



Challenges in corpus annotation of copulative perception verbs

Alon Fishman

The Open University of Israel

Overview

Copulative perception verbs (CPVs):=

- Specify perceptual modality
- Take perceived entity as subject
- Require predicative/clausal complement

(Viberg 2019)

English *sound*, Hebrew *nishma*, Russian *zvuchat'*,
Finnish *kuulostaa*, Polish *brzmieć*, Spanish *sonar*

Form

(1) a. She looks weird.

b. [Polish]

wygląda dziwnie.
looks.SG weirdly

c. [Hebrew]

nir'et muzar / -a.
looks.FSG weirdly / weird.FSG

d. [Finnish]

näyttää oudo-lta / -lle.
looks.SG weird.SG.ABL / SG.ALL

Function

(2) a. She has a weird look.

Attributary

b. Her look suggests
that she's weird.

Evidential

(Gisborne 2010)

Collostructional analysis

- Investigating form-function mappings by measuring "attraction" to the complement & subject slots
- Testing hypothesized cross-linguistic generalizations

(Fishman 2023)

Corpus annotation

Desiderata: reliability & generalizability

Challenge: polysemy

Spanish *sonar* 'sound' / 'play', Finnish *näyttää* 'look' / 'show',
Hebrew *nishma* 'sound' / 'be heard' / 'obey'

(3) a. ...taste the same [obj]. (UD_English-EWT)

b. The one that sounds the best [advmod].

c. [Spanish] (UD_Spanish-AnCora)

...aunque suene extraño [obj]...
although sounds.SG strange

'Although it sounds strange.'

d. [Finnish] (UD_Finnish-FTB)

Se ei kuulosta tarpeeksi hyvältä [nmod].
it not sounds.SG enough good.SG.ABL

'That doesn't sound good enough.'

Hebrew <i>nir'e</i> 'look' complements	Adverb obs. (exp.)	Adjective obs. (exp.)	Log odds ratio
<i>nifla</i> 'wonderful'	512 (122)	23 (413)	4.39
<i>nehedar</i> 'terrific'	571 (136)	27 (462)	4.34
<i>tov</i> 'good'	3482 (865)	324 (2941)	4.10
<i>metsuyan</i> 'excellent'	658 (162)	53 (549)	3.82
<i>ra</i> 'bad'	479 (120)	47 (406)	3.60
<i>raxok</i> 'far'	6 (106)	460 (360)	-3.13
<i>tamuha</i> 'perplexing'	4 (73)	314 (245)	-3.15
<i>gadol</i> 'big'	5 (91)	392 (306)	-3.15
<i>pashut</i> 'simple'	4 (110)	481 (375)	-3.58
<i>mat'im</i> 'fitting'	4 (113)	493 (384)	-3.61
Russian <i>vygljadit</i> 'look' complements	Adverb obs. (exp.)	Adjective obs. (exp.)	Log odds ratio
<i>starshiy</i> 'senior'	2007 (1477)	12 (542)	4.13
<i>velikolepny</i> 'magnificent'	3051 (2255)	32 (828)	3.57
<i>estetichnyy</i> 'aesthetic'	1288 (953)	15 (350)	3.46
<i>zamechatel'nyy</i> 'wonderful'	1200 (888)	14 (326)	3.45
<i>prevoskhodnyy</i> 'excellent'	1260 (936)	19 (343)	3.20
<i>uprugiy</i> 'elastic'	5 (90)	118 (33)	-4.17
<i>ochevidnyy</i> 'obvious'	9 (183)	241 (67)	-4.29
<i>mertvy</i> 'dead'	5 (114)	151 (42)	-4.41
<i>blestyashchiy</i> 'brilliant'	8 (204)	271 (75)	-4.53
<i>bol'noy</i> 'sick'	6 (200)	268 (74)	-4.81

Challenge: variability

English: ADJ [xcomp]

Polish: ADV [advmod]

Hebrew: ADJ [xcomp]

Hebrew: ADV N/A

Russian: ADJ [obl/xcomp]

Russian: ADV [advmod]

Finnish: ADJ [xcomp/nmod]

