The distribution of zero forms in inflectional morphology

Laura Becker

University of Freiburg

02/09/22, IMM 20, Budapest

Background

Zero marking in morphology

- ontological and theoretical status of zero forms / zero morphs (Garcia and van Putte 1989; Jakobson 1939; Lemaréchal 1997; Mel'cuk 2002)

- problems: assumption of more structure than we have evidence for (Anderson 1992; Blevins 2016; Pullum and Zwicky 1991)

Background

Zero marking in morphology

- ontological and theoretical status of zero forms / zero morphs (Garcia and van Putte 1989; Jakobson 1939; Lemaréchal 1997; Mel'cuk 2002)
- problems: assumption of more structure than we have evidence for (Anderson 1992; Blevins 2016; Pullum and Zwicky 1991)

Zero marking in typology

- **form-frequency effects** (Greenberg 1966; Croft 2003; Bybee 2011; Diessel 2019; Song 2018)
  more frequent expressions tend to have shorter or zero forms than comparable less frequent expressions
- ☞ **coding efficiency** (Haspelmath 2008a; Haspelmath 2008b; Haspelmath 2021)
- ☞ fairly weak crosslinguistic basis / evidence (Guzmán Naranjo and Becker 2021a; Seržant and Moroz 2022; Stolz and Levkovych 2019; Bickel et al. 2015; Cysouw 2003; Siewierska 2010)

Background

Zero marking in morphology

- ontological and theoretical status of zero forms / zero morphs (Garcia and van Putte 1989; Jakobson 1939; Lemaréchal 1997; Mel'cuk 2002)
- problems: assumption of more structure than we have evidence for (Anderson 1992; Blevins 2016; Pullum and Zwicky 1991)

Zero marking in typology

- **form-frequency effects** (Greenberg 1966; Croft 2003; Bybee 2011; Diessel 2019; Song 2018)
  more frequent expressions tend to have shorter or zero forms than comparable less frequent expressions
- ☞ **coding efficiency** (Haspelmath 2008a; Haspelmath 2008b; Haspelmath 2021)
- ☞ fairly weak crosslinguistic basis / evidence (Guzmán Naranjo and Becker 2021a; Seržant and Moroz 2022; Stolz and Levkovych 2019; Bickel et al. 2015; Cysouw 2003; Siewierska 2010)

Aim of this presentation

- assess the crosslinguistic distribution of zero forms in inflection morphology
- re-evaluate coding efficiency as an explanation of zero forms

Towards a definition of zero forms

- I will talk about zero forms (not markers)
- zero forms as a descriptive shorthand for the absence of material exponence of a given morphosyntactic function (Stolz and Levkovych 2019)
- ☞ the absence of an overt marker for a morphosyntactic function in opposition to overt marking in the same inflection paradigm

Towards a definition of zero forms

- I will talk about zero forms (not markers)
- zero forms as a descriptive shorthand for the absence of material exponence of a given morphosyntactic function (Stolz and Levkovych 2019)
☞ the absence of an overt marker for a morphosyntactic function in opposition to overt marking in the same inflection paradigm

- word and paradigm approach (no linear order problems)
- morphological analyses are based on the paradigmatic relation between different word forms, representing the different morphosyntactic functions a given word can have

Towards a definition of zero forms

stems as longest common substrings

- the phonetic material that is shared by all forms of a lexeme in its inflectional paradigm

markers as the remaining phonetic material

- the additional material of the form in that cell

zero as the use of stems in a specific morphosyntactic function

- if there is no phonetic material in addition to the stem in the form of a cell of the paradigm
- ☞ **zero form**

Dataset and segmentation

UniMorph database (McCarthy et al. 2020)

+

phonological transcription of some languages with *Epitran*
(Mortensen, Dalmia, and Littell 2018)

+

extraction of stems and markers with *Paradigma*
(Beniamine and Guzmán Naranjo 2021; Guzmán Naranjo and Becker 2021a)

Dataset
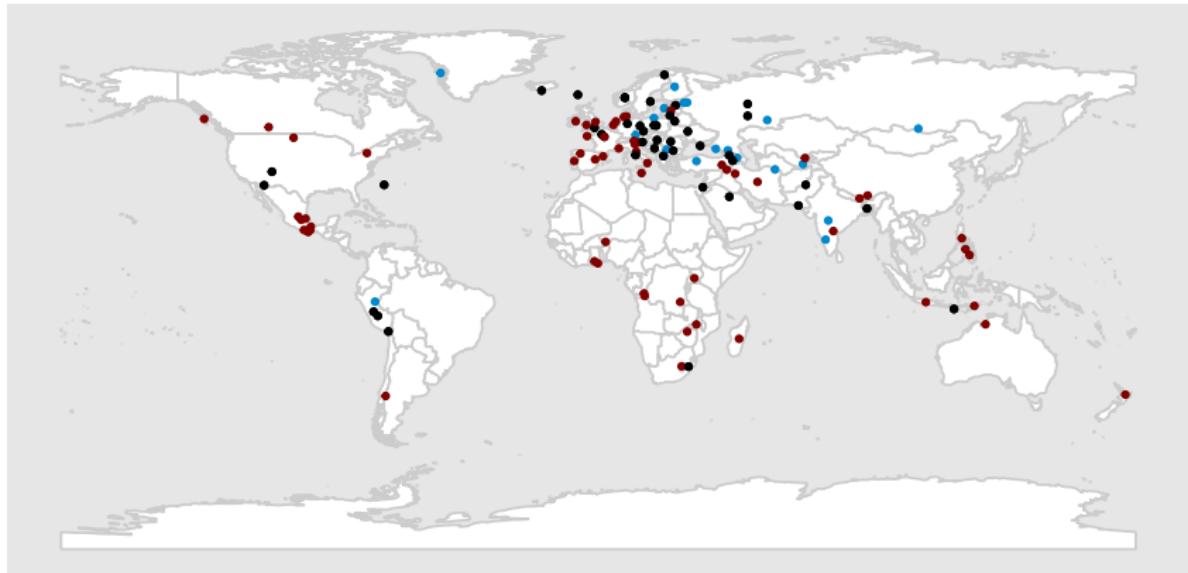
part of speech ● both ● noun ● verb

nouns : 61 languages   verbs : 104 languages   **141** languages **in total**

Examples: French

cell	form	stem	marker
PRS.IND.1SG	alym	alym	-
PRS.IND.2SG	alym	alym	-
PRS.IND.3SG	alym	alym	-
PRS.IND.1PL	alymɔn	alym	-ɔn
PRS.IND.2PL	alymɛz	alym	-ɛz
PRS.IND.3PL	alym	alym	-
PRS.COND.1SG	alymɛrɛ	alym	-ɛrɛ
PRS.COND.2SG	alymɛrɛ	alym	-ɛrɛ
PRS.COND.3SG	alymɛrɛ	alym	-ɛrɛ
PRS.COND.1PL	alymɛrjɔn	alym	-ɛrjɔn
PRS.COND.2PL	alymɛrjɛz	alym	-ɛrjɛz
PRS.COND.3PL	alymɛraj	alym	-ɛraj
PRS.SUBJ.1SG	alym	alym	-
PRS.SUBJ.2SG	alym	alym	-
PRS.SUBJ.3SG	alym	alym	-
PRS.SUBJ.1PL	alymjɔn	alym	-jɔn
PRS.SUBJ.2PL	alymjɛz	alym	-jɛz
PRS.SUBJ.3PL	alym	alym	-

Examples: French

cell	form	stem	marker
PRS.IND.1SG	**alym**	alym	-
PRS.IND.2SG	**alym**	alym	-
PRS.IND.3SG	**alym**	alym	-
PRS.IND.1PL	**alym**ɔn	alym	-ɔn
PRS.IND.2PL	**alym**ɛz	alym	-ɛz
PRS.IND.3PL	**alym**	alym	-
PRS.COND.1SG	**alym**ɛrɛ	alym	-ɛrɛ
PRS.COND.2SG	**alym**ɛrɛ	alym	-ɛrɛ
PRS.COND.3SG	**alym**ɛrɛ	alym	-ɛrɛ
PRS.COND.1PL	**alym**ɛrjɔn	alym	-ɛrjɔn
PRS.COND.2PL	**alym**ɛrjɛz	alym	-ɛrjɛz
PRS.COND.3PL	**alym**ɛraj	alym	-ɛraj
PRS.SUBJ.1SG	**alym**	alym	-
PRS.SUBJ.2SG	**alym**	alym	-
PRS.SUBJ.3SG	**alym**	alym	-
PRS.SUBJ.1PL	**alym**jɔn	alym	-jɔn
PRS.SUBJ.2PL	**alym**jɛz	alym	-jɛz
PRS.SUBJ.3PL	**alym**	alym	-

