The Semantics of Split-Intransitive Alignment Systems: A Multidimensional Scaling Analysis

Jens Van Gysel – University of New Mexico
Syntax of the World's Languages 8, 3-5 September 2018, Paris
jelvangysel@unm.edu

Structure

- Introduction: split-intransitive alignment
- Previous cross-linguistic research
- Data and methodology: Multidimensional Scaling
- Results and discussion
- Further research

Introduction: Split-Intransitive Alignment

- Split-intransitive alignment (also known as active-inactive, split-S)
  - (Usually) two large classes of intransitive verbs, whose arguments are marked in different ways
  - One class usually has the same marking as A arguments, the other as O arguments
- Example: Tupí-Guaranían languages (e.g. Tupinambá, Jensen 1990: 117):

a-só	a-i-nupã
1sg-go	1sg-3sg-hit
‘I go’	‘I hit it’

Introduction: Split-Intransitive Alignment

- Split-intransitive alignment (also known as active-inactive, split-S)
  - (Usually) two large classes of intransitive verbs, whose arguments are marked in different ways
  - One class usually has the same marking as A arguments, the other as O arguments
- Example: Tupí-Guaranían languages (e.g. Tupinambá, Jensen 1990: 117):

syé katú	syé nupã
1sg be.good	1sg hit
'I am good'	'He/she/it/they hit me'

Introduction: Split-Intransitive Alignment

- Within functional frameworks: essentially a semantic phenomenon
  - Van Valin (1990): splits based on Aktionsart (e.g. Georgian) or on agentivity (e.g. Acehnese)
  - Many studies on factors conditioning the split in individual languages/families (e.g. Klamer 2008)

- Lexicalisation and grammaticalisation can obscure semantic motivation (Mithun 1991)

Introduction: Split-Intransitive Alignment

- Even considering language-specific "weighting" of factors and idiosyncrasies: significant systematicities have been found
- This work builds on cross-linguistic accounts by Mithun (1991) and Croft (1998, 2012)
- Use Multidimensional Scaling to give a preliminary quantitative account of the factors cross-linguistically conditioning split-intransitive alignment

Previous Cross-Linguistic Research

- Mithun (1991): sample of five languages: Paraguayan Guaraní, Lakhota, Caddo, Central Pomo, Mohawk
  - Guaraní: Aktionsart is main conditioning factor
  - Events (achievements, accomplishments, activities, see Vendler 1967) take Sa marking
    - ‘get up’, ‘walk’, ‘fire a gun’, ‘work’, ‘fall’, ‘die’
  - States take So marking
    - ‘be sick’, ‘be sleepy’, ‘be short’, ‘have a cramp’

Previous Cross-Linguistic Research

- Mithun (1991): sample of five languages: Paraguayan Guaraní, Lakhota, Caddo, Central Pomo, Mohawk
  - North American languages in the sample: mostly force-dynamic factors
  - Lakhota: agency, instigation
    - 'be patient', 'walk' vs. 'die', 'be cold'
  - Central Pomo (and Caddo): control + affectedness
    - 'sneeze' would be Sa in Lakhota, but So in Pomo

=> interaction of Aktionsart, instigation, control and affectedness, which have a different weight in different languages

Previous Cross-Linguistic Research

- Croft (1998, 2012): same sample as Mithun + Tsova-Tush (Caucasian, Georgia)
  - MDS analysis => one-dimensional model based on only causal factors explains the data well:

controlled activities	More Sa-like
inactive actions	
inherent properties/dispositions	
bodily actions	
inchoatives	
uncontrolled activities/transitory states	More So-like

Data and Methodology

- The samples of aforementioned studies are small and rather concentrated geographically
  - E.g. importance of causal factors in Croft (2012) is not surprising, since for most of the sample, Mithun (1991) already shows this
- This preliminary study: six languages which have been described as showing active-inactive/split-intransitive alignment, geographically and genetically balanced

Data and Methodology

- Twenty languages were sampled, based on the WALS data for active-inactive alignment (Comrie 2013; Siewierska 2013), and other references in the literature

- Six were chosen, based on maximal geographical and genetic distance, and availability of sources, for this initial study

Data and Methodology

Language	Family	Region	Source(s)
Acehnese	Chamic (Austronesian)	Indonesia (Sumatra)	Durie (1985)
Beria	Saharan	Chad, Sudan	Jakobi & Crass (2004); Jakobi (2011)
Creek	Muskogean	USA (Oklahoma, Florida)	Martin & Mauldin (2000)
Pilagá	Guaykuruan	Argentina (Formosa)	Vidal (2001)
Rotokas	East Papuan	Papua New Guinea (Bougainville)	Robinson (2011)
Tsova-Tush	Caucasian	Georgia	Holisky (1987)

Data and Methodology

- For each language, the type of S-marking with as many intransitive predicates as possible was coded
- Twenty-four were predicates were then chosen for analysis based on:
  - Their semantic characteristics
  - Number of languages for which data for this predicate was found

Data and Methodology

Predicates studied		
Get up	Enter	Work
Be in a lying position	Be in a standing position	Be in a sitting position
Spit	Vomit	Breathe
Sneeze	Die	Fall
Spill	be bad	be quiet
be good	be short	be tall
Be big	Dry up	Get startled
Be cold	Be hungry	Be hot

