

UC Berkeley

UC Berkeley Previously Published Works

Title

Person as an inflectional category

Permalink

<https://escholarship.org/uc/item/3w33c23j>

Journal

Linguistic Typology, 21(3)

ISSN

1430-0532

Author

Nichols, J

Publication Date

2017-12-20

DOI

10.1515/lingty-2017-0010

License

<https://creativecommons.org/licenses/by-nc-nd/4.0/> 4.0

Peer reviewed

Author's final version. A number of errors and omissions were corrected by the editors and in the proofs, so **do not quote this version**; refer to the publication and quote from there:

Nichols, Johanna. 2017. Person as an inflectional category. *Linguistic Typology* 21:3.387-456.

Person as an inflectional category

Abstract

The category of person has both inflectional and lexical aspects, and the distinction provides a finely graduated grammatical trait, relatively stable in both families and areas, and revealing for both typology and linguistic geography. Inflectional behavior includes reference to speech-act roles, indexation of arguments, discreteness from other categories such as number or gender, assignment and/or placement in syntax, arrangement in paradigms, and general resemblance to closed-class items. Lexical behavior includes sharing categories and/or forms and/or syntactic behavior with major lexical classes (usually nouns) and generally resembling open-class items. Criteria are given here for typologizing person as more vs. less inflectional, some basic typological correlations are tested, and the worldwide linguistic-geographical distribution is mapped.

Keywords: person, inflection, morphology, linguistic geography, typology

Person as an inflectional category

1. Introduction

Person, probably more than any other inflectional category, originates in arguments but is strongly prone to end up as indexation on predicates and other phrase heads. As a result, person can be either lexical, figuring in independent arguments, or inflectional, figuring in indexation paradigms. This distinction, surveyed over a number of grammatical contexts, can be used to form a fine-grained composite notion which answers the current need for more such typological variables for use in tracking incremental diachronic developments and gradual spatial distributions, and it can integrate local relationships with long-distance and long-term ones on a single scale. At the high inflectional end of the scale, person marking is similar to head marking, a coarse-grained variable with few settings and one which can represent transitions only as large leaps and can capture only relatively gross differences.

The first order of business in a discussion of lexical vs. inflectional person is to define *person*, *inflection(al)*, and *lexical*, which is done in the next two subsections.

1.1. Person and reference

The focus of this paper is not the composition or semantics of person categories but their morphosyntactic behavior and patterning. Some of the properties that are important here have had little cross-linguistic or theoretical discussion, so some typological and methodological preliminaries need to be laid out briefly. (Technical finer points are in the footnotes, which readers are free to skip.)

Person is taken here to be the grammatical category that encodes the speech act roles (SAR's) of speaker, addressee or hearer, and non-locutor, and I take the grammatical category values of inclusive, first person, second person, and third person to be derived from SAR's along the lines proposed in Plank 1985 and Daniel 2005.¹ In this approach inclusive is an independent and separate person value, distinct from either first or second.²

¹ The grammaticalization follows this hierarchy, based on what SAR is essential to the meaning: speaker + hearer = inclusive, speaker = first person, hearer = second, non-locutor = third. The first stage (inclusive) is optional: only a minority of languages distinguish an inclusive. A small minority of languages, chiefly in the Americas and Australia-New Guinea, reorder the middle two steps, assigning second person before first; see Daniel 2005 for several examples. There are also languages that I would describe as having no grammatical value of third person at all, i.e. as omitting the fourth level of the hierarchy. These are languages that do not have dedicated third person pronouns in their independent pronoun series but recruit demonstratives instead, and which have zero marking in the third person singular and/or plural of verbs (if they have person indexation on verbs at all, as the majority of languages do). If such a language has an overt third person plural verbal index I would regard that as a marker of plural, not third person, and posit a person-based hierarchy for access to the person-number indexation slot: person markers have first access (coexponential person-number markers such as second person plural carry their number along with them), and plural marking has access only if there is no person value competing for access. That is, person > number.

² Theoretical work that I have seen is not explicit on this point, but I would define inclusive not as a conjunction of discrete speaker plus discrete hearer but as an undifferentiated monadic

Inclusive, first, and second persons can be collectively termed *speech act participants* (SAP's); speaker and hearer SAR's are *locutors*; and these contrast with respectively non-SAP's (third person) and non-locutor.³ This paper focuses primarily on the properties of person as a whole category and not on its different values (first person, second person, inclusive) or how they are related to SAP roles.

SAR's can be thought of as referential indexes, specifically shifter indexes where the reference is not to a fixed individual but to one defined by a role in the speech act. Person markers of all kinds (independent pronouns, verb indexation morphemes, nominal possessive morphemes, etc.) are morphemes that carry these referential indexes, and sometimes also other categories such as number or gender.⁴ Unalloyed person markers can only index and not register, but they can be coexponential with other categories such as number which in other contexts can register.⁵ An example of registration is the objective conjugation of Hungarian. The Hungarian subjective conjugation indexes the S/A by copying its person and number, while the objective conjugation indexes the person/number of the A and registers an O, indicating its presence but not its person or number:⁶

(1) Hungarian (Comrie 1988:466; SBJ = subjective conjugation, OBJ = objective)⁷

a	lát-ok	egy	kutyá-t
	see-1sgSBJ	one/a	dog-ACC
	'I see a dog'		
b	lát-om	a	kutyá-t
	see-1sgOBJ	the	dog-ACC

communicative space, with the individual identities of the two locutors entirely predictable to both of them. The same also applies to egophoric (a.k.a. conjunct-disjunct) reference. This definition informs the decisions about which grammatical behaviors count as inflectional (section 2).

³ Other important sources in the vast literature on person include Siewierska 2011, 2004, Filimonova ed. 2005, Cysouw 2003, Helmbrecht 2003, Plank 2003, Heath 1998, 1991, Zwicky 1977, Forchheimer 1953.

⁴ I follow Siewierska 2004, 2011 in using the term *person marker* for any morpheme signaling person: independent pronouns, weak pronouns, clitic pronominals, person affixes (and also reduplication and internal alternations such as ablaut if these mark person; no examples have come up in this survey).

⁵ Here and below I use *index(ation)* and *register (registration)* as defined in Nichols 1992:48-49: indexation refers to an argument and copies its features (e.g. person, number); registration does not copy features and does not refer but simply indicates that an argument is present. Indexation is variously referred to as agreement and pronominal argument(s); I lump these two together (the difference is never at issue here) and call them both *indexation* (or *agreement* where the copying of features is at issue).

⁶ If that O is definite or specific. While the definite conjugation is sometimes described as definiteness agreement (e.g. Coppock & Wechsler 2012), I regard definiteness as a condition on registration and not as a grammatical category of NP's that the verb agrees with.

⁷ Other abbreviations used here: ARTicle, ACCusative, DATive, GENitive case, INCLusive, OBLique, PF=perfective, POSSessive. Most interlinear terms are not abbreviated since examples come from different sources with different conventions.

'I see the dog'

- c lát-om
'I see (him/her/it)'

On this view, the referentiality of 'dog' or 'him/her/it' is contributed not by the verbal inflection but by the O itself: in (1b) 'dog' with its definite article is referential, while in (1c) there is arguably no O reference other than what is implicit in the larger context. The verb inflection signals only that there does exist an object but is no different in referentiality from the verb of (1a).

Person can index arguments on verbs (e.g. one argument in most Indo-European languages, two in most Quechuan varieties, three in Basque and West Caucasian languages); but some languages have no person indexation at all (e.g. Japanese or Dyirbal, with no argument indexation; Avar or Archi [Nakh-Daghestanian], with indexation of gender but not person). It can index adnominal dependents on head nouns (as in many Uralic and all Athabaskan languages), and in a few languages it can index head nouns on their modifiers (examples are below). It can be cumulative (fused, coexponential, portmanteau) with number (English and most Indo-European languages), or number can be separately marked as in Quechua (2), where plural pronouns have the same plural ending *-kuna* as plural nouns (cf. noun *wasi* 'house', pl. *wasi-kuna*):

- (2) Huallaga Quechua (Weber 1989:37)

<i>1sg</i>	<i>2sg</i>	<i>Incl</i>	<i>1pl</i>	<i>2pl</i>
noqa	qam	noqanchi(:)	noqa-kuna	qam-kuna

In some languages independent pronouns are much like nouns, with separate lexical roots for each person or person-number category and bearing the same categories and even forms of case and number marking as on nouns. The Quechua forms in (2) above are of this type. Indo-European pronouns are partly so, with largely separate roots and the same case and number categories as nouns have, though partly different case endings, e.g. Latin:

- (3) Latin
- | | <i>1sg</i> | <i>2sg</i> | <i>1pl</i> | <i>2pl</i> |
|------|------------|------------|------------|------------|
| Nom. | ego | tū | nōs | vōs |
| Acc. | mē | tē | " | " |
| Dat. | mihi | tibi | nōbis | vōbis |

At the opposite pole, pronouns can take the form of a generic pronominal base which takes inflectional person markers (such as subject or possessive affixes) so that person has no lexical representation at all, as in many North American languages:

- (4) Lakhota (Siouan; Rood & Taylor 1996:454, 458; stress omitted)

	<i>Emphatic</i>	<i>Contrastive</i>
1sg	mi-ye	mi-š

1pl	ʉki-ye	ʉki-š
2	ni-ye	ni-š

(5) Cree (Algonquian; Wolfart 1996:424, with colon for his raised dot):

<i>1sg</i>	<i>2sg</i>	<i>Incl</i>	<i>1pl</i>	<i>2pl</i>
ni:-ya	ki:-ya	ki:-ya:-naw	ni:-yanan	ki:-yawa:w

The Lakshota prefixes *mi-*, *ʉki-*, and *ni-* are possessive prefixes, and the independent pronouns consist of these prefixed to one of two generic bases *-ye* and *-š* which distinguish emphatic from contrastive semantics but have no person meaning at all. Similarly, Cree *ni-* and *ki-* are possessive prefixes and add the category of person to the generic base *-ya*. Generic pronominal bases are described in Nichols 2013.

Languages can carefully distinguish, with inclusive/exclusive forms, nonsingular first person forms involving speaker plus hearer, speaker plus non-locutor(s), or speaker plus hearer plus non-locutor(s) (Somali, Evenki, Bardi, Yurok, and many others with inclusive/exclusive distinctions). This amounts to keeping person values discrete from each other and from number (Cysouw 2003:69ff.), and such feature values are close to congruent with SAR's and keep the different SAR's discrete. In contrast, in a language like English with no inclusive/exclusive opposition the reference of a first person plural pronoun can include speaker and non-locutor (when it has exclusive meaning, i.e. 'me and others'), speaker and hearer (when it has inclusive meaning), or all three (inclusive plural meaning: 'you, me, and others').

Role indexation on verbs, in that majority of languages that have it, almost always involves person. Person may be fused with number in agreement (as in most Indo-European languages), or number can be separatively marked, often with a promiscuous plural morpheme (Leer 1991) that lets arguments compete for access to the plural category so that forms are either ambiguous or resolved by a referential or pragmatic hierarchy. An example of ambiguity is Pazar Laz, where the suffix *-t* indicates plurality of S, A, and/or O:

(6) Pazar Laz (Kartvelian; Turkey; Öztürk & Pöchtrager 2011:50)

ce-m-č-i-t
 PV-1.O-beat-PAST.2-PL
 'You (pl.) beat me', 'you (sg.) beat us', 'you (pl.) beat us'

Similarly in Crow (Siouan; Graczyk 2007:122-3):

(7) dii-waa-lit-úu
 2 1 hit PL
 'I hit you.PL', 'we hit you.SG', 'we hit you.PL'

Sometimes argument indexation is for number alone. For instance, in many Uralic languages verbs agree in person-number with the subject but only in number with the object, e.g. Tundra Nenets in (8):

(8) Partial paradigm of Tundra Nenets verb person suffixes (Salminen 2012:16) and three examples (Salminen 1997:96).⁸ Person-number combinations refer to the subject. ° = zero grade of schwa (in vowel-zero alternation). The examples show some morphophonemic alternations of the abstract forms in the paradigm.

	Subjective	Objective:	
		Singular	Dual-Plural
1sg	-t-əm	-m-ə	-n-ə
2sg	-n ~ -tə	-r-ə	-t-
1pl	-m-aq	-m-aq	-n-aq
2pl	-t-aq	-r-aq	-t-aq

Selected forms of *tonta-* 'cover':

tontaəd°m	'I covered (something)'
tontaəw°	'I covered it'
tonteyən°	'I covered them'

Only rarely is there indexation for gender alone with no person; this is found in western Nakh-Daghestanian languages where verbs agree in gender with the S/O but not in person, e.g. Hinuq (Forker 2013:466-467):

(9) Gender agreement in Hinuq with 'fall down' (a light verb construction with conjugated light verb *-iɬ'iyɔ* and lexical element *q'idir* 'fall').

de	q'idir	Ø-iɬ'iyɔ	I (masc.) fall down
de	q'idir	y-iɬ'iyɔ	I (fem.) fall down
me	q'idir	Ø-iɬ'iyɔ	You (masc.) fall down
me	q'idir	y-iɬ'iyɔ	You (fem.) fall down
uži	q'idir	Ø-iɬ'iyɔ	The boy falls down
ked	q'idir	y-iɬ'iyɔ	The girl falls down
k'et'u	q'idir	b-iɬ'iyɔ	The cat falls down
t'ek	q'idir	y-iɬ'iyɔ	The book falls down
t'oq	q'idir	r-iɬ'iyɔ	The knife falls down

Multiple exponence within phrases and words is relatively common for gender but infrequent for person; and if verb-argument agreement almost always involves person, agreement within NP's only rarely does. Here are some examples of the infrequent types. Lander 2017 regards (10) and (11) as involving a type of marking in which the possessor person is assigned to the phrase and shows up on every word of the phrase, while Nikolaeva 2005 regards (13) as having modifier-head agreement with 'white' agreeing in first person singular with 'reindeer'.

⁸ The objective conjugation is used when a definite, specific, or topical object is present, the subjective conjugation when there is no object or an indefinite/non-specific/nontopical object (Nikolaeva 2014:202-210).

- (10) beñɔ-m mɛ-m
 hand-2sg left-2sg
 'your left hand' (Iaai; Tryon 1968)
- (11) (møny) serako(-myi) te-myi
 1sg white-1sg reindeer-1sg
 'my white reindeer' (Tundra Nenets; Nikolaeva 2005)

The same construction is found in Tungusic languages (Evenki: Konstantinova 1968:75, Even: Benzing 1955:82, 115) and a similar one in Chukchi (Dunn 1999:150), so it appears to be a local north Asian trait. If it is viewed as agreement, it is agreement of a modifier with the head's syntactically assigned inflectional categories.

Less common is agreement of a modifier with the inherent person of the head noun, as in Bilua:

- (12) Bilua (isolate, Solomon Islands; Obata 2003: 87, 75)

kuleto=a=la	kobaka
first-LIGATURE-3sg.M	snail(Masc.)
'the first snail'	

Similarly in Cherokee (Iroquoian; Montgomery-Anderson 2015:166ff.). In both of these languages the modifier agrees in gender-number-person with the head noun, which is necessarily third person, so the gender contributes more information than the person. The gender-number-person marker is a portmanteau, so it is possible to regard this as primarily gender-number agreement, with person an automatic incidental consequence of marking gender.

This paper makes a first stab at placing these and other properties of person marking into a simple typology. It provides definitions and a questionnaire for the typology, surveys a worldwide genealogically and geographically diversified sample of 270 languages, and lays out some of the basic grammatical and geographical correlations. Grammatically, the status of person appears to be largely independent of other typological variables, thus a useful addition to the inventory of variables. Geographically and diachronically, it correlates interestingly with geography and contact in ways that call for explanation and have implications for language history and language evolution.

1.2. Inflection

The typology proposed here is based on determining for each language to what extent the category of person patterns like an inflectional category rather than a lexical one. Theoretical work defining inflection mostly focuses on distinguishing inflectional from derivational morphology (e.g. Anderson 1985, Bickel & Nichols 2007, Haspelmath & Sims 2010, Bybee 1985), while the focus here is on distinguishing inflectional properties from lexical ones. Primarily this means determining whether person shares more properties with closed-class grammatical morphemes or with lexemes, i.e. words of major lexical classes (primarily nouns). For instance, gender is an inherent property of nouns, and if pronouns

distinguish the same gender categories as nouns that is a lexical property. What follows is a plan for defining what is essential to inflection, not in contrast to derivation but in contrast to the structural, behavioral, and semantic properties of major-class lexemes.⁹

So defined, there are three properties I take to be specific or essential to inflectional categories (distilled from Corbett 2013, Corbett & Baerman 2006, Bickel & Nichols 2007, Anderson 1985, Zwicky 1985). This discussion mainly concerns person, occasionally mentioning other categories.

- **Inherent referentiality.** In the case of person markers, the referentiality in question concerns locutors. While an identifying property of nouns is their ability to bear a referential index, a person morpheme might be said not to bear but to *be* a referential index. More precisely, a locutor is pure referentiality, as is probably also true for non-locutor person markers (once semantic categories such as distal vs. proximal, natural gender, honorificity, etc. are abstracted away). Pure referentiality is presumably never a property of lexemes (other than names, which are not treated here). This property is specific to inflectional morphemes but not essential: tense, for instance, presumably has inherent referentiality but case does not.

- **Syntactic positioning.** Person markers are assigned and positioned in the syntax, chiefly by rules of agreement or indexation, which assign them to the head of a phrase or to a position relative to a head or to clause or phrase boundaries (clause-second, phrase-final, etc.). (Lander 2017 argues that all syntactically assigned marking is head marking, in some languages ending up overt only on the head and in some copied to other phrase or clause members, or positioned relative to phrase or clause boundaries.) Assignment in the syntax distinguishes inflectional from derivational morphemes or categories (Anderson 1985, 1988), but I believe it also distinguishes inflectional from lexical categories. The choice of lexemes is a matter of semantics and lexicon, not syntax. At first glance it may appear that the positioning of words by word-order rules is a matter of syntax applying to lexemes, but in fact what is positioned by word-order rules is not lexemes but argument types or terms of syntactic relations. Furthermore, word-order rules position them not relative to heads but to each other. Thus inflectional morphemes, but not derivational morphemes or lexical ones (whether roots, stems, or words), are assigned and positioned in the syntax. Note also that inflectional person categories are properties of arguments but end up marked (e.g. indexed) on other words, usually heads, and there is nothing analogous to this in word-order rules.

- **Paradigms.** Person markers have a strong cross-linguistic tendency to occur in paradigms, and consequently they are prone to display properties of small closed sets such as a shared phonotactic canon, other phonological properties such as assonance (discussed below), portmanteau coding, deponence, and syncretism.

Lexical person morphemes, in contrast, share essential properties with lexemes. The clearest examples of lexical person morphemes are independent pronouns, and these often share inflectional categories and/or inflectional classes with nouns. Their meanings often include not just locutor reference and not just person but formality, social status, generation level, sex, and other categories that are more lexicosemantic than grammatical. Though in

⁹ Case and valence can be governed or assigned by verbs, or otherwise provided for in the lexicon, and in this sense are often spoken of as lexical, but neither the fact of lexical determination nor the particular categories governed puts the categories themselves (or their markers) among the inherent lexical properties of the verbs.

some languages independent pronouns exhibit formal properties prototypical of small closed classes, cross-linguistically they do so much less frequently than inflectional person morphemes do.

Section 2 identifies a number of respects in which person markers do or do not pattern as inflection, and section 3 describes ways in which person does or does not correlate with various typological features.

2. Method

2.1. Criteria and definitions.

The notion of inflectional person used here rests on 42 different grammatical behaviors, each designed to distinguish one or another aspect of inflectional vs. lexical person as described just above. Entries for each datapoint are Yes, No, n/a (not applicable), and n.d. (no data, i.e. unknown). For most entries a Yes answer is an inflectional property; for some it is lexical. (They could be reworded so that for all of them a Yes answer means inflectional, but I have worded the questions so as to make them simpler and focus them on identifiable positive properties.) The type of a language can be expressed as the proportion or percent of Yes entries that are on the lexical side or on the inflectional side. The number of possible proportions for the 42 datapoints is very large, making the typology a finely divided continuum and enabling it to capture incremental differences and thus long-range gradients and long-term evolutionary trends. Alternatively, the total number of inflectional points can be used as the basis for the typology, without considering the lexical points; I do not use this definition here since the typology is focused on the strength of inflectional person in the overall context of person marking. In addition, the 42 properties themselves can be used as comparanda, defining a multidimensional space within which languages can be positioned.

This section describes each of the 42 survey points and justifies their categorization as inflectional or lexical. The full questionnaire, minus the discussion and examples here, is in Appendix 1. Items apply to only first and second persons (unless otherwise stated) and only singular and plural (if there are number distinctions); a dual, if present, is disregarded (though see item [16] below). Items apply to only main clauses (if dependent clauses are different). In order to keep illustrative example numbers straight from questionnaire item numbers, the latter are in square brackets.

Items [1]-[8] all have to do with indexation of person, in the form of person markers (most often affixes or clitics), on various words and phrases of the clause (or, for [7], a higher clause). The more places with person indexation, the more inflectional the category of person in a language.