Examples: French

cell	form	stem	marker
PRS.IND.1SG	alym	alym	-
PRS.IND.2SG	alym	alym	-
PRS.IND.3SG	alym	alym	-
PRS.IND.1PL	alym**ɔn**	alym	-ɔn
PRS.IND.2PL	alym**ɛz**	alym	-ɛz
PRS.IND.3PL	alym	alym	-
PRS.COND.1SG	alym**ɛrɛ**	alym	-ɛrɛ
PRS.COND.2SG	alym**ɛrɛ**	alym	-ɛrɛ
PRS.COND.3SG	alym**ɛrɛ**	alym	-ɛrɛ
PRS.COND.1PL	alym**ɛrjɔn**	alym	-ɛrjɔn
PRS.COND.2PL	alym**ɛrjɛz**	alym	-ɛrjɛz
PRS.COND.3PL	alym**ɛraj**	alym	-ɛraj
PRS.SUBJ.1SG	alym	alym	-
PRS.SUBJ.2SG	alym	alym	-
PRS.SUBJ.3SG	alym	alym	-
PRS.SUBJ.1PL	alym**jɔn**	alym	-jɔn
PRS.SUBJ.2PL	alym**jɛz**	alym	-jɛz
PRS.SUBJ.3PL	alym	alym	-

Examples: French

cell	form	stem	marker
PRS.IND.1SG	**alym**	alym	-
PRS.IND.2SG	**alym**	alym	-
PRS.IND.3SG	**alym**	alym	-
PRS.IND.1PL	alymɔn	alym	-ɔn
PRS.IND.2PL	alymɛz	alym	-ɛz
PRS.IND.3PL	**alym**	alym	-
PRS.COND.1SG	alymɛrɛ	alym	-ɛrɛ
PRS.COND.2SG	alymɛrɛ	alym	-ɛrɛ
PRS.COND.3SG	alymɛrɛ	alym	-ɛrɛ
PRS.COND.1PL	alymɛrjɔn	alym	-ɛrjɔn
PRS.COND.2PL	alymɛrjɛz	alym	-ɛrjɛz
PRS.COND.3PL	alymɛraj	alym	-ɛraj
PRS.SUBJ.1SG	**alym**	alym	-
PRS.SUBJ.2SG	**alym**	alym	-
PRS.SUBJ.3SG	**alym**	alym	-
PRS.SUBJ.1PL	alymjɔn	alym	-jɔn
PRS.SUBJ.2PL	alymjɛz	alym	-jɛz
PRS.SUBJ.3PL	**alym**	alym	-

Examples: German

| cell | *Klos* 'dumpling' | | | *Abend* 'evening' | | | *Kreuz* 'cross' | | |
	form	stem	marker	form	stem	marker	form	stem	marker
NOM.SG	klos	kls	-o-	abənt	abən	-t	kroyt͡s	kroyt͡s	-
ACC.SG	klos	kls	-o-	abənt	abən	-t	kroyt͡s	kroyt͡s	-
DAT.SG	klos	kls	-o-	abənt	abən	-t	kroyt͡s	kroyt͡s	-
GEN.SG	kloses	kls	-o-es	abəndes	abən	-des	kroyt͡ses	kroyt͡s	-es
NOM.PL	kløsə	kls	-ø-ə	abəndə	abən	-de	kroyt͡sə	kroyt͡s	-ə
ACC.PL	kløsə	kls	-ø-ə	abəndə	abən	-de	kroyt͡sə	kroyt͡s	-ə
DAT.PL	kløsən	kls	-ø-ən	abəndən	abən	-den	kroyt͡sən	kroyt͡s	-ən
GEN.PL	kløsə	kls	-ø-ə	abəndə	abən	-de	kroyt͡sə	kroyt͡s	-ə

Examples: German

cell	*Klos* 'dumpling'			*Abend* 'evening'			*Kreuz* 'cross'		
	form	stem	marker	form	stem	marker	form	stem	marker
NOM.SG	**klos**	kls	-o-	**abən**t	abən	-t	**kroy͡ts**	kroy͡ts	-
ACC.SG	**klos**	kls	-o-	**abən**t	abən	-t	**kroy͡ts**	kroy͡ts	-
DAT.SG	**klos**	kls	-o-	**abən**t	abən	-t	**kroy͡ts**	kroy͡ts	-
GEN.SG	**klos**es	kls	-o-es	**abən**des	abən	-des	**kroy͡ts**es	kroy͡ts	-es
NOM.PL	**kløsə**	kls	-ø-ə	**abən**də	abən	-de	**kroy͡ts**ə	kroy͡ts	-ə
ACC.PL	**kløsə**	kls	-ø-ə	**abən**də	abən	-de	**kroy͡ts**ə	kroy͡ts	-ə
DAT.PL	**kløsən**	kls	-ø-ən	**abən**dən	abən	-den	**kroy͡ts**ən	kroy͡ts	-ən
GEN.PL	**kløsə**	kls	-ø-ə	**abən**də	abən	-de	**kroy͡ts**ə	kroy͡ts	-ə

Examples: German

cell	*Klos* 'dumpling'			*Abend* 'evening'			*Kreuz* 'cross'		
	form	stem	marker	form	stem	marker	form	stem	marker
NOM.SG	kl**os**	kls	-o-	abə**nt**	abən	-t	kroy͡fs	kroy͡fs	-
ACC.SG	kl**os**	kls	-o-	abə**nt**	abən	-t	kroy͡fs	kroy͡fs	-
DAT.SG	kl**os**	kls	-o-	abə**nt**	abən	-t	kroy͡fs	kroy͡fs	-
GEN.SG	kl**oses**	kls	-o-es	abə**ndes**	abən	-des	kroy͡fs**es**	kroy͡fs	-es
NOM.PL	kl**øsə**	kls	-ø-ə	abə**ndə**	abən	-de	kroy͡fs**ə**	kroy͡fs	-ə
ACC.PL	kl**øsə**	kls	-ø-ə	abə**ndə**	abən	-de	kroy͡fs**ə**	kroy͡fs	-ə
DAT.PL	kl**øsən**	kls	-ø-ən	abə**ndən**	abən	-den	kroy͡fs**ən**	kroy͡fs	-ən
GEN.PL	kl**øsə**	kls	-ø-ə	abə**ndə**	abən	-de	kroy͡fs**ə**	kroy͡fs	-ə

Examples: Aymara

cell	*anu* 'dog'	*chaski* 'messenger'	*luk'ana* 'finger'
NOM.SG	anu	chaski	luk'ana
ACC.SG	an	chask	luk'an
GEN.SG	anuna	chaskina	luk'anana
COM.SG	anumpi	chaskimpi	luk'anampi
BEN.SG	anutaki	chaskitaki	luk'anataki
PRP.SG	anulayku	chaskilayku	luk'analayku
ABL.SG	anuta	chaskita	luk'anata
ALL.SG	anuru	chaskiru	luk'anaru
INESS.SG	anpacha	chaskpacha	luk'anpacha
EQTV.SG	anjama	chaskjama	luk'anjama
INTER.SG	anupura	chaskipura	luk'anapura
PROP.SG	anuni	chaskini	luk'anani
TERM.SG	anukama	chaskikama	luk'anakama
VERS.SG	anukata	chaskikata	luk'anakata

