

Predicting grammatical gender in Nakh languages

Three methods compared

Jesse Wichers Schreur¹, Marc Allasonnière-Tang², Kate Bellamy³,
Neige Rochant⁴

École Pratique des Hautes Études / Goethe University Frankfurt¹
National Museum of Natural History / CNRS / University Paris City²
KU Leuven³
Sorbonne Nouvelle University / CNRS: LACITO ~ LLACAN⁴

24/08/2022

55th Annual Meeting of the Societas Linguistica Europaea
University of Bucharest “Iorgu Iordan – Al. Rosetti” Institute of
Linguistics

Grammatical gender

- Means of identifying and tracking referents in discourse (Contini-Morava & Kilarski, 2013)

Grammatical gender

- Means of identifying and tracking referents in discourse (Contini-Morava & Kilarski, 2013)
- All nouns are assigned a gender, which is formally reflected in agreement markers in the clause (Hockett, 1958; Corbett, 1991)

Grammatical gender

- Means of identifying and tracking referents in discourse (Contini-Morava & Kilarski, 2013)
- All nouns are assigned a gender, which is formally reflected in agreement markers in the clause (Hockett, 1958; Corbett, 1991)
- Assignment on basis of semantic and/or formal (phonological and morphological) properties, with semantic features usually taking precedence (Corbett & Fraser 2000; Bellamy & Wichers Schreur 2021; Allasonnière-Tang et al. 2021)

Grammatical gender

- Means of identifying and tracking referents in discourse (Contini-Morava & Kilarski, 2013)
- All nouns are assigned a gender, which is formally reflected in agreement markers in the clause (Hockett, 1958; Corbett, 1991)
- Assignment on basis of semantic and/or formal (phonological and morphological) properties, with semantic features usually taking precedence (Corbett & Fraser 2000; Bellamy & Wichers Schreur 2021; Allasonnière-Tang et al. 2021)
- Some systems where assignment is completely transparent, others where it is more opaque

Our approach

- Question = What are the specific principles that underpin complex assignment systems?

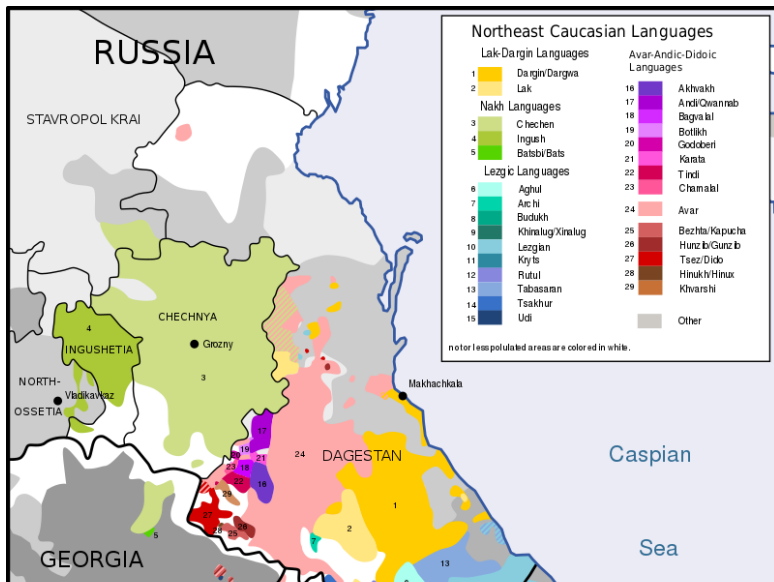
Our approach

- Question = What are the specific principles that underpin complex assignment systems?
- Three computational methods to test which factors most adequately predict gender in Chechen and Tsova-Tush

Studies concerning gender prediction

- Network Morphology approach (Corbett and Fraser (1993)) with a lexical knowledge representation language (DATR) (Fraser & Corbett, 1995; Evans & Gazdar, 1989a, 1989b, 1996)
- Polinsky and van Everbroeck (2003): study of the reanalysis of the Latin gender system as it transitioned to Old French, with frequency-based neural network
- Bateman and Polinsky (2010): decision trees to reconsider the gender system of Romanian as a two-value rather than the traditional three-value analysis
- Plaster, Polinsky and Harizanov (2012): decision tree model that predicts gender of 70% of nouns in Tsez (East-Caucasian). Result: semantic features are the most predictive
- Allasonnière-Tang, Brown and Fedden (2021): simple decision trees + random forest trees + neural networks to test claim that Mian (Ok) assigns gender predominantly on basis of semantics.

