

Using the DELPH-IN resources for second language acquisition studies: Overpassivization

Sanghoun Song
(Incheon National University, sanghoun@inu.ac.kr)

June 20, 2018

- 1 Introduction
- 2 Background
- 3 Corpus Study
- 4 Experimental Study
- 5 Findings

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**.

Goals

- This study 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
 - Interestingly, **ungrammatical passive unaccusatives** (e.g., **An accident was happened*) are frequently produced and judged as acceptable by learners from various L1 backgrounds.
 - By contrast, unergatives are rarely passivized.

- 1 Introduction
- 2 Background**
- 3 Corpus Study
- 4 Experimental Study
- 5 Findings

Overpassivization

- Overpassivization (see Yip, 1990) is a phenomenon defined as **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.
- Unlike unaccusatives, unergatives are rarely passivized.

Transitivity Hypothesis (Yip, 1990; 1995)

- Learners somehow interpret unaccusatives as **underlyingly transitive** because only transitive verbs allow passivization in English.
- 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**.
- 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.

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**.
 - Unaccusativity has been associated with **non-agentivity**, whereas unergativity has been mainly associated with agentivity.
 - Unaccusativity is mainly associated with **telicity**, whereas unergativity is associated with atelicity.

Animacy

- 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*).

- 1 Introduction
- 2 Background
- 3 Corpus Study**
- 4 Experimental Study
- 5 Findings

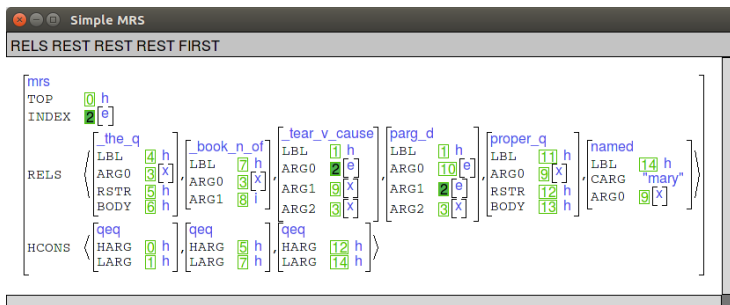
Corpora

- **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: $\text{ERG}(\text{MRS}) + \text{ACE}$
 - `parg_d`



- 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

*I was also **hurt** then because I loved the teacher.*

<p>10470/16462 (이진영)</p> <p>I was also hurt then because I loved the teacher.</p>	<p>Sentence Type:</p> <ul style="list-style-type: none"> <input type="radio"/> fragment <input type="radio"/> prop-or-ques <input type="radio"/> 평서문 <input type="radio"/> 의문문 <input type="radio"/> 명령문 <input type="radio"/> NA
<p>LEXEME: <input type="text" value="hurt"/></p>	<p>Tense:</p> <ul style="list-style-type: none"> <input type="radio"/> untensed <input type="radio"/> 현재 <input type="radio"/> 과거 <input type="radio"/> 미래 <input type="radio"/> NA
<p>I was also hurt then because I loved the teacher.</p>	<p>Aspect:</p> <ul style="list-style-type: none"> <input type="checkbox"/> 완료 <input type="checkbox"/> 진행
<p>I was also hurt then because I loved the teacher.</p>	<p>Voice:</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> 수동
<div style="border: 1px solid black; height: 140px;"></div>	<p>Modal:</p> <ul style="list-style-type: none"> <input type="checkbox"/> 조동사
	<p>etc:</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> 오류표현 <input type="checkbox"/> 해당없음 <input type="checkbox"/> 확인필요
	<p>메모:</p> <div style="border: 1px solid gray; height: 30px; margin-bottom: 5px;"></div> <div style="display: flex; justify-content: space-between;"> <div> <p>저장</p> <p>이전 문장</p> <p><input type="text" value=""/> 로 이동</p> </div> <div> <p>다음 문장</p> <p>맨뒤로</p> </div> </div>

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 (ERG1214)	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%

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
- 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.

