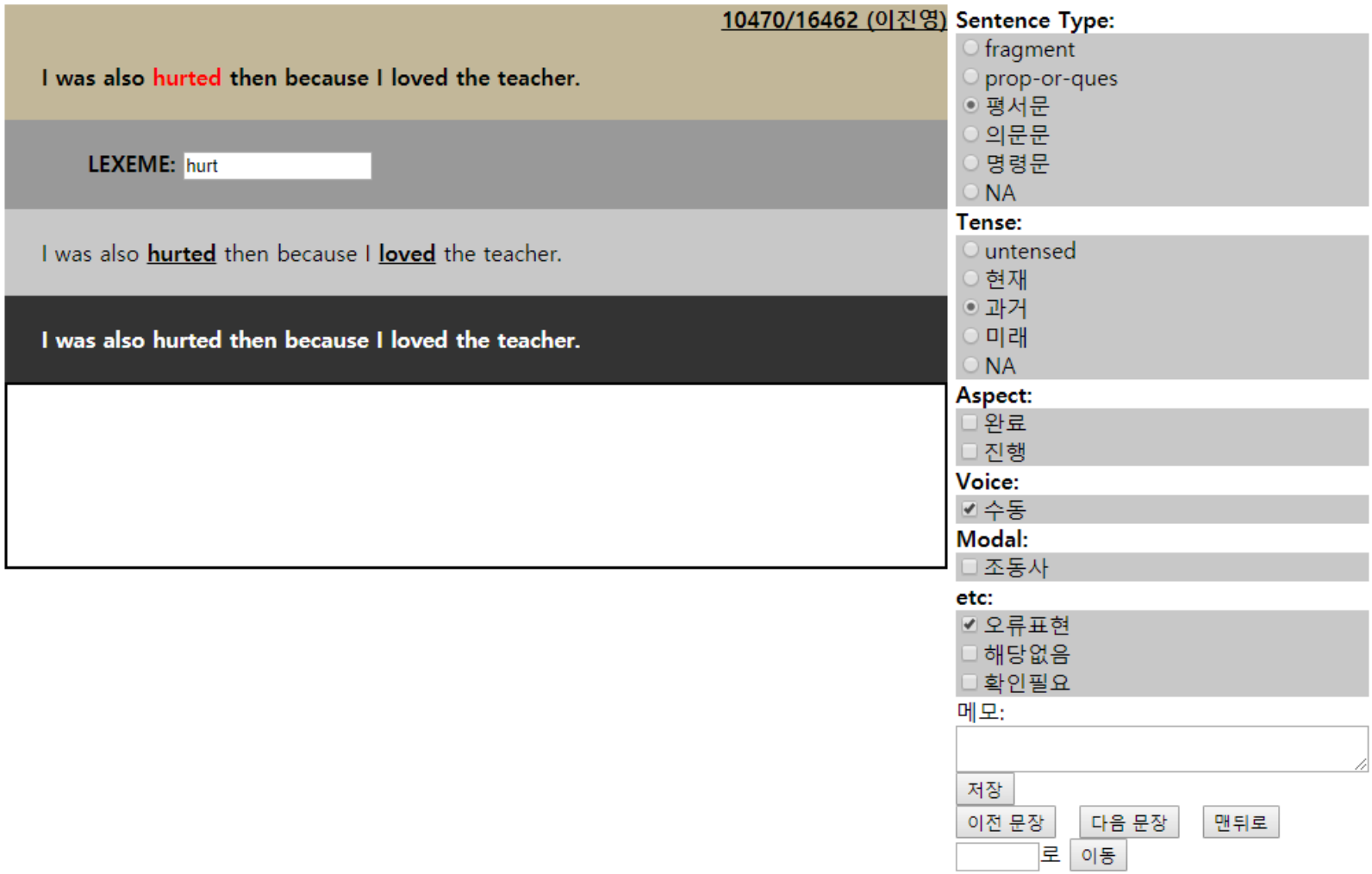
- manual post-tagging: online workbench
  - 10 annotators, four iterations

ERG(MRS)+ACE

- English Resource Grammar (Flickinger, 2000)
  - a broad-coverage precision HPSG for English
  - suitable for parsing, generation, and natural language understanding
  - *ver.* 1214
- Minimal Recursion Semantics (Copestake et al., 2005)
  - Meaning Representation System
- ACE (`http://sweaglesw.org/linguistics/ace`)
  - an efficient processor for DELPH-IN HPSG grammars
  - written in pure C and runs on the Linux and Mac OS X operating systems
  - distributed under the MIT License.

Online Workbench

Apache+PHP+MySQL



Data Points

	GLC	YELC	EFL	COCA
# of sentences	171,461	83,230	254,691	2,100,796
# of words	2,507,899	1,082,295	3,590,194	36,643,094
coverage (ERG)	71.44%	73.72%	72.03%	85.73%
# of finite verbs	258,244	106,927	365,171	1,968,523
# of passives	20,659	13,314	33,973	197,093
% of passives	8%	12.45%	9.37%	10.01%

Overpassivization

RANK	LEXEME	–log(p)	RANK	LEXEME	–log(p)
1	allow	Inf	16	prove	69.02206
2	bear	251.2714	17	influence	67.84252
3	develop	163.3705	18	form	66.21456
4	force	154.6765	19	fail	63.5936
5	**happen**	153.5315	20	**suffer**	58.30542
6	**die**	139.0201	21	**come**	57.5753
7	**appear**	125.0678	22	**exist**	56.28393
8	**occur**	112.2693	23	crowd	54.66814
9	**change**	108.4247	24	leak	52.93569
10	ban	105.1501	25	suppose	51.1637
11	**continue**	92.2981	26	**open**	49.23782
12	**go**	88.06512	27	relate	48.1266
13	**disappear**	84.08847	28	permit	48.07071
14	**break**	78.66791	29	**decline**	46.42306
15	**remain**	70.76212	30	**increase**	44.16889

Underpassivization

RANK	LEXEME	–log(p)	RANK	LEXEME	–log(p)
1	use	692.3265	16	commit	74.40425
2	thrill	276.3812	17	complete	71.48376
3	think	243.9063	18	find	69.96608
4	give	235.5362	19	understand	56.75307
5	drive	219.2896	20	study	56.40034
6	make	217.5622	21	eat	53.63739
7	know	204.7325	22	pay	51.62273
8	choose	125.9443	23	meet	47.08252
9	do	103.4137	24	kill	46.73978
10	ask	97.90421	25	follow	46.52597
11	frighten	97.48412	26	surprise	40.03886
12	excite	96.65247	27	select	39.77639
13	see	86.77317	28	copy	38.03153
14	amaze	76.61977	29	expect	37.77994
15	mean	75.021	30	shock	36.33821

Unergative Verbs

play, work, run, walk, cry, smile, sing, jump, swim, sweat, crawl, blush

- *play*: underpassivization
- *work*
  - RANK: 243
  - -log(p): 3.522283
- *cry*
  - RANK: 98
  - -log(p): 11.50295

by-Phrase

10.62% out of passives

LEXEME	FREQ	PASS	BY	LEXEME	FREQ	PASS	BY
happen	1,426	92	8	eat	4,699	28	5
die	761	72	8	love	2,239	45	16
occur	699	59	10	meet	1,648	18	1
appear	541	46	2	remember	667	18	1
exist	483	29	0	kill	614	107	42
suffer	329	28	6	catch	257	58	17
disappear	275	51	1	throw	236	21	3
remain	224	20	0	remove	155	20	2

by-Phrase (cont'd)

LEXEME	FREQ	PASS	BY	LEXEME	FREQ	PASS	BY
happen	1,426	92	8	work	1,642	21	3
die	761	72	8	run	581	22	8
occur	699	59	10	walk	479	6	0
appear	541	46	2	cry	275	10	0
exist	483	29	0	smile	232	1	0
suffer	329	28	6	jump	117	3	0
disappear	275	51	1	swim	71	0	0
remain	224	20	0	sweat	26	0	0

(4) a. This sitation was happened by the opposite people.