Data and Methodology

- Multidimensional scaling (see Poole 2005 for the maths)
  - A methodology for measuring (dis)similarity between entities
  - Dissimilarities between entities are represented as distances in geometric space
    - => In this case, the further away from each other two predicates are in the plot, the bigger the difference between them with regards to argument marking
    - => A mathematical implementation of the semantic map methodology (Croft & Poole 2008)

Data and Methodology

- Multidimensional scaling (see Poole 2005 for the maths)

	Creek Set I	Creek Set II	Creek Dative S	Tsova-Tush NOM	Tsova-Tush ERG
Work	1	6	6	6	1
Be lying	1	6	6	1	6
Be standing	1	6	6	1	1
Sit	1	6	6	1	1
Spit	1	1	6	9	9

Data and Methodology

- Multidimensional scaling (see Poole 2005 for the maths)
  - For each construction, a cutting point, line, or plane is fitted in space so that it divides the predicates into *yea* vs *nay* groups with as good a fit as possible
    - 23 constructions
    - 24 predicates
  - W-NOMINATE algorithm (Poole & Rosenthal 1985), implemented in a linguistics-friendly format by Timm (2017) for R (R Core Team 2018), calculates an ideal location for every point (= predicate) in space
    - I.e. where it is on the correct side of as many cutting points/lines as possible

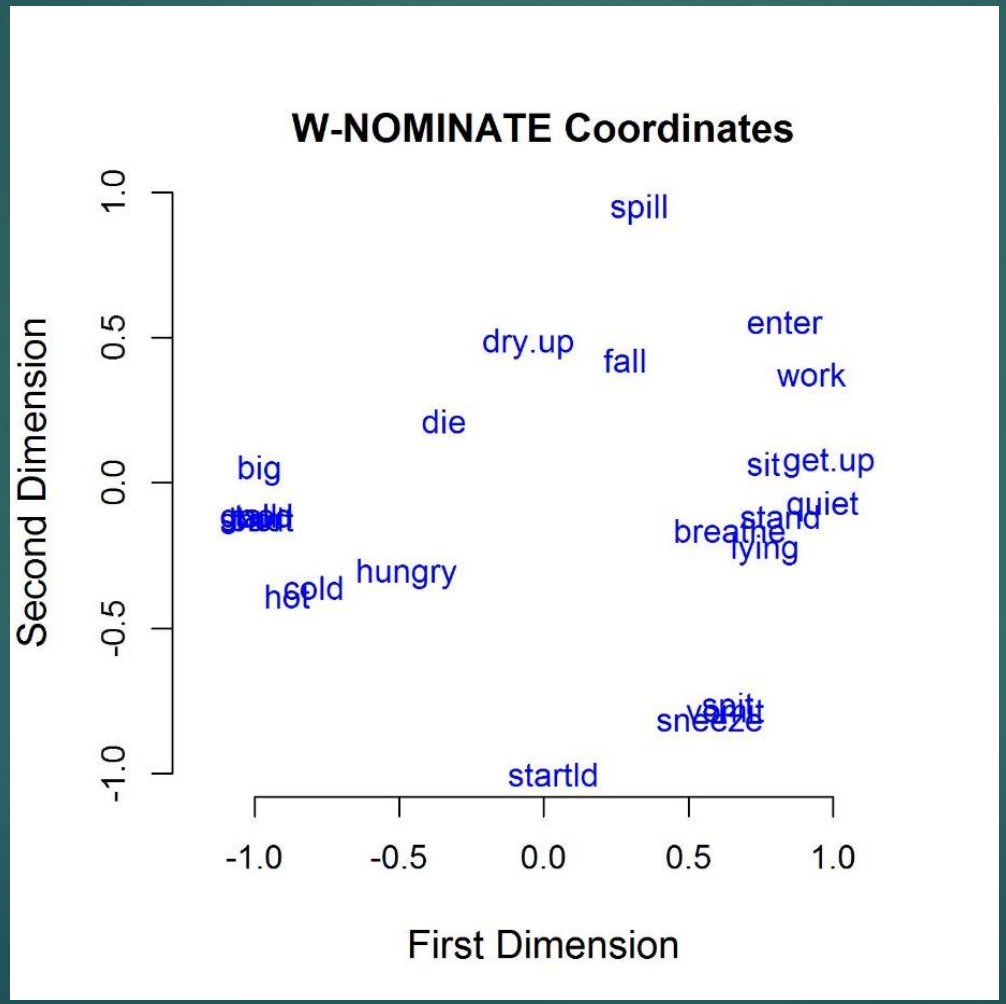
Data and Methodology

- Dimensionality of the model?