Locus of person marking

- [1] Person of A indexed on verb.
- [2] Person of O indexed on verb.

[1] and [2] refer to the default treatment in the default valence pattern for monotransitives. (Intransitives, and the treatment of S, are not counted in this survey.)

[3] Person of possessor indexed on possessum (head of possessive NP). This is possessive marking as in Hungarian:

(13) Hungarian

a	ház-am	a	ház-ad
ART	house-1sg	ART	house-2sg
	'my house'		'your house'

[4] Auto-person: The person of an independent pronoun is indexed on the pronoun itself, as a separate morpheme from the root. Examples are (14)-(15), repeated from (4-5) [above](#), where independent pronouns consist of a possessive affix on a generic base that lacks any lexical specification of person.

(14) Lakhota (Siouan; Rood & Taylor 1996:454, 458; stress omitted)

	<i>Emphatic</i>	<i>Contrastive</i>
1sg	mi-ye	mi-š
1pl	ʉki-ye	ʉki-š
2	ni-ye	ni-š

(15) Cree (Algonquian; Wolfart 1996:424, with colon for his raised dot):

<i>1sg</i>	<i>2sg</i>	<i>Incl</i>	<i>1pl</i>	<i>2pl</i>
ni:-ya	ki:-ya	ki:-ya:-naw	ni:-yanan	ki:-yawa:w

Occasionally person is indexed on non-generic, person-bearing pronoun stems, as in some oblique cases of some Uralic languages:

(16) Mansi (Uralic; Keresztes 1998:410, 413; Riese 2001:30-31)

	1sg	2sg
Nominative	am	nang
Accusative	aan-əm	nang-ən
	1sg-1sg	2sg-2sg
Dative	aan-əm-n	nang-ən-n
	1sg-1sg-DAT	2sg-2sg-DAT

(17) Komi (Uralic; Lytkin 1966:287, Hausenberg 1998:312-313)

Nominative	me	te
Genitive	men-am	ten-ad
	1sg-1sg	2sg-2sg

Ablative	men-śy-m	ten-śy-d
	1sg-ABL-1sg	2sg-ABL-2sg

Auto-person is frequent in the Americas, especially North America; see the map in Daniel 2013 (category: person-number affixes).

[5] Person of the possessor of an argument can be indexed on the verb. Siuslaw (Siuslawan, Oregon; Frachtenberg 1922b:189-190) has a suffix set that indexes the person of the possessor of an object (my interlinear and part retranscription):

(18) sînⁱxy-**ūl**-n hītsī'i
 like-**3obj:POSS**-1sg house
 'I like his house' (489)

hīn - ay - ūlts - anx	tl'xmīī
take.along - TAM? - 1-2obj:POSS -1sg>2sg	bow
'I take along your bow' (490)	

[6] Person of external possessor indexed on verb. In (19a) the verb indexes the possessor; in (19b) it does not.

(19) a. Creek (Muskogean; Martin 1999:230)

ifā-t	am-î:l-is
dog-NOM	1sg.DAT-die:sg:FGR-IND
My dog died (lit. 'the dog died for/to me')	

b. **am-ífa-t** î:l-is
 1sg.DAT-dog-NOM die:sg:FGR-IND
My dog died.

External possession is promotion of a possessor to argument status (Payne & Barshi eds. 1999), and the promoted possessor can be case-coded as an object, registered on the verb (e.g. in an applicative), and/or indexed on the verb as in (19b). The difference between criterion [5] illustrated in (18) and criterion [6] in (19) is that in [5] and (18) the possessor is not promoted to argument status.

[7] Possessive relative: The person of a gapped subject of a relative clause is indexed on the head noun, as in (20).

(20) Khalkha Mongolian; Ackerman & Nikolaeva 2014:261. ___ = gap (relativized noun); ... = null anaphora.

a. [... ___ bič-sen] nom-**min'**
 A (1sg) O write-PARTICIPLE book-1sg

'the book I wrote'

- b. [činii ög-sön] nom-ig-č**in'**
 2sg.GEN O give-PARTICIPLE book-ACC-2sg
 'the book you gave'

In (20a-b) the relativized object is 'book', which is third person, but the head noun agrees in person not with the relativized noun but with the subject of the relative clause. For this variable and its typological prerequisites see Ackerman & Nikolaeva 2014.

[8] Person is marked on or attracted to the negative morpheme, which is separate from the lexical verb. This is one form of asymmetrical negation (Miestamo 2005), in which positive and negative clause structures differ. In the following examples from Finnish, in the positive clause the verb agrees with the subject in person and number, while in the negative clause the negative morpheme takes the agreement (and also tense) and the lexical verb is an inert uninflected form (known as a *connegative* form). This behavior is inflectional because the person marker appears on the head of the phrase (the negative morpheme, which is the finite structural head of the clause). In Finnish, as in many languages with such constructions, the unquestionably inflectional category of tense behaves in the same way.

- (21) Finnish
- | | |
|------------------|-----------------|
| (minä) tiedän | (minä) en tiedä |
| 1sg.NOM know-1sg | NEG-1sg know |
| (sinä) tiedät | (sinä) et tiedä |
| 2sg.NOM know-2sg | NEG-2sg know |
| (hän) tietää | (hän) ei tiedä |
| 3sg.NOM know-3sg | NEG-3sg know |

For all of items [1]-[8], a Yes answer is an inflectional property.

Agreement in person on modifiers in an NP, as in (10)-(12) above, is so rare that I have not counted it in the survey. This might be regarded as the inflectional-person counterpart to gender and case agreement of modifiers in many Indo-European languages, but the examples I have found either have various restrictions or (as discussed above) appear to be primarily gender agreement, marked with a portmanteau person-number-gender affix.

Independent pronouns are lexemes and words, just as nouns are, and can usually head NP's or in other ways behave as NP's; but they may also have non-noun-like properties of structure or function. Items [9]-[17] capture ways in which independent pronouns behave unlike nouns and like affixes, and ways in which their person category is associated with an affix rather than with a lexical root. Here and below, an asterisk marks an item for which a Yes answer is a lexical property.

Independent pronouns as a word class (1st-2nd persons)

[9] Generic pronoun base. See (4)-(5), (16)-(17), and discussion of survey item [4] above. That item dealt with the person marker in such forms; this one applies to the pronominal base itself, which if generic does not contain a person category. See also the Abkhaz forms in (27) below and the discussion there.

[10]* Unique root per person or person-number, with person as a lexical property of the root. See the Latin pronouns in (3) above and the discussion there. English pronouns are also of this type: *I, me; you; we, us; you*, each a distinct root with no rhyme, alliteration, etc. between members of the paradigm.¹⁰

[11] Roots or stems rhyme, alliterate, etc. along person or number lines. Any form of assonance among pronoun forms marks them as members of a small closed set and a paradigm (Nichols 2001), an inflectional property. In the following examples person and number, while not marked with discrete morphemes, are factored out phonesthemically or paronomastically by rhyme, alliteration, or other assonance. This is common in Uralic, Turkic, and Avar-Andic languages.

(22) Finnish (Uralic)

<i>1sg</i>	<i>2sg</i>	<i>1pl</i>	<i>2pl</i>
minä	sinä	me	te

(23) Erzja Mordvin (Uralic)

<i>1sg</i>	<i>2sg</i>	<i>1pl</i>	<i>2pl</i>
mon	ton	miń	tiń

In Finnish the person forms begin with the same consonant, first person *m* and second person *t* (the alternation of *t* with *s* before *i* is regular in Finnish, so the identity of the underlying consonant is clear). Singular forms and plural forms rhyme; the *-nä* of the singular forms can be segmented off though it is not a morpheme (and certainly not a singularizing morpheme). In Mordvin a vowel alternation distinguishes the number paradigms; it is not a regular alternation in the language and certainly not a regular marker of number.

(24) Avar (Nakh-Daghestanian; Charachidzé 1981:71)

	<i>1sg</i>	<i>2sg</i>	<i>Incl</i>	<i>1pl</i>	<i>2pl</i>
Nom.	dun	mun	nił	niž	nuž
Gen.	di-r	du-r	neł-e-r	než-e-r	nuž-e-r
	1sg-GEN	2sg-GEN	INCL-OBL-GEN	1pl-OBL-GEN	2pl-OBL-GEN

¹⁰ The English general form *me* rhymes with subject form *we*, but these are not in the same paradigm position so this does not count as a partial root sharing.

Avar nominative singular forms rhyme, and oblique singular forms alliterate but do not rhyme; plural forms rhyme, first person plurals have the same vocalism, and first (exclusive) and second plural forms share a final *-ž* (a former plural, no longer a segmentable plural morpheme in Avar).

(25) Turkish

	<i>1sg</i>	<i>2sg</i>	<i>1pl</i>	<i>2pl</i>
	ben	sen	biz	siz

In the Turkish forms alliteration unites the person forms and the elements *-en*, *-iz* are shared by the singular and plural forms respectively, but none of these are obviously segmentable as morphemes. Plank 1999:187 treats Turkish *-iz* as a plural marker restricted to first and second person forms, a solution that would work for Finnish (22) as well (distinct number markers restricted to higher animacy levels are not uncommon cross-linguistically). By that token Finnish would then have a singulative marker for only first and second persons. Rather than take a stance on whether any of these are number morphemes I simply speak of number as *factored out* in such forms.

An interesting case comes from the West Caucasian languages, illustrated by Adyghe and Abkhaz:

(26) Person prefixes of verbs, and independent pronouns, in Adyghe (Arkadiev et al. 1999:45, Keraševa 1960:1076, retranscribed following Arkadiev et al.) and Abkhaz (Chirikba 2003:40, 32):

	<i>S/O</i>	<i>G</i>	<i>A</i>	<i>Independent</i>
1sg	sə-	s-	s-	se
2sg	wə-	w-/p-	w-/p-	we
1pl	tə-	t-	t-	te
2pl	š ^w ə-	š ^w -	š ^w -	š ^w e

(27) Abkhaz

	<i>S/O</i>	<i>G</i>	<i>A</i>	<i>Independent</i>
1sg	s(ə)-	s(ə)-	s(ə)- ~ z(ə)-	sa ~ sa-ra
2sg M	w(ə)-	w(ə)-	w(ə)-	wa ~ wa-ra
2sg F	b(ə)-	b(ə)-	b(ə)-	ba ~ ba-ra
1pl	h(a)-	h(a)-	h(a)-	ha ~ ha-ra
2pl	š ^w (ə)-	š ^w (ə)-	š ^w (ə)- ~ ž ^w (ə)-	š ^w a ~ š ^w a-ra

In both languages the independent form is nearly identical to the prefix form; the difference is that the affixes exhibit a vowel-zero alternation while the independent form has a stable vowel. The independent forms all rhyme, and they resemble a closed paradigm in the near-identity of prefixes and independent words. (The Abkhaz independent pronouns have an additional optional *-ra* which could be analyzed as a generic pronominal root, though all sources regard it as an affix. It adds no semantic or pragmatic meaning.)

[12]* There is no closed class of pronouns. This is typical of Southeast Asian languages, in which various nouns serve more or less conventionally, often with honorific or anti-honorific implications, in lieu of pronouns (Cooke 1968). Since such words are lexemes, this is a lexical property.

Independent pronouns: Inflectional paradigms (1st-2nd persons)

[13]* Pronouns have the same cases as nouns. Only the core grammatical cases are considered here: those that are default marking for A, S, O, G, T, and possessor (common case names: ergative, nominative, accusative, dative, genitive). In most languages of central and western Eurasia, if there are cases the nouns and pronouns distinguish the same set of cases. In languages such as the Pama-Nyungan family of Australia, however, it is common for nouns to distinguish core grammatical cases on an ergative pattern while pronouns are accusative or neutral (Silverstein 1976). E.g. subject and object cases in Pitjantjatjara (Western Desert; Australia; Bowe 1990:12, Dixon 1980:302, 331, 342; endings for nouns, whole forms for pronouns; T = /t/ in most contexts, but assimilating to some preceding consonants):

(28)	<i>A</i>	<i>S</i>	<i>O</i>	<i>Alignment:</i>
Common nouns	-Tu ~ -ngku	-Ø	-Ø	Ergative
Proper nouns	-Tu ~ -ngku	-nga ~ -nya	-nga ~ -nya	Ergative
Pronouns: 1sg	ngayu-lu	ngayu-lu	ngayu-nya	Accusative
2sg	nyuntu	nyuntu	nyuntu-nya	Accusative

A similar pattern is found in Adyghe and Kabardian (West Caucasian), where nouns make a nominative/oblique case distinction but pronouns do not. Another type of difference is that of English, where pronouns distinguish subject and object cases but nouns do not.

Displaying the same inventory of cases as nouns is a lexeme-like property in pronouns and counts as lexical.

[14]* Pronouns have the same case morphology as nouns. In Indo-European languages preserving the original case paradigms, the case endings of nouns and pronouns are different (though the case categories and functions are the same), as in (29) from Latin. In many other languages they are the same or largely so, as in (30).

(29) Cases in Latin (singular pronouns and three different noun declension classes in the singular). The pronoun genitives in *-ī* are used as objects and complements, and those in *-us* are used in possessive constructions; the possessive forms are adjectives and agree in gender, number, and case with the head noun.¹¹

<i>1sg</i>	<i>2sg</i>	<i>'water' (F)</i>	<i>'master' (M)</i>	<i>'name' (N)</i>
------------	------------	--------------------	---------------------	-------------------

¹¹ Here and below, cases are presented in an order that captures syncretisms in the language, or otherwise in this order: citation form (nominative), S, A, O, G, possessor. This may differ from the traditional order.

Nominative	ego	tu	aqua	dominus	nōmen
Accusative	mē	tē	aquam	dominum	nōmen
Dative	mihi	tibi	aquae	dominī	nōminī
Genitive	meī/meus	tuī/tuus	aquae	dominō	nōminis

(30) Cases in Finnish: pronouns and one noun (Fromm 1982). The partitive marks indefinite and/or quantified objects, subjects, and predicate nominals, contrasting with accusative and nominative, and has other functions.

	<i>1sg</i>	<i>2sg</i>	<i>1pl</i>	'lock'
Nominative	minä	sinä	me	lukko
Partitive	minu-a	sinu-a	meit-ä	lukko-a
Accusative	minu-t	sinu-t	meid-ä-t ¹²	luko-n
Allative (G)	minu-lle	sinu-lle	mei-lle	luko-lle
Genitive	minu-n	sinu-n	meid-ä-n	luko-n

The Finnish case endings are identical except that nouns have genitive-accusative syncretism in the singular. (If the core case paradigms differ by only one instance of syncretism I counted them as the same for purposes of item [14].) For alternations in the stems (consonant gradation, final vowel alternations) see item [15] below.

(31) Hup cases: singular pronouns, endings of nouns (Epps 2008:159, 166). Nonhuman nouns may omit the object case ending; inanimate nouns usually do. The possessive morpheme is a postpositional particle for nouns, a fused suffix for pronouns (224). Possessive pronoun forms are from the more transparent Umari Norte dialect.

	<i>1sg</i>	<i>2sg</i>	<i>Noun</i>
Subject (S, A)	ʔāh	ʔám	-Ø
Object (O, G)	ʔān	ʔám-ān	-ān
Oblique	ʔāh-āt	ʔám-āt	-Vt
Possessive	nīh	ʔam-nīh	nīh

[15]* Pronouns have the same root or stem flexivity as nouns. This item is designed to capture any paradigms where pronouns have the same ablaut, extensions, etc. as nouns, e.g. marking nominative vs. oblique cases. An example is Finnic and Saami consonant gradation shown for Finnish in (30) above, where the stem consonants *-kk-* of 'lock' and *-t-* of 1pl become *-k-* and *-d-* respectively before a closed syllable (the *-n-* of the singular pronouns is phonologically inelidable). Gradation applies to both both nouns and pronouns (and all other parts of speech) and is a partly morphologized but primarily phonological alternation rather than true ablaut. (The pronouns in (30) do not share all their stem flexivity with nouns, however. They undergo stem-final changes that are unique to pronouns: 1sg nom. *minä*, oblique *minu-*; 1pl *me*, *meit-*.) More common are languages where pronouns have suppletive stems (as in English *I* vs. *me* or Latin *ego* vs. *me*) but

¹² The *-ä-* in the accusative and genitive is probably best analyzed as epenthetic.

nouns do not, or nouns have oblique extensions but pronouns do not (as in many Nakh-Daghestanian languages: Kibrik 2003).

[16] Pronouns have different number categories from nouns. It is not uncommon for pronouns to distinguish number categories while nouns do not. In Günuna Küne (Chon; Argentina; Casamiquela 1983) and Mapudungun (isolate, Chile; Smeets 2008:72, 109-112) nouns do not distinguish number while pronouns distinguish singular/dual/plural;¹³ in Ainu (isolate, Japan; Tamura 2000) and Awa Pit (Barbacoan, Colombia; Curnow 1997), nouns make no number distinctions while pronouns distinguish singular/plural. Also common are languages in which only the highest-animacy nouns distinguish number (e.g. only human nouns, or only kin terms, or only a few of these) while all pronouns do, e.g. Great Andamanese (isolate; Abbi 2013) and Yurok (Algic; Robins 1958). Pronouns are of course high in animacy, so the set of words that distinguish number in such languages is a coherent one, but still the morphological observation can be made that for these languages number is a partial category for nouns but present in all pronouns. Several languages distinguish singular/dual/plural in pronouns but only singular/plural in nouns: Kunama (isolate; Bender 1996:12, 18), Taa (or !Xóǿ; Tuu or Ui-Taa) (Traill 1994: 27, 35), Saami (Uralic; Sammalahiti 1998:62), Belhare (Sino-Tibetan: Kiranti; Bickel 2003), Wintu (Wintun; California; Pitkin 1994). Combining these patterns, Hanis (Coos; Oregon; Frachtenberg 1922a:373-5) distinguishes singular/dual/plural in pronouns, only singular/plural in some human nouns, and no number in other nouns.

There are other ways in which nouns differentiate number more finely than pronouns do. In Bininj Gun-Wok (Gunwingguan; Evans 2003), nouns distinguish only singular/plural while pronouns have a minimal/augmented system. In Motuna (South Bougainville; Onishi 1994:4-5) pronouns distinguish only singular/plural while nouns have singular/dual/paucal/plural; similarly, in Kuniyanti (Bunaban; Australia; McGregor 1990), Nez Perce (Sahaptian, Oregon; Rude 1985:120, 76), and Hopi (Uto-Aztecan, Arizona; Hill 1998), pronouns distinguish singular/nonsingular and nouns singular/dual/plural. In Kiowa (Kiowa-Tanoan, Oklahoma; Watkins & McKenzie 1984), pronouns make no number distinctions while nouns distinguish either singular-dual vs. plural or dual-plural vs. singular with one of the members (the one including dual) basic and the other marked and an inverse suffix that switches number (while verbs distinguish all three of singular, dual, and plural) (Watkins & McKenzie 1984:78ff., Wonderly et al. 1954, Merrifield 1959). In Cherokee (Iroquoian; Montgomery-Anderson 2015) independent pronouns make no number distinctions while nouns take person prefixes that distinguish number.

Regardless of whether it is nouns or pronouns that distinguish number, or which of them makes more number distinctions, where number categories differ they are counted as an inflectional property in pronouns because the pronouns have their own inventory of categories distinct from those of nouns.

[17]* Pronouns have the same number markers as nouns (Daniel 2013, types 6 and 8). An example from Quechua is in (1) above, where plural pronouns are marked by the same

¹³ For most items in this survey duals are disregarded (see §2.1 above), but for this one the presence vs. absence of a dual is considered a difference in number categories.

plural suffix as nouns are. This kind of pattern is found in several other languages where number is separatively marked. By far the most common means of distinguishing number in independent pronouns is distinct roots or stems as in Latin 1sg nom. *ego*, acc. *me*, 1pl. nom.-acc. *nōs*. The singular and plural pronouns of Finnish (22 above) have the same initial consonant but otherwise have different stems; as noted there, the initial consonant is not a morpheme and not a pronominal root by itself.

Sharing number markers with nouns aligns pronouns with nouns and counts as a lexical property of pronouns.

Person markers: Morphological behavior. Items [18]-[22] are designed to capture some of the properties of inflectional morphology: indexation is distinct from the arguments themselves; inflectional morphemes are usually more peripheral than derivational and inherent morphology (e.g. Bybee 1985).

[18] Any multiple marking of person per argument. An argument itself occurs only once per clause, but an argument can be indexed or registered more than once per verb form. The most common form of multiple marking applies to number marking on verbs, where number of an argument can be signaled both in a person-number index and in a separate number marker as in (10-11) above. Multiple marking of person by itself is less common but does occur, e.g. Icarí Dargwa (Sumbatova & Mutalov 2003:70, 80-82) (person markers boldface):

(32) u ūq - u - **tt** - aj - ci - **di**
 2sg climb.over:M:PF-THEMATIC - **2** - SUBJUNCTIVE - ADJECTIVE - **2sg**
 You could cross (this canyon) (Sumbatova & Mutalov 2003:108)

The two person markers have different patterns of person and number syncretism, hence the different interlinears "2" and "2sg".