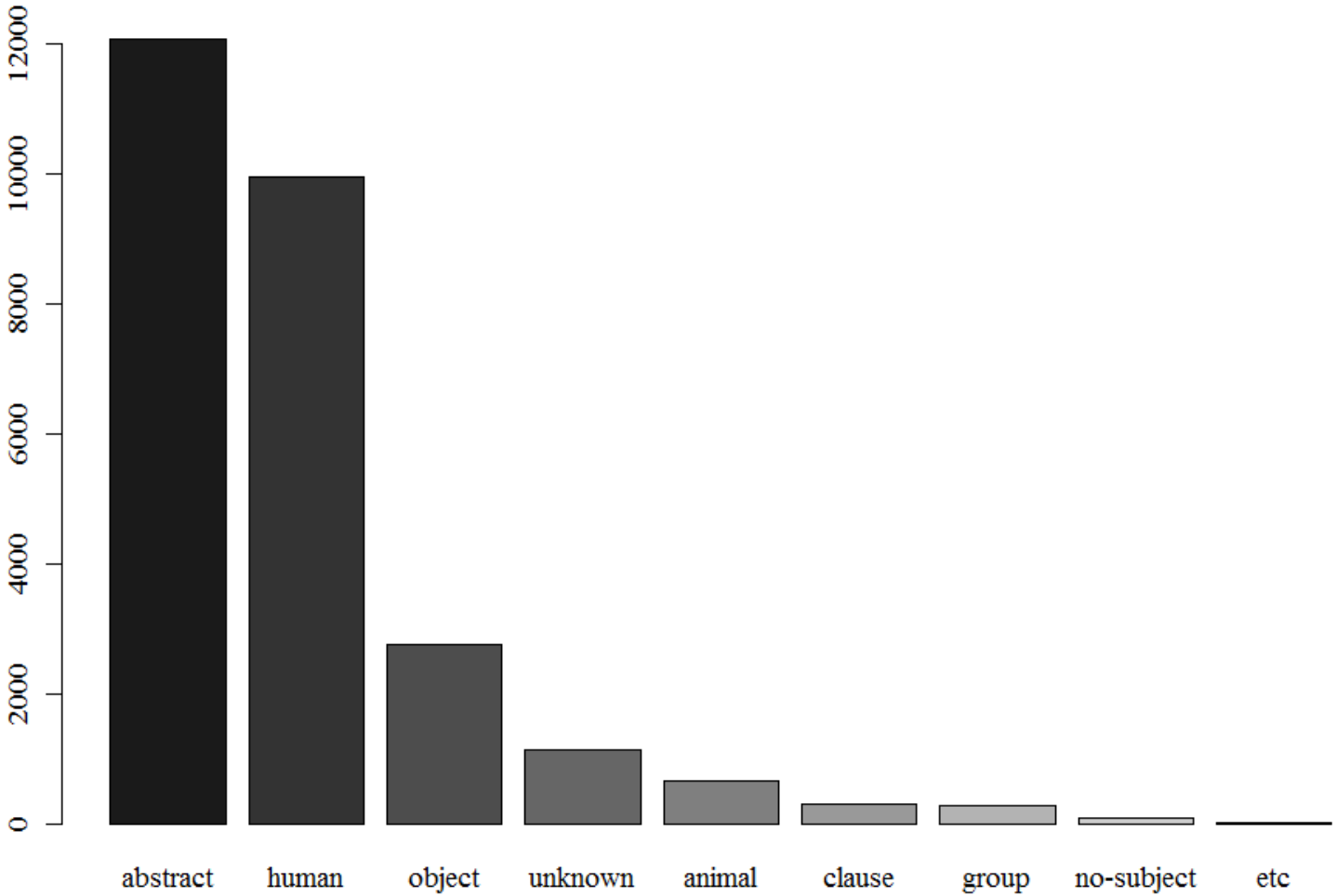
b. so they can be died by those diseases.

c. If so, traffic accidents that is occured by phoning driver will decrease.

d. Since I was young, I had been suffered by this for so long time.

e. It could be appeared by using fake name.

Types of Subjects



abstract	human	object	unknown	animal	clause	group
44.18%	36.38%	10.12%	4.20%	2.44%	1.17%	1.05%

Human Subjects

LEXEME	PASS	HS	%	LEXEME	PASS	HS	%
happen	92	3	3.26%	eat	28	4	14.29%
die	72	53	73.61%	love	45	25	55.56%
occur	59	1	1.69%	meet	18	12	66.67%
appear	46	2	4.35%	remember	18	2	11.11%
exist	29	3	10.34%	kill	107	74	69.16%
suffer	28	24	85.71%	catch	58	38	65.52%
disappear	51	5	9.8%	throw	21	5	23.81%
remain	20	1	5%	remove	20	2	10%

(5) a. We can be happened traffic accident especially on the highway.

b. many people have been died because of accidents.

c. the class is very beatiful place where problem-guys and rude children are not existed.

d. Because I have been suffered a hacking.

e. Because many beautiful women and handsome guys are appeared in TV.

Experimental Design

- A five-point Likert scale task (from 1 'least acceptable' to 5 'most acceptable')
- The order of presentation of test items was randomly assigned for each participant to avoid ordering effects.
- The toolkit used in the experiment was OpenSesame (Mathôt et al., 2012).
- Vocabulary translation task
  - To ensure participants' knowledge of the meanings of the test verbs.
  - Responses to those verbs that the participants correctly translated were included in the analysis.

Z-transformation

- The responses in the Likert scale task were Z-transformed per participant in order to alleviate the score bias.
- The Z-transformed values for each are mostly between –2 and 2 with a mean of 0 for each participant and a standard deviation of 1.
- A negative value means that the participant judges that the sentence sounds relatively unacceptable.

Participants

- 173 adult native Korean speakers
  - 31 beginners/59 low intermediate/52 high intermediate/ 31 advanced
  - Proficiency assessment: Michigan Test
- 27 native speakers of English served as controls

Test Items: Default

- Transitive

(6) a. A man built the house.
b. The boss reduced the cost.
c. A woman removed the file.
d. A boy caught the ball.
e. A man tested the product.
f. The singer released an album.
g. The woman collected stamps.

- Unergative

(7) a. The boy cried loudly.
b. The baker worked tirelessly.
c. The man walked slowly.
d. A kid ran quickly.
e. A baby crawled intently.
f. A student slept soundly.
g. A teacher smiled happily.

Test Items: Default (cont'd)

- Unaccusative (animate subjects)

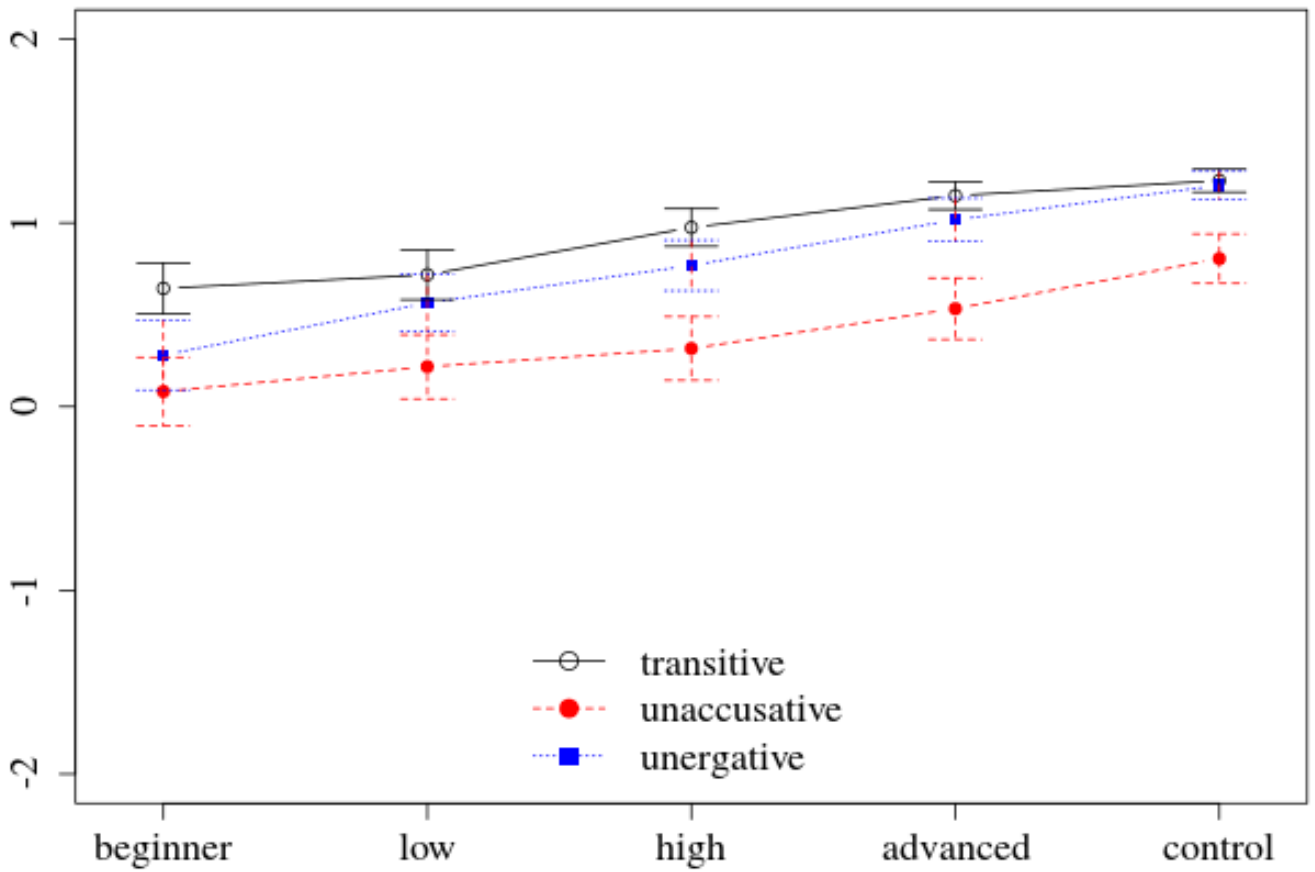
(8) a. The lady died painfully.
b. A boy appeared suddenly.
c. The man existed briefly.
d. A boy disappeared quickly.
e. The student remained stubbornly.
f. A worker vanished strangely.
g. The kid suffered tragically.

