Split-S systems, language endangerment and the languages of the Caucasus

James Baker, University of Cambridge

Third workshop on Romeyka and the languages of the eastern Black Sea

Round table *Emerging Ergativity as a Sprachbund feature*

Queens' College, Cambridge, 27th May 2016

1   Introduction

- Thanks and disclaimer

- *Split-S* (here): a system in which different intransitive predicates allow different case-/agreement-marking possibilities

    - Also known as *active* and by various other names (see Dixon 1994: 84).

(1)   Georgian (Harris 1981: 40, 147; Harris 1982: 293):

    a. glex-ma   datesa   simindi
       farmer-**ERG** he.sowed.it corn.**NOM**

       'the farmer sowed corn'

    b. Nino-m   daamtknara
       Nino-**ERG** she.yawned

       'Nino yawned'

    c. Rezo   gamoizarda
       Rezo.**NOM** he.grew.up

       'Rezo grew up'

- I include *fluid-S* languages under the general banner of split-S.

(2)   Tsova-Tush (Holisky 1987: 105):

    a. as   wože
       1SG.**ERG** fell

       'I fell' (it was my fault)

    b. so   wože
       1SG.**NOM** fell

       'I fell' (no implication that it was my fault)

- Outline:

  - Split-S: an endangered language type? (section 2)

  - Split-S in the Caucasus (section 3)

    * including *variation* in split-S systems

2   Split-S: an endangered language type?

- No!

- Split-S systems are rare?

    – Yes: about 2% of case systems (Comrie 2013) and 7% of agreement systems (Siewierska 2013) follow a split-S pattern according to the *World Atlas of Language Structures.*

    – But this does not by itself mean that the type is particularly endangered.

- Split-S systems are diachronically unstable?

    – This may account for their typological rarity.

    – Could instability of split-S systems contribute to increased likelihood of their extinction?

    – It's often the case that split-S languages are present in families alongside relatives that are *not* split-S, for example:

        * **Pomoan languages of California:**

            · 7 languages (one extinct), of which 4 split-/fluid-S, 1 nominative-accusative, 2 on which data unavailable.

            · It may be reasonable to reconstruct a split-S system for Proto-Pomo, but this has been lost in at least one daughter language.

        * **Basque:**

            · Different dialects have different degrees of split-S; some are very almost canonical ergative-absolutive (Aldai 2009 i.a.).

· Is this evidence that split-S is being lost?

· No: systems with more pronounced split-S behaviour are innovative, arising in the last couple of centuries (Creissels and Mounole Creissels and Mounole, Berro 2012).

* **Northeast Caucasian**:

· Split-S systems reported in: Ingush, Tabassaran, Tsova-Tush, Udi, Budukh, (Lezgian) (see next section for references).

· But not universal to the family, e.g. erg:abs case + agreement in Tsezic languages (Comrie et al. 2013) and Lak (Comrie 2013 and Siewierska 2013);

· Hunzib is reported to have erg:abs case (Comrie 2013) and nom:acc agreement (Siewierska 2013).

* **Kartvelian**:

· Georgian, Laz and Svan are (partially) split-S; but Mingrelian is fully nominative-accusative (Harris 1985: 57):

    (3)   Mingrelian (Harris 1985: 57):

        a. koč-k(i) doʔvilu   ɣeǰ-i
            man-ERG he.kill.it.II pig-NOM
            'the man killed a pig'

        b. ʒɣabi-k (ko)sxapu
            girl-ERG she.dance.II
            'the girl danced'

        c. koč-k   doɣuru
            man-ERG he.die.II
            'the man died'

· But Harris (1985) argues that split-S developed from ergative-absolutive in Proto-Kartvelian; Mingrelian nom:acc is a change from this earlier split-S system.

– Other examples of split-S systems arising in recent times: Hindi, Tibetan (Denwood 1991: 266).

– The continuing emergence of new split-S systems may mitigate against other factors endangering the type.

• *Individual* split-S languages are generally endangered?

– True to a large extent:

  * E.g. a high concentration of split-S languages in the Americas (see Fig. 1), where a very high proportion of languages are endangered (Whalen and Simons 2012).

– But not all split-S languages are endangered:

  * e.g. Hindi (258 million L1 speakers), Georgian (4.3 million speakers), Tibetan (1.1 million speakers) (Lewis et al. 2016).

