## A set-based representation of person features: Consequences for AGREE\*

## Christopher Hammerly

### University of Minnesota, Twin Cities

#### 1. Introduction

Since Silverstein (1976) the relationships between person CATEGORIES have been described by appealing to a *person-animacy hierarchy* (PAH), as in (1).

(1) FIRST (1) 
$$>$$
 SECOND (2)  $>$  THIRD (3)

This (partial) hierarchy encodes a ranking such that FIRST person is ranked above SECOND person, and both PARTICIPANTS are ranked above THIRD person. PAHs can be used to describe the typology of person effects in a wide variety of domains including ergative case marking (Silverstein 1976), differential object marking (Bossong 1991), omnivorous agreement (Nevins 2011), direct-inverse agreement (Dawe-Sheppard and Hewson 1990), and the person-case constraint (PCC; Farkas and Kazazis 1980).

However, hierarchies are a *second order* representation in that they *describe* relationships between person categories, but do not provide an *explanation* for how or why these relationships arise (see, e.g., Preminger 2014). To approach an explanation, categories are decomposed into sets of person FEATURES, as in (2).

(2) a. FIRST =  $\{\pi, [Participant], [Author]\}$ b. SECOND =  $\{\pi, [Participant]\}$ c. THIRD =  $\{\pi\}$ 

The categories therefore stand in the subset/superset relationships in (3), which can derive the hierarchy from (1).

#### (3) FIRST $\supset$ SECOND $\supset$ THIRD

<sup>\*</sup>Thanks to Rajesh Bhatt for indispensable feedback over the course of writing my dissertation. Much of the analysis here is adapted from there. Thanks also to Stefan Keine, Nico Baier, and Claire Halpert for discussion, and to audiences at NYU Syntax Brown Bag and NELS 51 for comments and feedback.

These subset/superset relationships do not appear from thin air—they are the consequence of hard-coded one-way implicational relationships between features, as represented by a *feature geometry* (Harley and Ritter 2002, Béjar 2003). The feature geometry for person stipulates that the presence of a more specific feature such as [Author] entails the presence of the less specific features [Participant] and  $\pi$ . While these implicational relationships are *conceptually* grounded in actual semantic entailment (i.e. being an *author* of a linguistic act entails being a *participant* in that act), like PAHs, feature geometries as formal objects are second order: they *describe* (and enforce) entailments via implicational relationships, but do not provide an explanation of how the entailments arise.

The goal of this paper is to introduce a new level of representation, person PRIMITIVES, and show that this addition allows for a deeper explanation of person-sensitive syntactic phenomena. Just as categories are composed of sets of features, the claim is that features are composed of sets of primitives, as shown in (4). For the features at hand, these primitives include the author I, the addressee U, and a finite number of non-participants  $O, O', \ldots, O^n$ .

```
(4) a. \pi = \{I, U, O, O', \dots, O^n\}
b. [Participant] = \{I, U\}
c. [Author] = \{I\}
```

As such, the features stand in the proper subset/superset relationships shown in (5).

(5) 
$$\pi \supset [Participant] \supset [Author]$$

I advance a model of AGREE where these containment relations directly derive the entailment relationships between features and the rankings between categories without appeal to second order representations. The result: both hierarchies and geometries can be eliminated from our theories. The proposal has both conceptual and empirical advantages over previous approaches.

#### 2. The empirical domain: PAH effects in agreement

As most recently detailed in Coon and Keine (2020), Deal (2020), and Hammerly (2020), there are a wide range of possible PAH effects in agreement. Given the limited space of this paper, the goal of this section is to simply summarize these possibilities without delving much into the empirical details.

A description of the range of possible effects can be derived from the scale in (6).

$$(6) \qquad \{1 > 2 \lor 2 > 1\} > 3$$

Given that (i) rankings between adjacent categories can be eliminated (e.g. we can have 1/2 > 3, collapsing the ranking between the participants) and (ii) either one or both of the rankings of the participants can be realized, seven possible classes of PAH effects are predicted by this scale. These are summarized in the table in (7), with the corresponding name as well as a language and phenomenon where the effect is attested.

