Assessing the typology of person portmanteaux

Sebastian Bank, Daniela Henze, Jochen Trommer & Eva Zimmermann – University of Leipzig

There is a widespread intuition in the literature that the distribution of portmanteaux affixes in transitive agreement paradigms is at least partially non-arbitrary and exhibits a strong cross-linguistic preference for 'local' contexts, i.e. clauses where both arguments are 1st or 2nd person (Heath 1991, 1998, Wunderlich 2006, Georgi 2012, Nevins 2012): the *Local Portmanteau Hypothesis*. A similar, but orthogonal claim is made by Lakämper & Wunderlich (1998) and Woolford (2010) who argue that person portmanteaux have a special affinity to 'direct' contexts, i.e., clauses where the subject is higher than the object on the person hierarchy $1 \succ 2 \succ 3$: the *Direct Portmanteau Hypothesis*. Woolford, Lakämper and Wunderlich draw their evidence for the Direct Portmanteau Hypothesis from only a handful of languages, but the most extensive empirical studies on the Local-Portmanteau Hypothesis, Heath (1991, 1998), cite an impressive inventory of local person portmanteaux from different languages. However, to establish the cross-linguistic affinity of portmanteau agreement and local scenarios, it is obviously not sufficient to demonstrate that this context exhibits many portmanteaux, but necessary to show that it exhibits significantly *more* portmanteaux than other paradigmatic contexts. In this talk, we report results of a typological pilot study which tests both hypotheses and related claims against a small, but crosslinguistically balanced language sample. As the identification of portmanteaux depends on understanding complete inflectional paradigms, which is underdetermined by the empirical data and prone to interference from alternating analytical biases, we use automatic computational procedures to segment and analyze unsegmented affix-paradigms. With this novel approach, the criteria for identifying (non-)portmanteaux are fully transparent and ensured to be uniform. Our results indicate that there is no significant cross-linguistic effect of the Local and Direct Portmanteau Hypothesis. We examine alternative learning strategies, and show that our results are independent of the specific evaluation metrics for affix segmentation.

Hypotheses on the distribution of portmanteaux Heath (1991, 1998) argues that due to the pragmatic awkwardness of transitive predications involving both the speaker and the hearer of a speech act, there is a strong crosslinguistic tendency to disguise this situation semantically, and more concretely to make it morphologically less transparent (opaque) in transitive pronominal agreement paradigms, e.g. by the neutralization of number features in this context, or by using allomorphs for the participants not employed elsewhere in the language. Under the assumption that portmanteaux are also morphosyntactically less transparent than marking a $1 \leftrightarrow 2$ predication by distinct affixes for subject and object, he concludes that using portmanteau affixes is also a strategy to the same end. Basically the same claim is made by Wunderlich (2006: 4) who states: "To express the combination I $\rightarrow$ you is a special communicative task, so it does not wonder that a portmanteau morpheme adapted to this special task is found in several languages."

(1) **The Heath-Wunderlich Prediction**

	1	**2**	**3**
1		P	
2	P		
3			

Woolford (2010) advocates a related claim on the distribution of person portmanteaux which implies that only specific local scenarios – those with 1st person subjects and 2nd person objects – show a tendency to exhibit portmanteaux. In fact she argues that the preference for portmanteaux in transitive agreement crosscuts the local/non-local distinction and reflects a person hierarchy in the sense that there portmanteaux affixes are preferred in constellations where the subject is higher in person than the object. This assumption is implemented by the constraint in (2):

(2) **Person Restriction on Portmanteau Agreement Formation:** (Woolford 2010: 24-25)
In a portmanteau agreement form, the person of the subject must be higher than or equal to the person of the object.

Under the assumption that there is a crosslinguistically uniform person hierarchy $1 \succ 2 \succ 3$, (2) predicts that $1 \rightarrow 2$ forms should exhibit more portmanteaux than $3 \rightarrow 2$ forms, but also $2 \rightarrow 1$ forms, and that the preference for portmanteaux should also be observable in $1 \rightarrow 3$ and $2 \rightarrow 3$ forms. Whereas Woolford's explicit claim thus goes beyond local scenarios, the evidence she cites to establish (2) is restricted to local forms. She argues based on the sample of languages cited in Heath (1998) that there are languages which have $1 \rightarrow 2$ portmanteaux, but no $2 \rightarrow 1$ portmanteaux, and languages which have portmanteaux in both contexts, but no languages with $2 \rightarrow 1$ portmanteaux that lack $1 \rightarrow 2$ portmanteaux, resulting in the typology in (3):

(3) **Woolford's Typology of Local Person Portmanteaux**

		2 → 1 Portmanteaux	
		+	−
1 → 2 Portmanteaux	+	attested	attested
	−	non-attested	attested

The same empirical predictions follow from a proposal by Lakämper & Wunderlich (1998). Strictly speaking, the Heath-Wunderlich prediction and the Woolford-Lakämper-Wunderlich prediction are not logically incompatible. Local contexts could show a preference for person portmanteaux over non-local contexts, and direct contexts over inverse ones predicting a distribution as in (4) where "PP" marks a particular preponderance of portmanteaux:

