

My Case, For an Adposition: Lexical Polysemy of Adpositions and Case Markers in Finnish and Latin

Daniel Chen, Mans Hulden

Creation of SNACS Corpus

- Adpositions and case markers cover wide range of semantic fields: grounded spatial relations, abstract causal relations
- Useful to create annotation guidelines for disambiguating different senses of tokens using context of sentence

SNACS - **S**emantic **N**etwork for **A**dposition and **C**ase **S**enses (Schneider et al 2018)

- Supersense hierarchy: four levels of granularity
 - Previous languages with existing SNACS corpora: English, German, Korean, Hindi, Mandarin Chinese
- New SNACS corpus for Finnish and Latin: languages with rich case systems
 - Source texts: translations of *Le Petit Prince*
 - ~30 unique scene roles for each language



Construal Analysis

Finnish

152 adposition + case marker annotations for *Pikku Pinssi*

Sentence (from <i>Pikku Prinssi</i> , Chapter IV)	Scene Role	Function Role
a. <i>Ystävä-ni lähti pois lampai-ne-en.</i> friend-1P.POSS went away sheep-COMITATIVE-3P.POSS My friend went away with his sheep.	ANCILLARY	ANCILLARY
b. <i>Erehdyn myös vähän pituude-ssa.</i> I err also a little length-INESSIVE I also err a little in/regarding length.	TOPIC	LOCUS

Latin

180 adposition + case marker annotations for *Pikku Pinssi*

Sentence (from <i>Regulus</i> , Chapter IV)	Scene Role	Function Role
a. <i>Quotientscumque cum eis</i> As_often_as with them.ABL.PL	RECIPIENT	ANCILLARY
<i>de nov-o amic-o loque-ris...</i> About new-2.ABL.SG friend-2.ABL.SG talk-2P.PRES.SG.ACT As often as you talk with them about the new friend...	TOPIC	TOPIC
b. <i>...mihi admiration-em magn-am movere non poterat</i> me.DATIVE admiration-3F.ACC great-3F.ACC to_move not was_able.3P.PST ...he was not able to move great admiration for me.	EXPERIENCER	GOAL

Results + Analysis

# Clusters	Cluster
50	##assa(36), ##uksessa(38), ##ssa(127), ##ssa(167), ##sina(192), ##ssa(233), ##na(268), ##ssa(314)
30	##ksi(22), ##assa(36), ##alla(53), ##oilla(93), ##lla(96), ##sena(111), ##ssa(127), ##a(136), ##ssa(167), ##lta(189), ##sina(192), ##lla(217), ##lla(223), ##a(227), ##ssa(233), ##lla(243), ##la(258), ##na(268), ##ssa(314)
15	##kella(17), ##ksi(22), ##ulla(25), ##kella(30), ##assa(36), ##uksessa(38), ##alla(53), ##oilla(93), ##lla(96), ##sena(111), ##ssa(127), ##a(136), ##a(155), ##uudessa(157), ##ssa(167), ##lta(189), ##sina(192), ##lla(217), ##lla(223), ##kill(226), ##a(227), ##ssa(233), ##lla(243), ##nna(248), ##llaan(253), ##lla(258), ##na(268), ##tajana(290), ##ssa(314)

1. Adpositions cluster well together, since they are separate words that typically make up the entirety of their own WordPiece token.
2. Adpositions are found in separate clusters from their case-marked noun objects => mBERT identifies no semantic similarity between them.
3. Multiword adpositions are not found in the same cluster.
4. Clusters do not group morphological alternations of the same case marker together.

