Thoughts on feature intersection

Tabea Reiner
LMU Munich
tabea.reiner@lmu.de

1. Feature intersection: the state of the art

1.1. What is feature intersection?

Feature intersection is a certain configuration in paradigms, which has been used as a criterion in the definition of canonical periphrasis (Spencer & Popova 2015: 214, 217). More precisely, the notion captures the following situation: a given cell is realized analytically although its two determining feature values are realized synthetically elsewhere in the paradigm(s) at hand (Brown et al. 2012: 250–252). Please note that despite the term *feature intersection* it is in fact the feature *values* that are relevant (Brown et al. 2012: 251). To get a clearer idea of the notion and its application to periphrasis, consider the potential paradigm in Table 1.

Table 1: Potential verbal paradigm of present tense (first part)

	active	
	singular	plural
1	s	a
2	a	s
3	a	s

Every *s* stands for a synthetic realization of the respective cell and every *a* stands for an analytic realization of the respective cell (the bold lines will become important later). I am abstracting away from any concrete linguistic material here since it would not add to the argument to be presented below; however, cf. section 2.2.2 for some data. Now consider the lightly shaded cell, that is to say {active, 3rd person, singular}. This cell is realized analytically although its two determining feature values, i.e. singular number and third person, are realized synthetically elsewhere in the paradigm: singular number has a synthetic realization in the 1st person and third person has a synthetic realization in the plural. Metaphorically speaking, you follow the two paths that cross in the cell under scrutiny and scan them for synthetic realizations. The result here is: yes, there is a pair of synthetic realizations, hence there is feature intersection. Please note that detecting this pair suffices for identifying feature intersection, irrespective of another cell "on the way", i.e. {active, 2nd person, singular}, having an analytic realization. Given feature intersection for the cell at hand now, we may count its analytic realization as unexpected or deviant, which matches the pre-theoretical intuition that this realization is a mere surrogate for a synthetic form. In other words: thanks to the notion of feature intersection, we can substantiate our intuition that this is a local exception to an overall synthetic pattern by simply observing a well-defined configuration. As an analytic exception to the synthetic rule such a form makes a good candidate for periphrasis and, in fact, feature intersection is one of the most popular criteria in defining the canonical ideal of periphrasis (cf. the references given above as well as Ackerman & Stump 2004, Bonami 2015; for an introduction to Canonical Typology cf. Corbett 2012, ch. 6).

However, it is not always as easy. Consider Table 2 below, in particular the dark shaded cell, i.e. {passive, 2nd person, plural}.

Table 2: Potential verbal paradigm of present (second part)

	passive	
	singular	plural
1	a	s
2	a	a
3	a	a

Following the same strategy as before, we do obtain one synthetic realization (for plural number) but fail to find a second one (for 2nd person). So can we safely conclude that, in this case, there is no feature intersection? No, we cannot. Although exactly two features have to be taken into account, the analyst is free to combine whichever two features fit the purpose. For example, keeping number constant we may look at voice and person. This means combining the bold-framed columns from the previous tables:

Table 3: Plural columns from tables 1 and 2

	active	passive
1	a	s
2	s	a
3	s	a

Now we see not only a synthetic realization of passive voice, viz. in the 1st person, but also a synthetic realization of 2nd person, viz. in the active voice. So, yes, there is feature intersection for {passive, 2nd person, plural} after all. All we had to do was merge paradigms (for a concrete example of such a solution cf. Brown et al. 2012: 252).

1.2. Where does feature intersection run into problems?

As indicated in the first subsection, the procedure for detecting feature intersection can become quite cumbersome. This is especially relevant when

- paradigms are many; e.g., voice, tense, aspect, and mood forms of one verb can hardly be squeezed into one 2x2 paradigm
- paradigms are large
- analytic realizations are scattered across cells
- or any blend of these conditions arises.

Furthermore, it is difficult to establish the absence of feature intersection. For proving this hypothesis systematically (s)he would have to go through each and every feature combination, each time showing that the combination does not reveal feature intersection.

Admittedly, all of these scenarios might represent rather unusual situations in morphologists' everyday work. However, morphological theory must respond also to the unusual. Against this background the following section intends to provide a more workable method of identifying of feature intersection, which boils down to using an alternative notation.

