

Visualizing morphological patterns in inflectional paradigms

Sebastian Bank, Daniela Henze, Jochen Trommer, Eva Zimmermann

University of Leipzig

DGfS 2013
Visualization Workshop
March 13

Background

Inflectional paradigms: organizing sets of inflectional forms

- arrange forms into tables such that adjacent cells correspond to natural classes

German verbal inflection

1SG	glaub-e		1SG	glaub-e		2SG	glaub-st
2SG	glaub-st		1PL	glaub-en		3SG	glaub-t
3SG	glaub-t		2SG	glaub-st		1SG	glaub-e
1PL	glaub-en	or	2PL	glaub-t	or	2PL	glaub-t
2PL	glaub-t		3SG	glaub-t		3PL	glaub-en
3PL	glaub-en		3PL	glaub-en		1PL	glaub-en

- identify full and partial syncretism
- correlate with invariant meaning
- more complex patterns: blocking, affix order, hierarchy effects

Background

Background: analyzing complex agreement morphologies

Hierarchy Effects in Kiranti and Broader Alpic
<http://proalki.uni-leipzig.de/project/>
The internal structure of person portmanteaus
<http://portmanteau.uni-leipzig.de/>

Agenda

data collection ... → pattern identification ... → automatic analysis

- keep things simple, bottom-up
- evolve data structures, formats, conventions, best practices
- assist manual analysis with interactive visualization tools
- implement pattern identification in learning algorithms
- gain theoretical and typological insights

Background

Organizing sets of inflectional forms

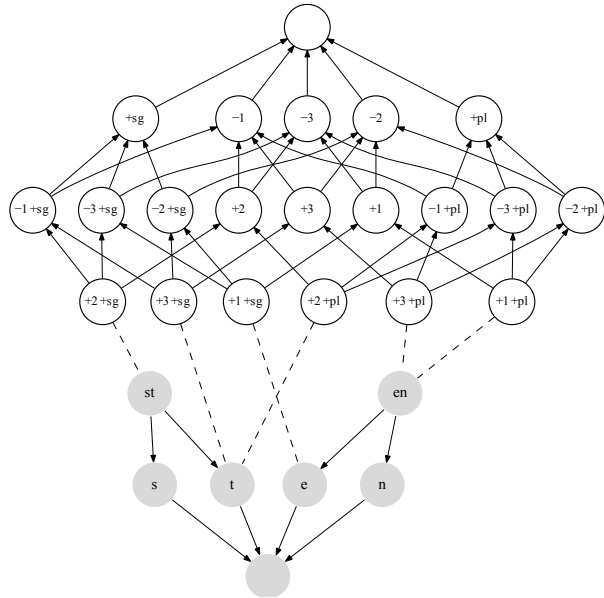
- maximize neighboring table cell form identity/similarity

German verbal inflection (present and past)

	SG	PL	SG	PL		SG	PL		
1	e	en	te	ten		e	te	en	ten
2	st	t	test	tet	or	st	test	t	tet
3	t	en	te	ten		t	te	en	ten
	PRESENT		PAST			PRS	PST	PRS	PST

- table approach has its limits

Visualizing paradigms as form-meaning relations



4 / 19

Data collection: paradigm representation and sources

- represent paradigms, morphemes, rules, etc. as structured documents
- additional information on relevant syntax, morphology, phonology
- kept as pages under the version control of a MediaWiki
- annotated for aggregation & search (Semantic MediaWiki)

- mostly collected from reference grammars with their segmentation
- underlying and surface forms, features, morpheme kind
- alternative segmentations and analyzes, comparison pages
- command-line tools (e.g. row/column reordering)

5 / 19

Data: (In)transitive verbal agreement paradigms

Kohi intransitive and transitive non-past paradigms.