Multiple exponence of person, discussed in connection with person agreement in NP's (14-15 above) but not surveyed here, and multiple exponence on verbs per item [18], can be viewed as discourse phenomena that have the effect of maximizing the text frequency of tokens of person in inflection. Discourse-based measures of lexical and inflectional person are not pursued here but will be covered in a future paper.

[19] A and O markers are formally identical. Where person markers have different forms for indexing different arguments, they are combining information about person with information about case or role and are therefore not pure person markers. When the forms are identical they distinguish only person (or person-number) and not role. (Where role is not marked segmentally or tonally it is indicated in some other way, typically by affix ordering or by hierarchically-based access to a single slot.)

Examples occur in West Caucasian languages, e.g. Adyghe and Abkhaz in (26-27) above. The only segmental difference between the A and O argument series is the schwa in the S/O set of Adyghe and the voicing alternation in fricatives in the A forms of Abkhaz. The alternation of Adyghe schwa is arguably predictable on an abstract phonological analysis, and the voicing in Abkhaz is sandhi. What primarily differentiates the three sets is that they occur in different prefix slots.

[20] Person is the outermost verb inflection. Inflectional morphemes are typically more peripheral than other morphemes, and seeking the outermost morpheme – the first prefix and/or the last suffix – is a conveniently determinate way of identifying an unquestionably peripheral morpheme. This count considers affixes and clitics together, so clitics are almost almost always outermost compared to affixes. I count person marking as outermost whether it is separative or coexponential with TAM. If more than one argument is marked, one outermost and one not, I count the outermost one. An example is (30) above, where person marking is counted as outermost though it also occurs internally.

Hengeveld 2012 shows that, where the distinction can be made, cross-reference person markers are located closer to the verb stem and agreement markers occur in more peripheral positions. Cross-reference markers are largely the same as what are also called pronominal arguments (e.g. Jelinek 1984, VanValin 1985), i.e. not agreement markers but the actual arguments themselves. These should be relatively lexical, as they are reduced or incorporated versions of independent pronouns, which can be lexical. Agreement markers should be more inflectional, and it is only the agreement markers that are expected to be in the outermost position. However, as Hengeveld notes, available descriptions for most languages do not make it possible to determine on an independent basis whether person indexes are cross-reference markers (pronominal arguments) or agreement markers, leaving the question open for further research.

[21] Person is the outermost noun inflection. Contrastive examples come from Finno-Ugric languages, which vary in the position of their person markers:

- (33) Finnish talo-ssa-ni
 house-INESSIVE-1sg
 in my house
- (34) Hungarian a ház-am-ban
 ART house-1sg-INESSIVE
 in my house

The person marker is final (outermost) in Finnish but not in Hungarian, so in this respect Finnish possessive marking is more inflectional than Hungarian. Criterial examples of items [20] and [21] arise mostly in languages with separative exponence of person from such categories as tense and case.

Items [22]-[29] identify some hallmarks of categorial discreteness of person.

Inclusive/exclusive distinguished in:

- [22] Independent pronouns
[23] Verb person indexes
[24] Noun possessive indexes
[25] Auto-person on pronouns

Inclusive/exclusive positions amount to isolating different persons relative to systems making no such distinction: as discussed in §1 above, the combination of first person plus second (inclusive) is distinguished from first plus third (exclusive). Alternatively and more precisely, they can be described as identifying or isolating SAR's and not the often broader person values. Most languages consistently either do or do not make an inclusive/exclusive distinction in all their person paradigms (Bickel & Nichols 2005:53), so for most languages it suffices to identify one context where the distinction is made. However, since not all languages index person in the same places, for survey purposes four different points are checked. Whether to count all four is a matter of weighting, and counting all four (as is done here) reflects the pervasiveness of inclusive/exclusive marking throughout the person system and the pervasiveness of person indexation throughout the morphology, allowing languages that mark inclusive/exclusive in more paradigms to rank as more inflectional. There are different kinds of inclusive/exclusive systems (see Bickel & Nichols 2005, Cysouw 2003, 2013), of which the minimal/augmented type (shown in (35)), treating inclusive on a par with the singulars of the other person values, maximally separates person from number and isolates pure person.

(35) Rembarrnga (Gunwingguan, northern Australia; Dixon 1980:352 following McKay 1978; cited from Bickel & Nichols 2005)

<i>Person</i>	<i>Minimal</i>	<i>Unit augmented</i>	<i>Augmented</i>
1 (Exclusive)	ngənə	yarr-parraʔ	yarr-ə
Inclusive	yəkkə	ngakorr-pparaʔ	ngakorr-ə
2	kə	nakorr-parraʔ	nakorr-ə

Person and number factored out (or no number distinctions made) in:

- [26] Independent pronouns
- [27] Verb person index for core arguments
 - 27a: A
 - 27b: O
- [28] Noun possessive indexes
- [29] Auto-person on pronouns

Examples for independent pronouns (criterion [26]) are in (22)-(25) above. Factoring out number distinctions, or marking number separately from person with an affix on pronouns, or not distinguishing number at all in pronouns, are different ways of isolating person from number, and all of them amount to treating person as a discrete inflectional category. Daniel 2013 finds that the single person-number stem – i.e. not factoring out person from number – represents the substantial plurality of the number-marking strategies in independent pronouns.

Noun genders distinguished in:

- [30]* Independent pronouns
- [31]* Verb person index for core arguments
 - 31a: A

- 31b: O
 [32]* 3rd person independent pronouns
 [33]* 3rd person verb index for core arguments
 33a: A
 33b: O
 [34]* Inherent (lexical) gender in pronominals

Gender of nouns, in languages that have it, is a lexical property of nouns. It may be semantically arbitrary,¹⁴ as in most Indo-European languages, or semantically transparent, as in the concord classes of many Bantu languages or in the gender systems of Avar and Andic (Nakh-Daghestanian) languages, which distinguish three genders with complete semantic predictability: male humans, female humans, and all else. To the extent that gender is predictable, the predictability usually rests on either the phonology or the semantics of the gender-bearing noun. These are lexical properties of nouns, and if pronouns or grammatical person markers distinguish the same gender categories as nouns, that behavior is more lexical than inflectional. Less clearly lexical, but still non-inflectional, is strictly pronominal natural gender as in English *he, she, it*. Here gender is not a lexical property of nouns, but a real-world property of referents, fixed with the referent regardless of the person of that referent (while the person is a shifter), so it is counted as lexical. Items [31]-[34] track lexical gender and [35]-[37] below track natural pronominal gender.

Items [30]-[33] refer to noun genders in languages in which all or most nouns belong to gender classes (e.g. Spanish, Russian, Arabic, Avar, Luganda). Pronouns usually have no inherent gender of their own but agree in gender with their anaphoric controller, but [34] describes the rare phenomenon whereby pronouns have their own inherent gender categories. So far I know of only three examples of this: (a) In three Nakh-Daghestanian languages, first and second person pronouns take different gender agreement in their plurals than third person pronouns do (Daniel in preparation, Corbett 2013), e.g. Ingush:¹⁵

- (36) Ingush (Nichols 2011a)
- | | | | |
|-----|-----|--------------------|-----------------------|
| txo | d.y | (we.EXCL D.is/are) | 'we are, it's us' |
| vai | d.y | (we.INCL D.is/are) | " " " " |
| shu | d.y | (you.PL D.is/are) | 'you are, it's you' |
| yzh | b.y | (they B.is/are) | 'they are, it's them' |

(b) In Uduk (Koman; Ethiopia; Killian 2015), all pronouns belong to gender 1, regardless of the sex of the referent or the gender of the noun which the pronoun anaphorically stands for. While in most languages with noun gender the gender of high-animacy nominals is

¹⁴ Except for human nouns, which in nearly all gender systems have predictable gender (Dahl 2000).

¹⁵ The most common take on the Nakh-Daghestanian pattern is that reported here: pronouns have their own gender categories, and the fact that first and second person behave differently from third further shows that the languages have a grammatical category of person. My own analysis is that – given that plural in Nakh-Daghestanian usually involves a change of gender marker – this is simply a distinctive type of number marking at the highest end of the hierarchy (for other examples of distinct high-animacy number marking see the discussion of item [16] above).

predictable, in Uduk high-animacy nouns are unpredictable and the threshold for predictability is set even higher, encompassing only proper names (all of which are gender 2 regardless of the sex of the referent) and pronouns. See Killian in press, 2015. (c) In Jarawara (Arawan; Dixon 2000:488) pronouns trigger feminine gender agreement on the verb regardless of the sex of the referent.¹⁶

Natural gender (not noun genders) distinguished in:

- [35]* Independent pronouns
- [36]* 3rd person independent pronouns
- [37]* Verb person index for A and/or O

Items [35]-[37] apply only to strictly pronominal gender in languages that have no noun gender, e.g. English. They do not apply to pronouns as in Spanish, German, or Russian, where the pronouns index the noun genders. (Since gender is almost always predictable for nouns referring to humans, in languages having noun gender it coincides with natural gender, except in the few languages just discussed that have inherent pronoun gender).

Integration into inflectional paradigms

- [38] Verb person index(es) cumulative with TAM
- [39] Person involved in hierarchical marking on verbs
- [40] Person involved in inverse marking on verbs
- [41] Egophoric (conjunct/disjunct) marking on verbs
- [42] Person determines access to (hierarchical, promiscuous) number marking on verbs

Inflectional categories often combine or interact with other inflectional categories. Where those categories are not lexical (as gender is; see just above) but purely inflectional, such as TAM or the person of other arguments, the interaction is an inflectional property.

Item [38] is inflectional behavior because person is fused or otherwise combined with the clearly inflectional category of TAM (the same criterion was used for attraction of person to negation, criterion [8] above). [39]-[42] have to do with person determining the status or configuration of other argument indexes.

Items [22-25] (inclusive/exclusive) and [41] (egophoric marking) are the only ones where the difference between SAR's and SAP's becomes relevant, and here it is SAR's and not SAP's that are the survey objects. These items share the fact that they involve the unseparated Speaker = Addressee value. Also, more theoretically, the whole question of how to assign and determine grammatical person values (SAP's) from SAR's is mostly driven by the need to explain how inclusive does and does not syncretize with first or second person plural, i.e. not just person itself but person-number combinations, and the referential structure of these is best explained in terms not of person values but of SAR's (in addition to the above sources see Cysouw 2003, whose concept of *pure person* is much the same as SAR).

2.2. Counting and typologizing

¹⁶ I thank Don Killian for clarification of the Uduk and Jarawara data.

Languages are typologized for inflectional person by calculating the proportion or percentage of their person marking that is inflectional. These range from zero to 100%, with a mean of about 70%. (The mean is high because there are more inflectional than lexical points in the survey items, and that is partly because the survey was designed to capture all surveyable inflectional properties and partly because grammatical morphemes show a greater variety of person-related morphosyntactic behavior than lexical items and in particular nouns do.) For this first exploration I have simply used the unweighted totals of points (one per numbered item above). What is counted is "yes" answers to the various items; "no" answers, "n/a" answers, and "n.d." answers made no contribution to the calculated proportion.¹⁷

3. Typology

3.1. Range of types

Languages with primarily inflectional person include several East African languages, the West Caucasian language family, and many languages of the Americas, e.g. languages of the Salish, Siouan, Algonquian, Kiowa-Tanoan, and Keresan families. As an example, Adyghe (West Caucasian) has indexation for person of three arguments as well as a fairly open-ended range of goal-like non-arguments or quasi-arguments (benefactive, goal of motion, etc.), indexation on heads of NP's of the person of a possessor (38), and personal pronouns which in their independent form consist of little more than the root consonant of the possessive index, with no lexical properties such as gender:

- (37) pšaše-m se s-jə-łeğh^w-ə-ğh
 girl-ERG 1sg-NOM 1sg.S/O-3sg.A-see-PAST
 'The girl saw me' (Adyghe; Arkadiev et al.1999:61, partly my transcription)

- (38) Adyghe possessive paradigm for 'father' (Arkadiev et al. 1999:56)

1sg	s-jate	1pl	t-jate
2sg	w-jate	2pl	š ^w -jate
3sg	jate	3pl	jate

See the Adyghe pronouns in (26) above. The pronouns make the same singular/plural distinction as nouns, but NP's distinguish core cases in that definite or specific nouns take a cliticized or suffixed article that is case-marked, while pronouns do not take this article and do not distinguish case.¹⁸

¹⁷ I experimented with a different count that counted "no" answers to lexical points as "yes" answers to inflectional points and vice versa, and another that counted only "yes" answers to inflectional points as a total of the items surveyed (letting "no" and "n/a" answers determine the total counted). Both made little difference to the overall outcome.

¹⁸ Nouns in Adyghe and closely related Kabardian also take instrumental and adverbial suffixes, which some sources regard as cases (e.g. Colarusso 1992:51, Arkadiev et al. 1999:53) but which do not mark core arguments so they are not considered here.

Adyghe, and the North American languages mentioned above, are highly complex and mostly polysynthetic languages, but some languages with relatively little morphology also have mostly or entirely inflectional person. An example is Ainu, with relatively simple inflectional morphology including person indexation for two arguments (39a), hierarchical patterning in the argument markers (39b), person indexation on heads of possessive NP's, personal pronouns consisting of the person index plus a generic base (40), and different inflectional paradigms for nouns and pronouns (nouns do not distinguish number; pronouns do).

(39) Ainu clause (Bugaeva 2012:473)

- a. cikap ku=Ø=nukar
bird 1sgA-3sgO=see
I saw a bird
- b. eci=nukar rusuy
1sgA>2sgO=see DESIDERATIVE
I was looking for you

(40) Ainu pronouns (Shibatani 1990:31, 28). The generic base *-ani* is a nominalized form of the verb 'be'.

1sg ku-ani
2sg e-ani
3sg ani

Another language with simple inflectional morphology and most of it inflectional person is Gbeya (Adamawa-Ubangi, Cameroon and Central African Republic). Yimas (Lower Sepik-Ramu, New Guinea) is polysynthetic but with fairly simple person morphology, most of it inflectional. Tzutujil (Mayan, Guatemala) has fairly simple morphology overall, much of it person and most of that inflectional.

Languages that entirely or almost entirely lack inflectional person include Hindi, Thai, Japanese, Dyirbal, and many Nakh-Daghestanian languages. Japanese and Dyirbal have no verb indexation at all, and most Nakh-Daghestanian languages have gender but not person indexation on verbs.¹⁹ These languages have moderate to complex inflectional morphology overall. Languages of the Panoan family (South America, chiefly Amazonia) have complex inflectional morphology and could almost be considered polysynthetic were it not for the fact that they have no person indexation at all and low proportions of inflectional person overall. Languages with simple inflectional morphology and little or no inflectional person include isolating languages of Africa, Southeast Asia, and the Pacific; e.g. Abun (isolate, New Guinea) has no inflectional person and little person-related morphology of any kind. Several languages of Southeast Asia have positive inflectional person scores because their pronoun structure typically factors out person from number,

¹⁹ This is a rarity, cross-linguistically, as noted above: the vast majority of languages with argument and/or possessor indexation use person, sometimes coexponential with gender or classification, but gender by itself as indexation is not common.

e.g. Mandarin (39) with a separate plural morpheme, though the language has no person indexation.

(41) Mandarin pronouns

	Singular	Plural
1	wǒ	wǒ-men
INCL		zán-men
2	nǐ	nǐ-men

I drew a sample of 270 languages which covers most of the world at standard densities for medium-sized samples but covers northern Eurasia more densely than other continents (in order to make the geographical description there as precise as possible). In the taxonomy of Miestamo et al. 2016 it is a genus-macroarea worldwide core sample with an extended sample of northern Eurasia based on family-internal diversity in the spirit of their Diversity Value method. In addition, the sample opportunistically chooses the occasional language because of its distinctive treatment of person. New Guinea and South America are somewhat undersampled relative to their total populations of families but still contribute large absolute numbers of languages. (See Table 2 below for language numbers per continent.) Of the 270 languages, ten had absolute total survey points so low that percentages veer dramatically if just one point is added or removed, so they are left out of counts based on percentages of inflectional person points. Counting only the inflectional points (rather than a proportion of total points) is occasionally used below to include those ten languages.

In the worldwide sample, inflectional percentages range from zero (Mande, Hunzib, Hindi, Pitjantjatjara, Shipibo-Conibo) to 100% (Aleut, Ainu, Nisgha, Kwak'wala, Hanis, Washo, Acoma, Kiowa, Lakhota, Creek, Pawnee, Huave, Cayuvava, Mapudungun) and the total absolute number ranges up to a maximum of 16.5 (Olutec). Lexical points range from zero (Ainu and Aleut) to a maximum of 7.5 (Russian). Figure 1 plots total inflectional vs. lexical points. Polysynthetic languages are at the upper left; languages with minimal person-related morphology are at the lower left (the extreme is Mandinka, with one lexical point and no inflectional points). The theoretical extreme of zero lexical and zero inflectional points is not met in the sample. (The closest are Mandinka, and Ainu and Aleut with zero lexical points and one inflectional.) The upper right quadrant is largely empty, suggesting that languages avoid having maximal elaboration along both dimensions. The middle of the plot is densely populated, showing that languages can easily mix the two types in more or less equal proportions and are evidently not under pressure to utilize complementarity and specialize in just one or the other person type.

Figure 1 about here

3.2. Correlations

I tested for correlations of inflectional person with five other typological variables of grammatical and/or geographical relevance to person: (a) Head-dependent marking, since head marking was expected to pattern closely with high inflectional person (which has a more fine-grained breakdown); (b) Verb-based vs. noun-based lexicon, the

proportion of items on a standard wordlist whose derivational paradigms have verbs vs. nouns as base (Nichols 2016, Foley in preparation), since preliminary studies had indicated that high inflectional person and verb-based lexicons both characterize the North Pacific Rim language population and North America more generally; (c) Lexical valence orientation, the preferred realization of the causative alternation based on an 18-pair standard wordlist (Nichols, Peterson, Barnes 2004), which has a similar geography (Fortescue 1998, Nichols 2011b, 2017a); (d) Inclusive/exclusive pronoun oppositions, which are grammatically related to inflectional person in that they factor out person from number and instantiate a pure person system (Cysouw 2003), and which have a Pacific Rim distribution similar to inflectional person (Bickel & Nichols 2005); and (e) hierarchical patterning, any indexation subsystem in which surface distributions are based on ranking persons rather than signaling argument roles, in this survey chiefly hierarchical and inverse clause argument marking (e.g. Zúñiga 2006). For comparison I also surveyed two stalwarts of typology, alignment and basic word order, neither of which has obvious grammatical connections to person but both of which are well described for many languages and have been extensively surveyed worldwide. For alignment I surveyed only accusative vs. ergative; for word order, verb-initial vs. other and verb-final vs. other. I also surveyed taxonomic, or inventory, complexity, i.e. a measure of the number of elements per subsystem in several subsystems of grammar (for this type of complexity see e.g. Sinnemäki 2011, Miestamo et al. eds. 2008, Dahl 2004), using an extended version of the database of Nichols 2009. The results are shown in Table 1.

Table 1 about here.

For all nine variables (inflectional person, the five related variables, and the three additional ones) I tested for simple correlations in the entire sample. (To avoid circularity, for head/dependent marking and inclusive/exclusive the counts for inflectional person excluded the survey items for argument indexation and inclusive/exclusive respectively.) For all of these comparisons I binned the inflectional percentage figures into high vs. low, splitting at the mean of 70%, and likewise for complexity. Thus all correlations are between two categorical variables (the others were defined categorically in the first place).

Table 1 shows which of the correlations are significant: a blank cell means a non-significant correlation, and the other entries show levels of significance or near-significance. I used available databases for the other variables, and none of them had data for all 270 the languages in the person survey. The *N* column in Table 1 shows, for each of the variables, the number of languages (where this differed among the variables it was compared with, the largest number is shown). All variables are well enough represented to make counts reliable, except noun-based/verb-based, a labor-intensive survey that is just getting underway (and even that one showed two near-significant correlations).

Most of the variables had significant or near-significant correlations with some of the others, summarized at the bottom of the table. Inflectional person has no greater, and in fact somewhat less, tendency to form correlations than the other variables. It correlates significantly with head marking, which as noted above is to be expected as the two are grammatically similar (even when obvious similarities are removed by excluding inflectional points for argument indexation). This is the strongest correlation on the table as shown in the highest level of significance. It also receives some continent-by-continent

support: five continents show the same skewing, with head marking favoring high inflectional person and/or dependent marking disfavoring it; five are inconclusive; none reverse the correlation.²⁰ (Margins and sample sizes are very small when the whole sample is broken down into ten continents. None of these continent-sized subsets reaches significance.)²¹

Inflectional person also correlates significantly with causativization, with which it has no obvious grammatical affinity but which has a similar geographical distribution (see §3.2). (Since only about half of the sample languages had data available on the causative alternations, the sample size was too small to support continent-by-continent testing.) Despite the grammatical affinity of inflectional person to inclusive/exclusive oppositions, there is no significant correlation (and in fact no appreciable skewing at all). Inflectional person might be expected to correlate with hierarchical patterning (in that, in a larger convenience sample of only that variable, hierarchical patterns always involve person and often involve only person), but it does not.