2. An alternative notation: two times one step

2.1. How does it work?

As before, the aim is to assess whether there is feature intersection (fi) and the point of departure is an analytically realized cell, generally: $c = \{f_x, g_y, h_z, \ldots.\}$, where f, g, h are features and x, y, z are their values. Also the procedure stays the same in principle; however, due to its formulation it is, I hope, more convenient to follow in the general case:

1. change exactly one of the feature values	$c = \{f_x, g_y, h_z, \ldots.\} \rightarrow d_1 = \{f_k, g_y, h_z, \ldots.\}$
2. repeat for all feature-value pairs in c	$c = \{f_x, g_y, h_z, \ldots.\} \rightarrow d_2 = \{f_x, g_m, h_z, \ldots.\}$
	$c = \{f_x, g_y, h_z, \ldots.\} \rightarrow d_3 = \{f_x, g_y, h_ß, \ldots.\}$
	…
3. check d_1, d_2, d_3,…	if <**2** ds with s-realization: no fi
	if ≥**2** ds with s-realization: fi

Metaphorically speaking, you walk through the paradigm(s) but take only **one** step at a time from your origin c. Then d_1, d_2, d_3,… are the cells you reach and check. Strictly speaking, they are groups of cells: since every step corresponds to the manipulation of one feature-value pair in the set, there are as many d_1s as there are values for that feature minus 1. For example, with a bivalent feature there is only one d_1, with a trivalent feature there are two d_1s etc. Needless to say, the same holds for d_2, d_3,… The only relevant question, however, is this: does any cell in the group have a synthetic realization? Recall the irrelevance of {active, 2nd person, singular} being realized analytically in table 1. So, what counts, again metaphorically speaking, is the single step as a whole. Of these steps at least **two** have to reach a synthetic realization if we want to diagnose feature intersection. From this perspective, the pivoting done by a basketball player might make a good metaphor.

Please note that a given cell may consist in any number of feature-value pairs, many more than can be shown graphically in a 2x2 paradigm. This limitation is already alluded to in Brown et al. (2012: 251–252) where the present proposal has its forerunner. However, the approach is not stated as explicitly there. Now that it is stated explicitly, it can be used for further purposes, as laid out in the following subsection (2.2).

2.2. Further applications

2.2.1. Quantifying the degree of canonicity

Canonical Typology is rightly proud of providing both the logical end point of phenomena, i.e. the canonical ideal, and a way to measure deviation from the ideal (Corbett 2012: 154). In practice this usually means: first, the canonical ideal is defined by several criteria; second, for any given pair of phenomena we let the criteria decide whether the two phenomena differ in canonicity and, if yes, which one is the more canonical (I am including here my own work within Canonical Typology, Reiner 2019). Thus, the criteria serve to rank the phenomena but, crucially, the ranks are not associated with any numerical scale, neither with respect to the whole set of criteria nor with respect to any single criterion. So strictly speaking the promise of measurability is not kept.[1]

However, with regard to feature intersection as a criterion for canonical periphrasis, such a scale is easily available if the criterion is conceptualized as suggested in the preceding

[1] However, cf. Round & Corbett (in press).

subsection (2.1). This is shown in Table 4 with #c = number of feature-value pairs in c, n $\in \mathbb{N}$ and #c – n ≥2.

Table 4: Canonicity scale (general)

number of ds with s-realization	degree of canonicity
#c	#c/#c = 1
#c – n	#c – n/#c

To be slightly more concrete, imagine that c consists of six feature-value pairs. Then the scale looks like Table 5.

Table 5: Canonicity scale (example)

number of ds with s-realization	degree of canonicity
6	6/6 = **1**
6 – 1 = **5**	(6 – 1)/6 = 5/6 = $\mathbf{0.8\overline{3}}$
6 – 2 = **4**	(6 – 2)/6 = 4/6 = $\mathbf{0.\overline{6}}$
6 – 3 = **3**	(6 – 3)/6 = 3/6 = **0.5**
6 – 4 = **2**	(6 – 4)/6 = 2/6 = $\mathbf{0.\overline{3}}$

Such a scale provides three kinds of useful quantifications. First, the canonical ideal is assigned "1", reflecting a situation where the number of direct steps to s-realized cells equals the highest possible number (i.e. the number of feature value-pairs in c) so that their ratio is #c/#c. In linguistic terms this is a situation where *every* feature value of the cell under scrutiny is realized synthetically somewhere else so that the given cell's analytic realization appears maximally unexpected. Second, the same logic applies to cases where the actual number of direct steps to s-realized cells is less than the maximally possible number: again the ratio between the former and the latter gives the degree of canonicity, this time below 1. Third, the tables suggest stopping at #c – n = 2, since below that there is no feature intersection (cf. 2.1). So a candidate periphrasis that fails to show feature intersection altogether would not be assigned any degree of canonicity (not even a low one) with respect to the criterion. This scenario refutes the occasional criticism that Canonical Typology does not specify lower bounds on the application of terms, so that, e.g., verbs end up being maximally non-canonical nouns.