• Conclusion: the split-S type as a whole is not obviously endangered, though many individual split-S languages are.

3   Split-S in the Caucasus

• Split-S seems to be something of an areal feature amongst languages of the Caucasus, though it is by no means universal (see section 2).
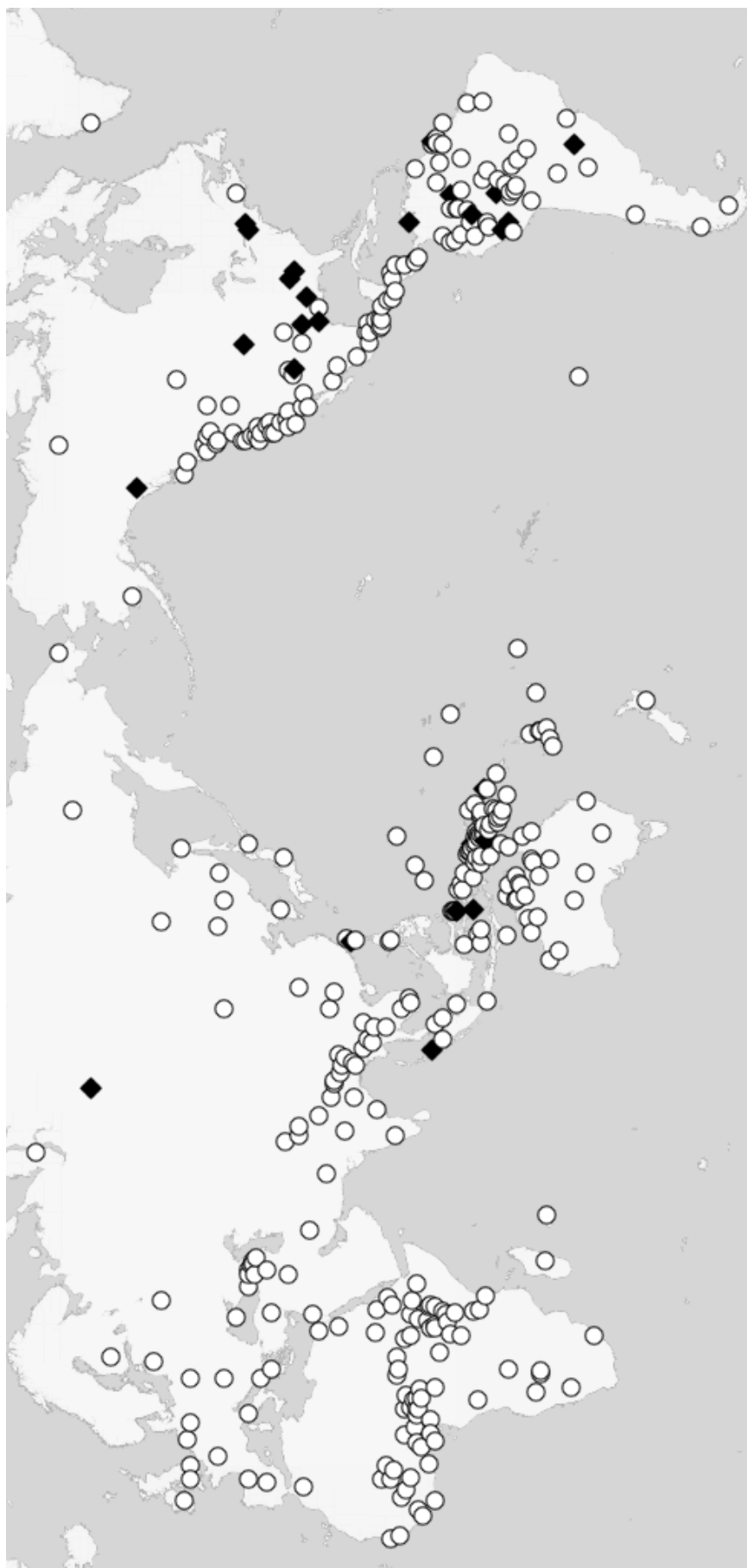
Figure 1: Agreement alignment in the world's language (Siewierska 2013). Black: split-S; white: other.