- Unaccusative (inanimate subjects)

(9) a. The tree died slowly.
b. A table appeared mysteriously.
c. The door existed previously.
d. A book disappeared occasionally.
e. The hat remained luckily.
f. A house vanished immediately.
g. The river suffered eventually.

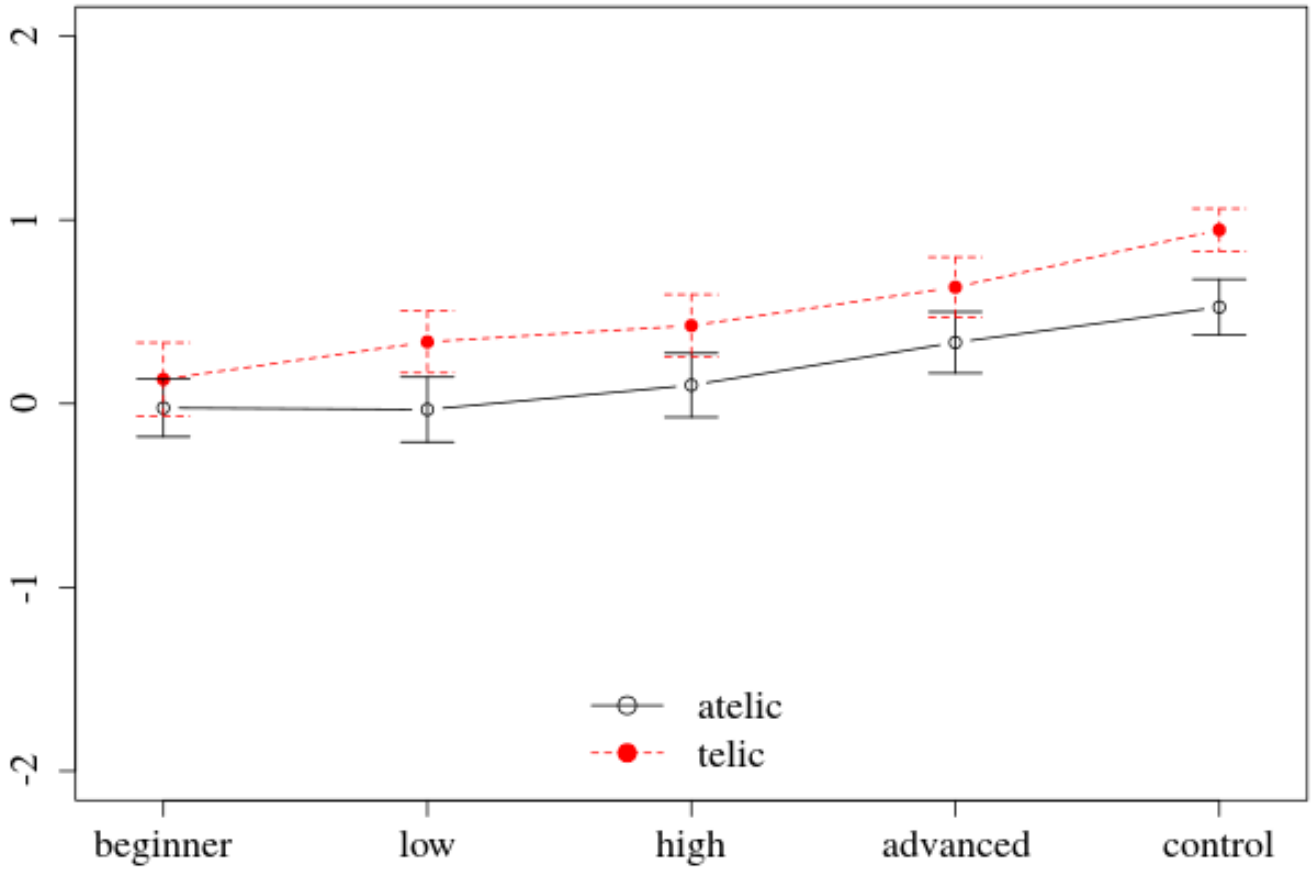
Default

- **TYPE:LEVEL** $F(4, 3394) = 2.345, p = 0.0525$



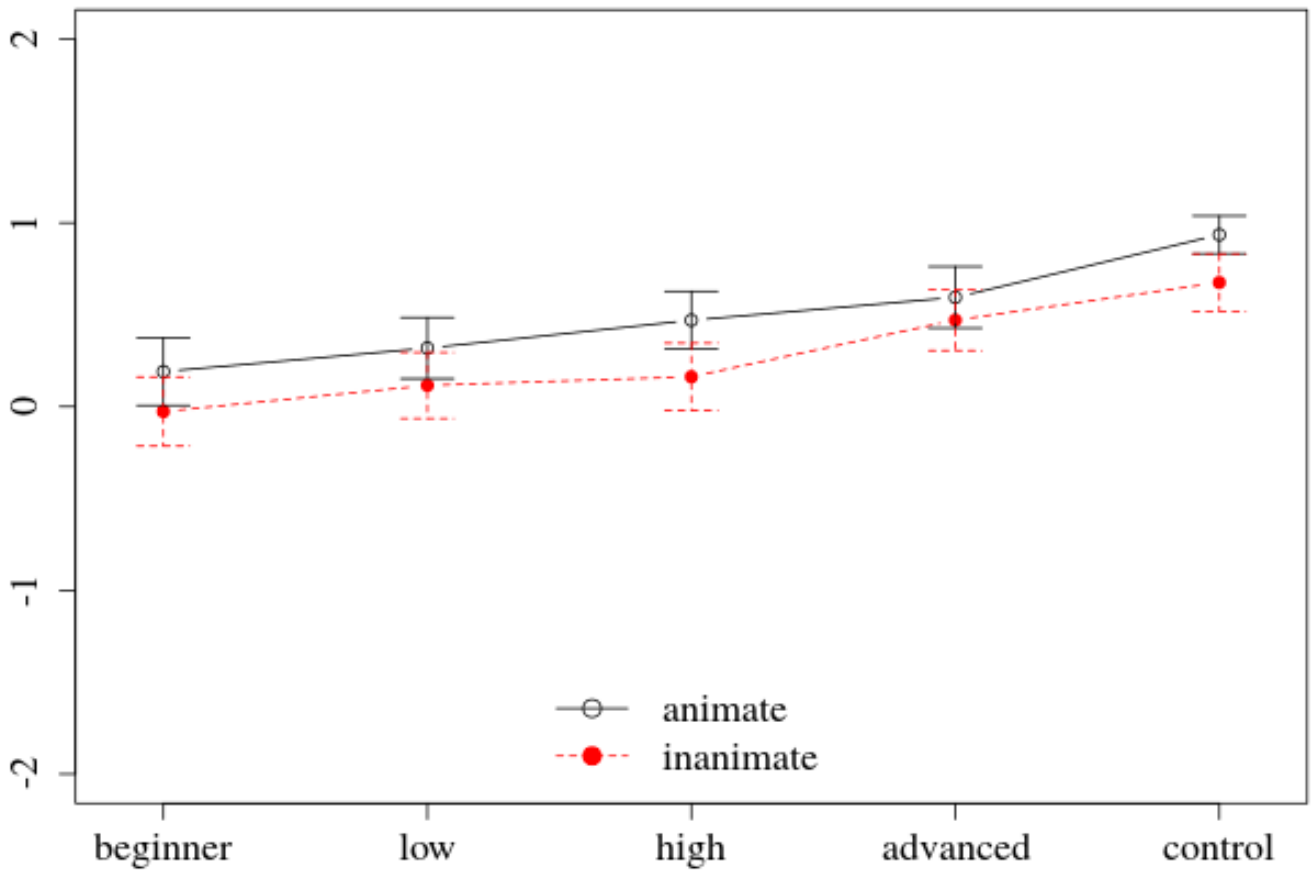
Default: Telicity

- **TELICITY:LEVEL** $F(4, 2198) = 0.922, p = 0.45$



Default: Animacy

- **ANIMACY:LEVEL** $F(4, 2198) = 0.726, p = 0.574$



Test Items: Passive

- Transitive

(10) a. A house was built.
b. The cost was reduced.
c. The file was removed.
d. The ball was caught.
e. The product was tested.
f. An album was released.
g. Stamps were collected.

- Unergative

(11) a. *A boy was cried.
b. *A baker was worked.
c. *A man was walked.
d. *A kid was run.
e. *A baby was crawled.
f. *A student was slept.
g. *A teacher was smiled.