In order to flesh out the scenario, consider the following hypothetical situation (table 6).

Table 6: no feature intersection

	active		passive	
	singular	singular	singular	plural
1	a	a	a	s
2	a	a	a	s
3	a	a	a	a

The method proceeds as follows:

c = {$\text{voice}_{\text{active}}$, person_3, $\text{number}_{\text{sg}}$}
d_1 = {$\text{voice}_{\text{passive}}$, person_3, $\text{number}_{\text{sg}}$} no s reached
d_{2a}= {$\text{voice}_{\text{active}}$, person_2, $\text{number}_{\text{sg}}$} no s reached
d_{2b}= {$\text{voice}_{\text{active}}$, person_1, $\text{number}_{\text{sg}}$} no s reached
d_3 = {$\text{voice}_{\text{active}}$, person_3, $\text{number}_{\text{pl}}$} no s reached

At this point, no more modifications are possible, since every feature has been modified in every possible way. Please recall that it is not permitted to change the values of more than one feature at a time. So for the hypothetical situation in table 6 we can safely conclude: there is no fi, hence the realization of c is definitely *not* periphrastic according to the criterion of feature intersection.

Coming back to measureability more generally, the measurement introduced above could (and should) be refined in future research by taking into account also *how many* s-realized cells are reached by a given single step, non-metaphorically: starting again from $c = \{f_x, g_y, h_z, \ldots\}$ the question is whether only $d_{1a} = \{f_k, g_y, h_z, \ldots\}$ is realized synthetically or also $d_{1b} = \{f_u, g_y, h_z, \ldots\}$, $d_{1c} = \{f_v, g_y, h_z, \ldots\}$, $d_{1d} = \{f_w, g_y, h_z, \ldots\}$, etc.

2.2.2. When the non-canonical is the usual: the case of German verbal paradigms

It is a truism within Canonical Typology that what we expect to find in natural languages is not the canonical ideal but rather a range of phenomena that deviate from the ideal to various extents. The question, partly answered in this paper, is how to *measure* the deviations precisely and when to stop applying the term associated with the canon. For example, German verbal paradigms are famous for displaying so many multi-word realizations that calling these realizations “periphrasis” might seem weird: do they really constitute the exception while their much rarer synthetic counterparts supply the rule? Applying the method from the previous subsection, this question can be answered. More precisely, we can determine the exact degree to which each of the realizations is a canonical periphrasis according to the criterion of feature intersection.

Let’s take as an example the verb *fangen* ‘catch’ and its form *hat gefangen*, which realizes $c = \{\text{voice}_{\text{active}}, \text{mood}_{\text{indicative}}, \text{tense}_{\text{perfect}}, \text{person}_3, \text{number}_{\text{sg}}\}$. Where do we get from here by taking one step at a time, i.e. which ds do we reach? Here is the list, beginning with a restatement of the starting point.

c	=	$\{\text{voice}_{\text{active}}, \text{mood}_{\text{indicative}}, \text{tense}_{\text{perfect}}, \text{person}_3, \text{number}_{\text{sg}}\}$	*hat gefangen*
d_1	=	$\{\text{voice}_{\textit{passive}}, \text{mood}_{\text{indicative}}, \text{tense}_{\text{perfect}}, \text{person}_3, \text{number}_{\text{sg}}\}$	*ist gefangen worden*
d_{2a}	=	$\{\text{voice}_{\text{active}}, \text{mood}_{\textit{conditional I}}, \text{tense}_{\text{perfect}}, \text{person}_3, \text{number}_{\text{sg}}\}$	*habe gefangen*
d_{2b}	=	$\{\text{voice}_{\text{active}}, \text{mood}_{\textit{conditional II}}, \text{tense}_{\text{perfect}}, \text{person}_3, \text{number}_{\text{sg}}\}$	*hätte gefangen*
d_{3a}	=	$\{\text{voice}_{\text{active}}, \text{mood}_{\text{indicative}}, \text{tense}_{\textit{present}}, \text{person}_3, \text{number}_{\text{sg}}\}$	***fängt***
d_{3b}	=	$\{\text{voice}_{\text{active}}, \text{mood}_{\text{indicative}}, \text{tense}_{\textit{preterite}}, \text{person}_3, \text{number}_{\text{sg}}\}$	***fing***
d_{3c}	=	$\{\text{voice}_{\text{active}}, \text{mood}_{\text{indicative}}, \text{tense}_{\textit{future}}, \text{person}_3, \text{number}_{\text{sg}}\}$	*wird fangen*
d_{3d}	=	$\{\text{voice}_{\text{active}}, \text{mood}_{\text{indicative}}, \text{tense}_{\textit{pluperfect}}, \text{person}_3, \text{number}_{\text{sg}}\}$	*hatte gefangen*
d_{3e}	=	$\{\text{voice}_{\text{active}}, \text{mood}_{\text{indicative}}, \text{tense}_{\textit{future perfect}}, \text{person}_3, \text{number}_{\text{sg}}\}$	*wird gefangen haben*
d_{4a}	=	$\{\text{voice}_{\text{active}}, \text{mood}_{\text{indicative}}, \text{tense}_{\text{perfect}}, \text{person}_{\mathit{1}}, \text{number}_{\text{sg}}\}$	*habe gefangen*
d_{4b}	=	$\{\text{voice}_{\text{active}}, \text{mood}_{\text{indicative}}, \text{tense}_{\text{perfect}}, \text{person}_{\mathit{2}}, \text{number}_{\text{sg}}\}$	*hast gefangen*
d_5	=	$\{\text{voice}_{\text{active}}, \text{mood}_{\text{indicative}}, \text{tense}_{\text{perfect}}, \text{person}_3, \text{number}_{\textit{pl}}\}$	*haben gefangen*[2]