	Correct classification	Average proportional reduction of error
1D	86,1%	0,46
2D	95,1%	0,81
3D	96,3%	0,86

- A 2D model provides a significant increase in correct classification, the increase from 2D to 3D does not warrant the use of extra statistical power

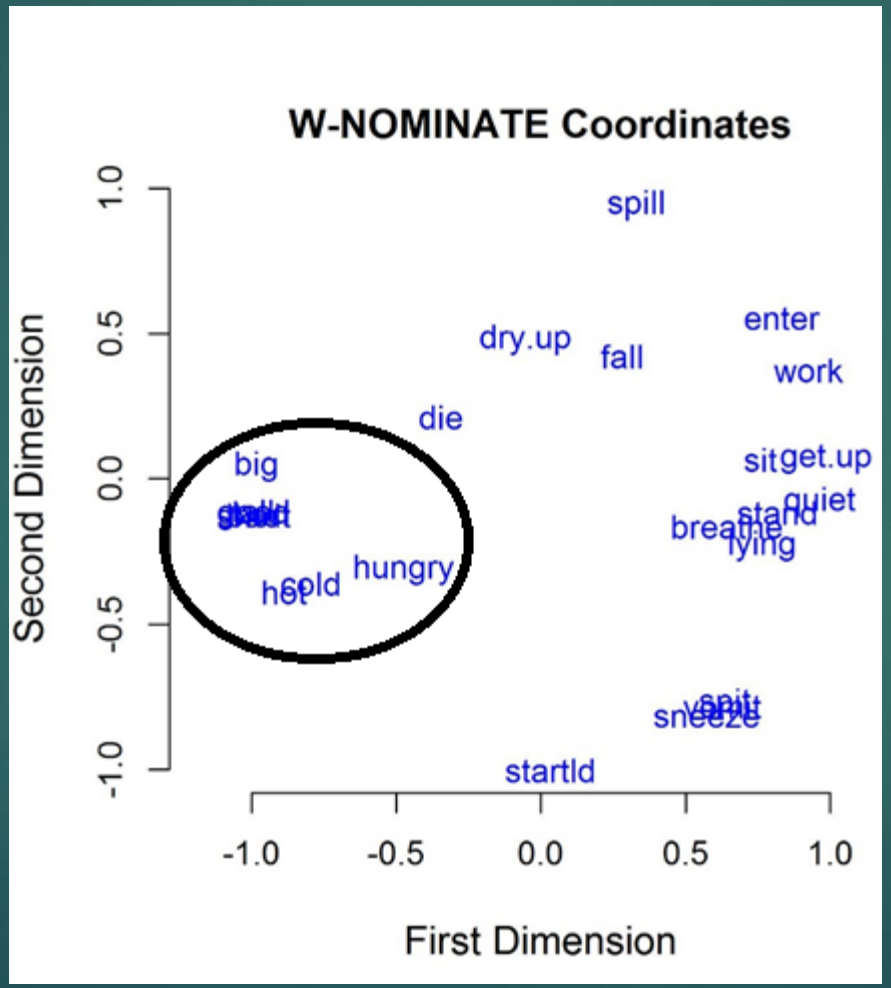
Results and Discussion



Results and Discussion

- The first dimension seems to correspond to causal factors:
  - On the left: non-agentive predicates, S is not in control
    - Non-controlled states, both inherent and transitory
    - S argument is significantly affected

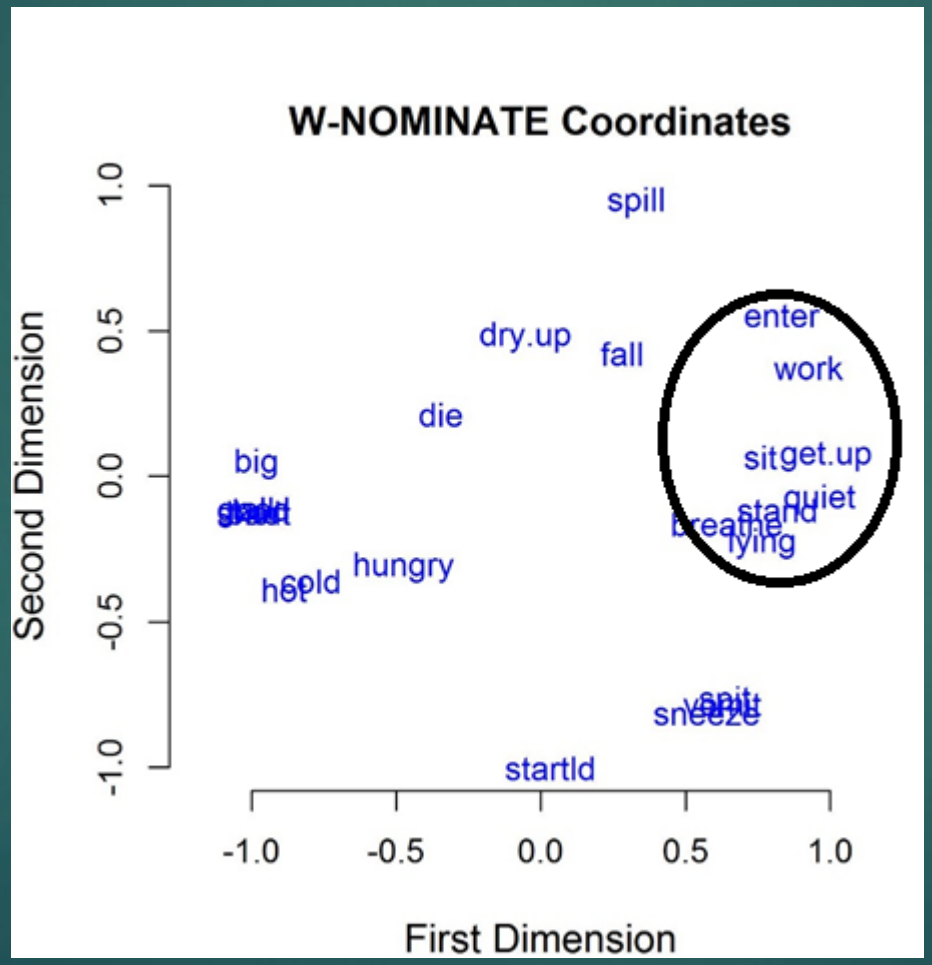
Results and Discussion



Results and Discussion

- The first dimension seems to correspond to causal factors:
  - On the right: agentive predicates, S is in control
    - Controlled activities
    - Controlled states: inactive actions, ‘be quiet’