Examples: Aymara

cell	*anu* 'dog'	*chaski* 'messenger'	*luk'ana* 'finger'
NOM.SG	**anu**	**chaski**	**luk'an**a
ACC.SG	**an**	**chask**	**luk'an**
GEN.SG	**an**una	**chask**ina	**luk'an**ana
COM.SG	**an**umpi	**chask**impi	**luk'an**ampi
BEN.SG	**an**utaki	**chask**itaki	**luk'an**ataki
PRP.SG	**an**ulayku	**chask**ilayku	**luk'an**alayku
ABL.SG	**an**uta	**chask**ita	**luk'an**ata
ALL.SG	**an**uru	**chask**iru	**luk'an**aru
INESS.SG	**an**pacha	**chask**pacha	**luk'an**pacha
EQTV.SG	**an**jama	**chask**jama	**luk'an**jama
INTER.SG	**an**upura	**chask**ipura	**luk'an**apura
PROP.SG	**an**uni	**chask**ini	**luk'an**ani
TERM.SG	**an**ukama	**chask**ikama	**luk'an**akama
VERS.SG	**an**ukata	**chask**ikata	**luk'an**akata

Examples: Aymara

cell	*anu* 'dog'	*chaski* 'messenger'	*luk'ana* 'finger'
NOM.SG	an**u**	chask**i**	luk'an**a**
ACC.SG	an	chask	luk'an
GEN.SG	an**una**	chask**ina**	luk'an**ana**
COM.SG	an**umpi**	chask**impi**	luk'an**ampi**
BEN.SG	an**utaki**	chask**itaki**	luk'an**ataki**
PRP.SG	an**ulayku**	chask**ilayku**	luk'an**alayku**
ABL.SG	an**uta**	chask**ita**	luk'an**ata**
ALL.SG	an**uru**	chask**iru**	luk'an**aru**
INESS.SG	an**pacha**	chask**pacha**	luk'an**pacha**
EQTV.SG	an**jama**	chask**jama**	luk'an**jama**
INTER.SG	an**upura**	chask**ipura**	luk'an**apura**
PROP.SG	an**uni**	chask**ini**	luk'an**ani**
TERM.SG	an**ukama**	chask**ikama**	luk'an**akama**
VERS.SG	an**ukata**	chask**ikata**	luk'an**akata**

Examples: Georgian

cell	t'exs 'break'	k'vecs 'cut off'	gaacnobs 'introduce'	ak'etebs 'make'
PRS.1SG	vt'ex	vk'vec	vacnob	vak'eteb
PRS.2SG	t'ex	k'vec	cnob	ak'eteb
PRS.3SG	t'exs	k'vecs	cnobs	ak'etebs
PRS.1PL	vt'ext	vk'vect	vcnobt	vak'etebt
PRS.2PL	t'ext	k'vect	cnobt	ak'etebt
PRS.3PL	t'exen	k'vecen	cnoben	ak'eteben
IMPERF.1SG	vt'exdi	vk'vecdi	vcnobdi	vak'etebdi
IMPERF.2SG	t'exdi	k'vecdi	cnobdi	ak'etebdi
IMPERF.3SG	t'exda	k'vecda	cnobda	ak'etebda
IMPERF.1PL	vt'exdit	vk'vecdit	vcnobdit	vak'etebdit
IMPERF.2PL	t'exdit	k'vecdit	cnobdit	ak'etebdit
IMPERF.3PL	t'exdnen	k'vecdnen	cnobden	ak'etebdnen
FUT.1SG	gavt'ex	ševk'vec	gavcnob	gavak'eteb
FUT.2SG	gat'ex	šek'vec	gacnob	gaak'eteb
FUT.3SG	gat'exs	šek'vecs	gacnobs	gaak'etebs
FUT.1PL	gavt'ext	ševk'vect	gavacnobt	gavak'etebt
FUT.2PL	gat'ext	šek'vect	gacnobt	gaak'etebt
FUT.3PL	gat'exen	šek'vecen	gacnoben	gaak'eteben
AOR.1SG	gavt'exe	ševk'vece	gavcne	gavak'ete
AOR.2SG	gat'exe	šek'vece	gacne	gaak'ete
AOR.3SG	gat'exa	šek'veca	gacna	gaak'eta
AOR.1PL	gavt'exet	ševk'vecet	gavcnet	gavak'etet
AOR.2PL	gat'exet	šek'vecet	gacnet	gaak'etet
AOR.3PL	gat'exes	šek'veces	gacnes	gaak'etes

Examples: Georgian

cell	t'exs 'break'	k'vecs 'cut off'	gaacnobs 'introduce'	ak'etebs 'make'
PRS.1SG	vt'ex	vk'vec	vacnob	vak'eteb
PRS.2SG	t'ex	k'vec	cnob	ak'eteb
PRS.3SG	t'exs	k'vecs	cnobs	ak'etebs
PRS.1PL	vt'ext	vk'vect	vcnobt	vak'etebt
PRS.2PL	t'ext	k'vect	cnobt	ak'etebt
PRS.3PL	t'exen	k'vecen	cnoben	ak'eteben
IMPERF.1SG	vt'exdi	vk'vecdi	vcnobdi	vak'etebdi
IMPERF.2SG	t'exdi	k'vecdi	cnobdi	ak'etebdi
IMPERF.3SG	t'exda	k'vecda	cnobda	ak'etebda
IMPERF.1PL	vt'exdit	vk'vecdit	vcnobdit	vak'etebdit
IMPERF.2PL	t'exdit	k'vecdit	cnobdit	ak'etebdit
IMPERF.3PL	t'exdnen	k'vecdnen	cnobden	ak'etebdnen
FUT.1SG	gavt'ex	ševk'vec	gavcnob	gavak'eteb
FUT.2SG	gat'ex	šek'vec	gacnob	gaak'eteb
FUT.3SG	gat'exs	šek'vecs	gacnobs	gaak'etebs
FUT.1PL	gavt'ext	ševk'vect	gavacnobt	gavak'etebt
FUT.2PL	gat'ext	šek'vect	gacnobt	gaak'etebt
FUT.3PL	gat'exen	šek'vecen	gacnoben	gaak'eteben
AOR.1SG	gavt'exe	ševk'vece	gavcne	gavak'ete
AOR.2SG	gat'exe	šek'vece	gacne	gaak'ete
AOR.3SG	gat'exa	šek'veca	gacna	gaak'eta
AOR.1PL	gavt'exet	ševk'vecet	gavcnet	gavak'etet
AOR.2PL	gat'exet	šek'vecet	gacnet	gaak'etet
AOR.3PL	gat'exes	šek'veces	gacnes	gaak'etes

Examples: Georgian

cell	t'exs 'break'	k'vecs 'cut off'	gaacnobs 'introduce'	ak'etebs 'make'
PRS.1SG	**v**t'ex	**v**k'vec	**va**cnob	**va**k'eteb
PRS.2SG	t'ex	k'vec	cnob	ak'eteb
PRS.3SG	t'ex**s**	k'vec**s**	cnob**s**	ak'eteb**s**
PRS.1PL	**v**t'ex**t**	**v**k'vec**t**	**v**cnob**t**	**v**ak'eteb**t**
PRS.2PL	t'ex**t**	k'vec**t**	cnob**t**	ak'eteb**t**
PRS.3PL	t'ex**en**	k'vec**en**	cnob**en**	ak'eteb**en**
IMPERF.1SG	**v**t'ex**di**	**v**k'vec**di**	**v**cnob**di**	**v**ak'eteb**di**
IMPERF.2SG	t'ex**di**	k'vec**di**	cnob**di**	ak'eteb**di**
IMPERF.3SG	t'ex**da**	k'vec**da**	cnob**da**	ak'eteb**da**
IMPERF.1PL	**v**t'ex**dit**	**v**k'vec**dit**	**v**cnob**dit**	**v**ak'eteb**dit**
IMPERF.2PL	t'ex**dit**	k'vec**dit**	cnob**dit**	ak'eteb**dit**
IMPERF.3PL	t'ex**dnen**	k'vec**dnen**	cnob**den**	ak'eteb**dnen**
FUT.1SG	**ga**vt'ex	**še**vk'vec	**ga**vcnob	**ga**vak'eteb
FUT.2SG	**ga**t'ex	**še**k'vec	**ga**cnob	**ga**ak'eteb
FUT.3SG	**ga**t'ex**s**	**še**k'vec**s**	**ga**cnob**s**	**ga**ak'eteb**s**
FUT.1PL	**ga**vt'ex**t**	**še**vk'vec**t**	**ga**vcnob**t**	**ga**vak'eteb**t**
FUT.2PL	**ga**t'ex**t**	**še**k'vec**t**	**ga**cnob**t**	**ga**ak'eteb**t**
FUT.3PL	**ga**t'ex**en**	**še**k'vec**en**	**ga**cnob**en**	**ga**ak'eteb**en**
AOR.1SG	**ga**vt'ex**e**	**še**vk'vec**e**	**ga**vcn**e**	**ga**vak'et**e**
AOR.2SG	**ga**t'ex**e**	**še**k'vec**e**	**ga**cn**e**	**ga**ak'et**e**
AOR.3SG	**ga**t'ex**a**	**še**k'vec**a**	**ga**cn**a**	**ga**ak'et**a**
AOR.1PL	**ga**vt'ex**et**	**še**vk'vec**et**	**ga**vcn**et**	**ga**vak'et**et**
AOR.2PL	**ga**t'ex**et**	**še**k'vec**et**	**ga**cn**et**	**ga**ak'et**et**
AOR.3PL	**ga**t'ex**es**	**še**k'vec**es**	**ga**cn**es**	**ga**ak'et**es**