(7) Summary and example of possible/predicted prominence effects

	1 11 1	1 33
Ultra Str. (Auth.)	1 > 2 > 3	Classical Arabic PCC (Nevins 2007)
Ultra Str. (Addr.)	2 > 1 > 3	Nez Perce C-agreement (Deal 2015)
Strong	$\{1 > 2 \land 2 > 1\} > 3$	Slovenian PCC (Stegovec 2020)
Weak	1/2 > 3	Italian PCC (Bianchi 2006)
Me-First	1 > 2/3	Romanian PCC (Nevins 2007)
You-First	2 > 1/3	Cuzco Quechua SMA (Myler 2017)
Flat	1/2/3	Moro PCC (Jenks and Rose 2015)

As an illustrative example, consider PCC effects: the finding that certain IO + DO clitic combinations are illicit. For example, in Basque, a first person IO clitic can combine with a third person DO clitic (8a), but the reverse alignment is ungrammatical (8b).

### (8) The PCC in Basque (examples from Coon and Keine 2020)

a. Zu-k ni-ri liburua saldu d-i-da-zu you-ERG me-DAT book.ABS sold 3ABS-aux-1DAT-2ERG 'You have sold the book to me.'

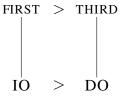
1DAT > 3ABS

b. \*Zu-k harakina-ri ni saldu n-(a)i-o-zu you-ERG butcher-DAT me.ABS sold 1ABS-aux-3DAT-2ERG Intended: 'You have sold me to the butcher.'

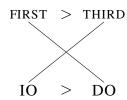
\*3DAT > 1ABS

The use of "alignment" above references the idea that having a higher ranked person category (e.g. first person) in a higher structural position (e.g. the IO position) and a lower ranked person (e.g. third person) in a lower syntactic position (e.g. the DO position) is DIRECT—that is, we have aligned high-to-high and low-to-low, as shown in (9a). In contrast, an INVERSE alignment like that in (9b) aligns high-to-low and low-to-high.





## b. INVERSE alignment



Across all of the different phenomena in the table in (7), inverse alignments are either ungrammatical (as in the PCC) or marked (as in Nez Perce C-agreement, where agreement reaches the object instead of stopping at the subject).

#### 3. Feature gluttony

Coon and Keine (2020) provide a general theory of the difference between direct and inverse alignments under an account they term *feature gluttony*. Gluttony occurs when a probe copies back features from multiple DPs. Following Deal (2015, 2020) I assume probes are defined by (i) satisfaction conditions, which tell a probe when to stop searching,

and (ii) interaction conditions, which tell a probe which features to copy back. Satisfaction conditions are the main focus of the paper.

Consider a probe with a set of satisfaction conditions [SAT:  $\{\pi, Part\}$ ]. As schematized in (10a) this probe is *fully* satisfied by interacting with either a first or second person DP, both of which are specified for  $\pi$  and [Part]. However, as shown in (10b), a probe can also be *partially* satisfied by interacting with a DP that matches a subset of its features. In the case of our [SAT:  $\{\pi, Part\}$ ] probe, a third person DP provides partial satisfaction, leaving [Part] unsatisfied. Under a cyclic model of AGREE (Béjar and Rezac 2009), this unsatisfied feature leads the probe to continue its search for a goal that can provide full satisfaction. If it finds such a goal (i.e. a first or second person), it agrees with that DP as well.

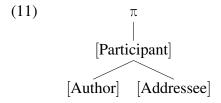
(10) a. 
$$[Probe_{[SAT: \{\pi, Part\}]}[...FIRST/SECOND[...THIRD]]]$$
 DIRECT b.  $[Probe_{[SAT: \{\pi, Part\}]}[...THIRD[...FIRST/SECOND]]]$  INVERSE

The second case gives rise to gluttony—an overload of features being copied back to the probe. As each matching DP is located, the features of the DP are copied. When agreement with the first encountered DP provides full satisfaction, a single set of features is copied. When the closer DP provides partial satisfaction, and a further DP provides additional satisfaction, two sets of features (one from each DP) are copied.

Ultimately, gluttony can lead to conflicts when it comes to generating clitics or inserting a vocabulary item. In some cases (e.g. the PCC) these conflicts result in ineffability (Coon and Keine 2020). In others (e.g. Nez Perce C-agreement) fission ameliorates the conflict (Deal 2015). In any case, we can derive the fact that inverse alignments are marked compared to direct. Gluttony occurs just in case the arguments are in an inverse alignment with respect to the probe—that is, a DP that provides only partial satisfaction intervenes between the probe and a DP that can provide full(er) satisfaction.