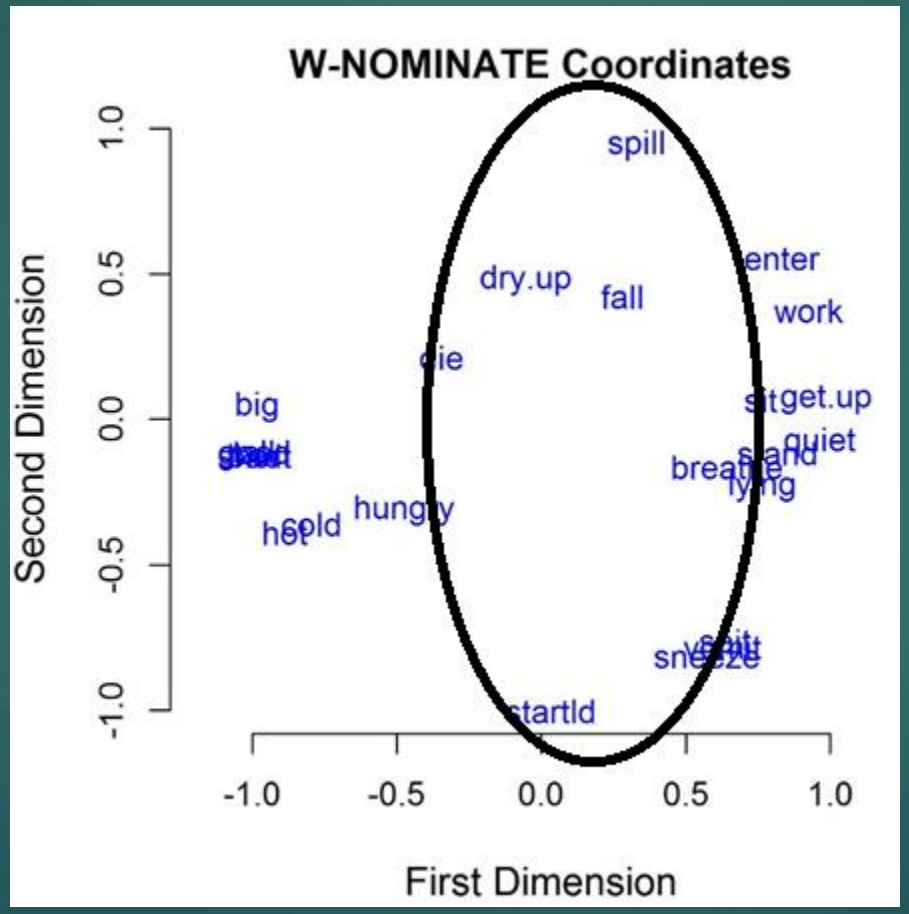
Results and Discussion



Results and Discussion

- The first dimension seems to correspond to causal factors:
  - In the middle: predicates where S can be either volitional or non-volitional
    - Uncontrolled activities
    - Bodily actions
  - Predicates with typically inanimate S, so less affected
    - ‘spill’, ‘dry up’