Examples: Georgian

cell	t'exs 'break'	k'vecs 'cut off'	gaacnobs 'introduce'	ak'etebs 'make'
PRS.1SG	vt'ex	vk'vec	vacnob	vak'eteb
PRS.2SG	**t'ex**	**k'vec**	cnob	ak'eteb
PRS.3SG	t'exs	k'vecs	cnobs	ak'etebs
PRS.1PL	vt'ext	vk'vect	vcnobt	vak'etebt
PRS.2PL	t'ext	k'vect	cnobt	ak'etebt
PRS.3PL	t'exen	k'vecen	cnoben	ak'eteben
IMPERF.1SG	vt'exdi	vk'vecdi	vcnobdi	vak'etebdi
IMPERF.2SG	t'exdi	k'vecdi	cnobdi	ak'etebdi
IMPERF.3SG	t'exda	k'vecda	cnobda	ak'etebda
IMPERF.1PL	vt'exdit	vk'vecdit	vcnobdit	vak'etebdit
IMPERF.2PL	t'exdit	k'vecdit	cnobdit	ak'etebdit
IMPERF.3PL	t'exdnen	k'vecdnen	cnobden	ak'etebdnen
FUT.1SG	gavt'ex	ševk'vec	gavcnob	gavak'eteb
FUT.2SG	gat'ex	šek'vec	gacnob	gaak'eteb
FUT.3SG	gat'exs	šek'vecs	gacnobs	gaak'etebs
FUT.1PL	gavt'ext	ševk'vect	gavacnobt	gavak'etebt
FUT.2PL	gat'ext	šek'vect	gacnobt	gaak'etebt
FUT.3PL	gat'exen	šek'vecen	gacnoben	gaak'eteben
AOR.1SG	gavt'exe	ševk'vece	gavcne	gavak'ete
AOR.2SG	gat'exe	šek'vece	gacne	gaak'ete
AOR.3SG	gat'exa	šek'veca	gacna	gaak'eta
AOR.1PL	gavt'exet	ševk'vecet	gavcnet	gavak'etet
AOR.2PL	gat'exet	šek'vecet	gacnet	gaak'etet
AOR.3PL	gat'exes	šek'veces	gacnes	gaak'etes

Zero forms in nominal paradigms
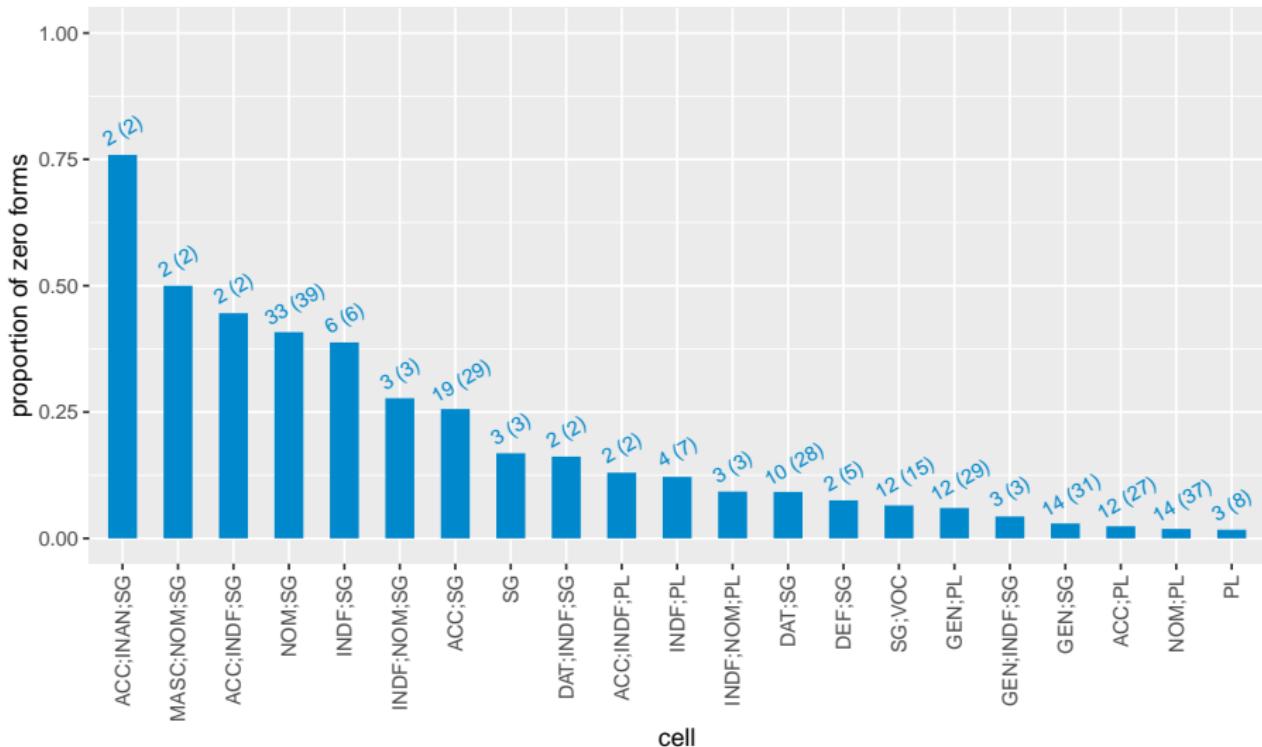
Cells with zero forms in nominal paradigms

Cells with zero forms in nominal paradigms

Conditional effects

Bayesian logistic regression + phylogenetic regression term as phylogenetic control
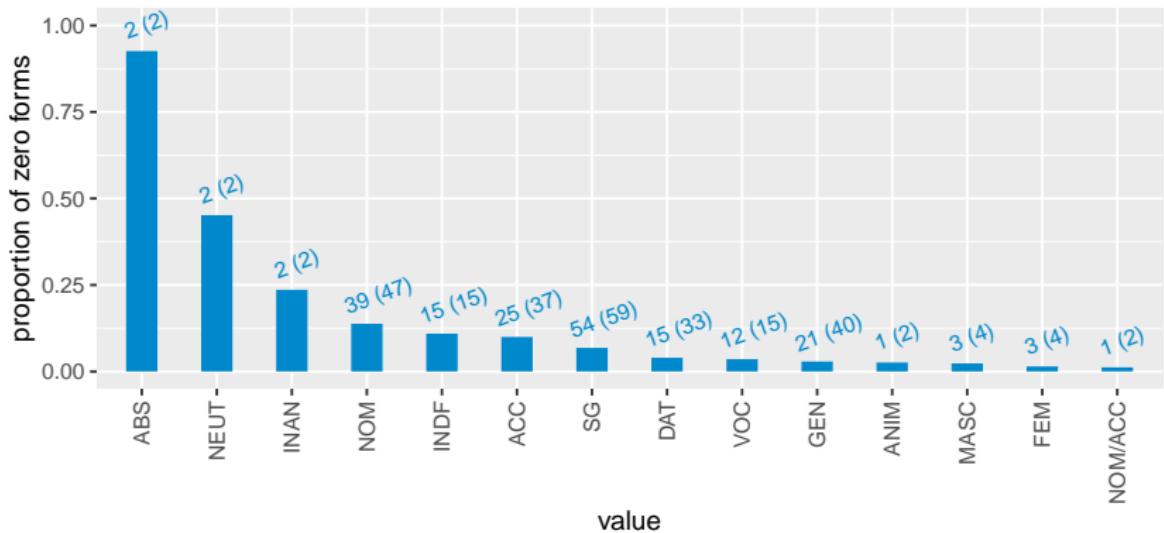
# Values with zero forms in nominal paradigms