3.1   Caucasian split-S languages

- Northeast Caucasian:

  - **Ingush** (323,000 speakers[1]; status: vulnerable[2]):[3]

    (4) a. Muusaaz   nab-ju
            Musa.(v).**ERG** sleep(**J**)-**J**.AUX.PRES
            'Musa is sleeping'

         b. yz   chy-v.axar
             3SG.**V** in-go.**V**.WP
             'he went home'

                                         (Nichols 2008: 58, 59)

  - **Tabassaran** (127,000 speakers; status: vulnerable):

    (5) a. uzi harʔ aqun-**zu**
             1SG forest end.up-1SG.A
             'I end up in a forest'

         b. uzu hariz Rušun-**za**
             1SG forest end.up-1SG.P
             'I enter the forest'

                                              (Kibrik 1985: 277)

  - **Tsova-Tush** (Bats, Batsbi; possibly 'far fewer than 3000' active
    speakers; status: severely endangered.):

    * see example (2) above.

---

[1]Numbers of speakers for each language taken from *Ethnologue* (Lewis et al. 2016)
[2]Endangerment statuses taken from Moseley (2010).
[3]V, J = gender classes; WP = witnessed past.

– **Udi** (6500 speakers; status: severely endangered):[4]

> (6) a. äyel-en   ọne-ne-xa
>   child-**ERG** crying-3S-LV.PRES
>   'the child is crying'
>
>   b. d'yel   kana-ne-bak-e
>   child.**ABS** big-3SG-BE-AORII
>   'the child grew up'

(Harris 2002: 252, 253)

– **Budukh** (200 speakers; status: severely endangered) also displays split-S behaviour (Harris 2002: 255).

– **Lezgian** (617,000 speakers; status: vulnerable) has slight split-S patterns (see below).

- Kartvelian:

  – **Georgian** (4.3 million speakers):

  > * see example (1) above.

  – **Laz** (22,000 speakers; status: definitely endangered):

  > (7) a. ǰoɣo-epe-k-ti   lales
  >   dog-PL-**NAR**-also bark
  >   'the dogs barked'
  >
  >   b. bee   dirdu
  >   child-**NOM** grow
  >   'the child grew'

(Asatiani 1974: 44, 82)

---

[4]LV = light verb.

– **Svan** (15,000 speakers; status: definitely endangered):

(8) a. māre-d   čwadmǝšie
man-**ERG** work.AOR
'the man has worked'

b. č'q'int'   kaloxgidda dīna-s
boy.**NOM** looked.AOR girl-DAT
'the boy looked at the girl'

(Sumbatova 1993: 258)

3.2 Variation in split-S systems

- Globally, there is a lot of variation amongst split-S systems: different languages employ different conditioning factors for the case/agreement split in intransitive predicates, for example:

  – Control/volition: Koasati (Kimball 1991), Eastern Pomo (McLendon 1978) …

  – Performance/effectedness/instigation: Lakhota (Mithun 1991).

  – Eventivity/stativity/dynamicity: Baniwa do Içana (Danielson and Granadillo 2008), Galela (Creissels 2008) …

  – Multiple factors may interact, e.g. control and perspective in Northern Pomo (Deal and O'Connor Deal and O'Connor) ; control, eventivity and affectedness in Central Pomo and Caddo (Mithun 1991).

  – Languages may also have some degree of apparently idiosyncratic lexical variation, e.g. Mohawk *-yeshų* 'smile' always takes patientive marking even where agentive might be expected: namely,

where it describes a voluntary act (Mithun 1991: 533).

- This variation is also apparent amongst languages of the Caucasus.

  – **Tsova-Tush:** fluid-S, broadly related to control/volition (Holisky 1987):

    * most intransitives allow either ergative or nominative marking (though some are more commonly found with one or the other).

        (9) Tsova-Tush (ibid.: 105):

            a. as   wože
               1SG.**ERG** fell
               'I fell (intentionally)'

            b. so   wože
               1SG.**NOM** fell
               'I fell (non-intentionally)'

    * some verbs (which denote necessarily *uncontrolled* events/states) are only accepted with nominative: e.g. *maicdar* 'be hungry', *q'erl'ar* 'be afraid', *dah" ğordar* 'freeze' etc. (ibid.: 109).