Intransitive		Transitive										
		1s	1de	1pe	1di	1pi	2s	2d	2p	3s	3d	3p
1s	Σ-tō						Σ-ne	Σ-ni-si	Σ-ni-ni	Σ-da	Σ-da-si	Σ-da-ni
1de	Σ-i-si						Σ-i-si	Σ-i-si	Σ-i-si	Σ-i-si	Σ-i-si	Σ-i-si
1pe	Σ-i-ko						Σ-i-ko	Σ-i-ko	Σ-i-ko	Σ-i-ko	Σ-i-ko	Σ-i-ko
1di	Σ-i-si								Σ-i-si	Σ-i-si	Σ-i-si	
1pi	Σ-i-ki								Σ-i-ki	Σ-i-ki	Σ-i-ki	
2s	Σ-e-na	Σ-tō-na	Σ-i-si-na	Σ-i-ko-na					Σ-da-na	Σ-da-si-na	Σ-da-na	
2d	Σ-i-si-na	Σ-tō-si-na	Σ-i-si-na	Σ-i-ko-na					Σ-i-si-na	Σ-i-si-na	Σ-i-si-na	
2p	Σ-i-ni	Σ-tō-na	Σ-eko-na	Σ-i-ko-na					Σ-i-ni	Σ-i-ni	Σ-i-ni	
3s	Σ-e	Σ-tō	Σ-i-si	Σ-i-ko	Σ-i-si	Σ-i-ki	Σ-e-na	Σ-i-si-na	Σ-i-ni	Σ-da	Σ-da-si	Σ-da-ni
3d	Σ-i-si	Σ-tō-si	Σ-i-si	Σ-i-ko	Σ-i-si	Σ-i-ki	Σ-e-na	Σ-i-si-na	Σ-i-ni	Σ-da-si	Σ-da-si	Σ-da-ni
3p	Σ-i-ni	Σ-tō-ni	Σ-i-si	Σ-i-ko	Σ-i-si	Σ-i-ki	Σ-e-na	Σ-i-si-na	Σ-i-ni	Σ-da-ni	Σ-da-ni	Σ-da-ni

role {S, A, P} person {1excl, 1incl, 2, 3} number {sg, du, pl}
 (11 + 11²) - 46 = 86 cells TAM { Nonpast, Past } 86 * 2 = 172 cells

6 / 19

Paradigm colorizer: implementation and user interface

- html generated by paradigm and comparison pages is annotated
- ParadigmColoriser.js offers in-browser visualization options

- underlying/surface forms, select/hide related information
- highlight morphemes based on different abstraction levels

7 / 19

Visualizing basic patterns with morpheme highlighting

Kohi (Kiranti, Nepal, Lahaussou 2009)

Intransitive Transitive

	1s	1de	1pe	1di	1pi	2s	2d	2p	3s	3d	3p
1s	Σ-ne					Σ-ne	Σ-ni-si	Σ-ni-ni	Σ-da	Σ-da-si	Σ-da-ni
1de	Σ-i-si	1de				Σ-i-si	Σ-i-si	Σ-i-si	Σ-i-si	Σ-i-si	Σ-i-si
1pe	Σ-i-ko	1pe				Σ-i-ko	Σ-i-ko	Σ-i-ko	Σ-i-ko	Σ-i-ko	Σ-i-ko
1di	Σ-i-si	1di									
1pi	Σ-i-si	1pi									
2s	Σ-e-na	2s	Σ-e-na	Σ-i-si-na	Σ-i-ko-na				Σ-da-na	Σ-da-si-na	Σ-da-na
2d	Σ-i-si-na	2d	Σ-i-si-na	Σ-i-si-na	Σ-i-ko-na				Σ-i-si-na	Σ-i-si-na	Σ-i-si-na
2p	Σ-i-ni	2p	Σ-i-na	Σ-eko-na	Σ-i-ko-na				Σ-i-ni	Σ-i-ni	Σ-i-ni
3s	Σ-e	3s	Σ-e	Σ-i-si	Σ-i-ko	Σ-i-si	Σ-i-na	Σ-i-ni	Σ-da	Σ-da-si	Σ-da-ni
3d	Σ-i-si	3d	Σ-i-si	Σ-i-si	Σ-i-ko	Σ-i-si	Σ-i-na	Σ-i-ni	Σ-da-si	Σ-da-si	Σ-da-ni
3p	Σ-i-ni	3p	Σ-i-si	Σ-i-si	Σ-i-ko	Σ-i-si	Σ-i-na	Σ-i-ni	Σ-da-ni	Σ-da-ni	Σ-da-ni

Hypotheses *-ki* ↔ [+1+2+pl] *-ko* ↔ [+1-2+pl] *-to* ↔ [Abs+1+sg]