(4) **The Woolford-Lakämper-Wunderlich Prediction and its combination with (1)**

	1	**2**	**3**
1		P	P
2			P
3			

	1	**2**	**3**
1		**PP**	P
2	P		P
3			

Methodology Methodologically our study applies a novel approach to one of the principle problems of morphological typology: the analytic underdetermination of the categories under investigation by empirical data. Thus it is by no means obvious whether a string of segments employed in transitive agreement morphology is a portmanteau or a combination of separable affixes. We tackle this problem by applying automatic decomposition algorithms to full affixal paradigms for regular verb stems. Thus the criteria for identifying (non-)portmanteau affixes become fully explicit and are applied homogeneously across the languages of our sample. Crucially, they are blind with respect to the local/non-local or hierarchical distinctions.

As the combination of segmentation options and meaning assignments results in combinatorial explosion even for smaller paradigms, the space of possible analyses is too large for a naive search algorithm. The procedure in (5) uses a greedy approach such that the analysis of a paradigm is conducted in individual learning steps. At each step, the learner searches for the best form-meaning-pair hypothesis that can be identified in the current state of the paradigm according to an OT-like optimization. Different analytic biases can be explored by manipulating the ranking of the optimization criteria. After the optimal hypothesis is found, it is added to the lexicon, its instances are removed from the paradigm and the algorithm proceeds to the next step until all material has been learned and the paradigm is empty:

(5) **Iterative Algorithm for Segmentation and Analysis**

a. (i) Build the set P of perfect hypotheses, i.e. all $\langle form, meaning \rangle$ pairs combining all free affix strings that do not have a free affix as substring with every meaning specification such that the meaning subsumes only and all the paradigm cells where the form occurrs (free or bound)

(ii) Identify the best marker hypotheses $O \subseteq P$ having ($\alpha > \beta > \gamma > \delta$):

- α maximal number of (free or bound) true positives
- β non-portmanteau > portmanteau
- γ maximal number of segments
- δ minimal number of blind cells

b. If $O = \emptyset$

(i) Build the set M of marker hypotheses with minimum 50% precision, i.e. all $\langle form, meaning \rangle$ pairs combining all free affix strings with every meaning specification such that at least one half of the cells subsumed by its meaning contain an occurrence of the form (free or bound)

(ii) Identify the best marker hypotheses $O \subseteq M$ having ($\alpha > \beta > \ldots \theta$):

- α maximal number of free true positives
- β non-portmanteau > portmanteau
- γ maximal number of bound true positives
- δ minimal number of false negatives
- ε minimal number of false positives
- ζ maximal number of segments
- η minimal number of blind cells

c. (i) Add some $\langle form, meaning \rangle \in O$ to the lexicon, let $O = \emptyset$ and remove a single occurrence of *form* from all paradigm cells subsumed by *meaning*

(ii) If any paradigm cell has a free occurrence of *form*, goto step b.

d. If any paradigm cell has a (non-empty) affix string, goto step a, else end.

Whenever the algorithm fails to identify a perfect hypothesis in (5a), it falls back to (5b) which also learns less accurate form-meaning pairs. This is crucial to heuristically distinguish portmanteaux from non-portmanteaux with imperfect surface distributions that often result from being blocked, possibly by portmanteaux. Consider for example the algorithm run result in (6):

(6) **Hixkaryana transitive agreement** (Derbyshire 1985), **categorized by (5)**

			1s	**1pe**	**1pi**	**2s**	**2p**	**3s**	**3p**
1s	k_1-	**1s**				k_1-	k_1-	ø-	ø-
1pe	n_2-	**1pe**				o-	o-	n_1-	n_1-
1pi	t-	**1pi**						t-	t-
2s	m-	**2s**	m-					m-	m-
2p	m-	**2p**	m-					m-	m-
3s	n_1-	**3s**	ro-	*	k_2-	o-	o-	n_1-	n_1-
3p	n_1-	**3p**	ro-	*	k_2-	o-	o-	n_1-	n_1-

t-	↔	SA[+1 +2]
ø-	↔	[+1 +sg]A→P[+3]
m-	↔	SA[–1 +2]
n_1-	↔	SP[+3]
n_2-	↔	S[+1 –2 +pl]
o-	↔	P[–1 +2]
ro-	↔	[+3]A→P[+1 +sg]
k_1-	↔	SA[+1 +sg]
k_2-	↔	P[+1 +2]

While *t-*:SA[+1 +2] and *m-*:SA[–1 +2] are 100% accurate, n_1-:SP[+3] has a 50% accurate person distribution such that it only occurs in 8 of the 16 cells that match its meaning possibly due to blocking by the perfect markers *m*-:SA[−1 +2], *t*-:SA[+1 +2], and ø-:[+1 +sg]A→P[+3]. For ø- on the other hand, neither [+1 +sg] nor [+3] make good enough meanings, both cannot easily reproduce the distribution by being blocked, so the portmanteau-meaning [+1 +sg]A→P[+3] remains the best choice.