Test Items: Passive (cont'd)

- Unaccusative (animate subjects)

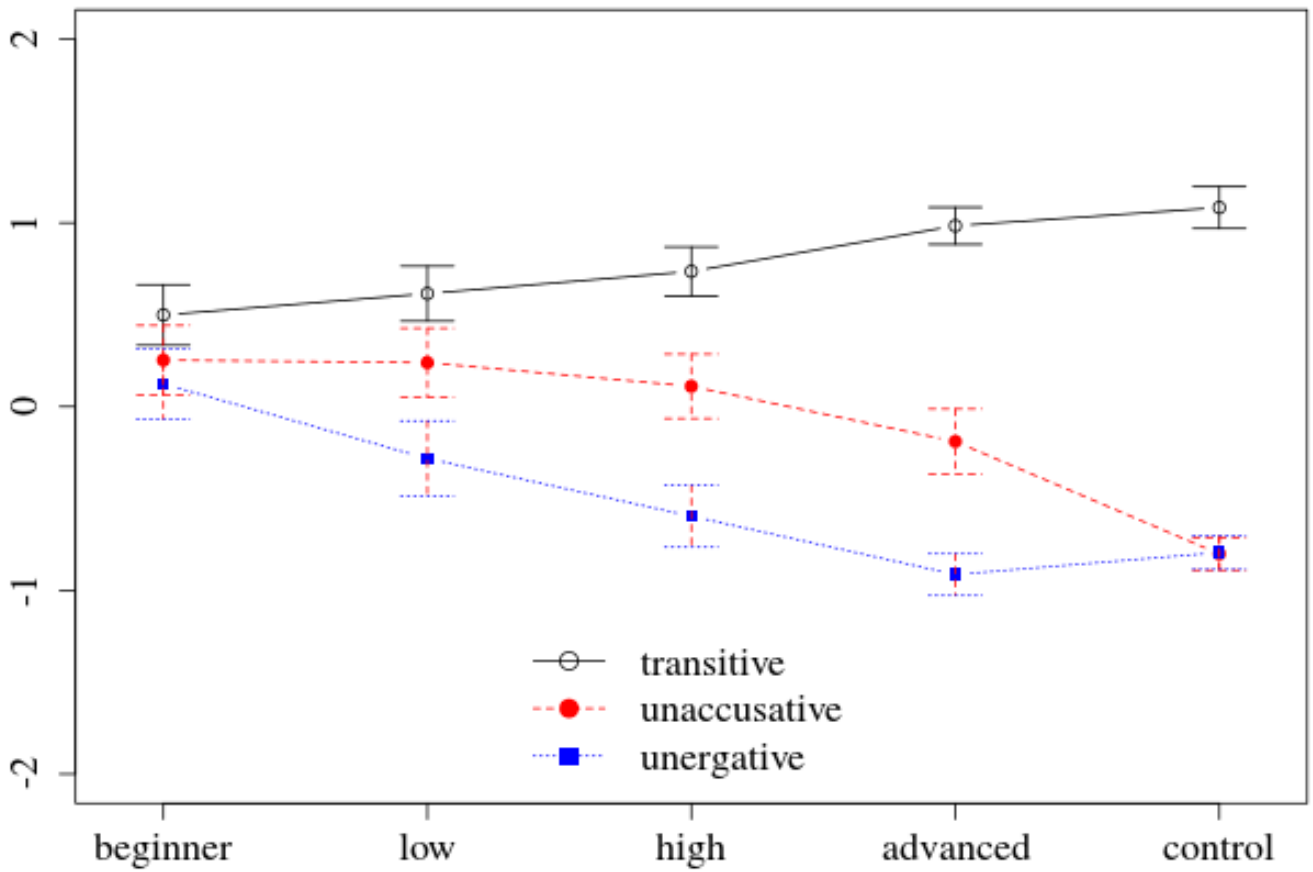
(12) a. *The lady was died.
b. *A boy was appeared.
c. *The man was existed.
d. *A boy was disappeared.
e. *The student was remained.
f. *A worker was vanished.
g. *The kid was suffered.

- Unaccusative (inanimate subjects)

(13) a. *The tree was died.
b. *A table was appeared.
c. *The door was existed.
d. *A book was disappeared.
e. *The hat was remained.
f. *A house was vanished.
g. *The river was suffered.