    * others are only accepted with ergative: e.g. dağar 'come', *lalar* 'walk, wander'; *dadar* 'swear', *axar* 'bark', *loca(d)dalar* 'say', *cerdal'ar* 'finish', *da:xar* 'live', *tešar* 'believe', *lap'c'ar* 'play' (ibid.: 113).

        · Nb. even in these cases the opposite marking may not be *categorically* impossible, just strongly disprefered (see ibid.: 115).

* Holisky attributes the split primarily to volitionality, but with some complications (1987: §3.7), for example (p. 116):

  > The verb *h"alO dopxdalar* 'get dressed', for example, is usually used with ergative marking, whether or not the subject acts agentively. However, in the unusual situation sketched above, where one gets dressed unintentionally, *if* the speaker wishes to emphasize the unusual nature of the situation, nominative marking is available to express non-agentive dressing. It must be stressed that use of nominative marking in this situation is optional.

  Holisky goes on to provide a more complex characterisation of the patterns (1987: 4.3).

– **Tabassaran:** similar patterns to Tsova-Tush, though marked in terms of *agreement* (Kibrik 1985: 277–278, see also Arkadiev 2008: 108–109):

  * some verbs only occur with agentive marking (*-za* in 1SG): e.g. *daqun-za* 'I lay down', *Rit'itXan-za* 'I flew away', *Rižun-za* 'I began to cry', *RiliXun-za* 'I worked (for a while)', *Rušun-za* 'I came' (Kibrik 1985: 278);

  * others only occur with patientive marking (*-zu* in 1SG): e.g. *k̄abqun-zu* 'I drowned' *RaRIun-zu* 'I swelled', *RarRun-zu* 'I froze', *kelXun-zu* 'I hung', *ergra-zu* 'I got tired' (ibid.);

  * still others occur with both, depending on whether the action is intentional or not e.g. *Ruʒun-za/-zu* 'I remained', *aqun-za/-zu* 'I fell', *hilirqun-za/-zu* 'I shook'.

  * control/volitional again seems to be the main factor in the

split.

- **Udi:**

  * Some intransitive verbs have ergative subjects (in certain dialects): *gölös-p-* 'dance', *maɣ-p-* 'sing', *ači-p-* 'play', *ayt-p-* 'speak', *ọne-p-* 'cry', *füt'-p-* 'whistle', *axšum-p-* 'laugh', *xuru-p-* 'snore'; *üzmiš-b-* 'swim', *fikir-b-* 'think'; *dava-sak-* 'fight' (Harris 2002: 252, 254) and all verbs with similar semantics with the exception of *t'it'* 'run' (ibid: 253).

    · Such verbs are historically formally transitive (except possibly *t'it'*).

  * Other intransitives have absolutive subjects: *t'it'* 'run', *bas-k'-* 'be lying down', *kala-bak-* 'grow (up)', *q'ari-bak* 'dry' (ibid.: 253, 278).

  * my tentative analysis: intransitives denoting *states* or *changes of state or location* assign ABS; other intransitives assign ERG.

    · This is similar to the patterns found in other languages, e.g. Chol (Mayan, Coon 2010), and also possibly Georgian (see below).

- **Lezgian:** intransitive verbs derived from N+*do* compounds take ergative subjects (Haspelmath 1993: 284):

  (10) ada   k'walax-zawa
        she.ERG work-IMPF
        'She was working'

        (cf. *ada k'walax iji-zwa* 'she.ERG work do-IMPF')

* Other examples include *c'uǧ-* 'howl' and *q'üler-* 'dance' (ibid.).

* There may also be some non-derived ergative-assigning intransitives (e.g. *xkadurun* 'run', *čukurun* 'chase')—but these are rare and their non-derived status is dubious (ibid.: 286).

* All other intransitives uniformly associated with absolutive case, e.g. *q̄uǧun* 'play', *qʰürün* 'laugh'; *ifin* 'become hot', *kusun* 'fall asleep' (ibid.: 271).

– **Georgian:**

* precise factors conditioning the split unclear:

  · Holisky (1981) suggests Class III verbs (intransitives which assign nominative not ergative) are *atelic* (or stative);

  · Cherchi (1997) connects the split to aspectual and agentive oppositions.