Of the other significant or near-significant correlations among these variables, complexity, alignment, and word order correlate with head/dependent marking but not with inflectional person. Word order correlates with causativization, perhaps due to similar geography (both are more common in the Americas and northern Asia than elsewhere) and/or to the tendency of both to be favored in certain contact situations (Fortescue 1998, Nichols 2011b). Inclusive/exclusive and word order may be due to similar geography. Hierarchical patterning and alignment (in which ergativity favors hierarchical patterns) is probably due to the general association of stative/active alignment and hierarchical patterns with base ergativity (Bickel 2011).

Overall, then, inflectional person behaves much like head/dependent marking and exhibits few other correlations apart from geographical cooccurrence and can therefore be useful as a fine-grained proxy or double-check for head marking which is grammatically independent of other variables. Head/dependent marking is not fine-grained: the only values for the variable are head, dependent, and neither, and the number of phrase types in which it can be surveyed cross-linguistically is small. It is designed to capture a simple bifurcation of structural patterns independent of the grammatical categories involved, while the finer-grained inflectional person involves various categories and various structural patterns.

All of these correlations involve only the languages surveyed here, though a number of them support correlations found with different samples elsewhere in the literature. In general, though individual variables differ considerably in frequency from continent to continent and language family to language family, correlations between these variables do

²⁰ For the breakdown into continents see §3.3.

²¹ All tests are binary and use Fisher's Exact Test (two-tailed, since, there was no advance expectation of either a positive or a negative correlation). For word order, each correlation was tested on three groupings of verb-initial vs. other, verb-final vs. other, and verb-medial vs. other; if any was significant that is reported. Similarly, head/dependent marking was tested on head vs. other, dependent vs. other, and double (head + dependent) vs. other. For these two variables, then, there was some amount of significance hunting (i.e. cherry-picking). For hierarchical patterns, absence of any mention in a grammar is counted as absence of the phenomenon; thus $N = 256$ when there were only 48 actual entries (37 Yes, 11 explicit No).

not seem to differ greatly from continent to continent. One that may be of note is a positive correlation of inflectional person with high complexity, only in northern Eurasia ($N=40$, $p \sim 0.11$ [Fisher 1-tailed]), not significant but borderline and interesting in such a small sample.

3.3. Linguistic geography

What is especially striking about inflectional person lies is its worldwide geographical distribution. Figure 2 plots longitude against inflectional person. In these longitudinal plots, western Europe and western Africa are at the left and eastern North and South America are at the right (northern and southern languages are intermingled, as the plot does not distinguish latitude). There is a worldwide slope, indicated by the trendline, and a fairly strong positive correlation, indicated by the R^2 value in the lower right corner (which measures the extent to which the individual points diverge from the trendline).²² The geography is revealed more clearly by breaking the world language population down into three latitudinal bands (Figures 3-5): northern higher latitudes (down to 40° N), northern mid latitudes (below 40° N), and southern continents (Africa, New Guinea-Australia-Oceania, and South America, all three included in their entirety though they all extend north of the equator). There is a very strong positive correlation between inflectional person and longitude in the northern higher latitudes (Figure 3) and a moderately strong one in the mid latitudes (Figure 4), but no correlation at all in the southern lands (Figure 5). The northern high-latitude one is the strongest large-scale clinal correlation I am aware of in linguistic geography.

Figures 2-5 about here

These distributions show that there is essential typological continuity across the entire Northern Hemisphere, especially its higher latitudes. The effect extends into the lower northern latitudes, partly because 40° N is an arbitrary cutoff point and partly because some originally northern languages have expanded southward (examples are Germanic, Indo-Iranian, Turkic, and Mongolic). Continents to the south that have not had northern influence do not show the longitudinal cline.²³

²² The slope of the trendline indicates something about the extent and intensity of the correlation – the difference between the low and high ends – but not its strength. A correlation can have a steep trendline slope but still be weak in the sense that many languages are far from the trendline. Fig. 7, discussed below, is somewhat like this, though in general it does happen that in the plots here the steeper slopes are those of the stronger correlations. Strength measures how close the languages are to the trendline, whatever their position along it. For these plots a correlation can be considered strong if it is above 30%; moderate if around 15-20%, perceptible if around 10%, and weak if much below 10%; anything below about 5% counts as no correlation. Note that these measures of strength and steepness describe tendencies in the sample population but are not measures of statistical significance (they do not measure the extent to which the sample population differs from what is expected).

²³ Post-1492 colonial influence is excluded from this claim, but in any case none is evident in this survey. Colonial languages have tended to cause extinction more than durable typological change.

No comparable correlation holds for latitude. Worldwide, there is no correlation between inflectional person and latitude (Figure 6, with a nearly level trendline and a very weak R^2 value). I separated the languages into three longitudinal bands: Africa-Europe, Asia-Australasia, and the Americas. There is a weak negative correlation in the west (Africa-Europe) and a weak positive correlation in the east (the Americas) (Figures 7-9). This must be the effect of the high-latitude east-west cline (low in Europe, high in North America) on the three bands.

Figures 6-9 about here

The overall situation, combining longitude and latitude, points to long-term macroareal continuity across the higher northern latitudes. The major language-family spreads in these latitudes echo the larger picture. Uralic, an old family with a large east-west spread in the far north of Eurasia, is conservative in its eastern reaches and has high inflectional person there. Uralic has spread mostly from east to west, and has evidently undergone Europeanization of its person indexation in the west: loss of person indexation in NP's, loss of inflectional person plus generic root structure of independent pronouns, same or similar declension of pronouns and nouns (these are items 2, 3, 4, 9, and 13-17 of the survey).²⁴ This is the only high-latitude language spread that has transgressed major longitudinal bands and remained in that latitudinal range. Indo-European originated in western Eurasia and historically spread far to the west and east, largely keeping its person type intact except in and near the lower latitudes (Indic, Armenian, Romance, Balkan languages); its easternmost expansion in the north, reaching the Mongolian steppe and the Altai region of Siberia in the form of early eastern Iranian and Tocharian, has not left modern survivors. More recent family spreads (Turkic, Mongolic, Athabaskan, Algonquian) have mostly remained within longitudinal ranges and appear to be consistent as to type, perhaps because they have not had time to change greatly, and/or perhaps because their spreads have not transgressed major typological boundaries, remaining mostly within eastern Eurasia or North America. (Athabaskan and Algonquian have not been sampled densely for this survey). The Dene-Yeniseian spread (Vajda 2010, other chapters in Kari & Potter eds. 2010), whatever its nature (spread of a family or diffusion of a verb morphology template), remains mostly in the Pacific Rim population and is true to that type; the westernmost representative, modern Ket, diverges somewhat with 62% inflectional person, lower (and closer to the western type) than Na-Dene to the east in North America (in this sample, Hupa at 91% and Slave at 90%).

For inflectional person the continent-by-continent mean frequencies reflect this same distribution. They are shown in Table 2.²⁵ Though not monotonic, there is a rise

²⁴ The Proto-Uralic homeland is still debated, but must have been not far from the Ural Mountains, either somewhat east (e.g. Janhunen 2009) or somewhat west (Kallio 2015) and north of the steppe. For the typology of Uralic pronouns see Kulonen 2001, Nichols 2013.

²⁵ Recall that in the language sample northern Eurasia (here, most of Western Eurasia and all of Northeastern Eurasia) is overrepresented and New Guinea and South America underrepresented, relative to the numbers of stocks in each continent. As long as there are enough languages per continent to guarantee reasonable accuracy of description, the imbalances have no impact on the findings here since the comparisons mostly involve continental totals or large-scale clines that do not depend on the numbers of languages per area (provided coverage gives reasonable accuracy). To double-check I drew a sparser

overall from lower figures in the west (Africa, Europe) to higher ones in the east (the Americas).

Table 2. Mean values of inflectional person per continent, using the macroareal breakdown of Nichols et al. 2013. Inflectional person is expressed as a proportion. * = One standard deviation or more from the overall mean: Western Eurasia is low; North America to Central America is high. (Western North America is actually not quite one standard deviation high; 0.81 would be required.) New Guinea is fairly high; all other areas are close to the mean.

<i>Continent</i>	<i>N</i>	<i>Mean % Inflectional</i>
Africa	22	0.65
W Eurasia	40	0.50 *
NE Eurasia	29	0.71
S & SE Asia	15	0.65
New Guinea & Oceania	28	0.77
Australia	16	0.69
W. North America	27	0.80 *
E. North America	20	0.87 *
Mexico-Central America	10	0.86 *
South America	36	0.71
Total and mean	243	0.70 (s.d. 0.11)

Figure 10 shows higher and lower proportions of inflectional person for the whole sample, plotted on a map. The higher incidence of low-inflectional languages (white symbols) in Africa and western Eurasia is quite visible.

Figure 10 about here

Causativization as a realization of the causative alternation shows a weaker version of the high-latitude east-west cline (Figures 11-14). These figures are for the nine pairs of verbs with prototypically animate S/O (Nichols, Peterson, Barnes 2004), which give the best historical typological profile; and for only the 123 languages surveyed here in the sample for which there is also data on the causative alternation. There is a perceptible east-west cline worldwide, with causativization less prevalent in the west than in the east. It is most evident in the northern high latitudes, where there is a moderately strong correlation; there is a perceptible correlation in the lower northern latitudes, and none at all in the southern continents. Causativization has its peak frequency among languages of several different families and local areas around the North Pacific Rim and languages that have spread from there (Fortescue 1998). High causativization is also found in several

sample of north Eurasian languages and compared the overall sample mean and standard deviation and the continental frequencies; the figures were identical or nearly so.

unconnected contact zones worldwide (the central Caucasus, Southeast Asia, Oceania), where it is apparently favored in certain sociolinguistic contexts (Nichols 2011b, Nichols & Bentz in press, Nichols 2017a).

Figures 11-14 about here

A similar distribution appears to be exhibited by verb-based vs. noun-based lexical types. This variable has so far been surveyed for only about 60 languages, but the verb-based type appears to be most frequent in northern Asia and North America. Another similar distribution is exhibited by basic event structure (Nichols 2017b), which has to do with whether predicates that include a state in their event structure (such as 'angry', 'asleep', 'red') are lexicalized, in their most basic or underlying grammatical form, as states, durative activities, etc. ('is angry', 'is sleeping') or as bounded or transition-based (telic, ingressive, etc.: 'got angry', 'fell asleep'). The transition-based type appears to peak in eastern Asia and North America. So far only about 86 languages have been surveyed.

All four of these variables (inflectional person, causativization, verb-based, transition-based) have finely subdivided or gradual ranges of types because all are based on proportions of values surveyed across a number of lexical items. It could be that many typological variables, if described as continua or finely divided scales, would exhibit a similar distribution; or it could be that these are four flukes and the overall worldwide distribution of typological variables is random. Figure 15 suggests that neither of these is the case; some variables have the clinal distribution and some do not. It plots the continental means for inflectional person (the same information as in Table 2) and four other variables that lend themselves to this kind of averaging. These are different from previous plots in that they are continental averages instead of geographical plots of individual languages, and the ordering of the continents from east to west schematizes the actual longitudinal locations. Where there is a clear slope and a high R^2 value there is an east-west gradient. The plots on the left show all continents; those on the right show only Eurasia and the Americas, which are at issue for the northern high-latitude clines.

Figure 15 about here

With this kind of grouping, only inflectional person and inclusive/exclusive distinctions show appreciable large-scale clines. Those two are strengthened by removing Africa, Australia, and New Guinea, while the others remain flat. The correlation for inflectional person is very strong either way. That for inclusive/exclusive is considerably stronger without the other three continents, mostly the effect of removing Australia, which has an atypically high frequency of the inclusive/exclusive opposition (it is visible as the high outlier in the middle of the graph on the left). Of the others, as was noted above, the coverage of the verb-based lexical type is too thin to give much confidence in the results.

One conclusion to be drawn is that the east-west cline for inflectional person is very strong and robust across various groupings and breakdowns, while the others are evident only under some conditions.

How to explain this geography? The worldwide high-latitude continuum cannot be explained by the linguistic and cultural impact of Roman or post-Roman Europe, colonial Europe, China, Mongolia, or the Central or South American empires. No empire had such a

span, the loanwords that would ordinarily accompany such an expansion are lacking, and no genetic or archaeological evidence suggests pre-colonial linguistic interaction at this scale. Furthermore, where we have evidence the typological distributions appear to be old. High proportions of causativization and inflectional person – an eastern profile – appear to be reconstructable for Proto-Uralic (over 4000 years old and perhaps as old as 6000 years), which originated in the eastern part of its present range; low causativization and low inflectional person – a western profile – characterized Proto-Indo-European, which originated on the western steppe. Both homelands are above 40° and reflect more eastern and western positions on the cline. Ancestral Turkic and Tungusic were evidently causativizing languages with high inflectional person. This is an eastern profile consistent with their homelands near the eastern steppe, and both are of Romance-like age or older. Their typological profiles are not obviously Chinese-influenced and they antedate imperial Mongolia by over a millennium. High inflectional person appears to be ancestral in several of the older North American families (e.g. Athabaskan and Algonquian). Though this picture is sketchy, it suggests that the present typological profile of the north Pacific Rim area is at least a few millennia old and not due to any identifiable historical development. For western Eurasia little can be said except that the typological profile of Indo-European is largely western, is 6000 years old, and has erased the pre-IE profile of western Eurasia.

One interpretation of the linguistic geography might then be that the far western low frequencies and eastern high frequencies reflect two prehistoric flukes: a disproportionate impact of a local type in early Europe and another in the vicinity of the north Pacific Rim or in North America. The western fluke could then have been the spread of Indo-European and the well-known millennia-long impact of Indo-European languages on their neighbors. The eastern fluke could have been some episode in the settlement history of the Americas that happened to give disproportionate demographic importance to one linguistic immigrant population.²⁶ Another possible interpretation is that the influential western fluke was not early Indo-European but the language of the first farmers to spread across Europe (there does appear to have been one language or one language family whose vocabulary was influential across the early Indo-European branches and Basque, and therefore probably across most of greater Europe: see Iverson & Kroonen 2017). The time frame for this development begins about 8000 years ago. Whether pre-agricultural Europe had any appreciable linguistic areality or linguistic influence is unknown but seems implausible, so for the western pole a long-range diffusion pattern is unlikely to be older than 8000 years. A global cline could conceivably have arisen from these two typologically different and geographically distant flukes, given enough time, enough local contact episodes, and gradual diffusion of structural patterns, which could eventually have formed a continuum between the poles. A weak point in any such account is that it is not obvious why the north Pacific Rim area and nearby should have been an influential center of diffusion or population expansion. One would expect any such linguistic diffusion pattern to be

²⁶ The relevant population, on this scenario, is unlikely to have been the very first entrant, so these are not founder effects for the whole hemisphere. Had it been the first entry we might expect to see the high pole not in high-latitude North America but in South America, with the cline continuing southward from North America. Or, alternatively, levels might be equally high throughout the Americas. Neither of these is the case; for all four variables frequencies are lower in South America than in North America.

accompanied by gene spread, so genetics is a possible source of support or counterevidence for the two-pole analysis.

Another interpretation is that the eastern center of linguistic diffusion and/or demographic expansion was not in northern Asia or North America but in Southeast Asia. The north Pacific Rim is the far periphery of this diffusion zone and retains an archaic typological profile as peripheries of spreads are expected to do; meanwhile, beginning at least with the development of paddy rice farming and statehood in Southeast Asia, the linguistic type of the centers of demographic expansion and grammatical diffusion changed, leaving the northern backwater to reflect the earlier type. (The eastern Himalayan highlands, which are isolated from the lowlands in various ways and represent a periphery to local and regional spreads, also appear to retain a typological profile closer to the north Pacific Rim type. They are undersampled here, however.) This account of prehistory finds support in the diffusion of genes from Southeast Asia northward and then westward across Eurasia, a process that began very long ago (Rootsi et al. 2007).

A third interpretation is that the cline is just one manifestation of the grammatical differences between the Pacific Rim linguistic population and the mainland Eurasian one (Bickel & Nichols 2006, Nichols et al. 2013). Why lexical/inflectional person and the correlated variables should have assumed a global distribution is not clear on this account. Perhaps it is just that they are stable enough and conspicuous enough to have remained detectable longer, and at greater distances, than others. Alternatively, it may be that composite definitions of variables are ideally designed to pick up and unify under a single label values of variables whose unity might otherwise escape detection.²⁷

The hypothesis of long-term gradual diffusion of linguistic variables can be tested on the linguistic side by typological comparison, though the work involved will be intricate and complex. If we can determine, even approximately, the likelihood of a spread of person marking to some new context, or a typological shift to or from (say) causativization or decausativization in this or that lexicosemantic verb class, as well as which direction of change is more likely (e.g. addition or loss of person marking, innovation or loss of causativization), and if we can also determine whether and how the other typological properties of the language favor or disfavor these changes, how often and readily they happen, and how easily they spread from language to language in what typological contexts, etc., and likewise for other variables, we should be able to determine how plausible a standing long-range diffusion is. The task, in short, is feasible though large.

All of these possibilities show that typology can raise hypotheses and shed light on historical and prehistoric linguistic population formation, but that much work remains to be done. The first question should be handled is whether the global cline is really that or the coincidental result of diffusion with isolation or drop-off by distance from two centers that happen to have opposite structural types but have no historical continuity. The more promising research hypothesis and the more interesting interpretation is that there is continuity between the two poles and the entire high-latitude northern hemisphere represents a single linguistic population.

3.4. Further applications

²⁷ A broader or automatically done survey might carry the risk of exaggerating minor points or seeing spurious resemblances, but this one was human-analyzed and limited to actual person markers, probably eliminating such risks.

Apart from its intriguing geography, inflectional person also offers some points of interest to the technical side of typological analysis. In informal statements it can capture something of the spirit of a language: describing person as mostly lexical in Russian or Czech but mostly inflectional in Hungarian or Kabardian says something informative about the language overall. Though this point has not been pursued here, it can also represent the type of an entire family. For example, Afroasiatic languages tend to have fairly balanced lexical and inflectional person, Indo-European languages somewhat higher lexical person, Uralic higher inflectional, Pama-Nyungan languages almost entirely lexical, and Athabaskan, Algonquian, and Salish high inflectional. Local areal patterns can be overlaid on these inherited preferences: consistent with the east-west cline, eastern Uralic languages and eastern Iranian (Indo-European) languages tend to have higher inflectional person than their more westerly sisters: in Appendix 3 compare Finnish and Estonian with Khanty and Selkup, or Waigali with Russian or Lithuanian. (The eastern pattern is conservative in Uralic and innovative in Indo-European.) If, as these examples suggest, the treatment of person within families can serve as an indicator of contact, it may help identify otherwise unsuspected contact episodes, and knowing something about rates of change in person types can tell us something about the antiquity or intensity of contact. Meanwhile, types appear to remain fairly stable within families despite local perturbances.

Areal typology is always in need of more variables that are independent of each other and can contribute to measuring typological distances between areas and determining whether different areas belong to the same larger population or not, whether principles such as isolation by distance obtain, or to what extent adjacent languages form an area. Is Southeast Asia typologically closer to New Guinea-Oceania or to mainland Asia? Is there a demonstrably distinct Pacific Rim linguistic population? Do Australia and New Guinea compose a single linguistic population? Is the northern east-west cline a matter of isolation by distance from two centers? Is the Caucasus a linguistic area? The more variables that can enter into such calculations, the firmer the answers to such questions, and since lexical and inflectional person are independent from each other and from the others surveyed here they can contribute two more variables to the mix.

Lexical and inflectional person, separately or in total, can also contribute to the measurement of linguistic complexity. Both of them and their sum are independent of inventory complexity as calculated here (see Table 1; additional plots not shown here), and since that measure includes nothing on pronouns and very little on person, bringing person in would expand the coverage of complexity measures and increase their overall accuracy.

4. Discussion and conclusions

I hope to have shown that analyzing person as relatively lexical vs. relatively inflectional is a useful addition to the linguistic toolkit which, since it is internally complex and can be expressed as a proportion or percentage, can capture fine-grained incremental differences and large-scale geographical distributions. It is independent of most other variables and can be employed in existing mass comparative measures such as population distance measures and measures of complexity. It holds promise for measuring rates and directions of change and thereby potentially for establishing past contacts, their time

frames, and perhaps even something about the sociolinguistics of contact episodes. Person by itself cannot tell us whether the northern east-west cline is a single phenomenon or the accidental result of isolation by distance from two typologically distant centers, but it has revealed its existence and extent more clearly than any of the several other variables with similar distributions.

