Empirical approaches to argument alternations

Tanja Samardžić and Paola Merlo

(1) a. Mary *broke* **the vase**. causative
b. **The vase** *broke*. anticausative

Availability:

	Causative	Anticausative
arrive, appear	+Japanese, +Salish, -**English**	+all languages
kill, cut	+all languages	+Greek, +Hindi, -**English**

Morphological marking:

	Causative	Anticausative
Mongolian	xajl-**uul**-ax	xajl-ax
	'melt'	'melt'
Russian	rasplavit'	rasplavit'-**sja**
	'melt'	'melt'
Japanese	atum-**eru**	atum-**aru**
	'gather'	'gather'

Table: Availability of the alternation and morphological marking for some verbs. (Alexiadou 2010, Haspelmath 1993)

The scale of spontaneous occurrence

Haspelmath (1993): 31 alternating verbs, 21 languages

$$\textbf{A/C(verb)} = \frac{\text{number of languages with marked anticausative}}{\text{number of languages with marked causative}}$$

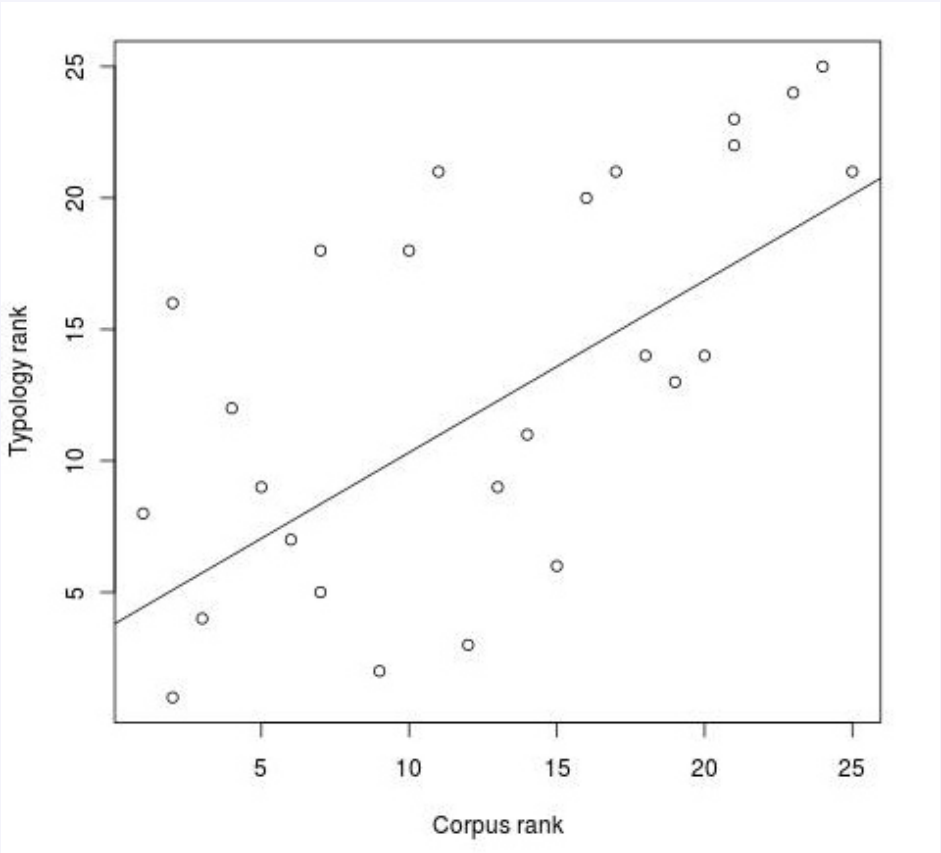
The scale of spontaneous occurrence:

freeze > *dry* > *melt* > ..... > *gather* > *open* > *break* > *close*

low A/C (spontaneous) high A/C (non-spontaneous)

Typology ranking and corpus data, 26 verbs

$$\textbf{C/A(verb)} = \frac{\text{frequency of causative uses}}{\text{frequency of anticausative uses}}$$



The correlation between the typology and corpus-based rankings: $r_s = 0.67$, $p < 0.01$.

Interpretation: The same semantic feature underlies both distributions.

Spontaneity: likelihood for an external causer to occur in an event described by a verb.