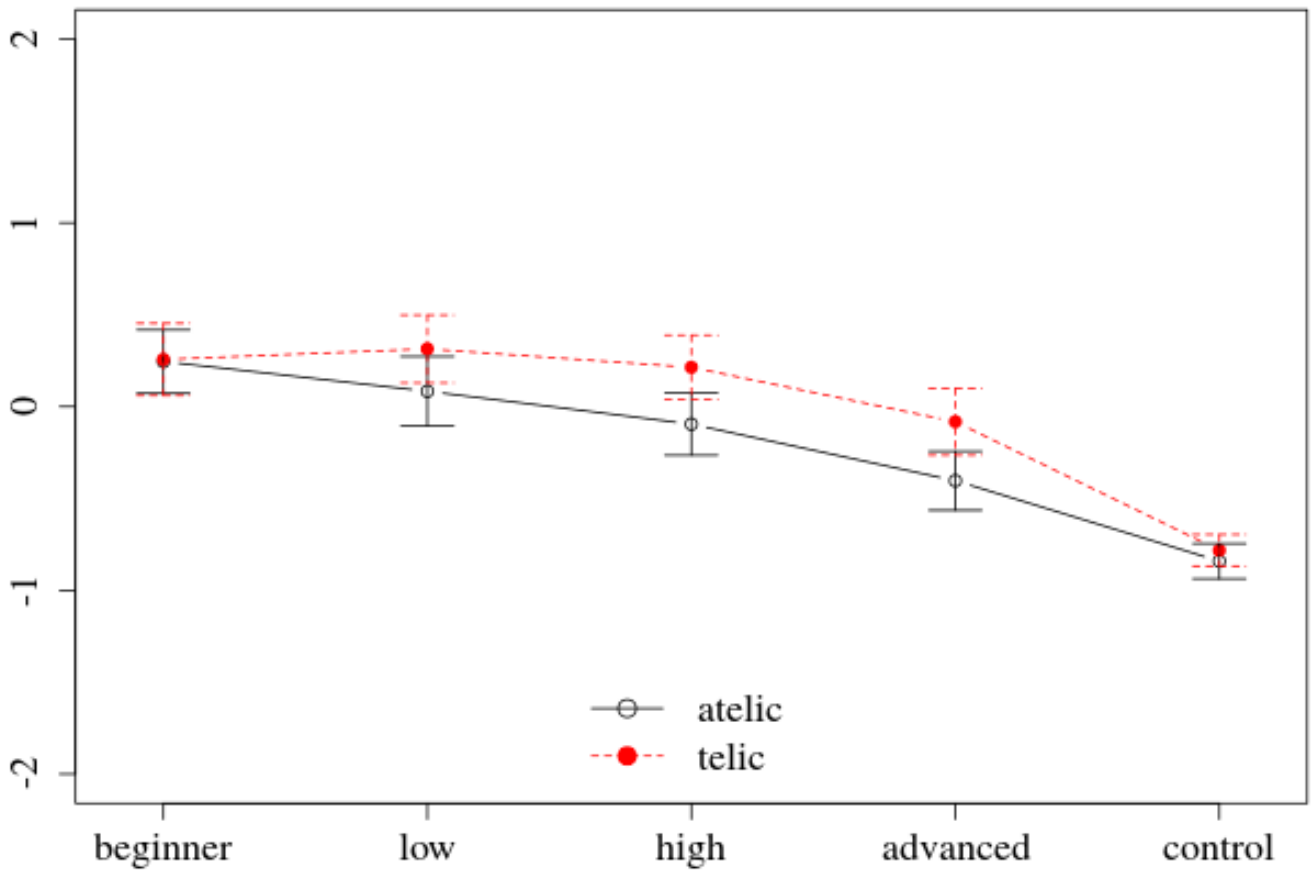
Passive

- **TYPE:LEVEL** $F(4, 3394) = 18.98, p < 0.001$



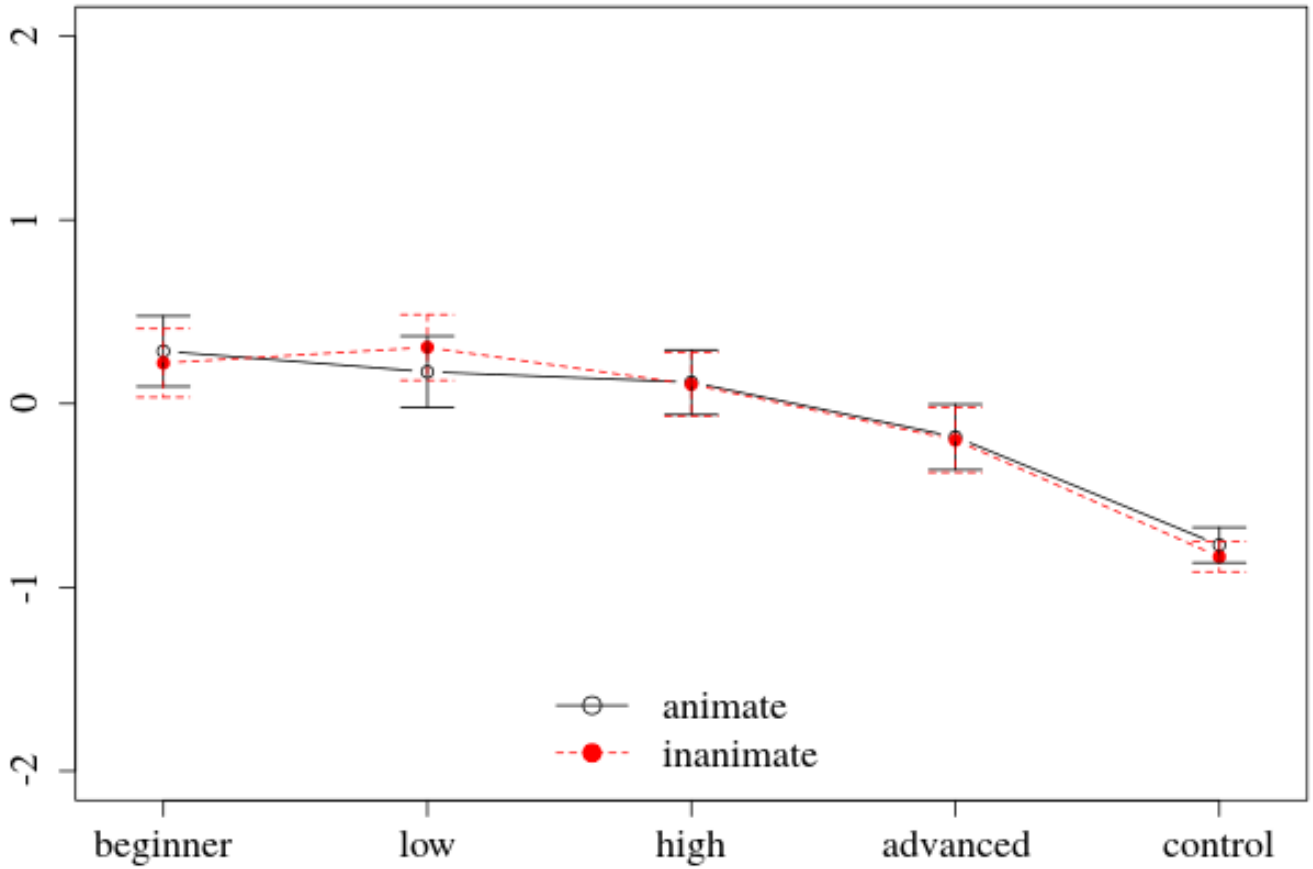
Passive: Telicity

- **TELICITY:LEVEL** $F(4, 2198) = 2.106, p = 0.0776$



Passive: Animacy

- **ANIMACY:LEVEL** $F(4, 2198) = 1.106, p = 0.352$



Test Items: Causative

- Unergative

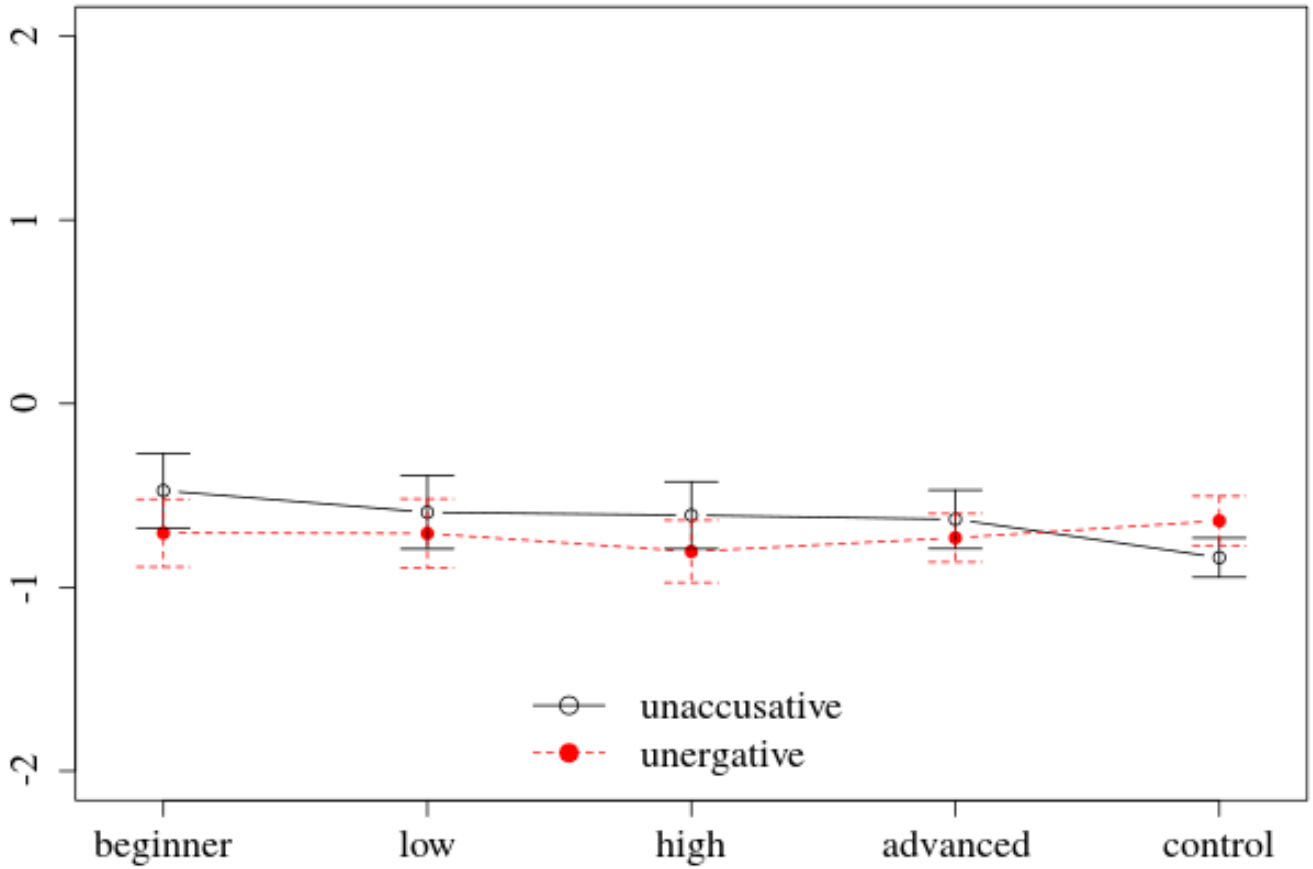
(14) a. *A boy cried a girl.
b. *A man worked a file.
c. *A woman walked a cart.
d. *The coach ran students.
e. *A mother crawled a baby.
f. *A teacher slept a student.
g. *A teacher smiled a class.

- Unaccusative

(15) a. *The man died a boy.
b. *The man appeared a rabbit.
c. *A man existed a book.
d. *A magician disappeared a bird.
e. *A woman remained the cookie.
f. *A boy vanished a house.
g. *A kid suffered a cat.