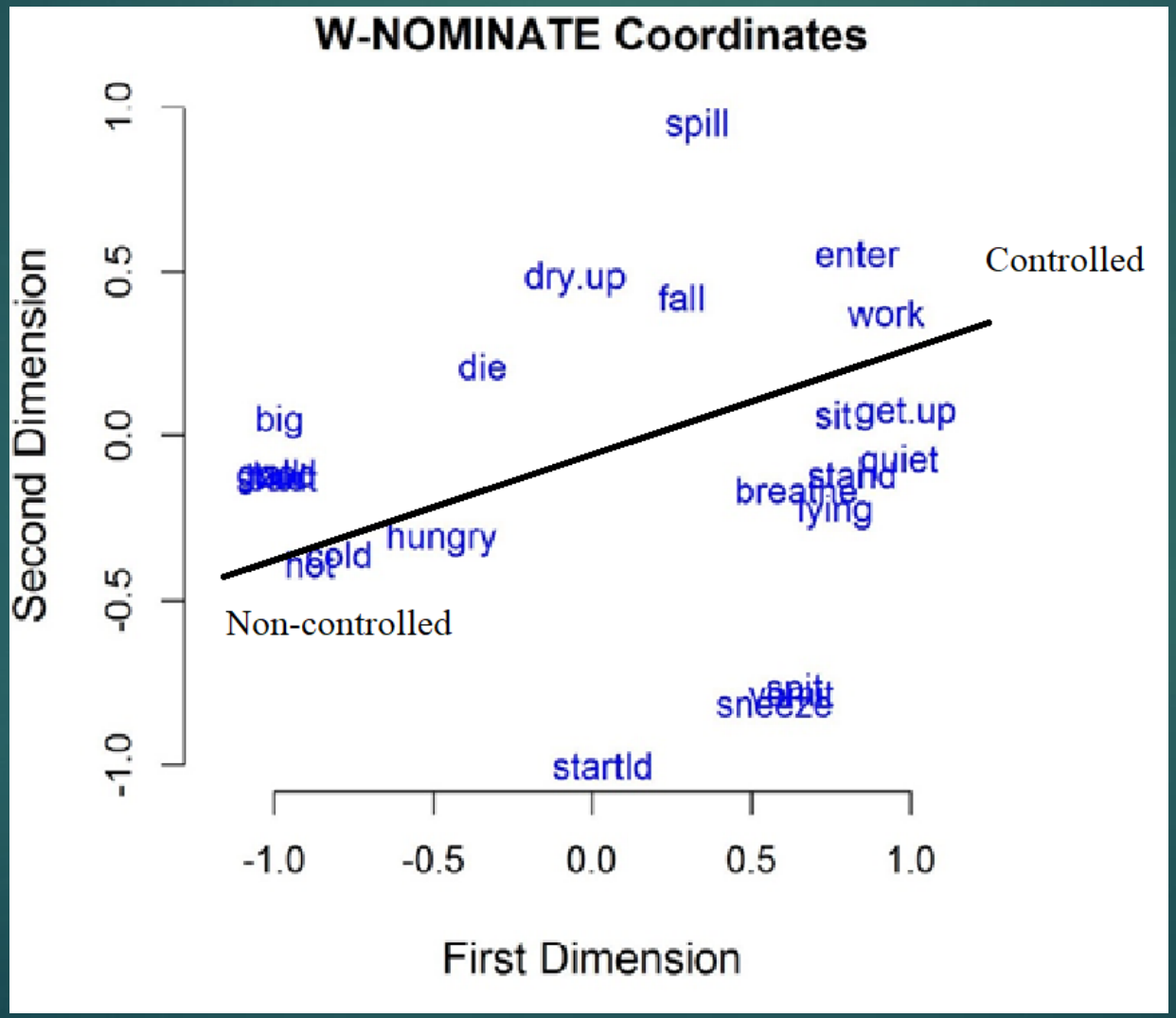
Results and Discussion



Results and Discussion

- The first dimension seems to correspond to causal factors:
  - Unexpected: inchoative-like predicate ‘get startled’ is found in the middle, even though the S is not in control and is affected

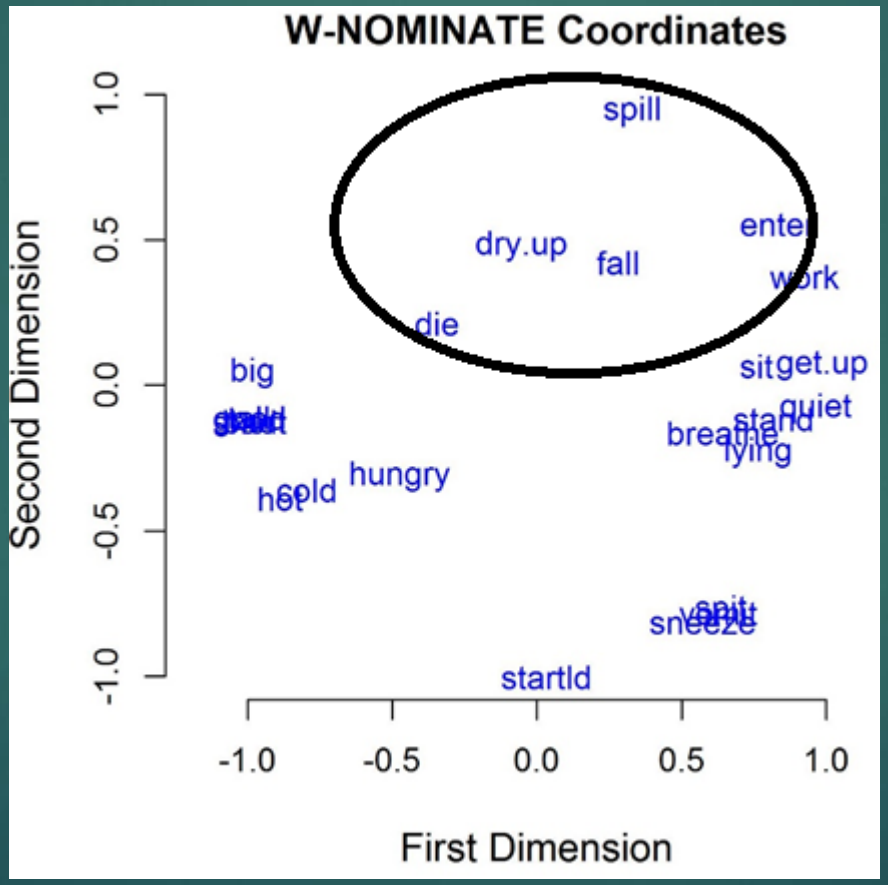
Results and Discussion



Results and Discussion

- The second dimension seems to correspond to aspectual factors:
  - Higher up: directed predicates, result in a change of state
    - Directed achievements
    - Directed accomplishments