Results For the pilot study, the learning procedure(s) have been applied to an areally and genealogically diverse sample of 26 languages. Only languages with obligatory agreement with A and P arguments on the verb were considered. Only in three of the languages marker occurrence locality interacts with portmanteau-status (Fisher's exact $p \leq .05$, cf. asterisks below), in only one (Tepehuan) in the predicted direction. As the number of cells and markers vary, and all languages have more non-local than local paradigm cells, we counted for every paradigm cell the ratio of portmanteau affixes from the total affixes in the cell. The Direct Portmanteau Hypothesis predicts that the mean portmanteau/affix ratio for local cells tends to be higher than the mean portmanteau/affix ratio for non-local cells. As shown in Figure 1, Ket, Maricopa, Reyesano, and Sahu where categorized to have no portmanteaus at all and therefore contribute no non-local vs. local distinction. The remaining languages did not show a significant correlation of non-local/local with low/high portmanteau ratio (point biserial correlation $p \not\leq .05$). In fact, 7 are analyzed to have portmanteaux only in non-local cells but only Lakhota to have them only in local cells.

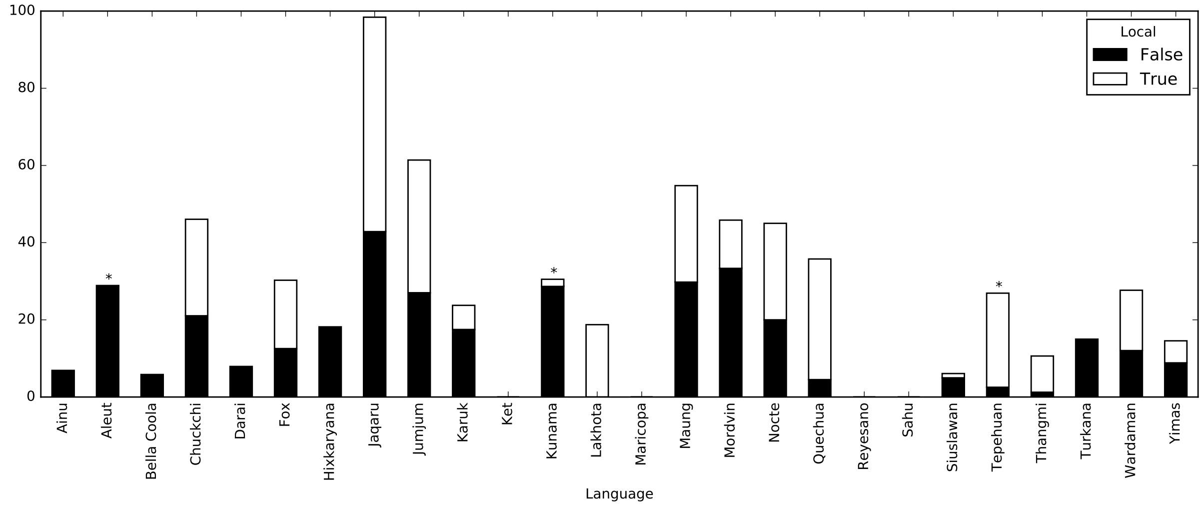

Figure 1: Mean percentage of portmanteau affixes per cell: non-local (black) vs. local (white)

References

Derbyshire, Desmond C. 1985. *Hixkaryana and Linguistic Typology*. Dallas: Summer Institute of Linguistics.

Georgi, Doreen. 2012. Deriving the Distribution of Person Portmanteaux by Relativized Probing. In *Proceedings of the North-Eastern Linguistic Society 42 (2011, Toronto)*, .

Heath, Jeffrey. 1991. Pragmatic disguise in pronominal-affix paradigms. In Frans Plank (ed.), *Economy of Inflection*, 75–89. Berlin: Mouton de Gruyter.

Heath, Jeffrey. 1998. Pragmatic Skewing in 1-2 pronominal combination in native American languages. *IJAL* 64(2). 83–104.

Lakämper, Renate & Dieter Wunderlich. 1998. Person marking in Quechua – A constraint-based minimalist analysis. *Lingua* 105. 113–148.

Nevins, Andrew. 2012. Haplological Dissimilation at Distinct Stages of Exponence. In Jochen Trommer (ed.), *The Morphology and Phonology of Exponence – The State-of-the-Art*, Oxford: Oxford University Press.

Woolford, Ellen. 2010. Active Stative Agreement in Choctaw and Lakota. In Gabriel de Ávila Othero & Sérgio de Moura Menuzzi (eds.), *Optimality Theoretic Syntax*, vol. 8:4 Special Issue of Revista Virtual de Estudos da Linguagem, 6–46. Revista Virtual de Estudos da Linguagem.

Wunderlich, Dieter. 2006. Towards a structural typology of verb classes. In Dieter Wunderlich (ed.), *Advances in the theory of the lexicon*, 57–166. Berlin: Mouton de Gruyter.