  · Another possible overall generalisation might be that verbs expressing states or changes of state or location assign nominative and others assign ergative …

  · but this doesn't work perfectly (e.g. *i-brjv-is* 'fight', *ṭir-is* 'cry', *tamaš-obs* 'play' associated with nominative—Holisky 1981: 172, 175).

– **Svan:** situation 'very much the same as Georgian' (Sumbatova 1993: 266).

* Nb. a degree of dialect variation (to be covered below).

* Intransitives which govern ergative in Series II include *limšie* 'work', *lizelāl* 'walk', *lic'k'ūli* 'mew, squeak' (Sumbatova 1993:

258–259); also *läigərgle* 'he talked', *ädkīzanāle* 'he laughed', *läṭwīliēle* 'he yelled', *läičirxāle* 'he sledded', *läicuzāle* 'he swam', *läičmuriāle* 'he ran' (Harris 1985: 45).

* Those associated with nominative include *lisgdi* 'look' and verbs meaning 'die', 'break', 'be born', 'get lost', 'go crazy' (Sumbatova 1993: 258, 266; Harris 1985: 44).

    · The small class of ablaut verbs are associated with nominative regardless of semantics (Sumbatova 1993: 266).

    · Also associated with nominative are verbs expressing reciprocal relations e.g. *liq'hāl* 'kiss', *limqlaräl* 'embrace', *liq'ärjāl* 'fight', *lišjāl* 'make war' (ibid.).

• Harris (1985: 123–4) discusses variation in case assignment in Series II amongst cognates in the split-S Kartvelian languages:

    – The reflexes of **qad/*qd* 'come, go' are associated with nominative in Svan and Old Georgian, but also occurs with ergative in Georgian dialects and Laz:

    (11)   a.   Svan: *eǰi anqad* (he.NOM came);

            b.   Laz: *ia moxtu* (he.NOM came);

                   cf. *ia koč-epe-k komoxtes* (that man-PL-ERG came) = 'those men came'.

– The reflexes of **gor/*gr* 'roll' are associated with ergative in Georgian but nominative in Laz:

    (12)   a. Georgian: *kac-ma igora* (man-ERG roll) 'the man rolled';

             b. Laz: *koči ingoru* (man-NOM roll) 'the man rolled'.

– The reflexes of **gza* 'travel' occur with ergative in Svan and Georgian but nominative in Laz.

– Cognates meaning 'talk' and 'play' each occur with ergative in some Svan dialects; but vary between ergative and nominative in other Svan dialects and Laz.

• Other variation between different Svan dialects (Harris 1985: 120–1, 123):

– In the Naḳra-Laxamula dialect but not elsewhere, some verbs may occur with either nominative and ergative in Series II, with no change in meaning:

    (13)   a. dina   ädšdiral
                 girl.NOM she.play.II
                 'she played'

             b. dina-d   ädšdirale
                 girl-ERG she.play.II
                 'she played'

      ∗ Other variable verbs include those meaning 'run', 'fight', 'whistle', 'yell', 'sing', 'sled', 'crawl', 'play', 'laugh', 'pass by', 'shoot a gun', 'limp', 'hurry', 'jump, spring'.

      ∗ These verbs all characterised by *-al* suffix.

– Further, the verb meaning 'sled' governs the ergative in the Upper Bal dialect but the nominative in Lašk (*läičirxāle* vs. *edčīrxān* 'he sledded').

• While the split-S type as a whole may not be endangered (see section 2), the endangerment of many split-S languages may threaten the *variety* of split-S systems we can observe.

– Consequences for our understanding of the ways in which split-S systems may vary

– Potentially further implications for our understanding of language / the human mind more generally

References

Aldai, G. (2009). Is Basque morphologically ergative?: Western Basque vs. Eastern Basque. *Studies in Language 33*(4), 783–831.

Arkadiev, P. (2008). Thematic roles, event structure, and argument encoding in semantically aligned languages. In M. Donohue and S. Wichmann (Eds.), *The Typology of Semantic Alignment.* Oxford: Oxford University Press.

Asatiani, I. (1974). *Č'anuri (Lazuri) t'ekst'ebi.* Tbilisi: Mecniereba.

Berro, A. (2012). Three levels of root insertion in Basque intransitive verbs. *Journal of Portuguese Linguistics 11*(1), 7–22.