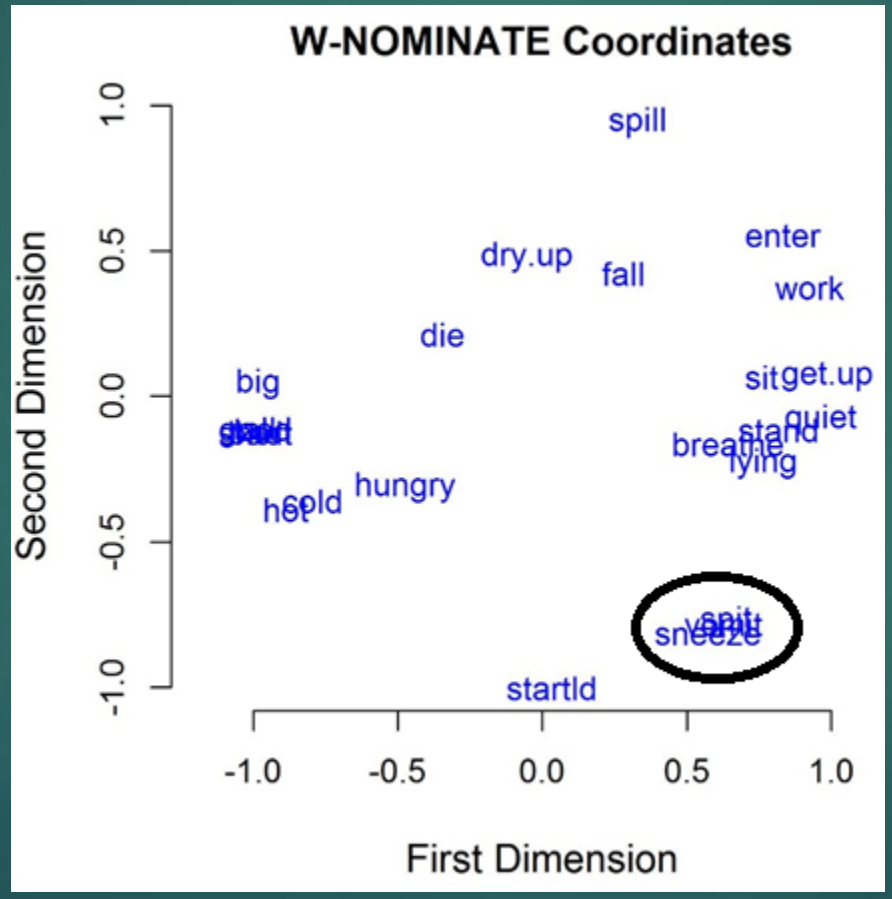
Results and Discussion



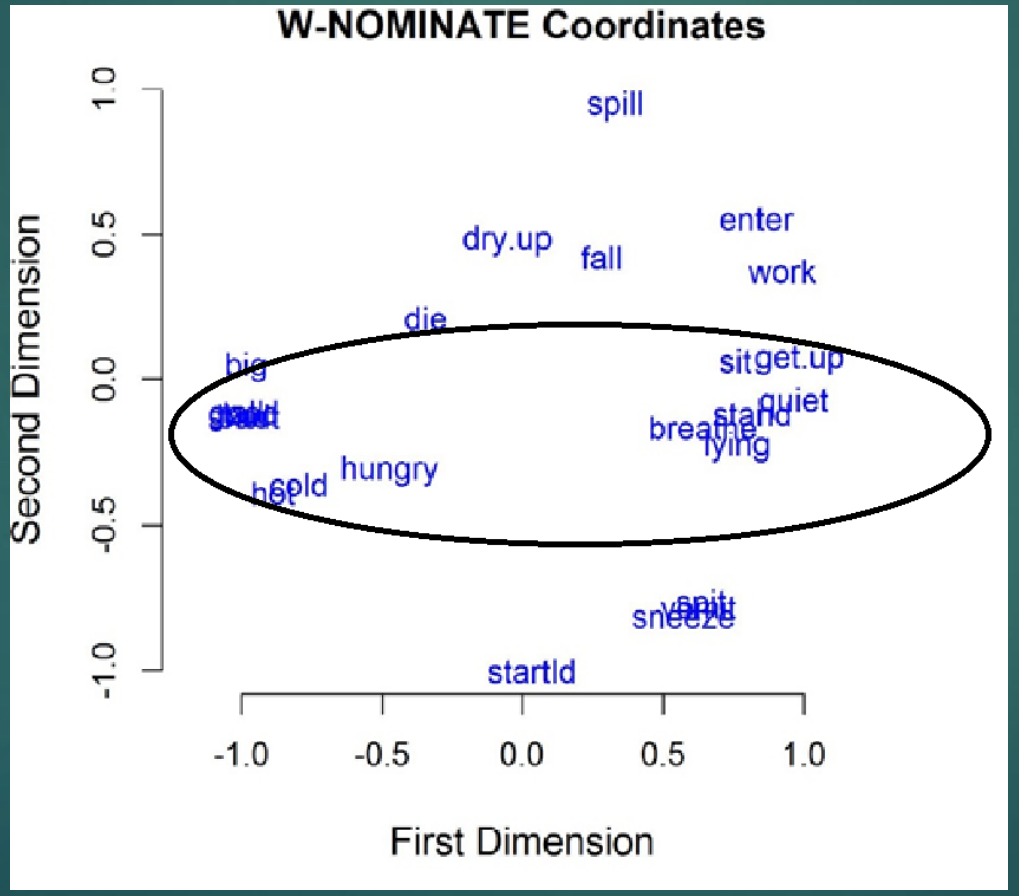
Results and Discussion

- The second dimension seems to correspond to aspectual factors:
  - Lower down: predicates that do not result in a change of state
    - Cyclic achievements
    - Undirected activity