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

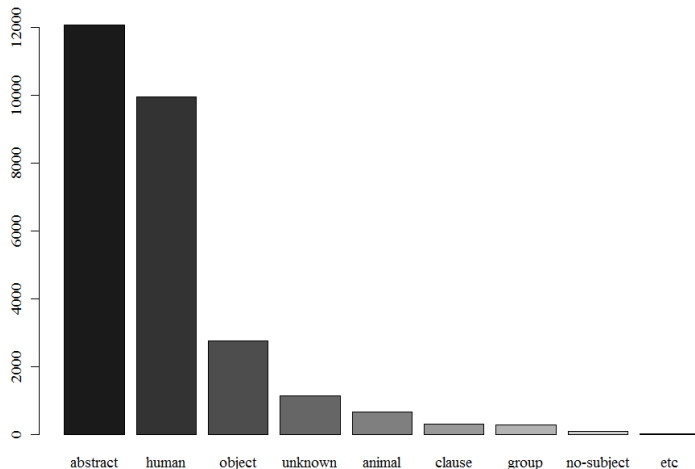
LEXEME	FREQ	PASS	BY	LEXEME	FREQ	PASS	BY
happen	1,426	92	8	eat	4,699	28	5
die	761	72	8	love	2,239	45	16
occur	699	59	10	meet	1,648	18	1
appear	541	46	2	remember	667	18	1
exist	483	29	0	kill	614	107	42
suffer	329	28	6	catch	257	58	17
disappear	275	51	1	throw	236	21	3
remain	224	20	0	remove	155	20	2

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

- (1) a. This situation was happened by the opposite people.
 b. so they can be died by those diseases.
 c. If so, traffic accidents that is occurred 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%

- (2) 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.

- 1 Introduction
- 2 Background
- 3 Corpus Study
- 4 Experimental Study**
- 5 Findings

Test Items: Passive

- Transitive

- (3)
- 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

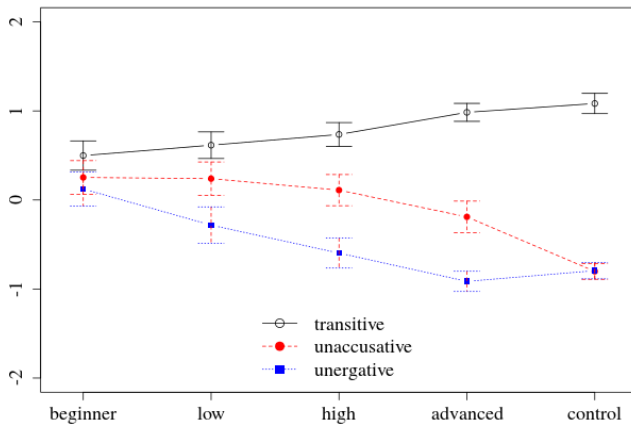
- (4)
- a. *A boy was cried.
 - b. *A baker was worked.
 - c. *A man was walked.
 - d. *A kid was run.
 - e. *A baby was crawled.
 - f. *A student was slept.
 - g. *A teacher was smiled.

Test Items: Passive (cont'd)

- Unaccusative (animate subjects)
 - (5) 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)
 - (6) 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.

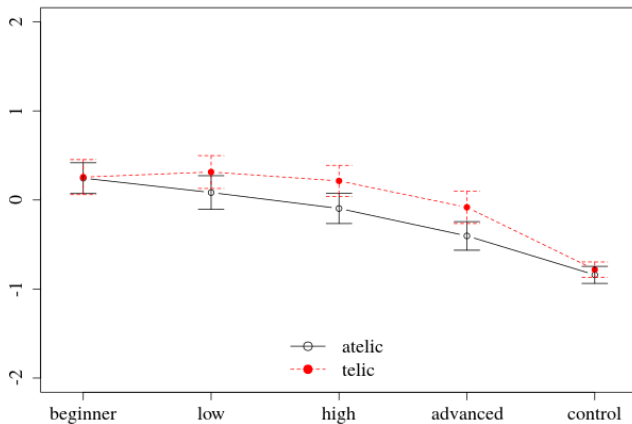
Passive

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



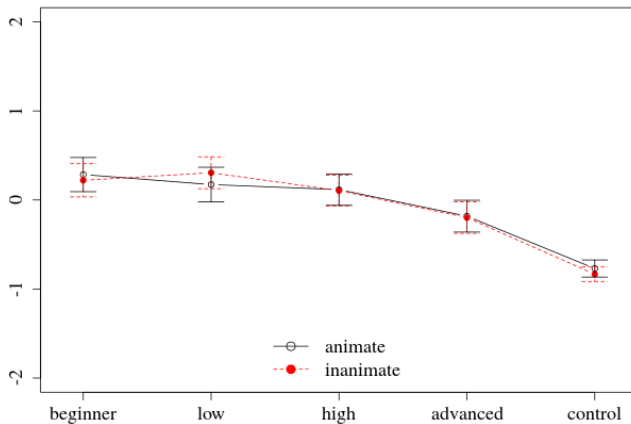
Passive: Telicity

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



Passive: Animacy

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



- 1 Introduction
- 2 Background
- 3 Corpus Study
- 4 Experimental Study
- 5 Findings**

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.
- ④ Using the resources that we have helps us to better characterize Korean speakers' knowledge of unaccusativity in English.