Cherchi, M. (1997). *Modern Georgian Morphosyntax.* Wiesbaden: Harrassowitz.

Comrie, B. (2013). Alignment of case marking of full noun phrases. In M. Dryer and M. Haspelmath (Eds.), *The World Atlas of Language Structures Online.* Leipzig: Max Planck Institute for Evolutionary Anthropology.

Comrie, B., D. Forker, and Z. Khalilova (2013). Alignment typology, reflexives, and reciprocals in Tsezic languages. In *Proceedings of the 37th Annual Meeting of the Berkeley Linguistics Society: Special Session on Languages of the Caucasus*, pp. 32–51.

Coon, J. (2010). *Complementation in Chol (Mayan): A Theory of Split Ergativity.* Ph. D. thesis, Massachusetts Institute of Technology.

Creissels, D. (2008). Remarks on split intransitivity and fluid intransitivity.

In O. Bonami and P. Cabredo Hofherr (Eds.), *Empirical Issues in Syntax and Semantics 7.*

Creissels, D. and C. Mounole. Non-canonical valency patterns in Basque, variation and evolution.

Danielson, S. and T. Granadillo (2008). Agreement in two Arawak languages: Baure and Kurripako. In M. Donohue and S. Wichmann (Eds.), *The Typology of Semantic Alignment.* Oxford: Oxford University Press.

Deal, A. R. and C. O'Connor. The perspectival basis of fluid-S case-marking in Northern Pomo. In S. Lima (Ed.), *Proceedings of the Fifth Conference on he Semantics of Under-represented Languages in the Americas.*

Denwood, P. (1991). *Tibetan.* Amsterdam: John Benjamins.

Dixon, R. (1994). *Ergativity.* Cambridge: Cambridge University Press.

Harris, A. C. (1981). *Georgian Syntax.* Cambridge: Cambridge University Press.

Harris, A. C. (1982). Georgian and the Unaccusative Hypothesis. *Language 58*(2), 290–306.

Harris, A. C. (1985). *Diachronic Syntax: the Kartvelian case*, Volume 18 of *Syntax and semantics.* Orlando: Academic Press.

Harris, A. C. (2002). *Endoclitics and the origins of Udi morphosyntax.* Oxford: Oxford University Press.

Haspelmath, M. (1993). *A Grammar of Lezgian.* Berlin: Mouton de Gruyter.

Holisky, D. (1981). *Aspect and Georgian Medial Verbs.* Delmar: Caravan Books.

Holisky, D. (1987). The case of the intransitive subject in Tsova-Tush (Batsbi). *Lingua 71*(1–4), 103–32.

Kibrik, A. E. (1985). Towards a typology of ergativity. In J. Nichols and A. C. Woodbury (Eds.), *Grammar inside and outside the clause.* Cambridge: Cambridge University Press.

Kimball, G. (1991). *Koasati Grammar.* Lincoln, Nebraska: University of Nebraska Press.

Lewis, M. P., G. F. Simons, and C. D. Fennig (2016). *Ethnologue: Languages of the World* (19th ed.). Dallas, Texas: SIL International.

McLendon, S. (1978). Ergativity, case and transitivity in Eastern Pomo. *International Journal of American Linguistics 44*(1), 1–9.

Mithun, M. (1991). Active/agentive case marking and its motivations. *Language 67*(3), 510–546.

Moseley, C. (2010). *Atlas of the World's Languages in Danger* (3 ed.). Paris: UNESCO.

Nichols, J. (2008). Case in Ingush syntax. In G. G. Corbett and M. Noonan (Eds.), *Case and grammatical relations: studies in honor of Bernard Comrie.* Amsterdam: John Benjamins.

Siewierska, A. (2013). Alignment of verbal person marking. In M. Dryer and M. Haspelmath (Eds.), *The World Atlas of Language Structures Online.* Leipzig: Max Planck Institute for Evolutionary Anthropology.

Sumbatova, N. R. (1993). Causative constructions in Svan: further evidence for role domination. In B. Comrie and M. Polinsky (Eds.), *Causatives and transitivity*. Amsterdam: John Benjamins.

Whalen, D. and G. F. Simons (2012). Endangered language families. *Language 88*(1), 155–73.