Scaling up from 26 to 354 verbs

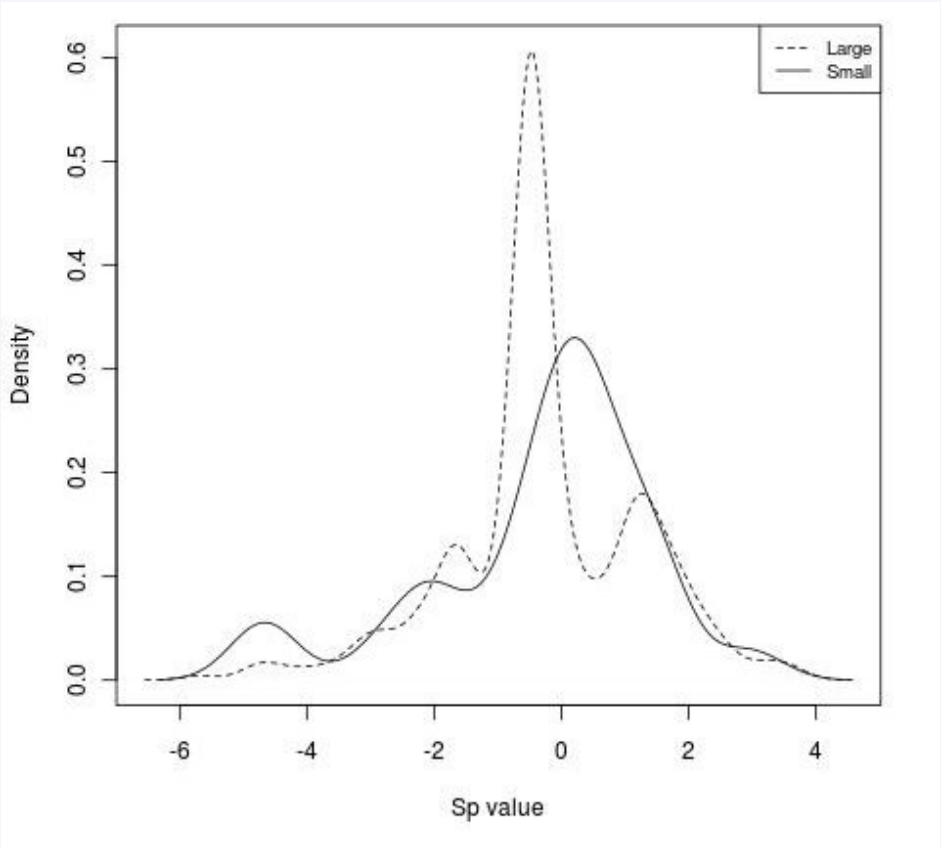


Figure: The distribution of the Sp-value over the small and the big sample of verbs, $t = -0.09$, $p = 0.93$.

Ratio $\rightarrow$ Sp-value

$$Sp = \ln\left(\frac{rate(v,caus)}{rate(v,acaus)}\right)$$

$$rate(form) = \frac{F(form,v)}{\sum_{form} F(form,v)}$$

Interpretation:

- 26 are an unbiased sample of 354 verbs
- Corpus-based measure applies to 354 verbs

Cross-linguistic data

Verb instances in parallel corpora:

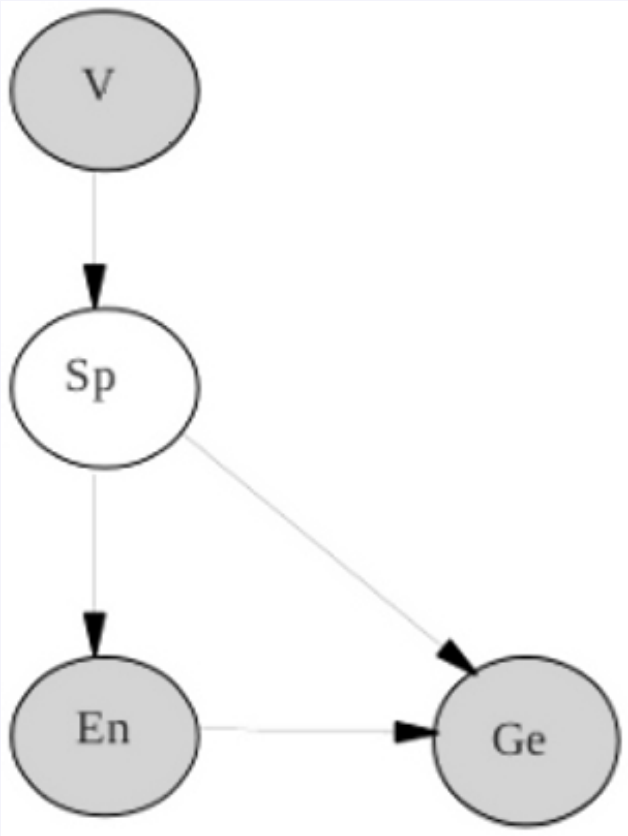
- Eliminating language biases
- Accounting for the cross-linguistic variation

.

V	En	Ge
move	pass	intrans
alter	trans	trans
improve	trans	trans
increase	pass	pass
improve	trans	trans
break	trans	pass
change	trans	intrans
grow	intrans	intrans
close	pass	intrans
split	intrans	trans

Table: Examples of 13033 cross-linguistic verb instances

Bayesian network:



$$P(v, sp, en, ge)$$

$$= P(v) \cdot P(sp|v) \cdot P(en|sp) \cdot P(ge|sp, en)$$

$$P(sp|v) = \frac{P(v,sp)}{P(v)}$$

$$= \frac{\alpha \sum_{en} \sum_{ge} P(v,sp,en,ge)}{\beta \sum_{sp} \sum_{en} \sum_{ge} P(v,en,ge)}$$

$$class(verb) = argmax_{sp} P(sp|v)$$

	Two classes		Three classes	
Verb	Typology	Model	Typology	Model
boil	a	a	a	a
dry	a	a	a	a
sink	a	a	a	a
melt	a	a	a	a
turn	a	a	m	m
dissolve	c	c	m	c
burn	c	c	m	c
fill	c	c	m	m
roll	c	c	m	m
improve	c	c	m	m
change	c	c	m	a
open	c	c	m	m
break	c	c	m	m
close	c	c	c	c
split	c	c	c	c

Table: Perfect agreement for two classes, 80% for three classes

Relevance for the cross-linguistic variation

Availability:

	Causative	Anticausative
arrive, appear	+Japanese, +Salish, **-English**	+all languages
kill, cut	+all languages	+Greek, +Hindi, **-English**

Relevance for the cross-linguistic variation

Availability:

	Causative	Anticausative
Low likelihood of occurrence of an external causer		
arrive, appear	+Japanese, +Salish, **-English**	+all languages
High likelihood of occurrence of an external causer		
kill, cut	+all languages	+Greek, +Hindi, **-English**

The likelihood can be automatically calculated for many verbs.