Results and Discussion



Results and Discussion

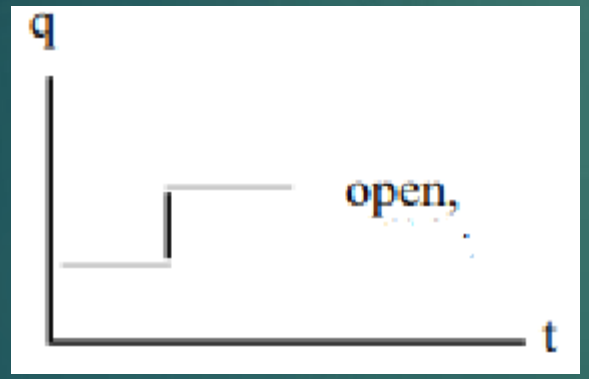


Results and Discussion

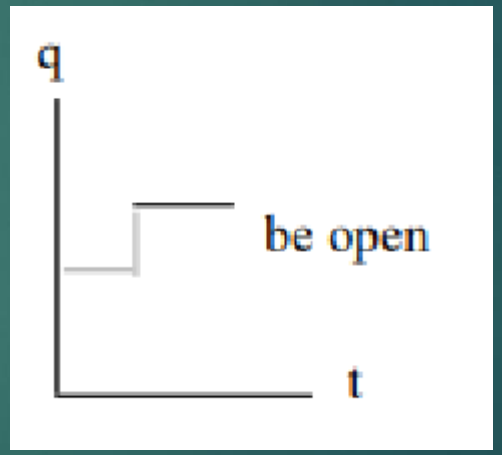
- The second dimension seems to correspond to aspectual factors:
  - In the middle
    - Transitory states
    - Permanent states
    - Inactive actions (can be construed as transitory states, see Croft 1998)
  - Does not seem to correspond to Mithun's (1991) dichotomy between event-like Sa predicates and state-like So predicates
  - Croft (2012: 170-1): many states can be analysed as having the same base as a directed achievement, but with a different profile

Results and Discussion

- The second dimension seems to correspond to aspectual factors:
  - In the middle: temporary and permanent states



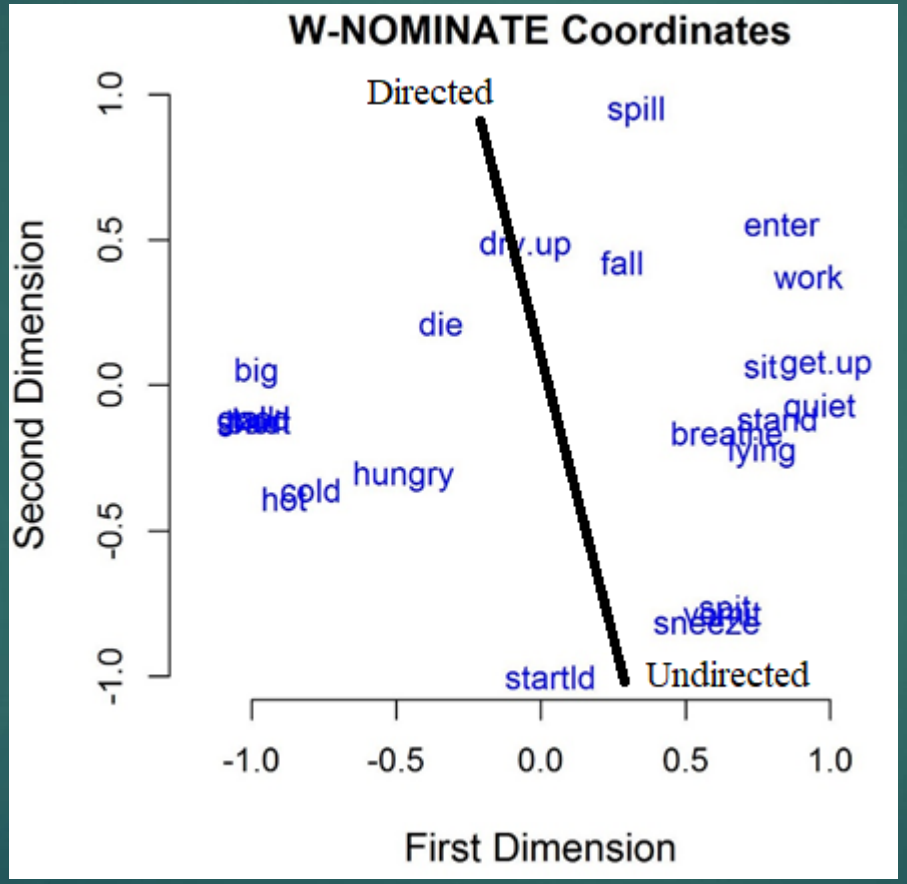
From Croft (2008)



Results and Discussion

- The second dimension seems to correspond to aspectual factors:
  - Unexpected: ‘get up’ and ‘get startled’ would be expected higher up with the directed predicates.