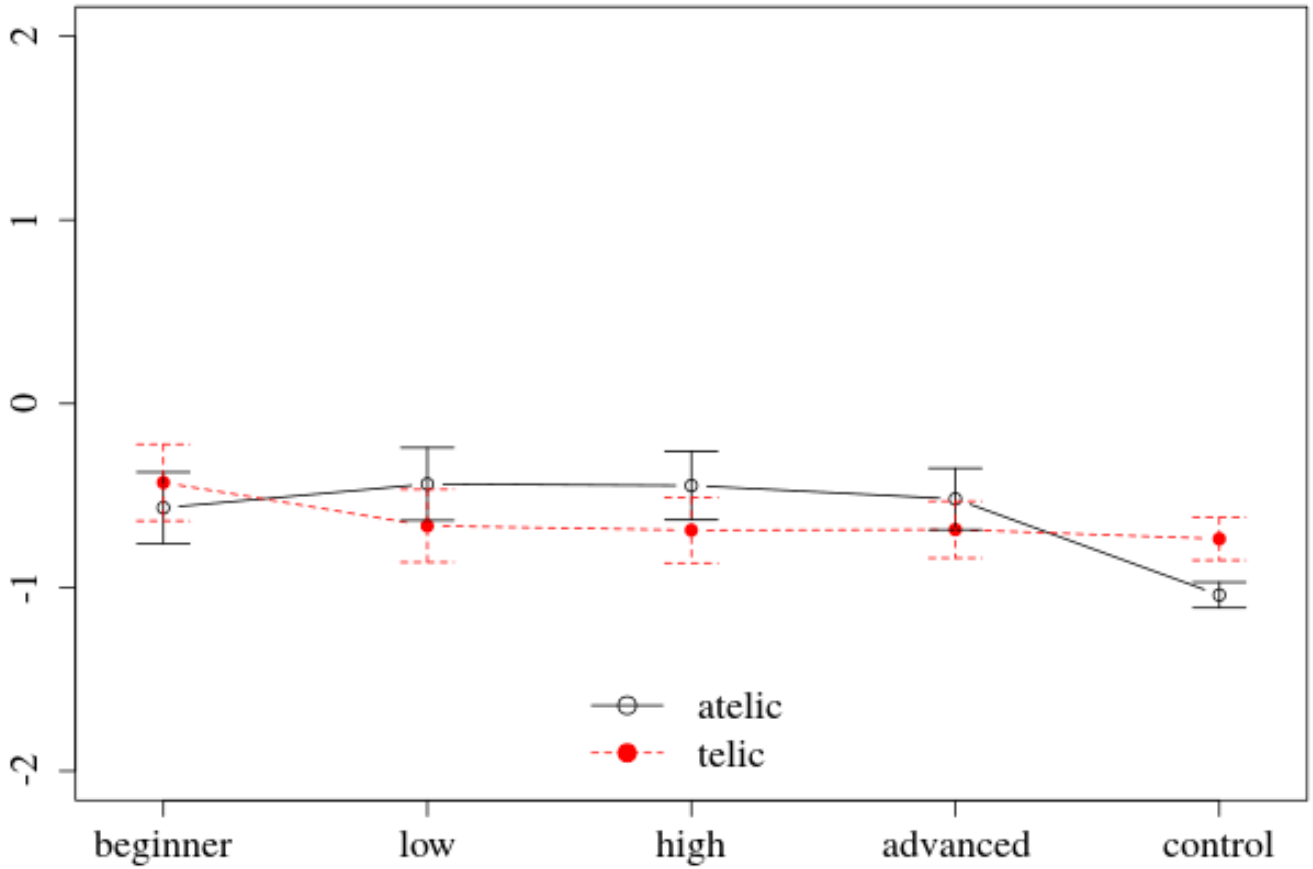
Causative

- **TYPE:LEVEL** $F(4, 2290) = 3.242, p = 0.01158$



Causative: Telicity

- **TELICITY:LEVEL** $F(4, 1094) = 3.152, p = 0.01372$



Test Items: Agent PP

- Unergative

(16) a. *A boy was cried by his older brother.
b. *A baker was worked by the boss.
c. *A man was walked by the doctor.
d. *A kid was run by the father.
e. *A baby was crawled by the mother.
f. *A student was slept by the teacher.
g. *A teacher was smiled by the student.

Test Items: Agent PP (cont'd)

- Unaccusative (animate subjects)

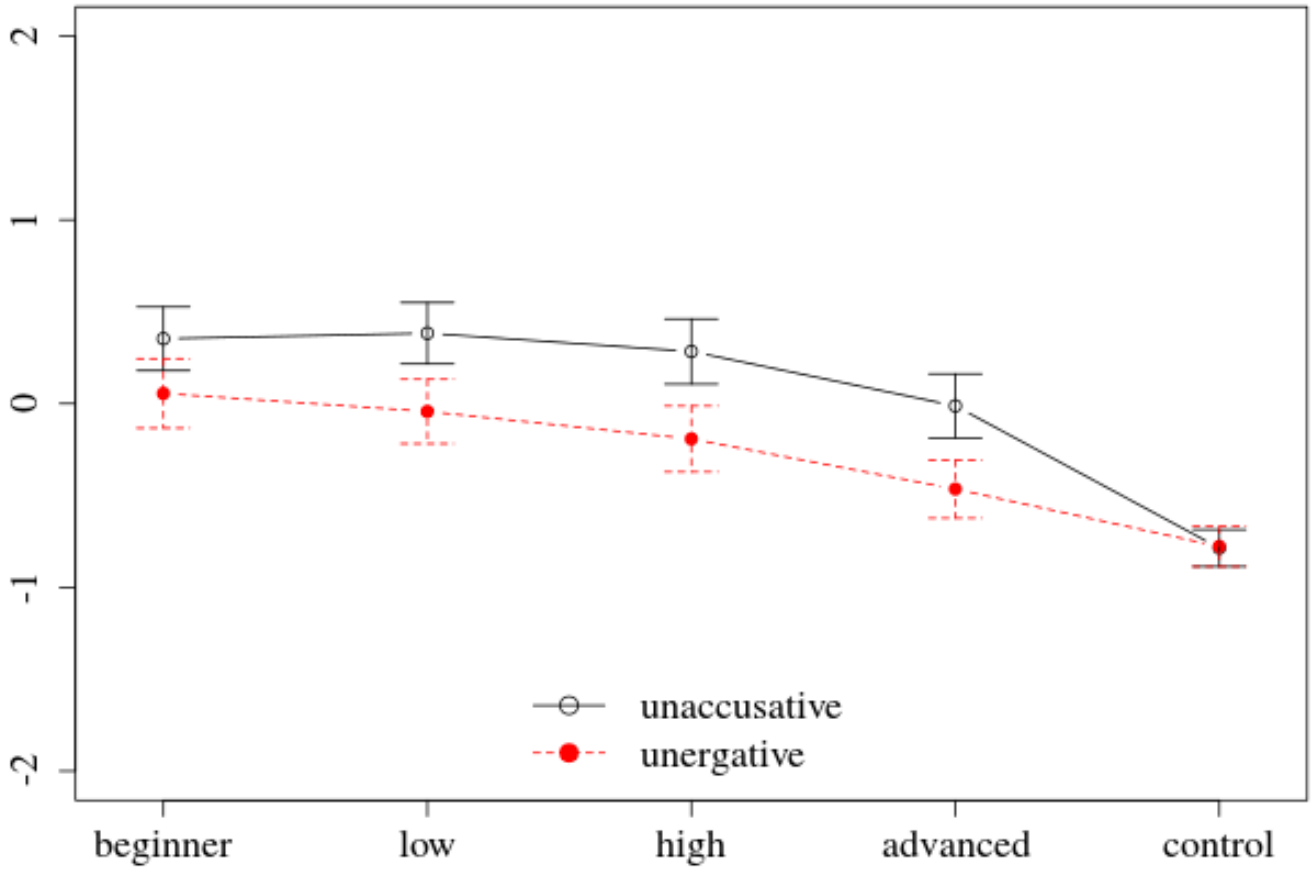
(17) a. *The lady was died by the disease.
b. *A boy was appeared by the father.
c. *The man was existed by his parents.
d. *A boy was disappeared by the magician.
e. *The student was remained by the teacher.
f. *A worker was vanished by the boss.
g. *The kid was suffered by the virus.

- Unaccusative (inanimate subjects)

(18) a. *The tree was died by the bugs.
b. *A table was appeared by the workers.
c. *The door was existed by the carpenter.
d. *A book was disappeared by the librarian.
e. *The hat was remained by the salesperson.
f. *A house was vanished by the construction company.
g. *The river was suffered by the pollution.