Analyzing imperfect morpheme distributions

Kohi

	1s	1de	1pe	1di	1pi	2s	2d	2p	3s	3d	3p
1s	Σ-to					Σ-ne	Σ-ni- ni	Σ-ni-ni	Σ-da	Σ-da- ni	Σ-da-ni
1de	Σ-i- ni	1de				Σ-i- ni	Σ-i- ni	Σ-i- ni	Σ-i- ni	Σ-i- ni	Σ-i- ni
1pe	Σ-i-ko	1pe				Σ-i-ko	Σ-i-ko	Σ-i-ko	Σ-i-ko	Σ-i-ko	Σ-i-ko
1di	Σ-i- ni	1di							Σ-i- ni	Σ-i- ni	Σ-i- ni
1pi	Σ-i-ki	1pi							Σ-i-ki	Σ-i-ki	Σ-i-ki
2s	Σ-e-na	2s	Σ-to-na	Σ-i- ni -na	Σ-i-ko-na				Σ-da-na	Σ-da- ni -na	Σ-da-na
2d	Σ-i- ni -na	2d	Σ-to- ni -na	Σ-i- ni -na	Σ-i-ko-na				Σ-i- ni -na	Σ-i- ni -na	Σ-i- ni -na
2p	Σ-i-ni	2p	Σ-to-na	Σ-eko-na	Σ-i-ko-na				Σ-i-ni	Σ-i-ni	Σ-i-ni
3s	Σ-e	3s	Σ-to	Σ-i- ni	Σ-i-ki	Σ-e-na	Σ-i- ni -na	Σ-i-ni	Σ-da	Σ-da- ni	Σ-da-ni
3d	Σ-i- ni	3d	Σ-to- ni	Σ-i- ni	Σ-i-ki	Σ-e-na	Σ-i- ni -na	Σ-i-ni	Σ-da- ni	Σ-da- ni	Σ-da-ni
3p	Σ-i-ni	3p	Σ-to-ni	Σ-i- ni	Σ-i-ki	Σ-e-na	Σ-i- ni -na	Σ-i-ni	Σ-da-ni	Σ-da-ni	Σ-da-ni

Hypothesis *-si* ↔ [+du]

Analyzing marker interactions with morpheme highlighting

Ainu (Isolate, Japan, Tamura 2000)

Intransitive Transitive

	1s	1p	2s	2p	3s	3p	x	
1s	ku-Σ		eci-Σ	eci-Σ	ku-Σ	ku-Σ	ku-i-Σ	
1p	Σ-as	1p	eci-Σ	eci-Σ	ci-Σ	ci-Σ	a-i-Σ	
2s	e-Σ	2s	en-Σ	un-Σ		e-Σ	e-Σ	e-i-Σ
2p	eci-Σ	2p	eci-en-Σ	eci-un-Σ		eci-Σ	eci-Σ	eci-i-Σ
3s	Σ	3s	en-Σ	un-Σ	e-Σ	eci-Σ	Σ	i-Σ
3p	Σ	3p	en-Σ	un-Σ	e-Σ	eci-Σ	Σ	i-Σ
x	i-Σ-an	x	a-en-Σ	a-un-Σ	a-e-Σ	a-eci-Σ	a-Σ	a-i-Σ

Hypotheses *eci-1* → [+2+pl] *eci-2* → [+1][+2+sg] *e* → [+2+sg]

Finding hierarchy effects with morpheme highlighting

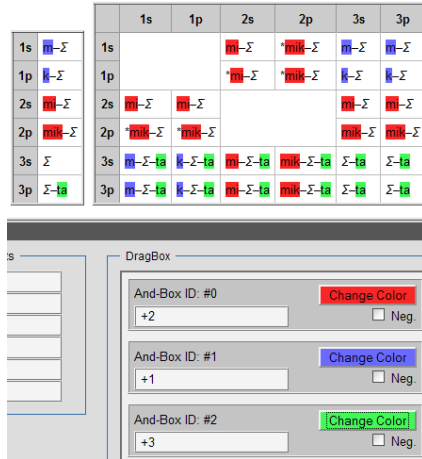
Kohi 2 ↔ 3, 3 → 3

	2d	2p	3s	3d	3p
2d			Σ-i-si-na	Σ-i-si-na	Σ-i-si-na
2p			Σ-i- ni	Σ-i- ni	Σ-i- ni
3s	Σ-i-si-na	Σ-i- ni	Σ-da	Σ-da-si	Σ-da- ni
3d	Σ-i-si-na	Σ-i- ni	Σ-da-si	Σ-da-si	Σ-da- ni
3p	Σ-i-si-na	Σ-i- ni	Σ-da- ni	Σ-da- ni	Σ-da- ni

Morpheme hypotheses *-si* ↔ [+du] *-ni* ↔ [+pl]
 Agreement hierarchies 2 > 3 pl > du