A technical question concerns the nature of a composite category like person as defined here, which consists of many different datapoints some of which form covariant clusters each of which is fairly discrete from the others. The survey items have been designed to be complementary so as to capture different facets of inflectional vs. lexical behavior of person. Some of the items are incompatible (e.g. [9] Generic pronoun base and [10] Unique root per person or person-number; or [9] Generic pronoun base and [11] Roots or stems rhyme, alliterate, etc.: for both sets, a language could have one or the other or neither but not both) and not all are present in all languages, so a systematic determination of harmonic correlations among the 42 subvariables will probably require a considerably larger sample. I tested for correlations between percentages of inflectional person within sets of related subvariables such as [1-8] (having to do with the locus of person marking), [9-21] (morphological structure and behavior), and [22-37] (categories and syntax): they proved to be extremely strong, worldwide and for smaller areas, indicating that the subvariables have some real grammatical unity. This means that inflectional/lexical person can be described as a superordinate variable or macrovariable subsuming a number of harmonic and implicationally linked subvariables.

Other typologically recognized phenomena have this sort of nature. Word order, to take a clear case, consists of individual patterns such as AOV, OAV, AVO, etc. which generalize to verb-initial, verb-medial, and verb-final or to AV vs. VA and OV vs. VO, which are mutually dependent (negatively correlated), and these form clusters such as clause word order, NP word order, PP word order, and others which may or may not form harmonic correlations. Alignment, in early work, was seen as a variable with a few values (accusative, ergative, neutral, three-way, split, etc.) separately ascertained for noun vs. verb inflection and main vs. subordinate clauses; but as other configurations, more splits, and more cases of hierarchically-driven splits were discovered it has evolved into a macrovariable subsuming a large number of variables that refer to different lexicosemantic classes of predicates, various syntactic constructions, and various morphological paradigms, per language (Witzlack-Makarevich 2010, Witzlack-Makarevich et al. 2016). We continue to speak of languages in approximate terms as head-final, ergative, etc. but the real descriptive and typological work goes on at the microlevels. If lexical vs. inflectional person proves more widely useful in linguistics it is likely to evolve in this direction, but meanwhile as a newly proposed distinction it is useful in its general and approximate form. Meanwhile, since the different specific properties all have different distributions across languages, the lumped general types make it possible to compare languages across a single scale and reach statistical significance with 270 languages, when many more would be required to cover the individual survey items evenly.

Finally, inflectional person is a convenient descriptive notion that can capture something deep-seated in the design of a language. It reflects one of the most striking differences between radically head-marking and radically dependent-marking languages – the appearance of person on many vs. few parts of speech – as well as the motivation for hierarchical and inverse patterning, and why those are apparently always based on person

and never on gender. It accounts for why polysynthetic and radically head-marking languages are likely to have pronouns with generic bases and auto-person. It may explain some of the distribution of inclusive/exclusive, egophoricity, and types of evidentiality, all of which have to do with the pragmatics of locutor relations.

I will close by pointing out some examples of two kinds of cases of historical interest involving person. One is contact situations where languages with marked differences of lexical vs. inflectional person abut against each other. An example is Australia, where most languages have absolutely low to rather low levels of lexical person but northern Australian languages of several families (Southern Daly, Wororan, Nyulnyulan, Iwaidjan, Gunwingguan in this sample) have high levels of inflectional person while languages of the widespread Pama-Nyungan family have relatively low inflectional person. There are various local contacts all along this frontier. Another is northern California, where languages with low levels of both kinds of person such as Yokuts (Yokuts-Utian), Wappo (Yuki-Wappo), and Pomoan languages are neighbors of high-inflectional languages such as Hupa (Athabaskan), Yurok (Algic), Miwokan languages, and Wintu (Wintun). Another is Japan, where low-inflectional and moderately low-lexical Japanese meets high-inflectional and zero-lexical Ainu. Another is Southeast Asia, where the isolating type, low in person overall but with more of the few person points at the lexical end (Cambodian, Thai, Mandarin), meets the high-inflectional type of the Himalayan foothills. Another is the Caucasus, which is divided between high-inflectional West Caucasian languages (Kabardian and Abkhaz in this sample) and low-inflectional Nakh-Daghestanian languages, buffered by the more balanced Kartvelian family (Georgian, Svan, Laz) in the south and Ossetic (Iranian) in the north. At the western periphery of the Eurasian steppe, high-inflectional languages of the Uralic and Turkic families have met low-inflectional Indo-European and Nakh-Daghestanian languages. All of these are promising arenas for investigating the behavior of person in different grammatical and sociolinguistic contexts.

Another situation of interest is where language families show differential evolution under apparent or possible contact influence. One such is Uto-Aztecan, with high-inflectional languages in the south (Pipil, Cora, and others in Central America) and low-inflectional Numic languages (Tümpisa Shoshone) in the north (the Great Basin). Another is Uralic, fairly high inflectional overall but less so in the west where there has been contact with Indo-European (Baltic, Germanic) for some time. Another is the Balkan Sprachbund, where languages from different Indo-European branches have increased their levels of inflectional person by adding clitic person indexation of objects, and have increased the text frequency of inflectional person by replacing inherited nonfinite verb forms with finite (person-conjugated) verbs plus subordinating conjunctions. All of these situations offer opportunities to observe how and at what rates and in what directions person marking evolves and adapts to adjust to new sociolinguistic and grammatical contexts.

Correspondence address:

Johanna Nichols
University of California, Berkeley
Slavic Languages 2979, 6303 Dwinelle
Berkeley, CA 94720, USA

johanna@berkeley.edu

*[If there is any affiliation other than the correspondence address, it should read:
University of California, Berkeley and Linguistic Convergence Laboratory, Higher
School of Economics, Moscow.]*

Acknowledgments:

Supported in part by a grant from the Russian Academic Excellence Project 5-100 to the Higher School of Economics, Moscow. I thank audiences at the Max Planck Institute for Evolutionary Anthropology, Leipzig, the LSA Linguistics Institute (University of Chicago, 2015), the University of Helsinki, and the University of Amsterdam for comments and questions on early versions of this work. Three anonymous reviewers helped greatly to improve the final version. Special thanks to Frans Plank for his patience with this paper, and more importantly for founding, building up, and maintaining *LT* as the venue of choice for a paper such as this one. And for writing Plank 1985, the intellectual cornerstone for work on person.

Appendix 1. The 42 survey items in questionnaire form. * = lexical property; the others are inflectional.

- [1] Person of A indexed on verb.
- [2] Person of O indexed on verb.
- [3] Person of possessor indexed on possessum (head of possessive NP).
- [4] Auto-person: The person of an independent pronoun is indexed on the pronoun itself, as a separate morpheme from the root.
- [5] Person of the possessor of an argument indexed on the verb.
- [6] Person of external possessor indexed on the verb.
- [7] Possessive relative: The person of a gapped subject of a relative clause is indexed on the head noun.
- [8] Person is marked on or attracted to the negative morpheme, which is separate from the lexical verb.
- [9] Generic pronoun base: the independent pronouns share the same root, which does not contain person as part of its meaning.
- [10]* Unique root per person or person-number, with person as a lexical property of the root.
- [11] Roots or stems rhyme, alliterate, etc. along person or number lines.
- [12]* There is no closed class of pronouns.
- [13]* Pronouns have the same cases (i.e. case categories) as nouns.
- [14]* Pronouns have the same case morphology as nouns.
- [15]* Pronouns have the same root or stem flexibility as nouns.
- [16] Pronouns have different number categories from nouns.
- [17]* Pronouns have the same number markers as nouns.
- [18] Any multiple marking of person per argument.
- [19] A and O markers are formally identical.
- [20] Person is the outermost verb inflection.
- [21] Person is the outermost noun inflection.
- [22] Inclusive/exclusive distinguished in independent pronouns
- [23] Inclusive/exclusive distinguished in verb person indexes
- [24] Inclusive/exclusive distinguished in noun possessive indexes
- [25] Inclusive/exclusive distinguished in auto-person on pronouns
- [26] Person and number discrete or factored out in independent pronouns
- [27a] Person and number discrete or factored out in verb person index for A
- [27b] Person and number discrete or factored out in verb person index for O
- [28] Person and number discrete or factored out in noun possessive indexes
- [29] Person and number discrete or factored out in auto-person on pronouns
- [30]* Noun genders distinguished in independent pronouns
- [31]* Noun genders distinguished in verb person index for A and/or O
- [32]* Noun genders distinguished in 3rd person independent pronouns
- [33]* Noun genders distinguished in 3rd person verb index for A and/or O
- [34]* Inherent (lexical) gender in pronominals
- [35]* Natural gender (not noun genders) distinguished in independent pronouns
- [36]* Natural gender (not noun genders) distinguished in 3rd person independent pronouns

- [37]* Natural gender (not noun genders) distinguished in verb person index for A and/or O
- [38] Verb person index(es) cumulative with TAM
- [39] Person involved in hierarchical marking on verbs
- [40] Person involved in inverse marking on verbs
- [41] Conjunct/disjunct marking on verbs
- [42] Person determines access to (hierarchical, promiscuous) number marking on verbs

Appendix 2. Survey item responses from 4 sample languages. Somali has both lexical and inflectional person, Russian mostly lexical, Mandarin little of either, and Cree mostly inflectional.

	Somali	Russian	Mandarin	Cree
1	1	1	0	1
2	1	0	0	1
3	1	0	0	1
4	0	0	0	1
5	nd	0	0	nd
6	nd	0	0	nd
7	0	0	0	0
8	0.5	0	0	0
9	0	0	0	1
10	1	1	1	0
11	0	0	0	1
12	0	0	0	0
13	1	1	0	0.5
14	1	0	0	0
15	nd	0	0	n/a
16	0	0	0.5	0
17	0	0.5	1	0
18	nd	0	0	nd
19	0	0	0	0
20	1	1	0	1
21	1	0	0	1
22	1	0	0	1
23	1	0	0	1
24	1	0	0	1
25	n/a	n/a	n/a	1
26	0	0	1	1
27	0.5	0	0	1
28	0	0	0	1
29	n/a	0	0	1
30	n/a	n/a	n/a	0
31	0	1	0	0
32	0	1	0	0
33	1	0	0	0
34	1	0	0	0
35	0	0	0	0
36	0	1	0	0
37	0	1	0	0
38	1	1	0	0
39	0	0	0	1
40	0	0	0	1
41	0	0	0	0
42	0	0	0	n/a

Appendix 3. Languages surveyed, by continent. Lex, Infl = total points for lexical and inflectional properties. %Infl = the percent of the sum of Lex and Infl that is inflectional (calculated only where the total is high enough to make percentages meaningful).

Continent	Language	Stock	Lex	Infl	%Infl
Africa	Tamazight	AA: Berber	5.0	6.5	0.57
	Arabic (MSA)	AA: Semitic	5.0	9.0	0.64
	Dongolese Nubian	Nubian	3.0	6.0	0.67
	Koyra Chiini	Songhai	2.0	1.0	
	Jamsay	Dogon	1.0	1.5	
	Mandinka	Mande	1.0	0.0	
	Kunama	Kunama	3.0	10.0	0.77
	Fula	North Atlantic	2.0	13.0	0.87
	Fur	Fur	3.0	1.5	0.33
	Hausa	AA: Chadic: Hausa	7.0	9.0	0.56
	Kanuri	Saharan	3.0	5.5	0.65
	Gumuz	Gumuz	3.0	7.0	0.70
	Uduk	Koman	2.0	6.0	0.75
	Goemai	AA:Chadic: Angas-Gerka	2.0	1.0	
	Yoruba	Benue-Congo	1.0	1.0	
	Ewe	Kwa	2.0	6.0	0.75
	Maale	Ta-Ne-Omotiic	2.0	0.0	
	Haro	Ta-Ne Omotiic	5.0	4.0	0.44
	Gbeya	Adamawa-Ubangi	1.0	4.0	0.80
	Ik	Kuliak	2.5	4.5	0.64
	Logbara	Central Sudanic: Moru-Madi	0.0	2.0	
	Turkana	E. Nilotic	4.0	2.0	0.33
	Somali	AA: Cushitic	5.0	10.0	0.67
	Lango	W. Nilotic	1.0	10.5	0.91
	Ngiti	E. Central Sudanic	1.5	9.0	0.86
	Luganda	Benue-Congo	3.0	6.0	0.67
	Dahalo	AA: Cushitic	5.0	4.5	0.47
	Sandawe	Sandawe	3.0	7.0	0.70
	!Kung	Ju	1.0	1.0	
	Nama	Kwadi-Khoe	7.0	5.0	0.42
	Taa (!Xóǀ)	Tuu (!Ui-Taa)	3.0	4.5	0.60
	Europe	English	IE: Germanic	2.0	2.0
Irish		IE: Celtic	2.5	4.5	0.64
French		IE: Italic	4.0	4.5	0.53
Italian		IE: Italic	4.0	3.5	0.47
Catalan		IE: Italic	3.0	8.0	0.73

German	IE: Germanic	5.0	3.0	0.38	
Norwegian	IE: Germanic	2.0	1.0	0.20	
Swedish	IE: Germanic	2.0	0.5	0.20	
Slovene	IE:Balto-Slavic	5.0	3.5	0.41	
BCS (Std.)	IE:Balto-Slavic	5.0	3.5	0.41	
Bulgarian	IE:Balto-Slavic	5.0	5.0	0.50	
Macedonian	IE:Balto-Slavic	5.0	5.0	0.50	
Romanian	IE: Italic	3.0	4.0	0.57	
Greek	IE: Greek	4.0	6.0	0.60	
Albanian	IE:Albanian	5.5	5.5	0.50	
Basque	Basque	6.0	5.5	0.48	
Ingush	ND: Nakh	6.0	3.0	0.33	
Avar	ND: Daghestanian	3.5	2.5	0.42	
Godoberi	ND: Daghestanian	5.5	2.0	0.27	
Botlikh	ND: Daghestanian	3.0	1.0	0.25	
Hunzib	ND: Daghestanian	3.5	0.0	0.00	
Lak	ND: Daghestanian	3.5	4.0	0.53	
Akusha/standard	ND: Daghestanian	3.0	6.5	0.68	
Icari	ND: Daghestanian	4.5	7.5	0.63	
Lezgi	ND: Daghestanian	2.0	1.0	0.33	
Tsakhur	ND: Daghestanian	3.0	1.0	0.25	
Kryz	ND: Daghestanian	2.0	2.0	0.50	
Archi	ND: Daghestanian	3.5	1.5	0.30	
Udi	ND: Daghestanian	2.0	1.0	0.33	
Xinalug	ND: Daghestanian	3.5	1.0	0.22	
Kabardian	WC	1.5	8.0	0.84	
Abkhaz	WC	3.5	8.5	0.71	
Georgian	Kartvelian	2.0	4.5	0.69	
Svan	Kartvelian	2.0	6.5	0.76	
Laz (Pazar)	Kartvelian	2.0	5.0	0.71	
Ossetic	IE: Iranian	3.0	4.0	0.57	
Persian	IE: Iranian	1.0	6.0	0.86	
E. Armenian	IE: Armenian	3.0	6.5	0.68	
Turkish	Turkic	3.0	8.0	0.73	
Hungarian	Uralic: Fi-U	2.0	5.0	0.71	
Lithuanian	IE:Balto-Slavic	5.0	3.0	0.38	
Russian	IE:Balto-Slavic	7.5	3.0	0.29	
Saami (generic)	Uralic: Fi-U	3.0	8.5	0.74	
Finnish	Uralic: Fi-U	3.0	7.0	0.70	
Estonian	Uralic: Fi-U	3.0	7.0	0.57	
N-C Asia	Erzja	Uralic: Fi-U	2.0	12.5	0.86
	Mari	Uralic: Fi-U	2.0	8.0	0.80

	Udmurt	Uralic: Fi-U			
	Komi	Uralic: Fi-U	3.0	9.0	0.75
	Mansi	Uralic: Fi-U	3.0	4.0	0.57
	Khanty (N)	Uralic: Fi-U	1.5	6.0	0.80
	Nganasan	Uralic: Samoyedic	2.0	9.0	0.82
	Tundra Nenets	Uralic: Samoyedic	2.5	7.0	0.74
	Selkup	Uralic: Samoyedic	1.5	8.5	0.85
	Ket	Yeniseian	5.0	8.0	0.62
	Yakut	Turkic	2.5	7.5	0.75
	Mongolian	Mongolic	2.0	8.0	0.80
	Manchu	Tungusic	3.0	3.0	0.50
	Nanai	Tungusic	3.0	8.5	0.74
	Evenki	Tungusic	2.5	14.0	0.85
	Even	Tungusic	2.5	14.0	0.85
	Yukagir	Yukagir	2.5	5.0	0.67
	Chukchi	Ch-Kam	3.5	8.0	0.70
	Itelmen	Ch-Kam	1.5	5.5	0.79
	Sib. Yup'ik	Eskimo-Aleut	2.5	8.5	0.77
	Aleut	Eskimo-Aleut	0.0	9.0	1.00
	Ainu	Ainu	0.0	12.5	1.00
	Nivkh	Nivkh	3.5	11.5	0.77
	Japanese	Japanese	3.5	1.0	0.22
S-SE Asia	Waigali	IE: Indo-Iranian	2.0	5.0	0.71
	Burushaski	Burushaski	5.0	7.0	0.58
	Palula	IE: I-I: "Dardic"	2.0	1.5	0.43
	Mandarin	ST: Sinitic	2.0	1.5	
	Hindi	IE: Indo-Iranian	1.5	0.0	0.00
	Hakha Lai	ST: Kuki-Chin	2.0	13.0	0.87
	Kharia	Austroasiatic: Munda	3.0	8.0	0.73
	Lhasa Tibetan	ST: Bodish	5.0	2.0	0.29
	Belhare	ST: Kiranti	4.0	11.0	0.73
	Brahui	Dravidian	2.0	3.5	0.64
	Paiwan	Austronesian	1.0	4.0	0.80
	Thai	Tai-Kadai	2.0	0.5	
	Tagalog	Austronesian	2.0	2.0	0.50
	Great Andamanese	Great Andamanese	1.0	8.0	0.89
	Cambodian	Austroasiatic	1.5	1.0	
	Kolami	Dravidian	5.0	4.0	0.44
	Acehnese	Austronesian	3.0	11.0	0.79
	Semelai	Austroasiatic	3.0	5.5	0.65
	Car Nicobarese	Austroasiatic	2.0	6.0	0.75

New Guinea	Maybrat	West Papuan	3.0	5.0	0.63
	Warembori	AN: Lower Mamberamo	1.0	10.5	0.91
	Meyah	East Bird's Head	1.0	9.0	0.90
	Abun	Abun	2.0	0.0	
	Hatam	Hatam	1.0	7.0	0.88
	Inanwatan/Suabo	Inanwatan	2.0	9.0	0.82
	Sentani	Sentani	1.0	5.5	0.85
	Kuot	Kuot	3.0	9.0	0.75
	Barupu	Macro-Skou	5.0	7.0	0.58
	Imonda	Border	3.5	2.0	0.36
	Dani	Dani	1.0	9.0	0.90
	Mali	Baining	5.0	7.5	0.60
	Alamblak	Sepik	3.0	5.5	0.65
	Yimas	Lower Sepik-Ramu	2.0	4.0	0.67
	Usan	Adelbert Range	1.0	7.0	0.88
	Mian	Macro-Ok	5.5	6.0	0.52
	Kobon	ENGH: Kalam	1.5	4.5	0.75
	Amele	Madang	2.0	5.5	0.73
	Tauya	Madang-Adelbert: Brahman	3.0	10.0	0.77
	Kombai	Macro-Ok	1.0	9.0	0.90
	Asmat	Macro-Ok	1.0	3.5	0.78
	Salt-Yui	Chimbu-Wahgi	2.0	6.0	0.75
	Hua	Eastern Highlands	2.0	6.5	0.76
	Kewa	Engan-Kewa	3.0	5.5	0.65
	Motuna	South Bougainville	4.0	12.0	0.75
	Kiwai	Kiwaian	1.0	6.5	0.87
	Abui	Timor-Alor-Pantar	1.0	13.0	0.93
	Menya	Angan	2.5	5.5	0.69
	Bilua	Bilua	3.0	9.0	0.75
	Teiwa	Alor-Pantar	1.0	9.0	0.90
	Lavukaleve	Lavukaleve	3.0	13.5	0.82
	Koiari (Grass)	Koiarian	0.0	1.0	1.00
Kambera	Austronesian	1.0	13.0	0.93	
Tawala	Austronesian	0.0	14.0	1.00	
Yelî Dnye	Yelî Dnye	2.0	9.5	0.83	
Rapanui	Austronesian	1.0	2.0		
Australia	Dyirbal	Pama-Nyungan	1.0	3.0	0.75
	Jingulu	Mirndi	2.5	7.0	0.74
	Kayardild	Tangkic	3.0	3.0	0.50
	Djapu	Pama-Nyungan	1.5	3.0	0.67
	BGW	Gunwingguan	3.0	12.5	0.81
	Uradhi	Pama-Nyungan	1.5	1.0	