## 4. Probe relativization and shortcomings of the feature geometry

The characterization of a configuration as direct versus inverse is not intrinsic to the alignment between DPs and syntactic positions, but rather on how DPs are configured with respect to a probe *with a particular set of satisfaction conditions*. Probes can be *relativized* to different degrees, leading to different types of PAH effects. For example, a probe with the satisfaction condition [SAT:  $\{\pi\}$ ] would be fully satisfied by any person, and would never create a gluttonous configuration (a *Flat* PAH effect; 1/2/3). Since Béjar (2003), the limits of probe relativization have been governed by feature geometries, as in (11).



## A set-based representation of person features

This geometry enforces the following implicational relationships: [Author] and [Addressee] imply the presence of [Participant] and  $\pi$ , and [Participant] implies the presence of  $\pi$ . Assuming that a person probe cannot lack  $\pi$ , the probes in (12) are generated, with their respective PAH effects indicated.

(12) Correspondence between possible feature-geometric  $\pi$ -probes and the PAH

This leaves the two possible PAH effects in (13) unaccounted for. Both of the probes that would capture the *Me-First* and *You-First* effects include either [Author] or [Addressee], but lack the implied [Participant] feature, breaking the implicational relationships required by the feature geometry (see also footnote 22 of Coon and Keine 2020).

(13) Impossible  $\pi$ -probes under the feature geometry

# 5. A set-based representation of categories and probes

To take stock, we have two motivations for abandoning feature geometries. On the conceptual front, they stipulate rather than derive the entailment relationships between features. On the empirical front, the implicational relationships are too strong—they do not allow for the generation of probes to capture the *Me-First* and *You-First* classes of PAH effects.

A solution to both issues can be found by adding a new level of representation that I call PRIMITIVES. For the core persons, the possible set of primitives includes I, U, and a finite number of O's  $(O, O', \ldots, O^n)$ . I take these primitives to be the syntactic analogues of the semantic *person ontology* proposed by Harbour (2016). That is, I is interpreted as i (the author), U is interpreted as u (the addressee), and the O's are interpreted as o's (the others).

What is the status of these primitives? To take an analogy, just as molecules consist of atoms and atoms of particles, categories consist of (sets of) features and features of (sets of) primitives. These correspondences are given in (14).

```
(14) a. [Author] = \{I\}
b. [Addressee] = \{U\}
c. [Participant] = \{I, U\}
d. \pi = \{I, U, O, O', \dots, O^n\}
```

Features can appear in one of three possible states: positively valued, negatively valued, or unvalued, as shown in (15).

(15) a. *Positive*: [+F] b. *Negative*: [-F] c. *Unvalued*: [F]

The positive/negative variants define *goals* by restricting the particular set of primitives that define a DP. The head  $\pi$  defines the maximal set of possible primitives, and the features [ $\pm$ Author] and [ $\pm$ Participant] restrict this set according to their values.<sup>1</sup> This set is formed at the DP level, making it visible to probes along the verbal spine (for details, see Hammerly 2020). Formal definitions of these values are given in (16). Positive values find the set intersection—all elements shared between F and  $\pi$ . Negative values find the set difference—those elements in  $\pi$  that are not in F.

(16) Syntactic definition of feature values (Hammerly 2020)

a. 
$$+F(\pi) = \pi \cap F = \{x : x \in \pi \land x \in F\}$$
  
b.  $-F(\pi) = \pi \backslash F = \{x : x \in \pi \land x \notin F\}$ 

As a result, the correspondences in (17) hold between categories, sets, and features.