In prose and partly metaphorical: there are five feature-value pairs in c, so there are five directions to take a single step and five chances to reach s-realized cells; however, in actual fact only the third step is successful (bold face)—so the realization under scrutiny, i.e. *hat gefangen*, stays below the threshold of 2/5 and does not qualify as a periphrasis, not even as a mildly canonical one. Please note that *fängt* and *fing* count as one since they are reached by the same step, i.e. by manipulation of the same feature-value pair. Also note that the additional synthetic form *finge* does not count since it realizes a set of feature-value pairs that

---

[2] Complete paradigms can be found in Helbig & Buscha (2001: 23-25).

differs from c in not just one but two places, i.e. {$voice_{active}$, $mood_{conditional\ II}$, $tense_{present}$,[3] $person_{3}$, $number_{sg}$}. So it is not "two times one step but one time two steps".

3. Conclusion

In this paper, I have promoted an approach to feature intersection that makes the notion at the same time more accessible and more generally applicable, viz. applicable to arbitrarily large sets of features. In particular, it provides a real measure for the canonicity of candidate periphrases with respect to an important criterion, i.e. feature intersection. Unfortunately, not all criteria for all canonical ideals are as easily quantifiable. Maybe, generally, canonical ideals could be presented as sets of equally weighted constraints so that deviations could be measured in an OT-fashion (Kuhn 2003).

Acknowledgements

Sections 1 and 2.1 of the present paper originate from a poster presentation, during which somebody asked "Are you from a Computational Linguistics background?". I am not but this question made me think that, in fact, there is something to be computed here, so the question inspired section 2.2.
Moreover, thanks to Jenny Audring for helpful comments on the first version of this paper, especially the suggestion to discuss the situation in table 6.

References

Ackerman, F. & G. T. Stump. 2004. Paradigms and periphrastic expression. A study in realization-based lexicalism. In L. Sadler & A. Spencer (Eds.), *Projecting Morphology*. Stanford: CSLI (Stanford Studies in Morphology and the Lexicon 1), 111–157.

Bonami, O. 2015. Periphrasis as collocation. *Morphology* 25(1): 63–110.

Brown, D., M. Chumakina, G. G. Corbett, G. Popova & A. Spencer. 2012. Defining 'periphrasis'. Key notions. *Morphology* 22(2): 233–275.

Corbett, G. G. 2012. *Features*. Cambridge: Cambridge University Press (Cambridge Textbooks in Linguistics).

Helbig, G. & J. Buscha. 2001. *Deutsche Grammatik. Ein Handbuch für den Ausländerunterricht*. Berlin: Langenscheidt.

Kuhn, J. 2003. *Optimality-theoretic syntax*. *A declarative approach*. Stanford: CSLI (Studies in Constraint-Based Lexicalism).

Reiner, T. 2019. Variation in non-finiteness and temporality from a canonical perspective. In: A. Dammel & O. Schallert (Eds.), *Morphological variation*. *Theoretical and empirical perspectives*. Amsterdam: Benjamins (Studies in Language Companion Series 207), 283–310.

Round, E.R. & Corbett, G. G. in press. Comparability and measurement in typological science. The bright future for linguistics. To appear in *Linguistic Typology* (2020).

Spencer, A. & Popova, G. 2015. Periphrasis and inflection. In M. Baerman (Ed.), *The Oxford Handbook of Inflection*. Oxford: Oxford University Press, 197–230.

[3] Or preterite, depending on one's stance towards the combinatorics of mood and tense forms in German; in any case it is not perfect anymore and this is what is relevant here.