Conditional effects

Bayesian logistic regression + phylogenetic regression term as phylogenetic control

# Zero forms in verbal paradigms
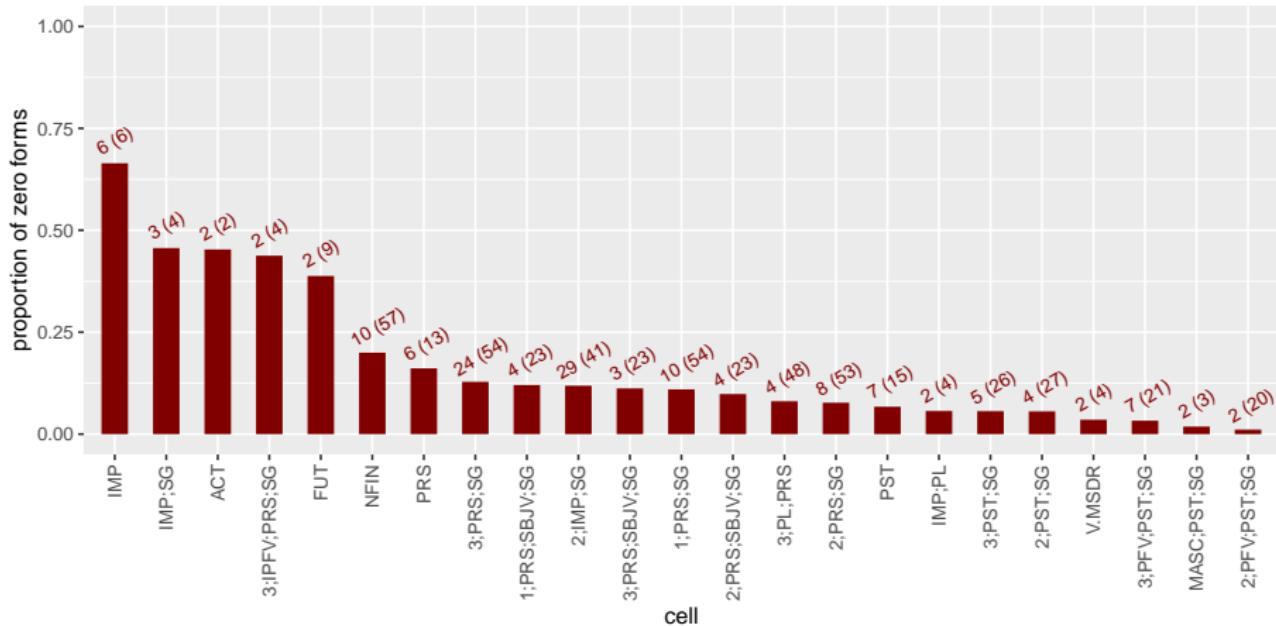
Cells with zero forms in verbal paradigms

Cells with zero forms in verbal paradigms

Conditional effects

Bayesian logistic regression + phylogenetic regression term as phylogenetic control
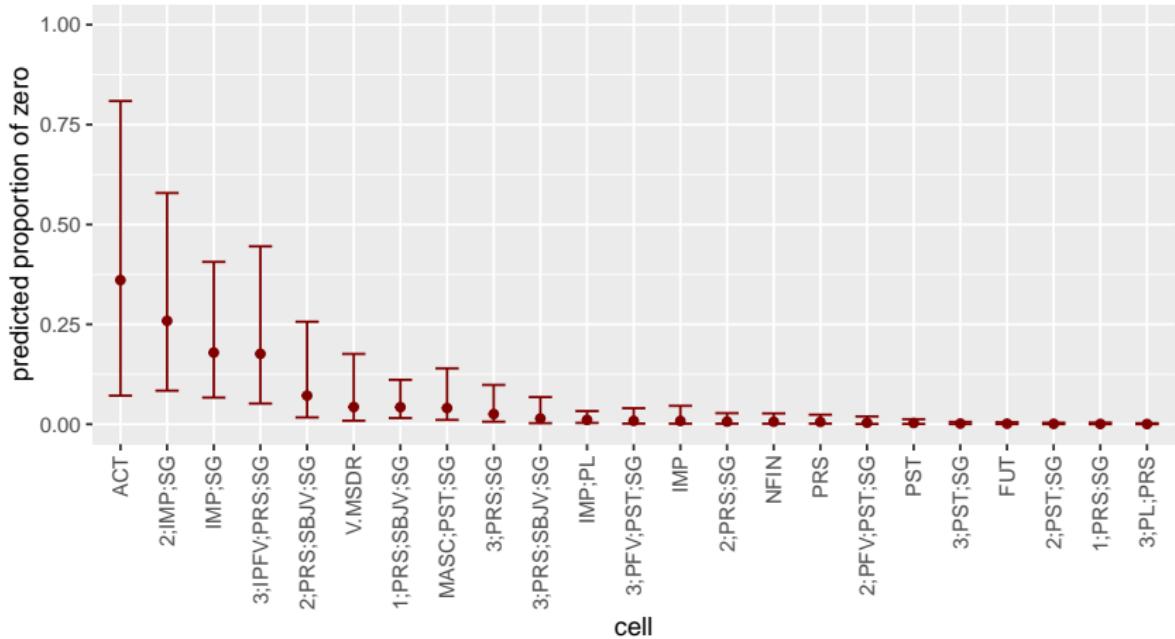
Values with zero forms in verbal paradigms
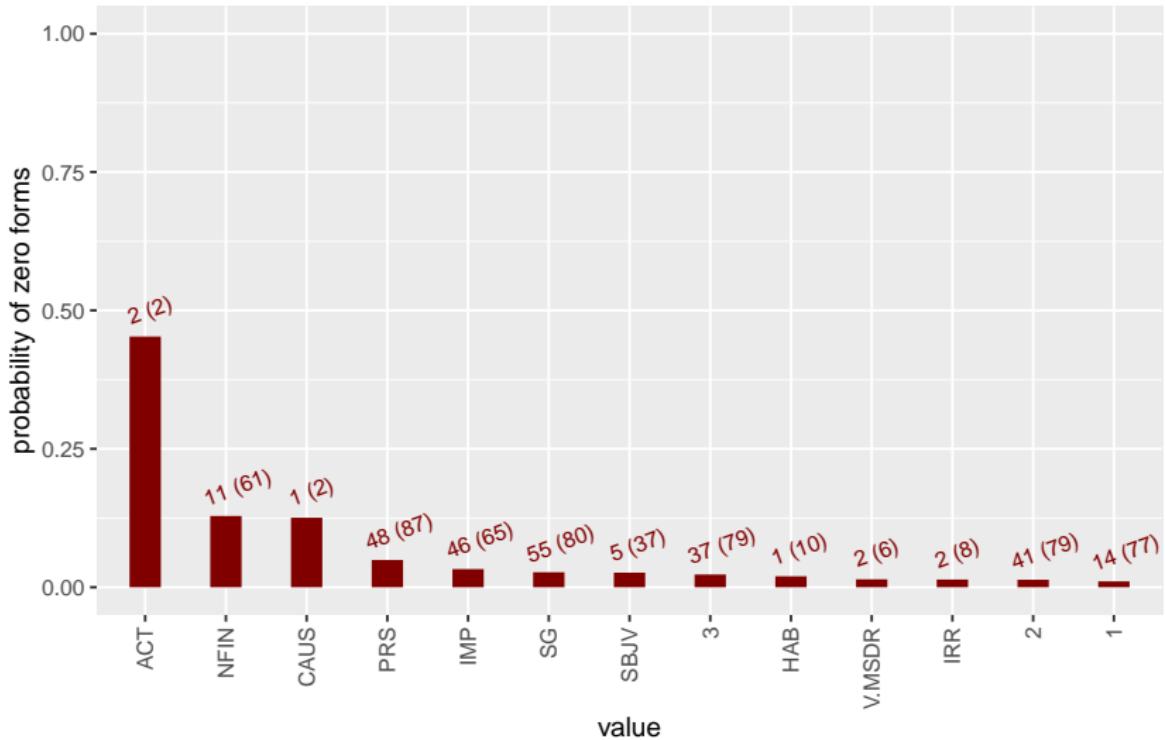
Values with zero forms in verbal paradigms