(17) Representation of person categories under the set-based theory

Category	Features	Set
FIRST	$\{+Auth, +Part\}(\pi)$	$\{I\}$
SECOND	$\{-Auth, +Part\}(\pi)$	$\{U\}$
THIRD	$\{\pm Auth, -Part\}(\pi)$	$\{O,O',\ldots,O^n\}$

Unvalued features are used to define the interaction/satisfaction conditions of a probe. Given that there are no restrictions on possible feature combinations (modulo the self-evident restriction that a person probe include  $\pi$ ), eight possible probes are generated, capturing all seven possible PAH effects (there are two probes that generate the *Strong* effects—a redundancy, but not a problem). Critically, freedom from the implicational relationships of the feature geometry allows probes that are specified for [Author] and/or [Addressee] but lack [Participant] to be generated, as shown in (18c,d).

```
(18)
         a.
               [SAT: \{\pi\}]
                                                                                           Flat: 1/2/3
               [SAT: \{\pi, Part\}]
                                                                                       Weak: 1/2 > 3
         b.
               [SAT: \{\pi, Auth\}]
                                                                                  Me-First: 1 > 2/3
         c.
              [SAT: \{\pi, Addr\}]
                                                                                  You-First: 2 > 1/3
         d.
                                                                   Ultra Strong (Auth.): 1 > 2 > 3
               [SAT: \{\pi, \text{Part}, \text{Auth}\}]
         e.
         f.
               [SAT: \{\pi, Part, Addr\}]
                                                                   Ultra Strong (Addr.): 2 > 1 > 3
               [SAT: \{\pi, Addr, Auth\}]
                                                                     Strong: \{1 > 2 \land 2 > 1\} > 3
         g.
               [SAT: \{\pi, Part, Addr, Auth\}]
                                                                     Strong: \{1 > 2 \land 2 > 1\} > 3
         h.
```

<sup>&</sup>lt;sup>1</sup>The role of [Addressee], which allows for an inclusive/exclusive distinction to be made (cf. Harley and Ritter 2002, McGinnis 2005) gets rather thorny, so will not be covered here. For details, see the discussion of the *Addressee Asymmetry* in Hammerly (2020).

In the next section, I advance a model of AGREE that operates with respect to primitives to form the link between the probes and PAH effects stated in (18).

## 6. Defining AGREE

The operation AGREE can be broken down into four components, *Search*, *Match*, *Copy*, and *Satisfy*, formalized in (19).

- (19) Sub-components of AGREE (cf. Deal 2015, Hammerly 2021b)
  - a. Search: A probe with a set S of satisfaction conditions and a set I of interaction conditions Searches its locality-restricted c-command domain for the (next) closest goal with a set of primitives G
  - b. *Match*: Match is evaluated via set intersection between two sets of primitives F (from the interaction/satisfaction conditions of the probe) and G (from the goal). Match holds iff  $F \cap G \neq \emptyset$
  - c. Copy: Evaluate Match between each  $\iota \in I$  and G. Match holds iff  $\iota \cap G \neq \emptyset$ . If Match holds, then Copy G to the probe
  - d. Satisfy: Evaluate Match between each  $\sigma \in S$  and G. Match holds iff  $\sigma \cap G$   $\neq \emptyset$ . If Match holds, then  $\sigma$  is Satisfied. Search is halted if every  $\sigma \in S$  is Satisfied (or all goals have been Searched)

Relevant here is the definition of Satisfy. On the definition in (19d), Satisfaction is evaluated by comparing (via Match) the sets of primitives that define each feature of the probe with the set of primitives that define the goal. If there is overlap between the two sets of primitives (i.e. the intersection of the two sets is non-null, so Match holds), then the feature is Satisfied. From this, we achieve the relations in (20).

(20) *Match/Satisfaction relations between person categories and features* 

```
a. \checkmark \pi \cap \text{THIRD} = \{I, U, O, O', \dots, O^n\} \cap \{O, \dots\} = \{O, \dots\}
\checkmark \pi \cap \text{SECOND} = \{I, U, O, O', \dots, O^n\} \cap \{U\} = \{U\}
\checkmark \pi \cap \text{FIRST} = \{I, U, O, O', \dots, O^n\} \cap \{I\} = \{I\}
b. \checkmark [\text{Part}] \cap \text{THIRD} = \{I, U\} \cap \{O, \dots\} = \emptyset
\checkmark [\text{Part}] \cap \text{SECOND} = \{I, U\} \cap \{U\} = \{U\}
\checkmark [\text{Part}] \cap \text{FIRST} = \{I, U\} \cap \{I\} = \{I\}
c. \checkmark [\text{Auth}] \cap \text{THIRD} = \{I\} \cap \{O, \dots\} = \emptyset
\checkmark [\text{Auth}] \cap \text{SECOND} = \{I\} \cap \{U\} = \emptyset
\checkmark [\text{Auth}] \cap \text{FIRST} = \{I\} \cap \{I\} = \{I\}
d. \checkmark [\text{Addr}] \cap \text{THIRD} = \{U\} \cap \{O, \dots\} = \emptyset
\checkmark [\text{Addr}] \cap \text{SECOND} = \{U\} \cap \{U\} = \{U\}
\checkmark [\text{Addr}] \cap \text{FIRST} = \{U\} \cap \{I\} = \emptyset
```