	Mawng	Iwaidjan	3.0	10.0	0.77
	Tiwi	Tiwi	3.0	6.0	0.67
	Marrithiyel	W. Daly	4.0	6.5	0.62
	Ngan'gityemerri	S. Daly	6.0	10.0	0.63
	Ungarinjin	Wororan	2.0	8.5	0.81
	Bardi	Nyulnyulan	3.0	12.5	0.81
	Kuniyanti	Bunuban	1.0	8.0	0.89
	Warlpiri	Pama-Nyungan	4.0	8.0	0.67
	Martuthunira	Pama-Nyungan	2.0	1.0	0.33
	Pitjantjatjara	Pama-Nyungan	2.0	0.0	
	Diyari	Pama-Nyungan	2.0	2.0	0.50
	Wembawemba	Pama-Nyungan	1.0	10.0	0.91
W N America	C Alaskan Yup'ik	Eskimo-Aleut	3.0	4.0	0.57
	Haida	Haida	1.0	5.5	0.85
	Nisgha	Tsimshian	0.0	7.0	1.00
	Kwak'wala	Wakashan	0.0	15.0	1.00
	Nuuchahnulth	Wakashan	1.0	8.5	0.89
	Yurok	Algic	1.0	8.0	0.89
	Hupa	Athabaskan	0.5	5.0	0.91
	Chimariko	Chimariko	2.0	8.0	0.80
	Karuk	Karuk	1.0	6.5	0.87
	Yuki	Yuki-Wappo	3.5	5.0	0.59
	E Pomo	Pomoan	2.0	2.5	0.56
	Kashaya Pomo	Pomoan	4.5	3.5	0.44
	Wappo	Yuki-Wappo	2.0	1.0	
	Wintu	Wintun	2.0	8.0	0.80
	Klamath	Klamath-Modoc	3.0	3.0	0.50
	Takelma	Takelma-Kalapuya	1.0	8.0	0.89
	Hanis Coos	Coos	0.0	13.0	1.00
	Wishram	Chinookan	3.0	15.0	0.83
	Klallam	Salish	1.0	7.5	0.88
	Halkomelem	Salish	1.0	7.0	0.88
	Thompson	Salish	1.0	5.0	0.83
	Maidu	Maidun	2.0	7.0	0.78
	Ineseño Chumash	Chumashan	1.0	7.0	0.88
	SSM	Miwok-Costanoan	1.0	9.5	0.90
	Washo	Washo	0.0	15.0	1.00
	Yokuts	Yokuts	2.0	4.0	0.67
	Kumeyaay	Yuman	2.0	10.5	0.84
	Cupeño	Uto-Aztecan	4.0	12.0	0.75
	Tümpisa	Uto-Aztecan	2.0	3.0	0.60

E N America	Slave	Athabaskan	1.0	9.0	0.90	
	NezPerce	Sahaptin	1.0	8.0	0.89	
	Hopi	Uto-Aztecan	1.5	4.5	0.75	
	Zuni	Zuni	0.5	3.0	0.86	
	Acoma	Keresan	0.0	13.0	1.00	
	Kiowa	Kiowa-Tanoan	0.0	9.0	1.00	
	Wichita	Caddoan	2.0	3.0	0.60	
	Lakhota	Siouan	0.0	8.0	1.00	
	Tonkawa	Tonkawa	3.0	6.5	0.68	
	Chitimacha	Chitimacha	1.0	4.5	0.82	
	Tunica	Tunica	4.0	11.0	0.73	
	Atakapa	Atakapa	1.0	6.0	0.86	
	Timucua	Timucua	2.0	9.0	0.82	
	Koasati	Muskogean	2.0	8.0	0.80	
	Creek	Muskogean	0.0	6.0	1.00	
	Seneca	Iroquoian	2.0	11.0	0.85	
	Cherokee	Iroquoian	1.0	12.0	0.92	
	Cree	Algonquian	0.5	17.0	0.97	
	Pawnee	Caddoan	0.0	15.5	1.00	
	Natchez	Natchez	0.0	6.5	1.00	
Mexico-C Am.	Seri	Seri	1.0	7.0	0.88	
	Rarámuri	Uto-Aztecan	2.0	5.0	0.71	
	Yaqui	Uto-Aztecan	2.0	7.5	0.79	
	Cora	Uto-Aztecan	1.0	8.5	0.89	
	Chichimec	Otomanguean	1.0	12.0	0.92	
	Totonac (Filomeno Mata)	Totonac-Tepehuan	1.0	12.0	0.92	
	Purépecha	Purépecha	3.0	6.0	0.67	
	Olutec	Mixe-Zoque	0.5	16.5	0.97	
	Chalcatongo Mixtec	Otomanguean	4.5	11.0	0.71	
	Highland Chontal	Tequistlatecan	1.0	9.0	0.90	
	Huave	Huave	0.0	11.0	1.00	
	Tzutujil	Mayan	1.0	4.0	0.80	
	Pipil	Uto-Aztecan	2.0	7.5	0.79	
	S America	Canela-Kraho	Macro-Ge	1.0	8.0	0.89
		Pirahã	Muran	2.0	2.0	0.50
Sanuma		Yanomaman	3.0	12.0	0.80	
Hup		Nadahup	3.0	2.5	0.45	
Hixkaryana		Cariban	2.0	12.0	0.86	
Apurinã		Arawakan	3.0	6.0	0.67	
Shuar		Jivaroan	2.0	5.5	0.73	

Paez	Paesan	5.0	4.0	0.44
Yagua	Yaguan	2.0	10.0	0.83
Macushi	Cariban	4.0	13.5	0.77
Urarina	Urarina	1.5	8.0	0.84
Shipibo-Conibo	Pano-Tacanan: Panoan	2.0	0.0	0.00
Iskonawa	Pano-Tacanan: Panoan	2.0	2.0	0.50
Awa Pit	Barbacoan	1.5	7.5	0.83
Mochica	Mochica	3.0	3.0	0.50
Cholon	Cholonan	4.0	9.0	0.69
Karajá	Macro-Ge	3.5	8.5	0.71
Wari	Chapakuran	4.0	8.0	0.67
Kwaza	Kwaza	2.5	5.5	0.69
Nambikuara	Nambikuara	2.0	10.0	0.83
Jaqaru	Aymaran	2.0	10.5	0.84
Cavineña	Pano-Tacanan: Tacanan	2.0	6.0	0.75
Cayuvava	Cayuvava	0.0	7.5	1.00
Movima	Movima	3.0	8.0	0.73
Mosetén	Mosetén	6.0	7.0	0.54
Yurakare	Yurakare	1.0	5.0	0.83
Aymara	Aymaran	4.0	12.0	0.75
Alto Perené	Arawakan	4.0	7.5	0.65
Yanesha'	Arawakan	1.0	8.0	0.89
Huallaga Quechua	Quechuan	3.0	11.5	0.79
Chipaya	Uru-Chipaya	6.0	5.0	0.45
Kadiwéu	Guaycuruan	1.0	5.5	0.85
Mataco	Matacoan	1.0	11.5	0.92
Xokleng	Macro-Ge	4.0	1.0	0.20
Guaraní	Tupian	2.0	11.0	0.85
Mapudungun	Mapudungun	0.0	11.0	1.00
Gününa Küne	Chon	0.0	14.0	1.00

References

- Abbi, Anvita. 2013. *A Grammar of the Great Andamanese Language. An Ethnolinguistic Study*. Leiden: Brill.
- Ackerman, Farrell, and Irina Nikolaeva. 2014. *Descriptive Typology and Linguistic Theory: A Study in the Morphosyntax of Relative Clauses*. Stanford: CSLI.
- Anderson, Stephen R. 1985. Inflectional morphology. Tim Shopen, ed., *Language Typology and Syntactic Description, III: Grammatical categories and the lexicon*, 3-56. Cambridge: Cambridge University Press.
- Arkadiev, P. M., Ju. A. Lander, A. B. Letučij, N.R. Sumbatova, and Ja. G. Testeleťs. 2009. Vvedenie. Osnovnye svedenija ob adygejskom jazyke. P. M. Arkad'ev, Ju. A. Lander, A. B. Letučij, N.R. Sumbatova, and Ja. G. Testeleťs, eds., *Aspekty polisintetizma: Očerki po grammatike adygejskogo jazyka*, 17-120. Moscow: Rossijskij gosudarstvennyj gumanitarnyj universitet.
- Bender, Lionel. 1996. *Kunama*. Munich: Lincom Europa.
- Benzing, Johannes. 1955. *Lamutische Grammatik*. Wiesbaden: Franz Steiner.
- Bickel, Balthasar. 2003. Belhare. In *The Sino-Tibetan Languages*, ed. Graham Thurgood, and Randy J. LaPolla. London: Routledge.
- Bickel, Balthasar. 2011. Grammatical relations typology. In Jae Jung Song, ed., *The Oxford Handbook of Linguistic Typology*, 399-444. Oxford: Oxford University Press.
- Bickel, Balthasar and Johanna Nichols. 2007. Inflectional morphology. Tim Shopen, ed., *Language Typology and Syntactic Description, 3*, 169-240. Cambridge: Cambridge University Press.
- Bickel, Balthasar and Johanna Nichols. 2006. Oceania, the Pacific Rim, and the theory of linguistic areas. Zhenya Antić, Charles B. Chang, Clare S. Sandy and Maziar Toosarvandani, ed., *Proceedings of the 32nd Annual Meeting: Special Session on the Languages and Linguistics of Oceania*, 3-15. Berkeley: Berkeley Linguistics Society.
- Bickel, Balthasar, and Johanna Nichols. 2005. Inclusive/exclusive as person vs. number categories worldwide. Elena Filimonova, ed., *Clusivity*, 47-70. Amsterdam: Benjamins.
- Bickel, Balthasar, Johanna Nichols, Taras Zakharko, Alena Witzlack-Makarevich, Kristine Hildebrandt, Michael Riessler, Lennart Bierkandt, Fernando Zúñiga and John B. Lowe. 2017. The Autotyp database, release version 0.1.0.
- Bowe, Heather. 1990. *Categories, Constituents, and Constituent Order in Pitjantjatjara: An Aboriginal Language of Australia*. London: Routledge.
- Bugaeva, Anna. 2012. Southern Hokkaido Ainu. In Nicolas Tranter, ed., *The Languages of Japan and Korea*, 461-507. Milton Park: Routledge.
- Bybee, Joan. 1985. *Morphology: A Study of the Relation Between Meaning and Form*. Amsterdam: Benjamins.
- Casamiquela, Rodolfo M. 1983. *Nociones de gramática del gūnūna kūne. Présentation de la langue des tehuelche septentrionaux australs (Patagonie continentale)*. Paris: CNRS.
- Charachidzé, Georges. 1981. *Grammaire De La Langue Avare*. Paris: Editions Jean-Favard.

- Chirikba, Viacheslav A. 2003. *Abkhaz*. Munich: Lincom Europa.
- Colarusso, John. 1992. *A Grammar of the Kabardian Language*. Calgary: University of Calgary Press.
- Comrie, Bernard. 1988. General features of the Uralic languages. Denis Sinor, ed., *The Uralic Languages: Description, History, and Foreign Influences*, 451-477. Leiden: Brill.
- Cooke, J. R. 1968. *Pronominal Reference in Thai, Burmese, and Vietnamese*. (UCPL 52.) Berkeley-Los Angeles: University of California Press.
- Coppock, Elizabeth and Stephen Wechsler. 2012. The objective conjugation in Hungarian: agreement without phi-features. *NLLT* 30: 699-740.
- Corbett, Greville G. 2013. Canonical morphosyntactic features. In *Canonical Morphology and Syntax*, ed. Dunstan Brown, Marina Chumakina, and Greville Corbett. Oxford: Oxford University Press.
- Corbett, Greville and Matthew Baerman. 2006. Prolegomena to a typology of morphological features. *Morphology* 16: 231-246.
- Curnow, Timothy J. 1997. *A Grammar of Awa Pit (Cuaiquer): An Indigenous Language of Southwestern Colombia*. Ph.D. dissertation, Australian National University.
- Cysouw, Michael. 2003. *The Paradigmatic Structure of Person Marking*. Oxford: Oxford University Press.
- Cysouw, Michael. 2013. Inclusive/exclusive distinction in independent pronouns. Inclusive/exclusive distinction in verbal inflection. Matthew S. Dryer and Martin Haspelmath, eds., *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
<http://wals.info/chapter/39> and <http://wals.info/chapter/40>
- Dahl, Östen. 2004. *The Growth and Maintenance of Linguistic Complexity*. Amsterdam: Benjamins.
- Dahl, Östen. 2000. Animacy and the notion of semantic gender. In *Gender in Grammar and Cognition*, ed. Barbara Unterbeck, and Matti Rissanen. Berlin: Mouton de Gruyter.
- Daniel, Michael. In preparation. Person in Archi revisited. MS, HSE, Moscow.
- Daniel, Michael. 2013. Plurality in independent personal pronouns. Matthew S. Dryer and Martin Haspelmath eds., *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
 (Available online at <http://wals.info/chapter/35>. Accessed on 2017-01-13.)
- Dixon, R. M.W. 2000. Categories of the noun phrase in Jarawara. *Journal of Linguistics* 36:3.487-510.
- Dixon, R. M. W. 1980. *The Languages of Australia*. Cambridge: Cambridge University Press.
- Dryer, Matthew. 1992. The Greenbergian word order correlations. *Language* 68:81-138.
- Dunn, Michael John, 1999. *A Grammar of Chukchi*. Ph.D. dissertation, Australian National University.
- Epps, Patience. 2008. *A Grammar of Hup*. Berlin: Mouton de Gruyter.
- Evans, Nicholas. 2003. *Bininj Gun-Wok: A Pan-Dialectal Grammar of Mayali, Kunwinjku and Kune*. Canberra: Research School of Pacific and Asian Studies, Australian National University.
- Filimonova, Elena, ed. 2005. *Clusivity*. Amsterdam/Philadelphia: Benjamins.

- Foley, William A. In progress. *The Epidemiology of Language: The Evolution of Word Class Categorization in the Austronesian Languages*.
- Forchheimer, Paul. 1953. *The Category of Person in Language*. Berlin: de Gruyter.
- Forker, Diana. 2013. *A Grammar of Hinuq*. Berlin: de Gruyter Mouton.
- Fortescue, Michael. 1998. *Language Relations Across Bering Strait*. London and New York: Cassell.
- Frachtenberg, Leo J. 1922a. Coos. In *Handbook of American Indian Languages, Part 2*, ed. Franz Boas. Washington, DC: Government Printing Office.
- Frachtenberg, Leo J. 1922b. Siuslawan (Lower Umpqua). In *Handbook of American Indian Languages, Part 2*, ed. Franz Boas. Washington, DC: Government Printing Office.
- Fromm, Hans. 1982. *Finnische Grammatik*. Heidelberg: Carl Winter.
- Graczyk, Randolph. 2007. *A Grammar of Crow*. Lincoln: University of Nebraska Press.
- Grünthal, Riho and Johanna Nichols. In press. Transitivity/detransitivizing typology and language family history. *Lingua Posnaniensis* 58:2.
- Haspelmath, Martin. 2013. Ditransitive constructions: the verb 'give'. Matthew Dryer and Martin Haspelmath, eds., *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/105>
- Haspelmath, Martin and Andrea Sims. 2010. *Understanding Morphology*. London: Hodder Education.
- Hausenberg, Anu-Reet. 1998. Komi. Daniel Abondolo, ed., *The Uralic Languages*, 305-326. London: Routledge.
- Hawkins, John A. 2007. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
- Heath, Jeffrey. 1998. Pragmatic skewing in 1 < > 2 pronominal combinations in Native American Languages. *IJAL* 64:2:83-104.
- Heath, Jeffrey. 1991. Pragmatic disguise in pronominal-affix paradigms. In *Paradigms: The Economy of Inflection*, ed. Frans Plank. Berlin: Mouton de Gruyter.
- Helmbrecht, Johannes. 2003. Politeness distinctions in second person pronouns. Friedrich Lenz, ed., *Deictic Conceptualization of Space, Time, and Person*, 185-202. Amsterdam: Benjamins.
- Hill, Kenneth C., ed. 1998. *Hopi Dictionary / Hopiikwa Lavaytutuveni: A Hopi-English Dictionary of the Third Mesa Dialect*. Tucson: University of Arizona Press.
- Iversen, Rune and Guus Kroonen. 2017. Talking Neolithic: Linguistic and Archaeological Perspectives on How Indo-European Was Implemented in Southern Scandinavia. *American Journal of Archaeology* 121:4511-525.
- Janhunen, Juha. 2009. Proto-Uralic – what, where, and when? In *The Quasiquicentennial of the Finno-Ugric Society*, Helsinki: Suomalais-Ugrilainen Seura.
- Jelinek, Eloise. 1984. Empty categories, case, and configurationality. *Natural Language and Linguistic Theory* 2:39-76.
- Kallio, Petri. 2015. The language contact situation in prehistoric Northeastern Europe. Robert Mailhammer, Theo Vennemann and Birgit Anette Olsen, eds., *The Linguistic Roots of Europe*, 77-102. Copenhagen: Museum Tusulanum Press, University of Copenhagen.
- Kari, James, and Ben A. Potter, eds. 2010. *The Dene-Yeniseian Connection* Fairbanks:

- Alaska Native Language Center.
- Keraševa, Z. I. 1960. Kratkij grammatičeskij očerk adygejskogo jazyka. In *Russko-adygejskij slovar'*, ed. X. D. Vodoždokova. Moscow: Gosudarstvennoe Izdatel'stvo Inostrannyx i Nacional'nyx Slovarnej.
- Keresztes, László. 1998. Mansi. Daniel Abondolo, ed., *The Uralic Languages*. London: Routledge.
- Kibrik, A. E. 2003. Nominal inflection galore: Daghestanian, with side glances at Europe and the world. In *Noun Phrase Structure in the Languages of Europe*, ed. Frans Plank. Berlin: Mouton de Gruyter.
- Killian, Don. In press. Uduk gender from a Koman perspective. Francesca di Garbo and Bernhard Wälchli, eds., *Grammatical Gender and Linguistic Complexity*. Berlin: Language Science Press.
- Killian, Don. 2015. Topics in Uduk Phonology and Morphosyntax. Ph.D. dissertation, University of Helsinki.
- Konstantinova, O. A. 1968. Èvenkijskij jazyk. P. Ja. Skorik, ed., *Jazyki narodov SSSR, 5: Mongol'skie, tnuguso-man'čžurskie i paleoaziatskie jazyki*, 68-87. Leningrad: Nauka.
- Kulonen, Ulla-Maija. 2001. Über die Deklination der Personalpronomina in der finnisch-ugrischen Grundsprache. Anu Turk, Triinu Palo, and Tõnu Seilenthal, eds., *Congressus Nonus Internationalis Fenno-Ugristarum, Pars 6*, 178-182.
- Lander, Ju. A. 2017. Reformulating head/dependent marking. To be presented at ALT 12, Canberra, December.
- Leer, Jeff. 1991. Evidence for a Northern Northwest Coast language area: Promiscuous number marking and periphrastic possessive constructions in Haida, Eyak, and Aleut. *IJAL* 57:2: 158-193.
- Lytkin, V. I. 1966. Komi-zyrjanskij jazyk. V. I. Lytkin and K. E. Majtinskaja, eds. *Jazyki narodov SSSR 3: Finno-ugorskie i samodijskie jazyki*. Leningrad: Nauka.
- Martin, Jack. 1999. External possession in Creek. Payne & Barshi eds, 228-250.
- McGregor, William. 1990. *A Functional Grammar of Gooniyandi*. Amsterdam: Benjamins.
- McKay, Graham R. 1978. Pronominal person and number categories in Rembarrnga and Djebbana. *Oceanic Linguistics* 17:27-37.
- Merrifield, William R. 1959. Classification of Kiowa nouns. *IJAL* 25:269-71.
- Miestamo, Matti. 2005. *Standard Negation: The Negation of Declarative Verbal Main Clauses in Typological Perspective*. Berlin: Mouton de Gruyter.
- Miestamo, Matti, Kaius Sinnemäki and Fred Karlsson, eds. 2008. *Language Complexity: Typology, Contact, Change*. Amsterdam: Benjamins.
- Miestamo, Matti, Dik Bakker and Antti Arppe. 2016. Sampling for variety. *Linguistic Typology* 20:2: 233-296.
- Montgomery-Anderson, Brad. 2015. *Cherokee Reference Grammar*. Norman: University of Oklahoma Press.
- Nichols, Johanna. 2017a. Causativization as non-diversity: Linguistic and non-linguistic causes. Presented at SLE Annual Meeting, Zurich.
- Nichols, Johanna. 2017b. Basic event structure: Phases and transitions. Presented at University of Amsterdam, June.