Conditional effects

Bayesian logistic regression + phylogenetic regression term as phylogenetic control
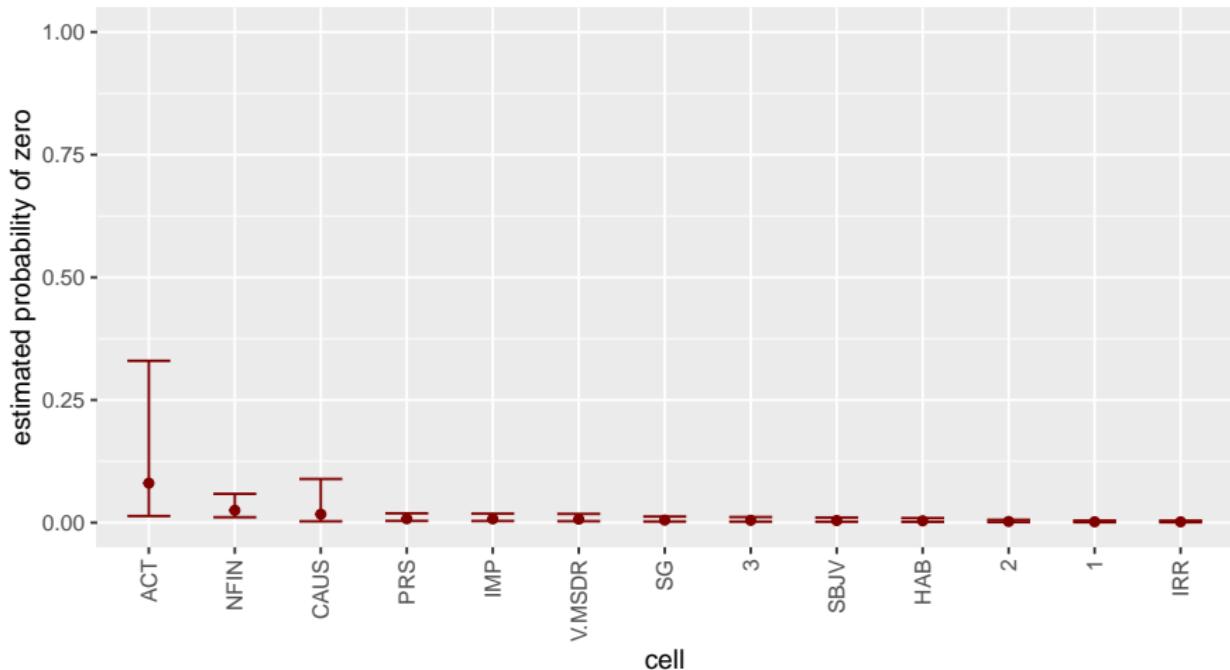
Implications

Zero forms and coding efficiency

Coding efficiency does not fully explain the distribution of zero forms.

Zero forms and coding efficiency

Coding efficiency does not fully explain the distribution of zero forms.

- zero forms are not preferred in any morphosyntactic functions

Zero forms and coding efficiency

Coding efficiency does not fully explain the distribution of zero forms.

- zero forms are not preferred in any morphosyntactic functions

- especially for the cells and values with higher estimated probabilities of zero, uncertainty intervals were very large

☞ high degree of variation
  - across lexemes within languages
  - across languages

- (despite zero forms being overall very common in inflectional morphology)

Zero forms and coding efficiency

Coding efficiency does not fully explain the distribution of zero forms.

- zero forms are not preferred in any morphosyntactic functions

- especially for the cells and values with higher estimated probabilities of zero, uncertainty intervals were very large

☞ high degree of variation
  - across lexemes within languages
  - across languages

- (despite zero forms being overall very common in inflectional morphology)

- in line with previous work: Bickel et al. (2015), Cysouw (2003), Guzmán Naranjo and Becker (2021a), Seržant and Moroz (2022), Siewierska (2010), and Stolz and Levkovych (2019)

No form-frequency effects at all?

- the values with the strongest association with zero forms tend to be more frequent than other values of the feature

Nominal features in the Universal Dependencies					
		N value	N total	N langs	prop value
number	singular	3789088	5144632	80	0.74
	plural	1347473	5152931	88	0.26
	dual	5472	1080628	11	0.01
case	nominative	887053	3166314	57	0.28
	accusative	708587	2868514	46	0.25
	absolutive	9801	32220	5	0.30
	ergative	1984	22465	3	0.09
	dative	234759	3182335	44	0.09
	genitive	862330	3070696	56	0.28
	vocative	6205	1849520	24	0.003
definiteness	definite	384650	971411	17	0.40
	indefinite	458271	957243	16	0.48

No form-frequency effects at all?

- the values with the strongest association with zero forms tend to be more frequent than other values of the feature

Nominal features in the Universal Dependencies		N value	N total	N langs	prop value
number	singular	3789088	5144632	80	**0.74**
	plural	1347473	5152931	88	**0.26**
	dual	5472	1080628	11	**0.01**
case	nominative	887053	3166314	57	0.28
	accusative	708587	2868514	46	0.25
	absolutive	9801	32220	5	0.30
	ergative	1984	22465	3	0.09
	dative	234759	3182335	44	0.09
	genitive	862330	3070696	56	0.28
	vocative	6205	1849520	24	0.003
definiteness	definite	384650	971411	17	0.40
	indefinite	458271	957243	16	0.48

No form-frequency effects at all?

- the values with the strongest association with zero forms tend to be more frequent than other values of the feature

Nominal features in the Universal Dependencies		N value	N total	N langs	prop value
number	singular	3789088	5144632	80	0.74
	plural	1347473	5152931	88	0.26
	dual	5472	1080628	11	0.01
case	nominative	887053	3166314	57	**0.28**
	accusative	708587	2868514	46	**0.25**
	absolutive	9801	32220	5	0.30
	ergative	1984	22465	3	0.09
	dative	234759	3182335	44	0.09
	genitive	862330	3070696	56	**0.28**
	vocative	6205	1849520	24	0.003
definiteness	definite	384650	971411	17	0.40
	indefinite	458271	957243	16	0.48

No form-frequency effects at all?

- the values with the strongest association with zero forms tend to be more frequent than other values of the feature

Nominal features in the Universal Dependencies		N value	N total	N langs	prop value
number	singular	3789088	5144632	80	0.74
	plural	1347473	5152931	88	0.26
	dual	5472	1080628	11	0.01
case	nominative	887053	3166314	57	0.28
	accusative	708587	2868514	46	0.25
	absolutive	9801	32220	5	**0.30**
	ergative	1984	22465	3	**0.09**
	dative	234759	3182335	44	0.09
	genitive	862330	3070696	56	0.28
	vocative	6205	1849520	24	0.003
definiteness	definite	384650	971411	17	0.40
	indefinite	458271	957243	16	0.48

No form-frequency effects at all?