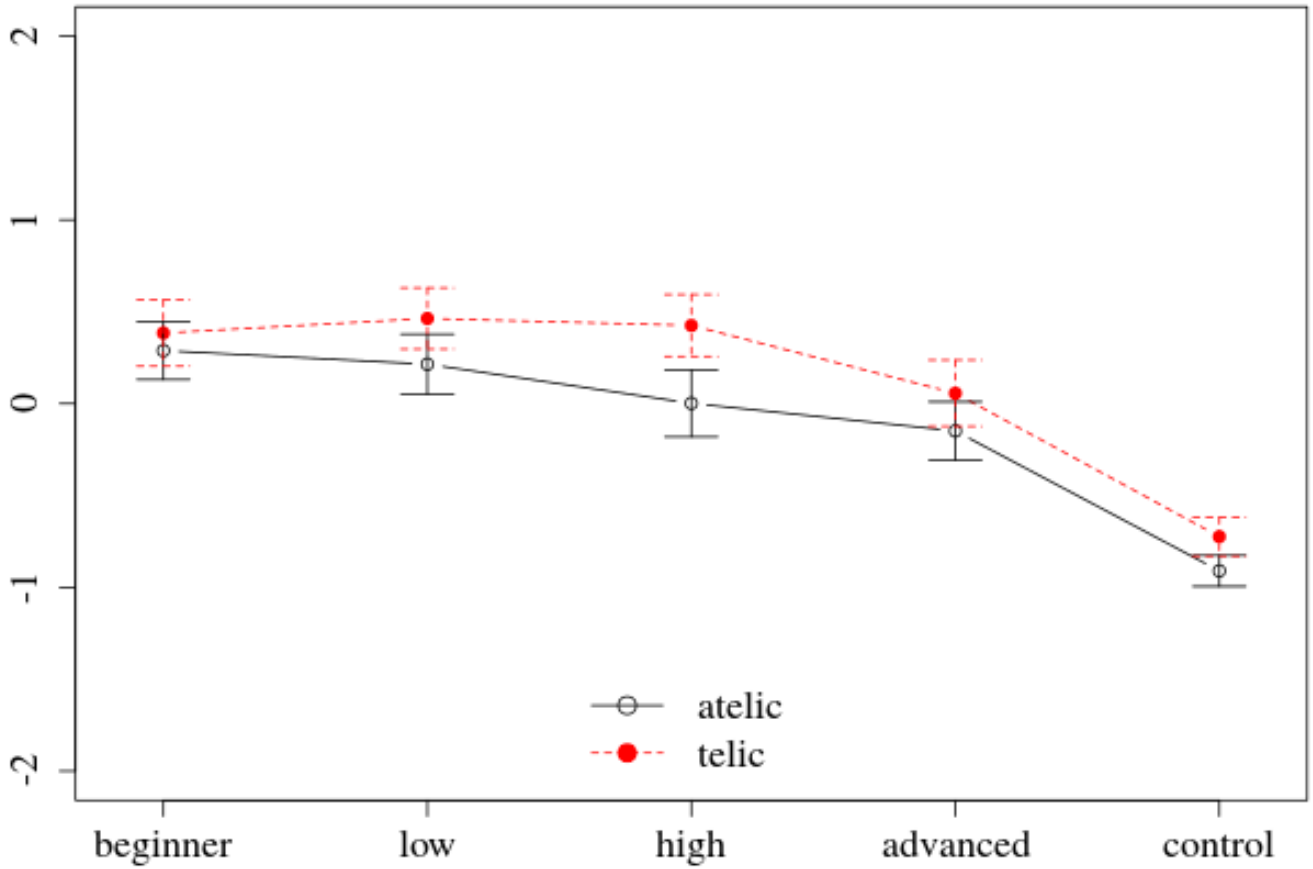
Agent PP

- **TYPE:LEVEL** $F(4, 3394) = 6.942, p < 0.001$



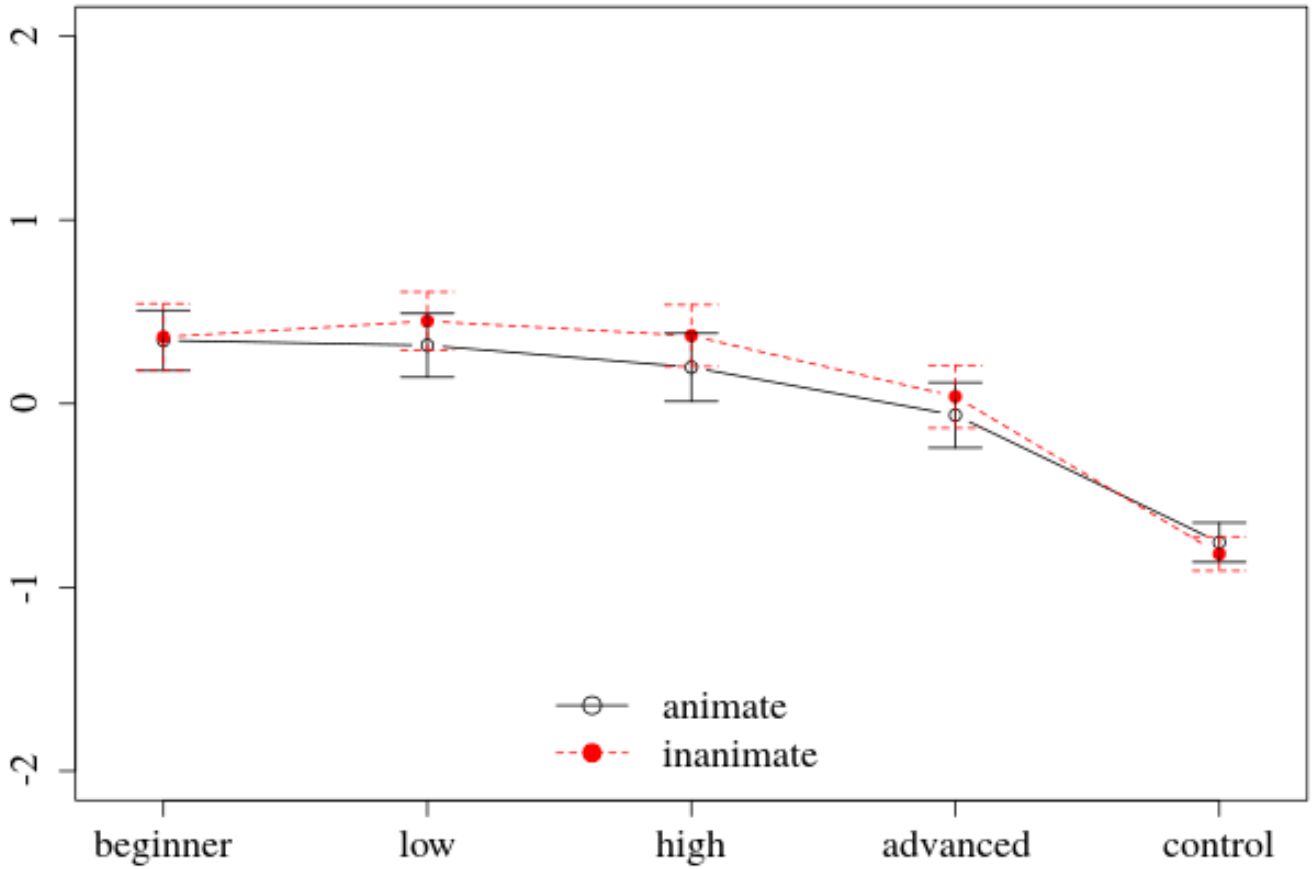
Agent PP: Telicity

- **TELICITY:LEVEL** $F(4, 2198) = 2.13, p = 0.0746$



Agent PP: Animacy

- **ANIMACY:LEVEL** $F(4, 2198) = 1.251, p = 0.28707$



Test Items: Purpose

- Unergative

(19) a. *A boy was cried to get his bottle.
b. *A baker was worked to finish the wedding cake.
c. *A man was walked to go to the store.
d. *A kid was run to catch the ball.
e. *A baby was crawled to play with a toy.
f. *A student was slept to rest for his soccer game.
g. *A teacher was smiled to make the students feel better.

Test Items: Purpose (cont'd)

- Unaccusative (animate subjects)

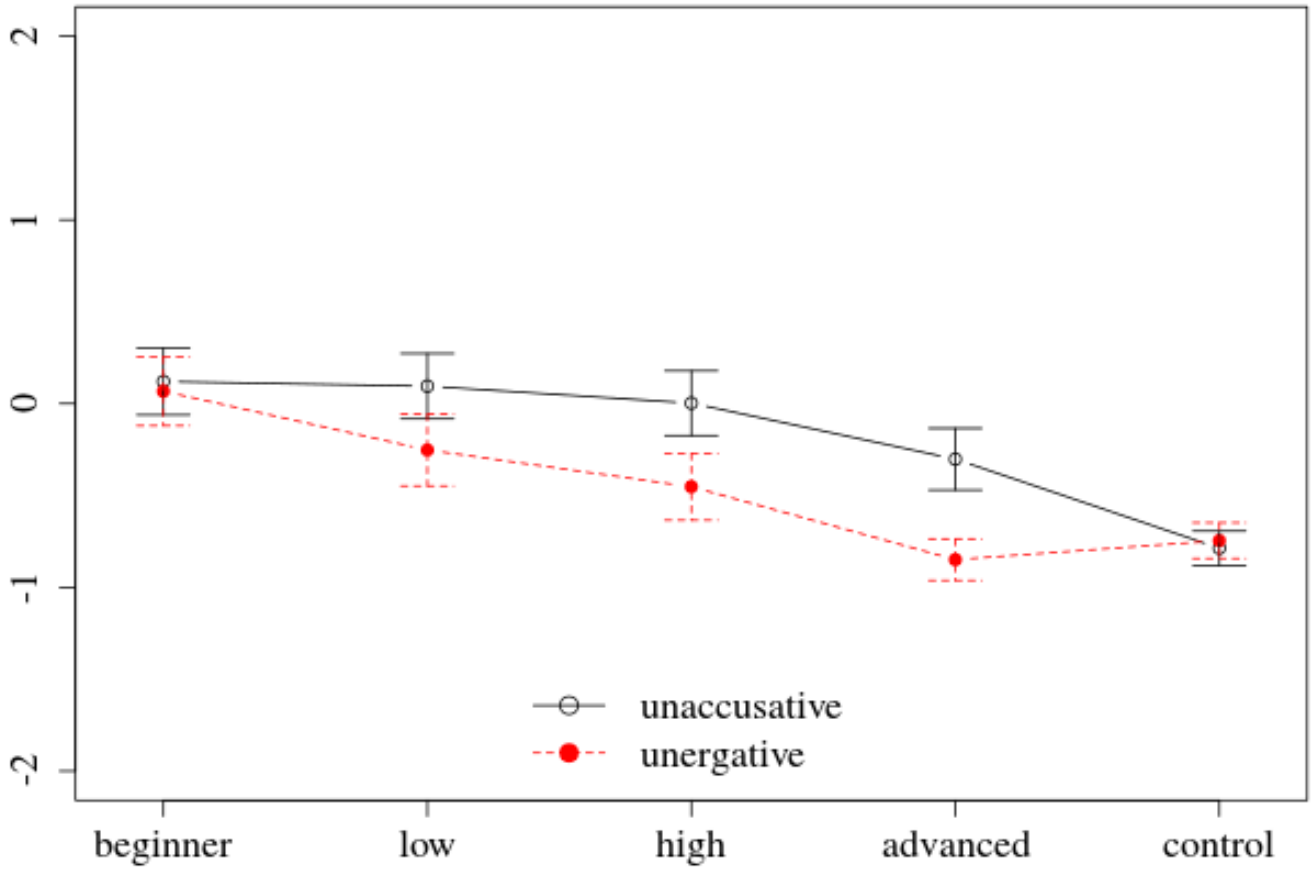
(20) a. *The lady was died to protect the family.
b. *A boy was appeared to eat a snack.
c. *The man was existed to serve his country.
d. *A boy was disappeared to play a game.
e. *The student was remained to study for the test.
f. *A worker was vanished to eat her lunch.
g. *The kid was suffered to punish his bad behavior.