Let us see how this plays out with a *Weak* probe to derive a 1/2 > 3 PAH effect. When the structurally closer DP is a first person, as in (21a), both features within the satisfaction

conditions of the probe are Matched with the goal and Satisfied (a second person would give way to the same result). Since all features are Satisfied, the probe is Satisfied, so the structurally further third person DP does not enter into an agreement relation with the probe. This contrasts with the configuration in (21b), where the initial agreement relation with the closer third person (21b-i) leads to partial Satisfaction, with [Participant] remaining Unmatched and Unsatisfied. The probe therefore enters into an agreement relation with the structurally further first person DP (21b-ii), providing full Satisfaction.

(21) *Match with a Weak probe:* 1/2 > 3

From here, the same issues that arise under the original feature gluttony account of Coon and Keine (2020) outlined in §3 can be maintained. In (21a) a single set of features is copied back the probe, while in (21b) a gluttonous set arises.<sup>2</sup>

Returning to the general properties of the account, as noted in the introduction, the sets of primitives that define the features stand in proper containment relationships, now repeated in full in (22).

(22) 
$$\pi \supset [Participant] \supset [Author], [Addressee]$$

A look back at the correspondences in (20) reveals the critical facet of the proposal: The containment relationships in (22) can also describe the possible Matching relations for each feature. That is,  $\pi$  Matches with all three persons, [Participant] Matches with a proper subset of these (the first and second persons), and [Author] and [Addressee] a subset of these (but these two features do not stand in containment relations to each other).

This captures the core benefit of the feature geometry in theories of AGREE that motivated its widespread adoption : If a goal "checks" a more specific feature such as [Author], it will "check" [Participant] and  $\pi$  too. Under the feature geometry, this was due to the implicational relations in the specification of features—if a goal was specified for [Author], then it was necessarily specified for [Participant] and  $\pi$  as well. Under the current theory, these entailments flow from the fact that [Participant] and  $\pi$  include I, the member of the set that defines [Author]. As a result, if a goal Matches and Satisfies [Author] it will also Match

<sup>&</sup>lt;sup>2</sup>Given (19c), it is technically G, the set of *primitives* of the goal, that is Copied back to the probe when Match holds rather than a set of features. However, there is a transparent mapping between sets of primitives and features and their values as outlined in (17), so this difference is immaterial.

#### A set-based representation of person features

and Satisfy [Participant] and  $\pi$ . Furthermore, upon interpretation and without further ado, the containment relations between the sets give rise to the appropriate semantic entailments that motivated the feature geometry in the first place—for example, [Participant] denotes the set  $\{i,u\}$ , which contains the set  $\{i\}$  that [Author] denotes (see also Harbour 2016:74). We have thus obviated the need for an appeal to geometries and have derived both syntactic and semantic entailments between person features from first principles of set theory.

#### 7. Outlook and extensions

The current paper showed that the dominant view of probe relativization—that possible and impossible probes are defined by the feature geometry—has both conceptual and empirical shortcomings, motivating the adopted account. However, in Harley and Ritter (2002), the paper that introduced and motivated the original syntactic feature geometry, the empirical focus was on distinguishing patterns of possible person and number categories. For example, the fact that some but not all languages split the first person category into exclusive and inclusive. What of the application of the geometry to capture possible person categories? This too has been argued to be empirically and conceptually suspect, perhaps most forcefully by Harbour (2016), but also in my own work that extends and adapts Harbour's framework to capture animacy and obviation in Ojibwe (Hammerly 2020, 2021a). The conclusion: feature geometries do not pass muster in either the domain of probe relativization or possible person categories, and should be eliminated from our theoretical toolbox.