- the values with the strongest association with zero forms tend to be more frequent than other values of the feature

Nominal features in the Universal Dependencies		N value	N total	N langs	prop value
number	singular	3789088	5144632	80	0.74
	plural	1347473	5152931	88	0.26
	dual	5472	1080628	11	0.01
case	nominative	887053	3166314	57	0.28
	accusative	708587	2868514	46	0.25
	absolutive	9801	32220	5	0.30
	ergative	1984	22465	3	0.09
	dative	234759	3182335	44	0.09
	genitive	862330	3070696	56	0.28
	vocative	6205	1849520	24	0.003
definiteness	definite	384650	971411	17	**0.40**
	indefinite	458271	957243	16	**0.48**

No form-frequency effects at all?

- the values with the strongest association with zero forms tend to be more frequent than other values of the feature

Verbal features in the Universal Dependencies		N value	N total	N langs	prop value
tense	past	739008	1666938	80	0.44
	present	857398	1667680	77	0.51
	future	47114	1049995	44	0.04
mood	indicative	1216937	1329281	75	0.92
	imperative	37822	1331878	77	0.03
person	1	132245	1232404	78	0.11
	2	68328	1232234	73	0.06
	3	1021191	1231610	76	0.83
voice	active	1032873	1227679	40	0.84
	passive	144708	1011323	52	0.14

No form-frequency effects at all?

- the values with the strongest association with zero forms tend to be more frequent than other values of the feature

Verbal features in the Universal Dependencies				
		N value	N total	N langs	prop value
tense	past	739008	1666938	80	**0.44**
	present	857398	1667680	77	**0.51**
	future	47114	1049995	44	**0.04**
mood	indicative	1216937	1329281	75	0.92
	imperative	37822	1331878	77	0.03
person	1	132245	1232404	78	0.11
	2	68328	1232234	73	0.06
	3	1021191	1231610	76	0.83
voice	active	1032873	1227679	40	0.84
	passive	144708	1011323	52	0.14

No form-frequency effects at all?

- the values with the strongest association with zero forms tend to be more frequent than other values of the feature

Verbal features in the Universal Dependencies		N value	N total	N langs	prop value
tense	past	739008	1666938	80	0.44
	present	857398	1667680	77	0.51
	future	47114	1049995	44	0.04
mood	indicative	1216937	1329281	75	**0.92**
	imperative	37822	1331878	77	**0.03**
person	1	132245	1232404	78	0.11
	2	68328	1232234	73	0.06
	3	1021191	1231610	76	0.83
voice	active	1032873	1227679	40	0.84
	passive	144708	1011323	52	0.14

No form-frequency effects at all?

- the values with the strongest association with zero forms tend to be more frequent than other values of the feature

Verbal features in the Universal Dependencies		N value	N total	N langs	prop value
tense	past	739008	1666938	80	0.44
	present	857398	1667680	77	0.51
	future	47114	1049995	44	0.04
mood	indicative	1216937	1329281	75	0.92
	imperative	37822	1331878	77	0.03
person	1	132245	1232404	78	**0.11**
	2	68328	1232234	73	**0.06**
	3	1021191	1231610	76	**0.83**
voice	active	1032873	1227679	40	0.84
	passive	144708	1011323	52	0.14

No form-frequency effects at all?

- the values with the strongest association with zero forms tend to be more frequent than other values of the feature

Verbal features in the Universal Dependencies		N value	N total	N langs	prop value
tense	past	739008	1666938	80	0.44
	present	857398	1667680	77	0.51
	future	47114	1049995	44	0.04
mood	indicative	1216937	1329281	75	0.92
	imperative	37822	1331878	77	0.03
person	1	132245	1232404	78	0.11
	2	68328	1232234	73	0.06
	3	1021191	1231610	76	0.83
voice	active	1032873	1227679	40	**0.84**
	passive	144708	1011323	52	**0.14**

Zero forms as a diachronic by-product

Processes that lead to the development of zero forms

- differential phonetic reduction
- differential non-development
- differential morphosyntactic reduction
- reanalysis

- zero forms rather seem to be a diachronic by-product of other processes (cf. Cristofaro 2019; Cristofaro 2021)
- those processes may be accounted for by extra-linguistic factors such as efficiency
- but they may depend on very language-specific and construction-specific factors

Conclusion & outlook

- although zero forms are common in both nominal and verbal inflection morphology, no strong robust patterns emerge across lexemes and languages
- there is no morphosyntactic cell or value that would show a robust preference for zero marking
- this is not compatible with an efficiency-based explanation commonly used in typology
- zero forms rather appear to be the by-product of various independent diachronic processes

Conclusion & outlook

- although zero forms are common in both nominal and verbal inflection morphology, no strong robust patterns emerge across lexemes and languages
- there is no morphosyntactic cell or value that would show a robust preference for zero marking
- this is not compatible with an efficiency-based explanation commonly used in typology
- zero forms rather appear to be the by-product of various independent diachronic processes

To do

- test the distribution of zero forms in naturalistic corpus data
- ☞ DoReCo data
- better definition of zero forms:
  with usage frequencies, forms can be weighted for determining the longest common substring

Thank you!

References I

Anderson, Stephen R. (1992). *A-Morphous Morphology.* Cambridge: Cambridge University Press.

Beckman, Jill (1998). "Positional Faithfulness." PhD thesis. Amherst: University of Massachusetts.

Beniamine, Sacha and Matías Guzmán Naranjo (2021). "Multiple Alignments of Inflectional Paradigms." In: *Proceedings of the Society for Computation in Linguistics* 4, Article 21.

Bickel, Balthasar et al. (2015). "Exploring Diachronic Universals of Agreement: Alignment Patterns and Zero Marking across Person Categories." In: *Agreement from a Diachronic Perspective.* Ed. by Jürg Fleischer, Elisabeth Rieken, and Paul Widmer. Berlin: De Gruyter, pp. 29–52.

Blevins, James P. (2016). *Word and Paradigm Morphology.* Oxford: Oxford University Press.

Bybee, Joan (2011). "Markedness." In: *The Oxford Handbook of Typology.* Ed. by Jae Jung Song. Oxford: Oxford University Press, pp. 1–11.

Bybee, Joan, William Pagliuca, and Revere Perkins (1990). "On the Asymmetries in the Affixation of Grammatical Material." In: *Studies in Typology and Diachrony. Papers Presented to Joseph H. Greenberg on His 75th Birthday.* Ed. by William Croft, Suzanne Kemmer, and Keith Denning. Amsterdam: Benjamins, pp. 1–42.

Cho, Taehong and Patricia Keating (2001). "Articulatory and Acoustic Studies on Domain-Initial Strengthening in Korean." In: *Journal of Phonetics* 29.2, pp. 155–190.

Cho, Taehong and Patricia Keating (2009). "Effects of Initial Position versus Prominence in English." In: *Journal of Phonetics* 37.4, pp. 466–485.

Cho, Taehong, James McQueen, and Ethan Cox (2007). "Prosodically Driven Phonetic Detail in Speech Processing: The Case of Domain-Initial Strengthening in English." In: *Journal of Phonetics* 35, pp. 210–243.

Cristofaro, Sonia (2019). "Taking Diachronic Evidence Seriously: Result-oriented vs. Source-Oriented Explanations of Typological Universals." In: *Explanation in Typology: Diachronic Sources, Functional Motivations and the Nature of the Evidence.* Ed. by Karsten Schmidtke-Bode et al. Conceptual Foundations of Language Science. Berlin: Language Science Press, pp. 25–46.

Cristofaro, Sonia (Aug. 2021). "Typological Explanations in Synchrony and Diachrony: On the Origins of Third Person Zeroes in Bound Person Paradigms." In: *Folia Linguistica* 55.s42-s1, pp. 25–48.

Croft, William (2003). *Typology and Universals.* 2nd edn. Cambridge: Cambridge University Press.

References II

Cysouw, Michael (2003). *The Paradigmatic Structure of Person Marking*. Oxford: Oxford University Press.

Cysouw, Michael (2009). "The Asymmetry of Affixation." In: *Snippets* 20, pp. 10–144.

Diessel, Holger (2019). *The Grammar Network: How Linguistic Structure Is Shaped by Language Use.* Cambridge: Cambridge University Press.

Fougeron, Cécile (2001). "Articulatory Properties of Initial Segments in Several Prosodic Constituents in French." In: *Journal of Phonetics* 29.2, pp. 109–135.