- Unaccusative (inanimate subjects)

(21) a. *The tree was died to build a house.
b. *A table was appeared to serve the meal.
c. *The door was existed to connect the hotel rooms.
d. *A book was disappeared to clean off the desk.
e. *The hat was remained to decorate the snowman.
f. *A house was vanished to build a highway.
g. *The river was suffered to build a dam.

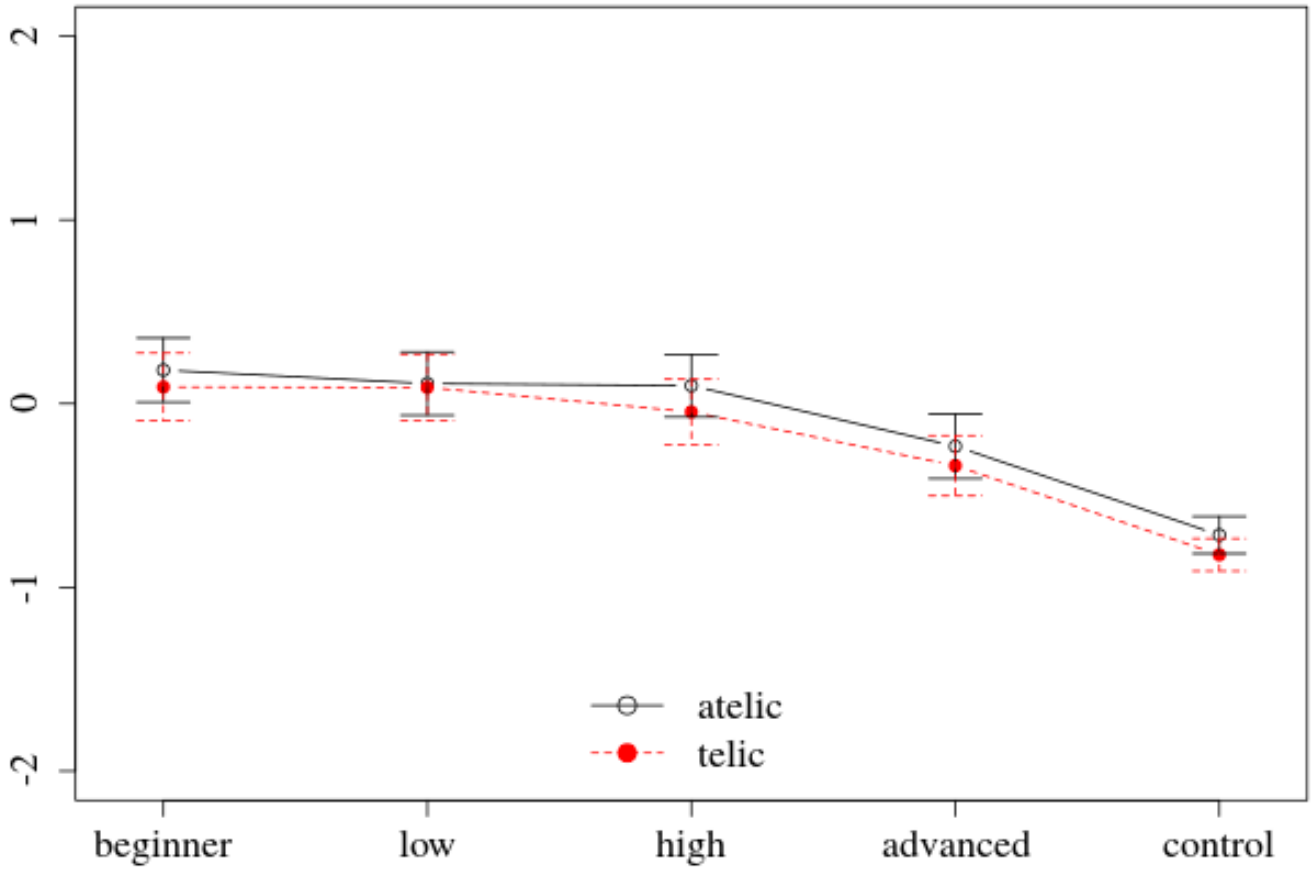
Purpose

- **TYPE:LEVEL** $F(4, 3394) = 11.01, p < 0.001$



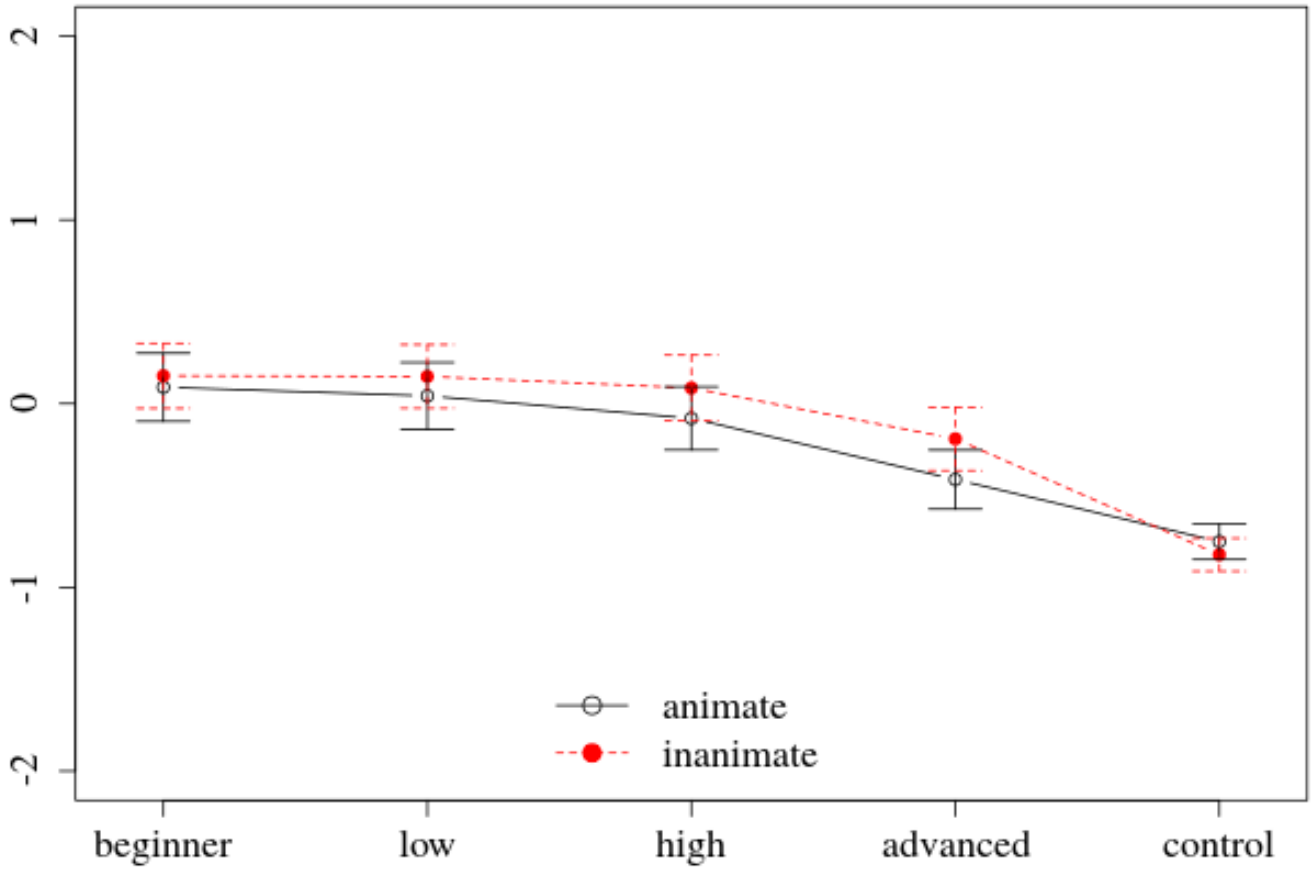
Purpose: Telicity

- **TELICITY:LEVEL** $F(4, 2198) = 0.380, p = 0.8230$



Purpose: Animacy

- **ANIMACY:LEVEL** $F(4, 2198) = 1.534, p = 0.18972$



Summary: ANOVA (Interaction)

	Default	Passive	Causative	Agent PP	Purpose
Type	.	***	*	***	***
Telicity		.	*	.	
Animacy			NA		

Findings

1. Korean participants managed to distinguish unaccusatives from unergatives in English.
2. Korean participants seemed to differentiate between transitives and unaccusatives, which suggests that they are unlikely to perceive unaccusatives as underlying transitives.
3. Both experimental and corpus results seem to indicate that telicity, not animacy, is one semantic factor, which guides Korean participants' acquisition of unaccusativity in English.

Conclusion

- Experimental study, coupled with corpus study, helps us to better characterize Korean speakers' knowledge of unaccusativity in English.
- Future work should investigate the role of telicity in the process of acquiring unaccusativity in English by Korean speakers.
- More research is needed to thoroughly explore how transitivity and unaccusativity are differentiated by Korean speakers.