- Nichols, Johanna. 2016. Noun-based and verb-based languages. Poster, SLE Annual Meeting, Naples.
- Nichols, Johanna. 2013. The origin and evolution of case-suppletive pronouns: Eurasian *Siewierska*, ed. Dik Bakker, and Martin Haspelmath. Berlin: De Gruyter Mouton.
- Nichols, Johanna. 2011a. *Ingush Grammar*. Berkeley-Los Angeles: University of California Press.
- Nichols, Johanna. 2011b. Causativization and contact in Nakh-Daghestanian. *BLS* 37: 68-80.
- Nichols, Johanna. 2001. Why “me” and “thee”? Laurel J. Brinton, ed., *Historical Linguistics 1999*, 253-276. Amsterdam-Philadelphia: Benjamins.
- Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press.
- Nichols, Johanna. 1986. Head-marking and dependent-marking grammar. *Language* 62:1:56-119.
- Nichols, Johanna and Christian Bentz. In press. Morphological complexity of languages reflects the settlement history of the Americas. Katerina Harvati, Gerhard Jäger and Hugo Reyes-Centano, ed., *New Perspectives on the Peopling of the Americas*, Tübingen: Kerns.
- Nichols, Johanna, and Balthasar Bickel. 2013. Locus of marking in the clause. Matthew S. Dryer and Martin Haspelmath, eds., *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/23> (Accessed on 2017-01-13.)
- Nichols, Johanna, and Balthasar Bickel. 2013. Locus of marking: Whole-language typology. Matthew S. Dryer and Martin Haspelmath, eds., *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/25> (Accessed 2017-01-13.)
- Nichols, Johanna, and Balthasar Bickel. 2013. Possessive Classification. Matthew S. Dryer and Martin Haspelmath, eds., *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/59> (Accessed 2017-01-13.)
- Nichols, Johanna, David A. Peterson, and Jonathan Barnes. 2004. Transitivity and detransitivizing languages. *Linguistic Typology* 8:2:149-211.
- Nichols, Johanna, Alena Witzlack-Makarevich, and Balthasar Bickel. 2013. The Autotyp genealogy and geography. www.spw.uzh.ch/autotyp/.
- Nikolaeva, Irina. 2014. *A Grammar of Tundra Nenets*. Berlin: De Gruyter Mouton.
- Nikolaeva, Irina. 2005. Modifier-head person concord. In *Morphology and Linguistic Typology. Proceedings of the Fourth Mediterranean Morphology Meeting*, ed. Geert et al. Boon. Bologna: Università degli Studi di Bologna.
- Obata, Kazuko. 2003. *A Grammar of Bilua*. Canberra: Research School of Pacific and Asian Studies, Australian National University.
- Onishi, Masayuki. 1994. A Grammar of Motuna (Bougainville, Papua New Guinea). Ph.D. dissertation, Australian National University
- Öztürk, Balkiz, and Markus A. Pöchtrager, eds. 2011. *Pazar Laz*. Munich: Lincom Europa.
- Payne, Doris L., and Immanuel Barshi, eds. 1999. *External Possession*. Amsterdam: Benjamins.

- Pitkin, Harvey. 1984. *Wintu Grammar*. Berkeley-Los Angeles: University of California Press.
- Plank, Frans. 2003. The selective elaboration of nominal or pronominal inflection. In *Noun Phrase Structure in the Languages of Europe*, ed. Frans Plank. Berlin: Mouton de Gruyter.
- Plank, Frans. 1999. Split morphology: How agglutination and flection mix. *Linguistic Typology* 3:3:279-340.
- Plank, Frans. 1985. Die Ordnung der Personen. *Folia Linguistica* 119: 111-176.
- Riese, Timothy. 2001. *Vogul*. Munich: Lincom Europa.
- Robins, R. H. 1958. *The Yurok Language: Gramamr, Texts, Lexicon*. Berkeley-Los Angeles: University of California Press.
- Rood, David S., and Allan R. Taylor. 1996. Sketch of Lakhota, a Siouan language. In *Handbook of North American Indians, vol. 17: Languages*, ed. Ives Goddard. Washington: Smithsonian Institution.
- Rootsi, Siiri, Lev A. Zhivotovsky and et al. 2007. A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia toward Europe. *European Journal of Human Genetics* 15:2: 204-211.
- Rude, Noel E. 1985. *Studies in Nez Perce Grammar and Discourse*. Ph.D. dissertation, University of Oregon.
- Salminen, Tapani. 2012. *Tundra Nenets*. (Revised version of Salminen 1998.) MS, University of Helsinki.
- Salminen, Tapani. 1998. Nenets. In *The Uralic Languages*, ed. Daniel Abondolo. London: Routledge.
- Salminen, Tapani. 1997. *Tundra Nenets Inflection*. Helsinki: Suomalais-Ugrilainen Seura.
- Sammallahti, Pekka. 1998. Saamic. In *The Uralic Languages*, ed. Daniel Abondolo. London: Routledge.
- Shibatani, Masayoshi. 1990. *The Languages of Japan*. Cambridge: Cambridge University Press.
- Siewierska, Anna. 2011. Person marking. In *The Oxford Handbook of Linguistic Typology*, ed. Jae Jung Song. Oxford: Oxford University Press.
- Siewierska, Anna. 2004. *Person*. Cambridge: Cambridge University Press.
- Silverstein, Michael. 1976. Hierarchy of features and ergativity. In *Grammatical Categories in Australian Languages*, ed. R. M. W. Dixon. Canberra; Atlantic Highlands, NJ: Australian Institute of Aboriginal Studies; Humanities Press.
- Sinnemäki, Kaius. 2011. *Language Universals and Linguistic Complexity: Three case studies in core argument marking*. Ph.D. dissertation, Helsinki University.
- Smeets, Ineke. 2008. *A Mapuche Grammar*. Berlin: Mouton de Gruyter.
- Sumbatova, Nina R., and Rasul O. Mutalov. 2003. *A Grammar of Icari Dargwa*. Munich: Lincom.
- Tamura, Suzuko. 2000. *The Ainu Language*. Tokyo: Sanseido.
- Traill, Anthony. 1994. *A !Xóõ Dictionary*. Koeln: Rüdiger Köppe Verlag.
- Tryon, Darrell T. 1968. *Iai Grammar*. (Pacific Linguistics B 8.) Canberra: Australian National University.
- Vajda, Edward. 2010. A Siberian link with Na-Dene languages. In *The Dene-Yeniseian Connection*, ed. James Kari, and Ben A. Potter. Fairbanks: Alaska Native Languages Center.

- Van Valin, Robert D., Jr. 1985. Case marking and the study of the Lakhota clause. In *Grammar Inside and Outside the Clause: Approaches to Theory from the Field*, ed. Johanna Nichols, and Anthony C. Woodbury. Cambridge: Cambridge University Press.
- Watkins, Laurel J., and Parker McKenzie. 1984. *A Grammar of Kiowa*. Lincoln: University of Nebraska Press.
- Weber, David John. 1989. *A Grammar of Huallaga (Huanuco) Quechua*. Berkeley-Los Angeles: University of California Press.
- Witzlack-Makarevich, Alena. 2010. Typological Variation in Grammatical Relations. Ph.D. dissertation, University of Leipzig.
- Witzlack-Makarevich, Alena, Taras Zakharko, Lennart Bierkandt, Fernando Zúñiga, and Balthasar Bickel. 2016. Decomposing hierarchical alignment: Co-arguments as conditions on alignment and the limits of referential hierarchies as explanations in verb agreement. *Linguistics*
- Wolfart, H. Christoph. 1996. Sketch of Cree, an Algonquian language. In *Handbook of North American Indians, 17: Languages*, ed. Ives Goddard. Washington, DC: Smithsonian Institution.
- Wonderly, William, Lorna P. Gibson, and Paul L. Kirk. 1954. Number in Kiowa: nouns, demonstratives, and adjectives. *IJAL* 20:1-7.
- Zúñiga, Fernando. 2006. *Deixis and Alignment: Inverse Systems in Indigenous Languages of the Americas*. Amsterdam: Benjamins.
- Zwicky, Arnold M. 1977. Hierarchies of person. *Papers from the Thirteenth Annual Meeting, Chicago Linguistic Society*, 714-733. Chicago: Chicago Linguistic Society.
- Zwicky, Arnold M. 1985. How to describe inflection. *BLS* 19: 372-385.

Figures

Figure 1

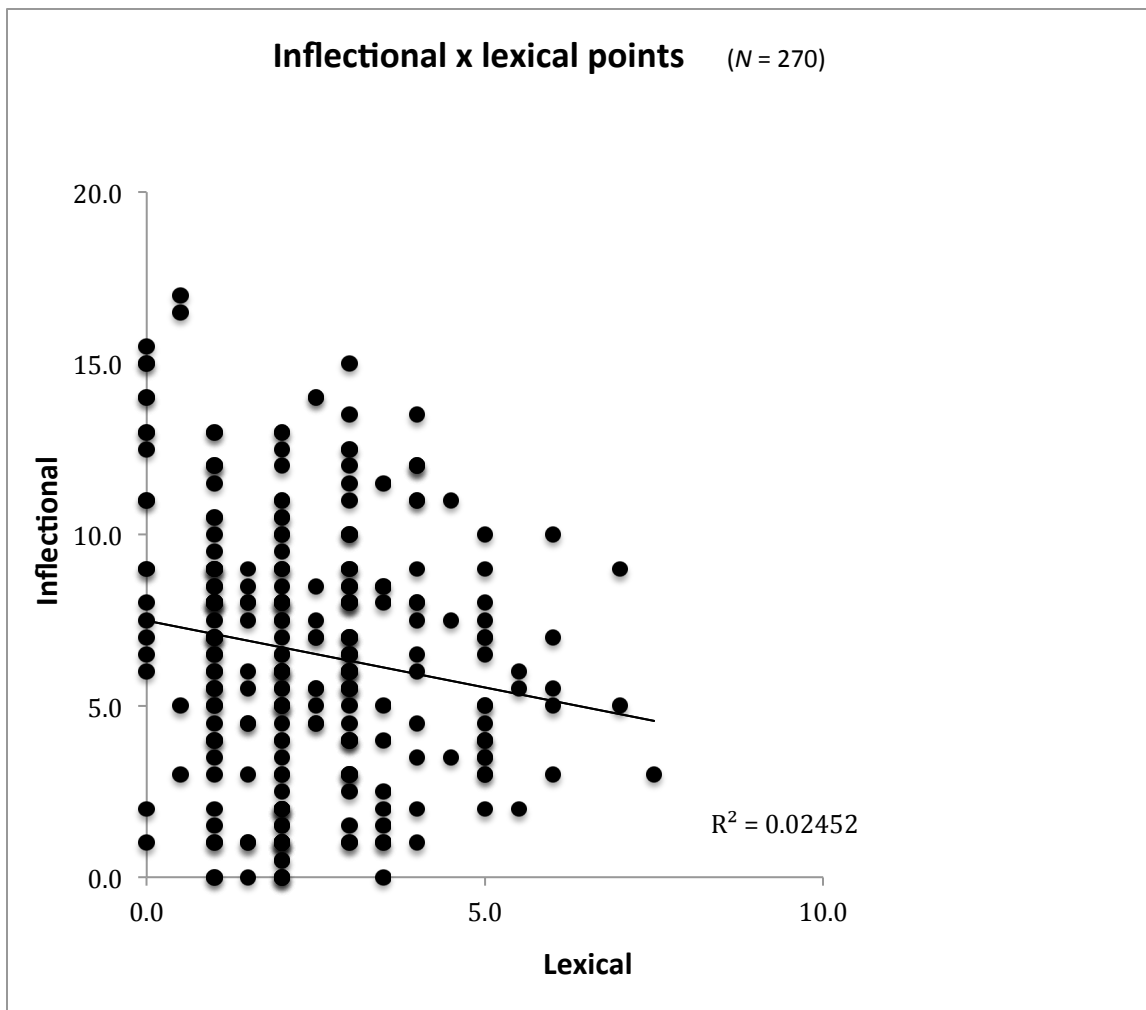
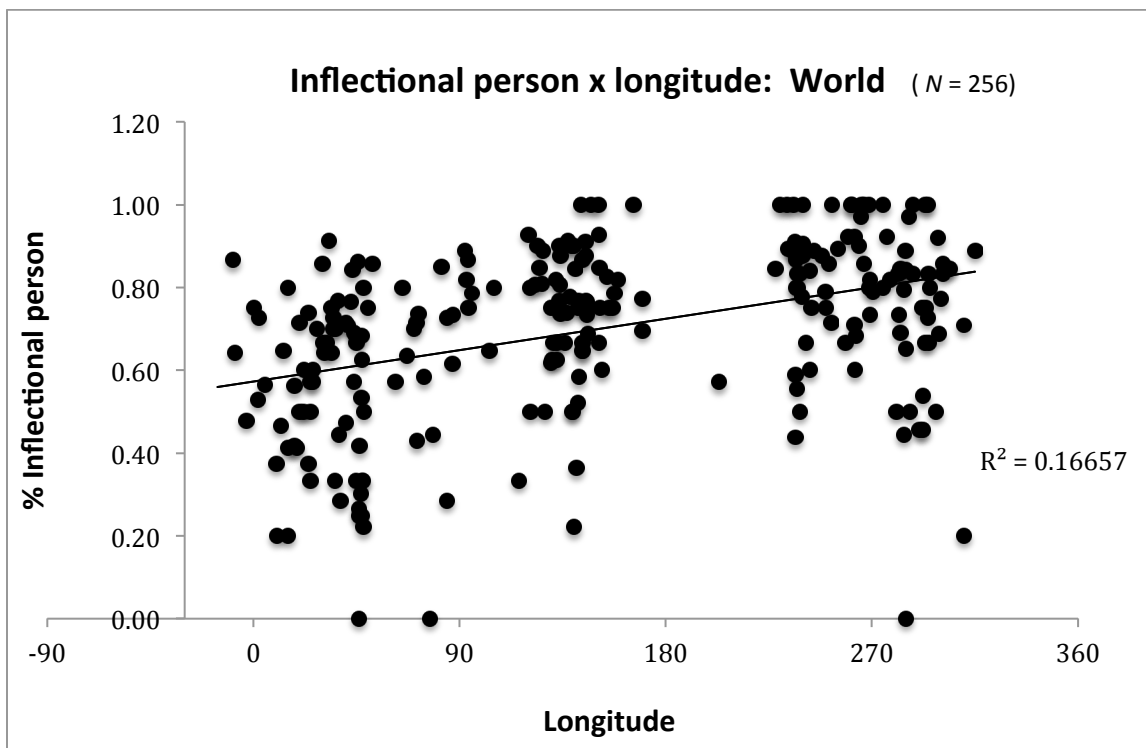


Figure 2. Plot of inflectional person by longitude. Western Europe and west Africa are at the left, eastern North and South America at the right.



Figures 3, 4, 5. Inflectional person by longitude, for three latitudinal bands.

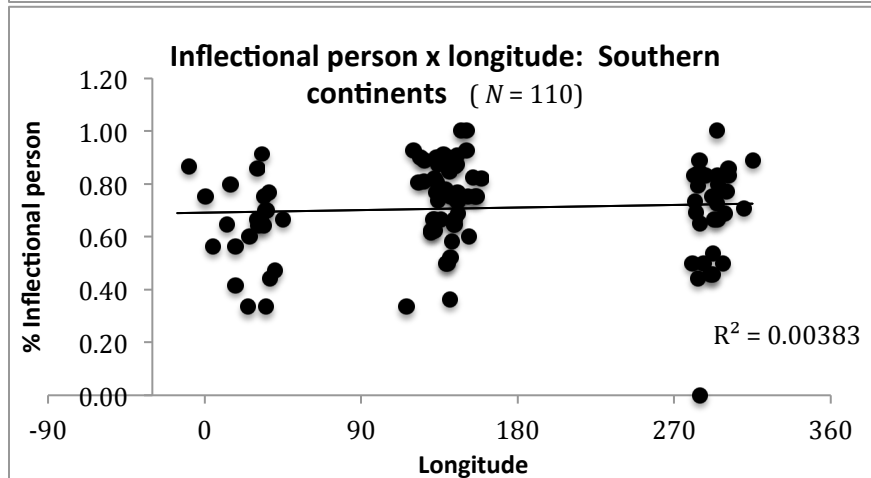
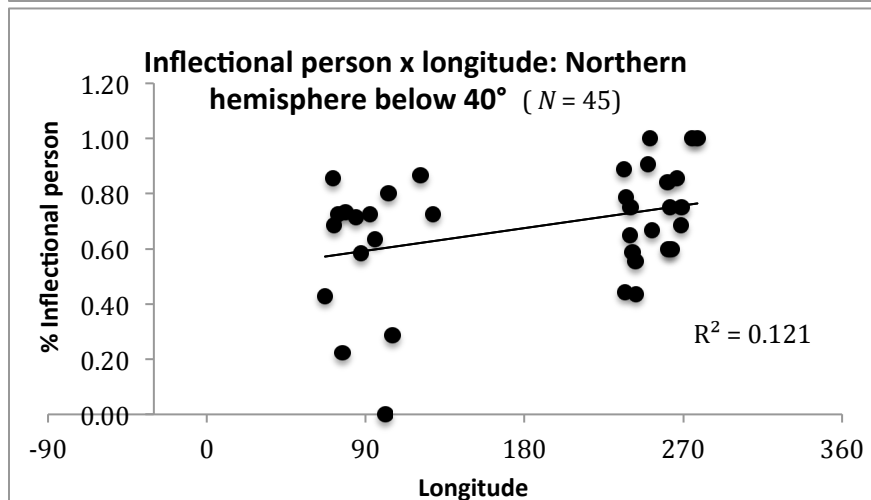
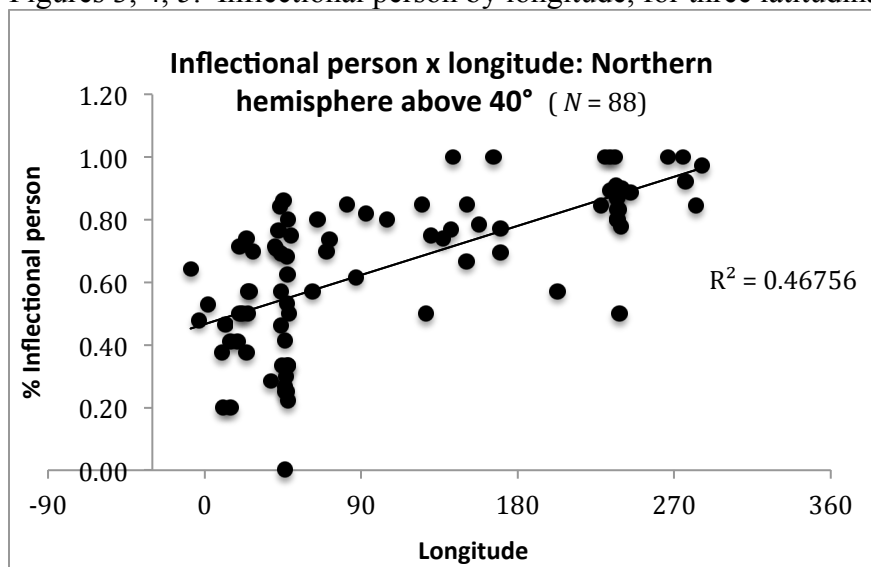
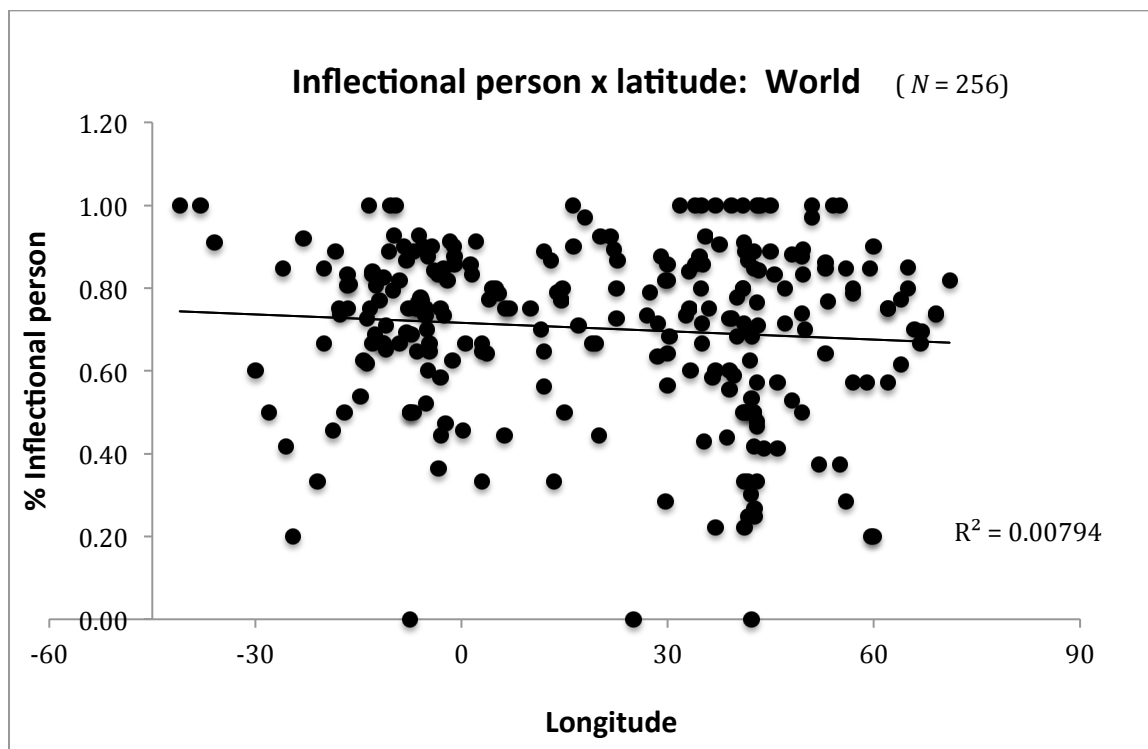


Figure 6. Inflectional person by latitude.