This paper also had the somewhat narrow focus of considering person, but not the other  $\phi$ -features (number, noun classification, and obviation). Extending the theory to these domains is of critical importance. Foley and Toosarvandani (2021) show that noun classification systems grounded in animacy, humanness, and social status exhibit PAH effects akin to the PCC, couching their analysis in feature geometric terms. Similarly, the feature geometry has been used in Algonquian languages to place obviation within a PAH, accounting for patterns of direct-inverse agreement (Bliss and Jesney 2005, Oxford 2019). In Hammerly (2020, 2021a) I show how the current system captures PAH effects with noun classification and obviation with the addition of new primitives and features that sit in proper containment relations. These features also have the advantage of correctly predicting the typology of possible categories within these domains.

#### References

Béjar, Susana. 2003. Phi-syntax: A theory of agreement. Doctoral dissertation, University of Toronto.

Béjar, Susana, and Milan Rezac. 2009. Cyclic agree. Linguistic Inquiry 40:35-73.

Bianchi, Valentina. 2006. On the syntax of personal arguments. Lingua 116:2023–2067.

Bliss, Heather, and Karen Jesney. 2005. Resolving hierarchy conflict: Local obviation in Blackfoot. *Calgary Papers in Linguistics* 26:92–116.

Bossong, Georg. 1991. Differential object marking in Romance and beyond. *New analyses in Romance linguistics* 143–170.

- Coon, Jessica, and Stefan Keine. 2020. Feature gluttony. *Linguistic Inquiry* Early Access. Dawe-Sheppard, Audrey, and John Hewson. 1990. Person and gender hierarchies in Micmac. *Journal of the Atlantic Provinces Linguistic Association* 1–12.
- Deal, Amy Rose. 2015. Interaction and satisfaction in  $\varphi$ -agreement. In *Proceedings of NELS*, volume 45.
- Deal, Amy Rose. 2020. Interaction, satisfaction, and the PCC. Manuscript, UC Berkeley.
- Farkas, Donka, and Kostas Kazazis. 1980. Clitic pronouns and topicality in Rumanian. In *Papers from the 16th Regional Meeting, Chicago Linguistics Society*, 16, 75–82.
- Foley, Steven, and Maziar Toosarvandani. 2021. Extending the person-case constraint to gender: Agreement, locality, and the syntax of pronouns. *Linguistic Inquiry* Early Access.
- Hammerly, Christopher. 2020. Person-based prominence in Ojibwe. Doctoral dissertation, University of Massachusetts Amherst.
- Hammerly, Christopher. 2021a. A set-based semantics for obviation and animacy. Manuscript, University of Minnesota.
- Hammerly, Christopher. 2021b. A verb-raising analysis of the Ojibwe VOS/VSO alternation: Lessons for feature copying and movement. Manuscript, University of Minnesota.
- Harbour, Daniel. 2016. Impossible persons. MIT Press.
- Harley, Heidi, and Elizabeth Ritter. 2002. Person and number in pronouns: A feature-geometric analysis. *Language* 78:482–526.
- Jenks, Peter, and Sharon Rose. 2015. Mobile object markers in Moro: The role of tone. *Language* 269–307.
- McGinnis, Martha. 2005. On markedness asymmetries in person and number. *Language* 699–718.
- Myler, Neil. 2017. Cliticization feeds agreement: a view from Quechua. *Natural Language & Linguistic Theory* 35:751–800.
- Nevins, Andrew. 2007. The representation of third person and its consequences for person-case effects. *Natural Language & Linguistic Theory* 25:273–313.
- Nevins, Andrew. 2011. Multiple agree with clitics: Person complementarity vs. omnivorous number. *Natural Language & Linguistic Theory* 29:939–971.
- Oxford, Will. 2019. Inverse marking and multiple agree in Algonquin. *Natural Language & Linguistic Theory* 37:955–996.
- Preminger, Omer. 2014. Agreement and its failures. MIT Press.
- Silverstein, Michael. 1976. Hierarchy of features and ergativity. In *Grammatical cate-gories in Australian languages*, ed. by Robert M.W. Dixon. Atlantic Highlands, NJ: Humanities Press.
- Stegovec, Adrian. 2020. Taking case out of the person-case constraint. *Natural Language & Linguistic Theory* 38:261–311.

Christopher Hammerly hamme381@umn.edu