Results and Discussion



Conclusions and Further Research

- MDS plot indicates general cross-linguistic trends in the semantics of agentive-inagentive alignment systems
  - Quite clear aspectual and force-dynamic dimensions
  - Directed predicates with controlling S => more prototypically Sa
  - Note: Aktionsart and agentivity are not always independent

- More geographically balanced sample shows the presence of aspectual factors, next to the causal factors found by Croft (2012)
- However, the most important aspectual factor seems to be directedness, rather than eventhood (as proposed by Mithun 1991)

Conclusions and Further Research

- Croft (2012: 166): "the general effect of MDS is to detect broad patterns in complex and messy data"
  - We always expect noise, especially with relatively small datasets
- To reduce noise: increase number of languages under study
  - More datapoints for the predicates studied here improves their location
  - More cutting lines allows the inclusion of more predicates in order to find subtler distinctions
- Bayesian MDS models are not yet very widespread (see e.g. Okada & Shigemasu 2009) but could improve the model
  - Provide more accurate estimates of uncertainty in the data
  - Gives a direct indication of optimal dimensionality

References

Comrie, Bernard. 2013. Alignment of case marking of full noun phrases. The World Atlas of Language Structures Online, eds. Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Croft, William. 1998. Event structure in argument linking. The Projection of Arguments: Lexical and Compositional Factors, eds. William Geuder and Miriam Butt, 21-64. Stanford: CSLI.

Croft, William. 2008. Aspectual and causal structure in event representations. Routes to Language Development: In Honor of Melissa Bowerman, ed. Virginia Gathercole, pp. 139-66. Mahwah, NJ, Lawrence Erlbaum.

Croft, William. 2012. Verbs: Aspect and Causal Structure. Oxford: Oxford University Press.

Croft, William and Keith T. Poole. 2008. Inferring universals from grammatical variation: Multidimensional scaling for typological analysis. *Theoretical Linguistics* 34 (1):1-37.

Durie, Mark. 1985. A Grammar of Acehnese on the Basis of a Dialect of North Aceh. Dordrecht: Foris.

Holisky, Dee Ann. 1987. The case of the intransitive subject in Tsova-Tush (Batsbi). Lingua 71: 103-32.

Jakobi, Angelika. 2011. Split-S in Beria. Kanuri, Borno and Beyond: Current Studies on the Lake Chad Region, eds. Doris Löhr, Eva Rothmaler and Georg Ziegelmeyer, pp. 87-109. Cologne: Rüdiger Köppe Verlag.

References

Jakobi, Angelika and Joachim Crass. 2004. Grammaire du Beria (Langue Saharienne). Avec un Glossaire Français-Beria. Cologne: Rüdiger Köppe Verlag.

Jensen, Cheryl. 1990. Cross-referencing changes in some Tupí-Guaraní languages. Amazonian Linguistics: Studies in Lowland South American Languages, ed. Doris L. Payne, pp. 117-158. Austin: University of Texas Press.

Klamer, Marian. 2008. The semantics of semantic alignment in Eastern Indonesia. The Typology of Semantic Alignment, eds. Mark Donohue and Søren Wichmann, pp. 221-51. Oxford: Oxford University Press.

Martin, Jack B. and Margaret M. Mauldin. 2000. A Dictionary of Creek/Muskogee. With Notes on the Florida and Oklahoma Seminole Dialects of Creek. Lincoln/London: University of Nebraska Press.

Mithun, Marianne. 1991. Active/agentive case marking and its motivations. Language 67 (3):510-46.

Okada, Kensuke and Kazuo Shigemasu. 2009. BMDS: A collection of R functions for Bayesian multidimensional scaling. Applied Psychological Measurement 33 (7):570-1.

Poole, Keith T. 2005. Spatial Models of Parliamentary Voting. Cambridge: Cambridge University Press.

Poole, Keith T. and Howard Rosenthal. 1985. A spatial model for legislative roll call analysis. American Journal of Political Science 29:357–84.

References

R Core Team. 2018. R 3.5: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org.

Robinson, Stuart. 2011. Split intransitivity in Rotokas, a Papuan language of Bougainville. Doctoral dissertation: Radboud Universiteit Nijmegen.

Siewierska, Anna. 2013. Alignment of verbal person marking. The World Atlas of Language Structures Online, eds. Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Van Valin, Robert D. Jr. 1990. Semantic parameters of split intransitivity. *Language* 66 (2):221-60.

Timm, Jason. 2017. MDS_for_Linguists. https://github.com/jaytimm/MDS_for_Linguists/blob/master/resources/MDS_Script.R.

Vendler, Zeno. 1967. Linguistics in Philosophy. Ithaca: Cornell University Press.

Vidal, Alejandra. 2001. Pilagá grammar (Guaykuruan family, Argentina). Doctoral dissertation, University of Oregon.