Garcia, Erica C. and Florimon van Putte (Jan. 1989). "Forms Are Silver, Nothing Is Gold." In: *Folia Linguistica Historica* 21, pp. 365–384.

Greenberg, Joseph Harold (1966). *Language Universals: With Special Reference to Feature Hierarchies.* The Hague: Mouton.

Guzmán Naranjo, Matías and Laura Becker (2021a). "Coding Efficiency in Nominal Inflection: Expectedness and Type Frequency Effects." In: *Linguistics Vanguard* 7.s3.

Guzmán Naranjo, Matías and Laura Becker (2021b). "Statistical Bias Control in Typology." In: *Linguistic Typology*.

Hall, Christopher (1988). "Integrating Diachronic and Processing Principles in Explaining the Suffixing Preference." In: *Explaining Language Universals.* Ed. by John Hawkins. London: Basil Blackwell, pp. 321–349.

Haspelmath, Martin (2008a). "A Frequentist Explanation of Some Universals of Reflexive Marking." In: *Linguistic Discovery* 6.1.

Haspelmath, Martin (2008b). "Frequency vs. Iconicity in Explaining Grammatical Asymmetries." In: *Cognitive Linguistics* 19.1, pp. 1–33.

Haspelmath, Martin (2021). "Explaining Grammatical Coding Asymmetries: Form–Frequency Correspondences and Predictability." In: *Journal of Linguistics*, pp. 1–29.

Jakobson, Roman (1939). "Signe Zéro." In: *Mélanges de Linguistique Offerts à Charles Bally*. Genève: Georg, pp. 143–152.

Keating, Patricia et al. (2003). "Domain-Initial Strengthening in Four Languages." In: *Phonetic Interpretation: Papers in Laboratory Phonology* 6.

Kim, Sahyang (2004). "The Role of Prosodic Phrasing in Korean Word Segmentation." PhD thesis. University of California, Los Angeles.

Lemaréchal, Alain (1997). *Zéro(s).* Paris: Presses Universitaires de France.

References III

McCarthy, Arya D. et al. (May 2020). "UniMorph 3.0: Universal Morphology." In: *Proceedings of the 12th Language Resources and Evaluation Conference.* Marseille, France: European Language Resources Association, pp. 3922–3931.

Mel'cuk, Igor (2002). "Towards a Formal Concept Zero Linguistic Sign: Applications in Typology." In: *Morphology 2000: Selected Papers from the 9th Morphology Meeting, Vienna, 24–28 February 2000.* Ed. by Sabrina Bendjaballah et al. Amsterdam: Benjamins, pp. 241–258.

Mortensen, David R, Siddharth Dalmia, and Patrick Littell (2018). "Epitran: Precision G2P for Many Languages." In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).*

Pullum, Geoffrey and Arnold Zwicky (1991). "A Misconceived Approach to Morphology." In: *Proceedings of the West Coast Conference on Formal Linguistics* 10.

Seržant, Ilja A. and George Moroz (Feb. 2022). "Universal Attractors in Language Evolution Provide Evidence for the Kinds of Efficiency Pressures Involved." In: *Humanities and Social Sciences Communications* 9.1, pp. 1–9.

Siewierska, Anna (2010). "Person Asymmetries in Zero Expression and Grammatical Functions." In: *Essais de Linguistique Generale et de Typologie Linguistique Offerts Au Professeur Denis Creissels à l'occasion de Ses 65 Ans.* Ed. by Franck Floricic. Paris: Presses de l'École Normale Supérieure, pp. 425–438.

Smith, Jennifer (2005). *Phonological Augmentation in Prominent Positions.* Oxfordshire: Taylor & Francis.

Song, Jae Jung (2018). *Linguistic Typology.* Oxford: Oxford University Press.

Stolz, Thomas and Nataliya Levkovych (Sept. 2019). "Absence of Material Exponence." In: *Language Typology and Universals* 72.3, pp. 373–400.

White, Laurence, Silvia Benavides-Varela, and Katalin Mády (2020). "Are Initial-Consonant Lengthening and Final-Vowel Lengthening Both Universal Word Segmentation Cues?" In: *Journal of Phonetics* 81, p. 100982.

The probability of zero forms

- a priori, we would not expect the probability of zero forms to differ in nominal vs. verbal paradigms

- a priori, we would not expect the probability of zero forms to differ in nominal vs. verbal paradigms

- but:
nouns	57	out of	61 languages	(90%)
verbs	76	out of	104 languages	(73%)

# The probability of zero forms

- a priori, we would not expect the probability of zero forms to differ in nominal vs. verbal paradigms

- but:

nouns	57	out of	61 languages	(90%)
verbs	76	out of	104 languages	(73%)

- confounding factors:
  - **number of lexemes** for which paradigms are available in a language
  - **number of cells** a paradigm has
 - **number of features** encoded in the cells of a paradigm

The probability of zero forms

- a priori, we would not expect the probability of zero forms to differ in nominal vs. verbal paradigms

- but:

nouns	57	out of	61 languages	(90%)
verbs	76	out of	104 languages	(73%)

- confounding factors:
  - **number of lexemes** for which paradigms are available in a language
  - **number of cells** a paradigm has
  - **number of features** encoded in the cells of a paradigm

- I fitted a series of Bayesian logistic regression models with phylogenetic regression term as a phylogenetic control (cf. Guzmán Naranjo and Becker 2021b)

- the best-performing model includes pos and N(lexemes) as fixed effects

The probability of zero forms

Conditional effects
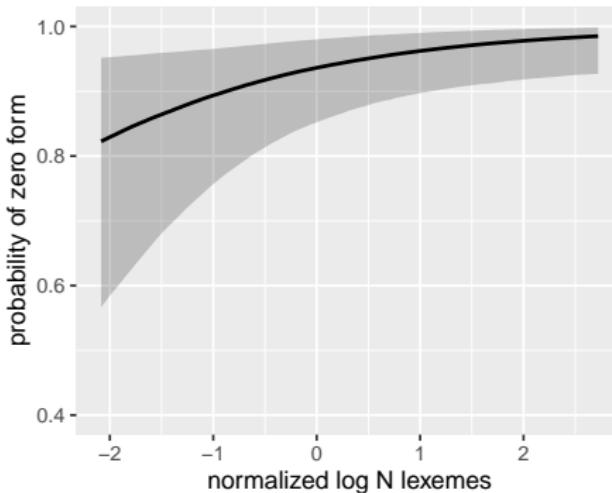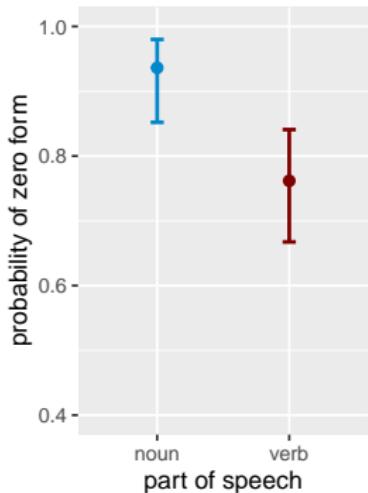
- zero forms are indeed more likely to occur in nominal paradigms than in verbal paradigms

Zero forms in nominal vs. verbal paradigms

Suffixing vs. prefixing preference?

- nouns have a stronger suffixing preference than verbs (Bybee, Pagliuca, and Perkins 1990)
- verbal agreement markers with a smaller number of markers (≤ 6) have a prefixing preference (Cysouw 2009)
- phonetic material may be reduced at higher rates at the end of words (Bybee, Pagliuca, and Perkins 1990; Hall 1988)
- word-initial syllables are more prominent than other syllables (Beckman 1998; Cho, McQueen, and Cox 2007; Keating et al. 2003; Kim 2004; Smith 2005)
- word-initial consonants tend to be strengthened and lengthened (Cho and Keating 2001; Cho and Keating 2009; Fougeron 2001; White, Benavides-Varela, and Mády 2020)
- prefixes may be more time-stable than suffixes (Bybee, Pagliuca, and Perkins 1990)