Visualizing hierarchy effects with feature highlighting

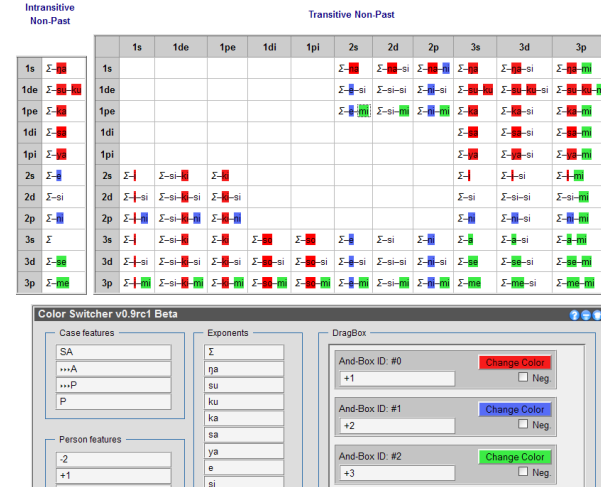
Reyesano (Tacanan, Bolivia, Guillaume 2009)



Prefix hierarchy $2 > 1$
 Suffix hypothesis $-ta \leftrightarrow [\text{Nom} +3]$

Finding feature based linearization with feature highlighting

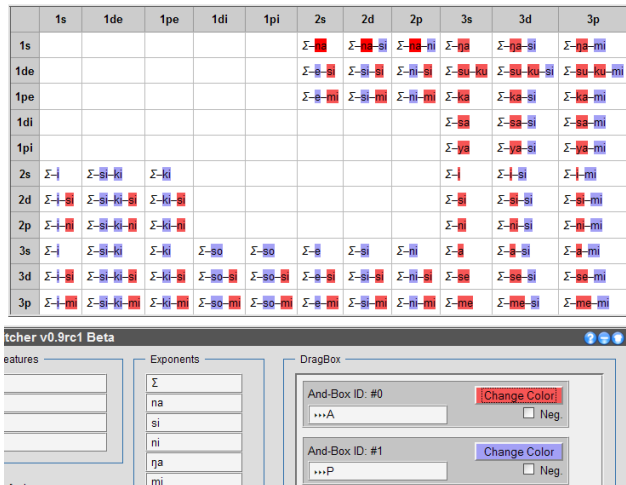
Older Bahing (Kiranti, Nepal, Houghton 1858)



Linearization hierarchy $1 > 2 > 3$

Finding feature based linearization with feature highlighting

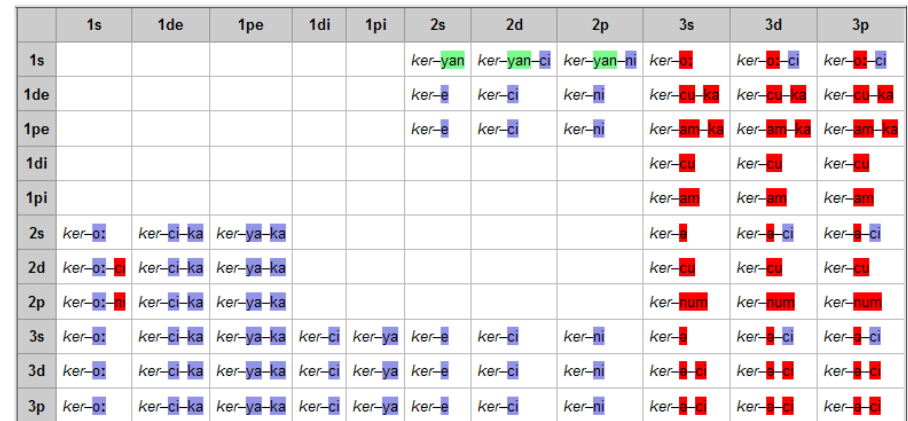
Older Bahing, agreed with argument



Linearizations $A > P$ and $P > A$

Determining marker alignment with argument highlighting

Kulung, agreed with argument (Kiranti, Nepal, Tolsma 2006)



- single argument marking, not when agreed with argument is singular
- generic SP-markers, special A-markers for $X \rightarrow 3$, $1s \rightarrow 2$ portmanteau

Finding hierarchical alignment with argument highlighting

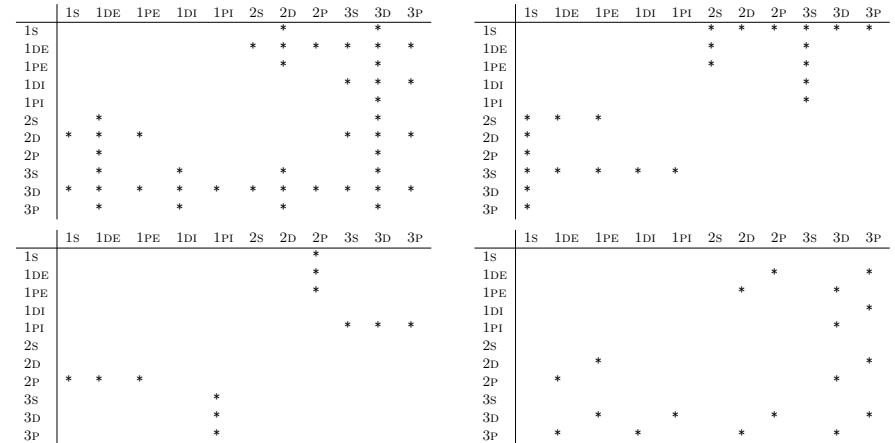
Hayu, agreed with argument (Kiranti, Nepal, Michailovsky 1981)