Figures 7, 8, 9. Inflectional person by latitude for three longitudinal bands.

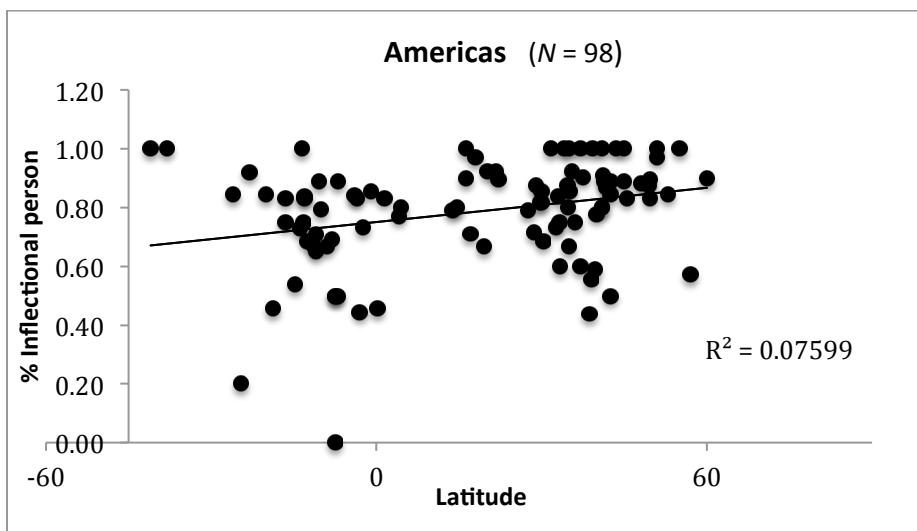
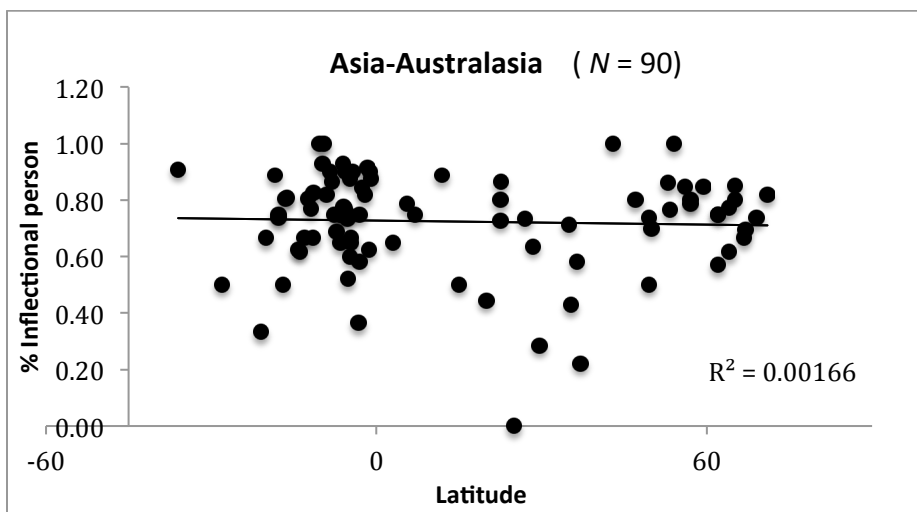
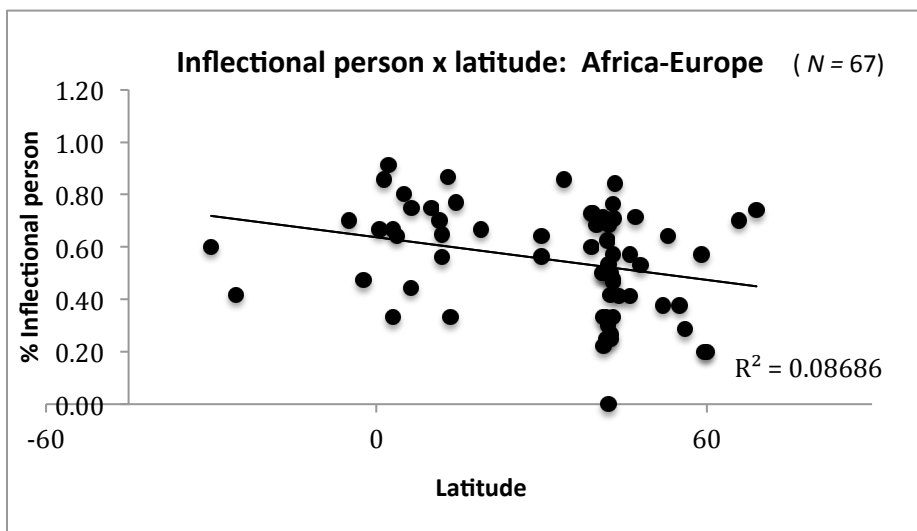


Figure 10. Top: Languages in sample for which proportions were calculated ($N = 251$). Circles show languages below (white) and above (black) the worldwide mean of 70% for inflectional person points. Bottom: All languages in sample ($N = 270$); absolute values below (white) and above (black) the worldwide mean of 6.6 inflectional points.

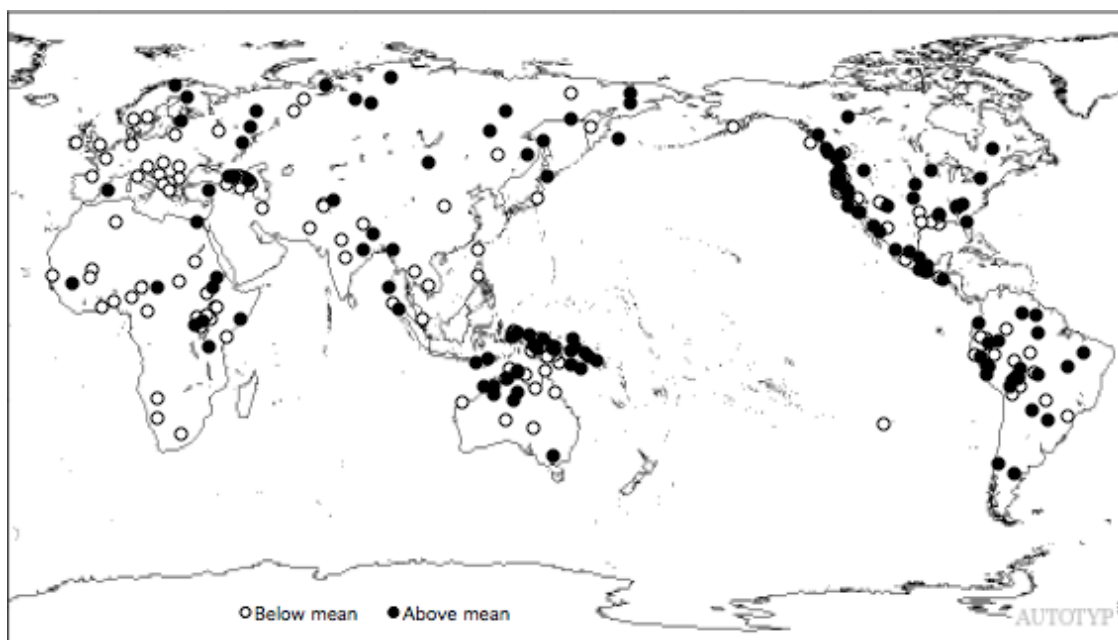
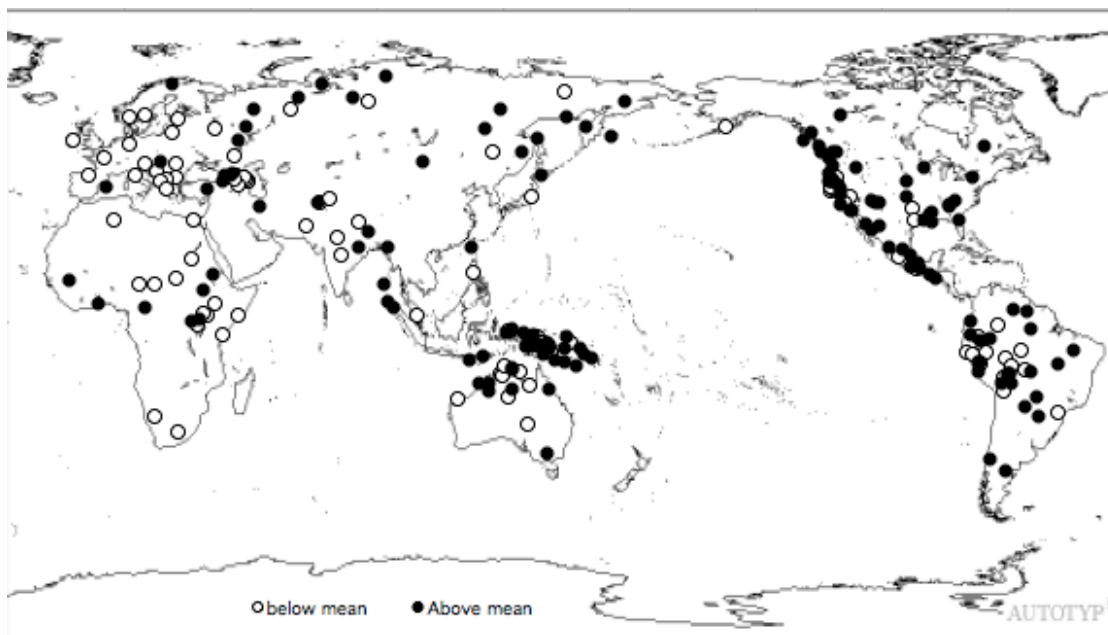
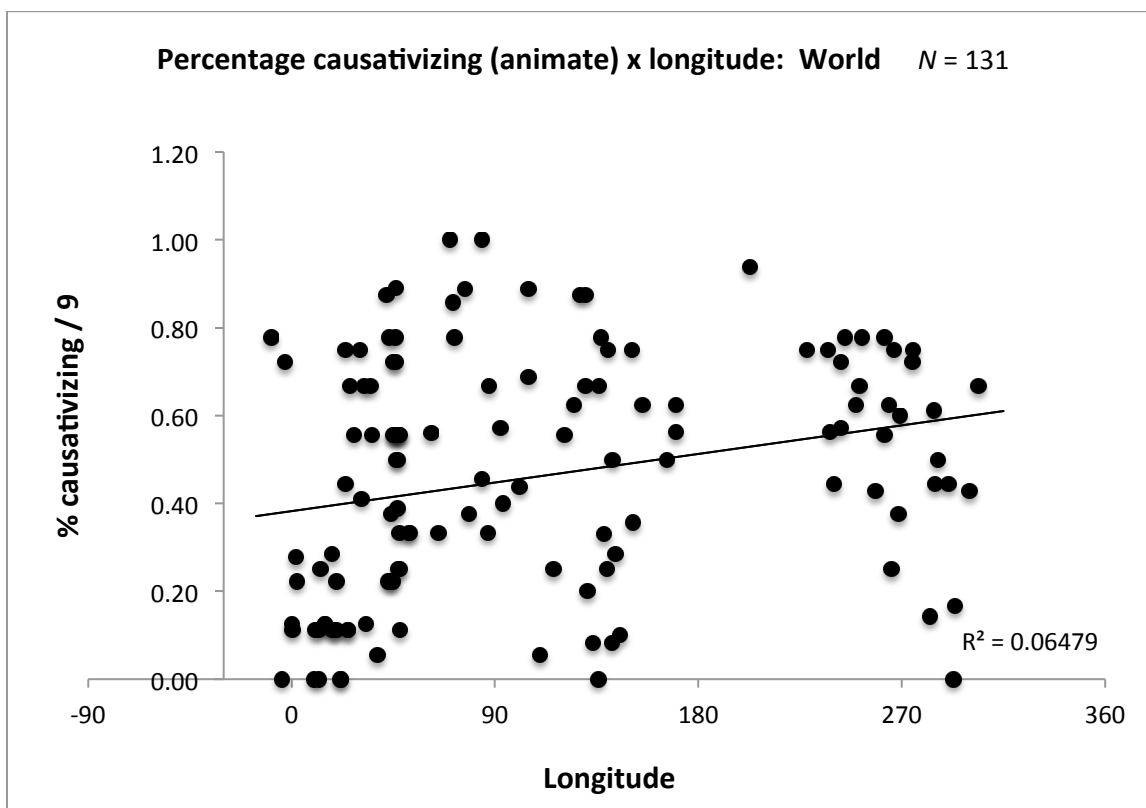
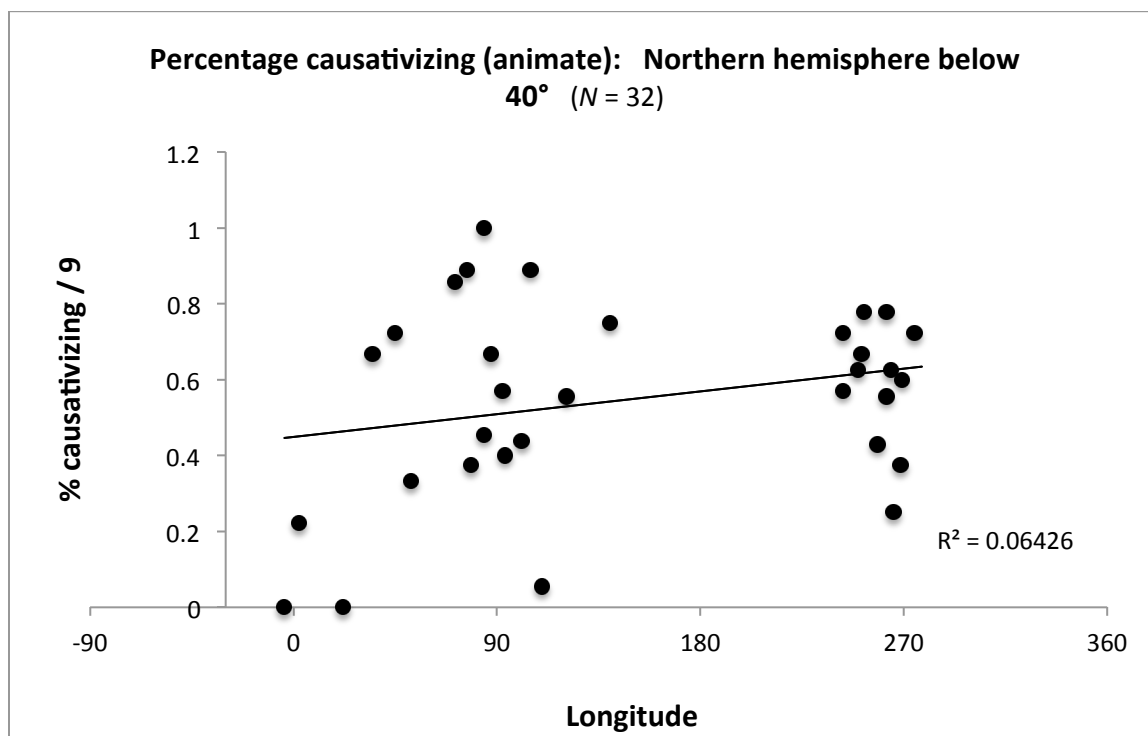
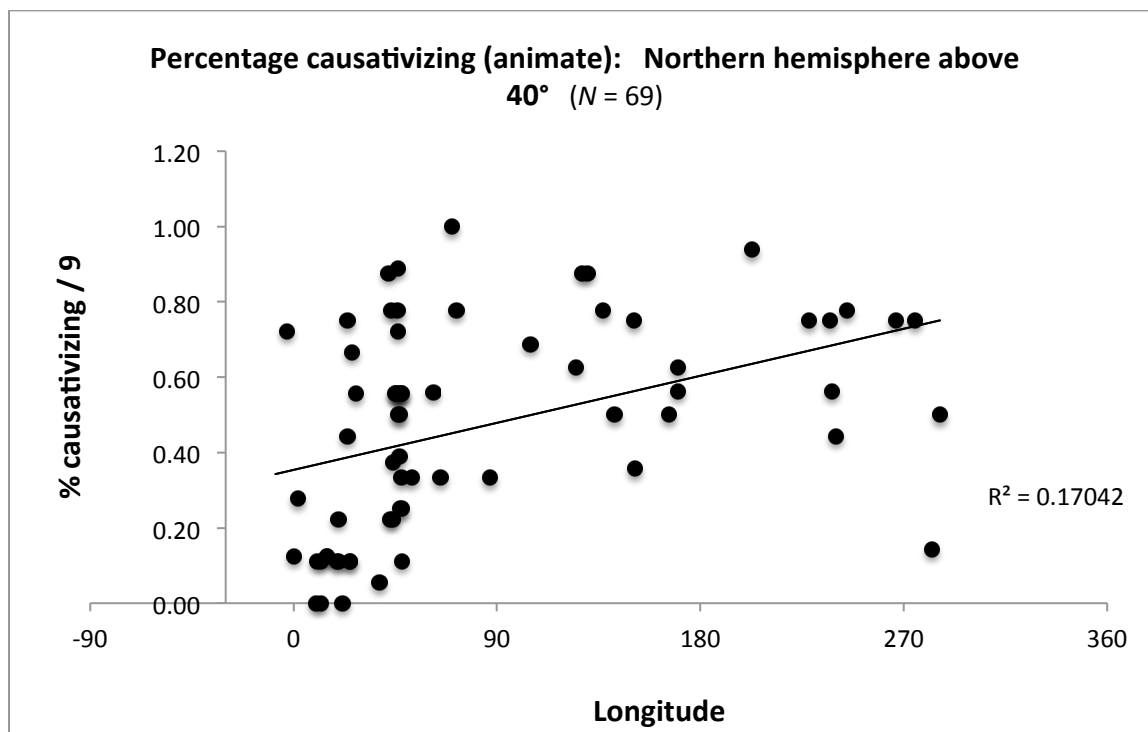


Figure 11. Percent of verbs using causativization as realization in the causative alternation, by longitude.



Figures 12, 13, 14. Percent of verbs using causativization as realization in the causative alternation, by longitude, in three latitudinal bands.



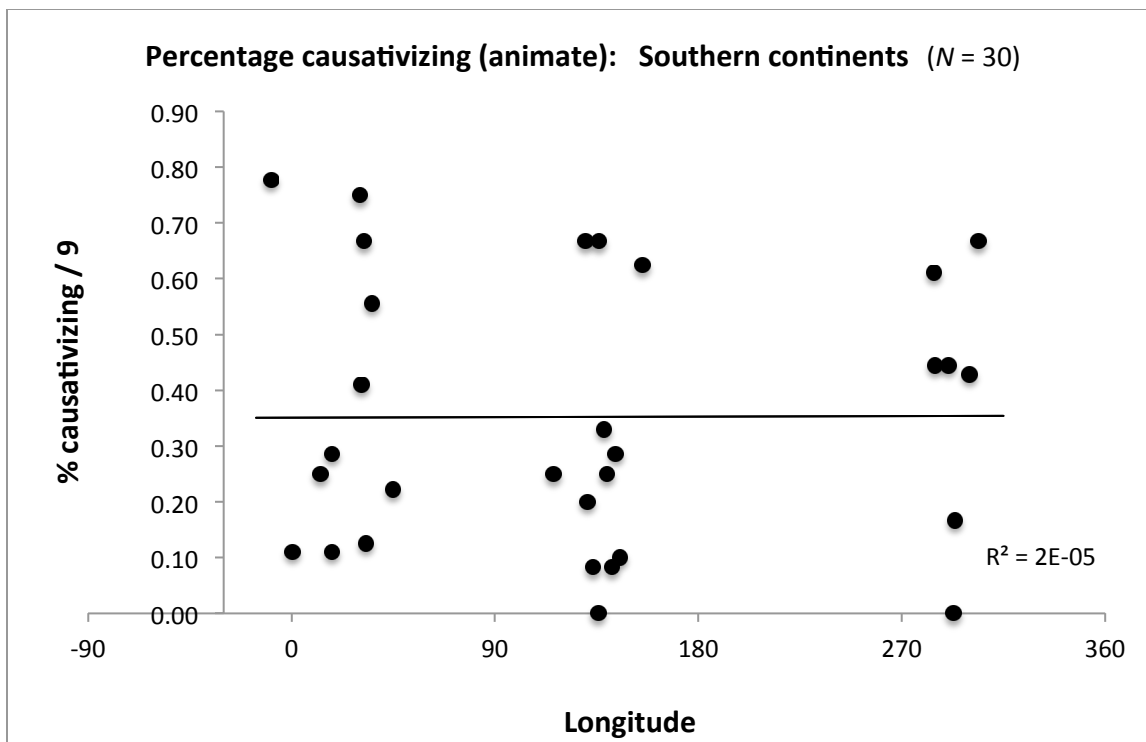


Figure 15. Continental frequencies for inflectional person and four other variables. Left: all continents. Right: Eurasia and Americas only. Continents are ordered from west (left) to east: 1. Africa, 2. Europe, 3. Inner and northern Asia, 4. S & SE Asia, 5. New Guinea, 6. Australia, 7. W. N America, 8. E. N America, 9. Mexico-Central America, 10. S. America.