	1s	1de	1pe	1di	1pi	2s	2d	2p	3s	3d	3p
1s						pUk- ne	pUk- ne-tshik	pUk- ne	pUk- ne	pUk- ne-tshik	pUk- ne
1de						pUk- shok	pUk- shok	pUk- shok	pUk- shok	pUk- shok	pUk- shok
1pe						pUk- kok	pUk- kok	pUk- kok	pUk- kok	pUk- kok	pUk- kok
1di									pUk- tshik	pUk- tshik	pUk- tshik
1pi									pUk- ka	pUk- ka	pUk- ka
2s	pUk- ne	pUk- tshok	pUk- kok						pUk	pUk	pUk- me
2d	pUk- ne-tshik	pUk- tshok	pUk- kok						pUk- tshik	pUk- tshik	pUk- tshik
2p	pUk- ne	pUk- tshok	pUk- kok						pUk- ne	pUk- ne	pUk- ne
3s	pUk- ne	pUk- tshok	pUk- kok	pUk- tshik	pUk- ka	pUk	pUk- tshik	pUk- ne	pUk	pUk- tshik	pUk- me
3d	pUk- ne-tshik	pUk- tshok	pUk- kok	pUk- tshik	pUk- ka	pUk	pUk- tshik	pUk- ne	pUk- tshik	pUk- tshik	pUk- me
3p	pUk- ne	pUk- tshok	pUk- kok	pUk- tshik	pUk- ka	pUk- ne	pUk- tshik	pUk- ne	pUk- ne	pUk- ne	pUk- ne

- single argument marking, not when agreed with argument is singular
- 1 > 2 hierarchy, 1s → 2 portmanteau

Limitations

- distributional pattern of pure number markers is still hard to see
- underspecified portmanteau patterns (ambiguous exponents)



Outlook

- identify and visualize blocking relations between markers
- visualize linear order relations between markers (slots, template)

	1s	1p	2s	2p	3s	3p	x
1s			eci-∑	eci-∑	ku-∑	ku-∑	ku-i-∑
1p			eci-∑	eci-∑	ci-∑	ci-∑	a-i-∑
2s	en-∑	un-∑			e-∑	e-∑	e-i-∑
2p	eci-en-∑	eci-un-∑			eci-∑	eci-∑	eci-i-∑
3s	en-∑	un-∑	e-∑	eci-∑	∑	∑	i-∑
3p	en-∑	un-∑	e-∑	eci-∑	∑	∑	i-∑
x	a-en-∑	a-un-∑	a-e-∑	a-eci-∑	a-∑	a-∑	a-i-∑

Slot Analysis

-3	-2	-1	0
a/1	a/2	ci	stem
e	en		
eci	i		
ku	un		

- identify hierarchical agreement patterns
- subanalyzing paradigm learning algorithms
- typical location of portmanteaus in paradigms

Conclusion

- restructuring paradigm tables helps with small paradigms
- highlighting of morphemes helps to identify patterns, reduce errors and make aware analysis strategies and options
- different abstraction levels (morphemes, features, agreed with argument) are needed to identify more complex patterns
- visual strategies can be transferred to learning algorithms

- Guillaume, Antoine (2009), Hierarchical Agreement and Split Intransitivity in Reyesano. *International Journal of American Linguistics* 75(1): 29–48.
- Hodgson, Brian Houghton (1858), Comparative vocabulary of the languages of the broken Tribes of Nepal (part II). *Journal of the Asiatic Society of Bengal* 27:393–456.
- Lahaussois, Aimée (2009), Koyi Rai: An initial grammatical sketch. *Himalayan Linguistics Archive* 4:1-33.
- Michailovsky, Boyd (1981), Grammaire de la langue hayu (Népal). Dissertation or doctoral thesis. Berkeley, University of California
- Tamura, Suzuko (2000), *The Ainu Language*. Tokyo: Sanseido.
- Tolsma, Gerard Jacobus (2006), A grammar of Kulung. Languages of the greater Himalayan region. Brill Academic Publishers.