Nakh languages



Chechen and Tsova-Tush

Chechen	Tsova-Tush
<ul style="list-style-type: none">● Chechnya (RF)● 95% of Chechnya● Official● Bilingualism with Russian● Not endangered	<ul style="list-style-type: none">● aka Bats, aka Batsbi● Georgia > Zemo Alvani● 500 people● Not official● Bilingualism with Georgian● Severely endangered

Gender in Chechen and Tsova-Tush

- 5 genders

Gender in Chechen and Tsova-Tush

- 5 genders
- Agreement is marked by the same 4 consonantal prefixes

Gender in Chechen and Tsova-Tush

- 5 genders
- Agreement is marked by the same 4 consonantal prefixes
- Agreement targets=

Gender in Chechen and Tsova-Tush

- 5 genders
- Agreement is marked by the same 4 consonantal prefixes
- Agreement targets=
 - 1/3 of all underived verbs

Gender in Chechen and Tsova-Tush

- 5 genders
- Agreement is marked by the same 4 consonantal prefixes
- Agreement targets=
 - 1/3 of all underived verbs
 - small number of adverbs

Gender in Chechen and Tsova-Tush

- 5 genders
- Agreement is marked by the same 4 consonantal prefixes
- Agreement targets=
 - 1/3 of all underived verbs
 - small number of adverbs } Agreement with NOM

Gender in Chechen and Tsova-Tush

- 5 genders
 - Agreement is marked by the same 4 consonantal prefixes
 - Agreement targets=
 - 1/3 of all underived verbs
 - small number of adverbs
 - ten underived adjectives
- } Agreement with NOM

Gender in Chechen and Tsova-Tush

- 5 genders
 - Agreement is marked by the same 4 consonantal prefixes
 - Agreement targets=
 - 1/3 of all underived verbs
 - small number of adverbs
 - ten underived adjectives
 - the numeral 'four'
- } Agreement with NOM

Gender in Chechen and Tsova-Tush

- 5 genders
- Agreement is marked by the same 4 consonantal prefixes
- Agreement targets=
 - 1/3 of all underived verbs } Agreement with NOM
 - small number of adverbs }
 - ten underived adjectives } Agreement with head noun
 - the numeral 'four' }

Gender in Chechen and Tsova-Tush

- (1) Chechen (Nichols, 1994: 37)

ča jiett āra b-ēl-ir.
 one cow(B) out B-go-AOR

'One cow went out.'

- (2) Tsova-Tush (Kadagidze, 2009: 170)

se herc'ō d-ux d-erc'-d-∅-eb so !
 1SG.GEN pot(D) D-back D-return-D-TR-IMP 1SG.DAT

'Give me back my pot!'

Gender in Chechen and Tsova-Tush

Gender	Semantics	Markers (SG/PL)	
		Tsova-Tush	Chechen
M	male rationals	v- / b-	v- / b-
F	female rationals	j- / d-	j- / b-
B	animals, inanimates	b- / d-	b- / d-
J	animals, inanimates	j- / j-	j- / j-
D	animals, inanimates	d- / d-	d- / d-
Bb	animals, inanimates	b- / b- (6 items)	b- / b- (22 items)
Bj	body parts, 'step', 'kick'	b- / j- (17 items)	-
Dj	body parts	d- / j- (6 items)	-

Last 3 = exceptions aka 'inquorate' genders, (Corbett, 1991: 170–175), excluded from this study

Gender assignment in Chechen and Tsova-Tush

- M & F = male vs. female rationals
- vs. 3 non-fully predictable non-rational genders

Gender assignment in Chechen and Tsova-Tush

Table: Tendencies which predict gender of small portion of non-rationals

Chechen (Nichols 2007)	Tsova-Tush (Wichers Schreur 2021)
	Some semantically-based clusterings of genders → 15% of non-rationals predicted
Some derived abstract nouns → D/J depending on degree of abstractness	All verbal nouns → D class Most de-adjectival abstract nouns → J class
	60-65% of b-, d- and j- nouns → B, D and J genders (auto-gender + alliterative concord)

Gender assignment in Chechen and Tsova-Tush

Table: Special cases

	Chechen	Tsova-Tush
Loanwords	New loanwords often given gender of near synonym or immediate generic (Nichols 2007), but all recent Russian loans have gender J	Same set of semantic and phonological tendencies as native nouns (Wichers Schreur 2021)

Data

Two lexical databases

- Chechen: Erwin Komen - combined dictionaries of Matsiev (1961) and Jamalkhanov and Aliroev (1991)
- Tsova-Tush: Kadagidze and Kadagidze (1984)

Language	B	D	J	M	F
Chechen	320	868	1370	88	27
Tsova-Tush	462	968	836	104	30

Data

Two lexical databases

- Chechen: Erwin Komen - combined dictionaries of Matsiev (1961) and Jamalkhanov and Aliroev (1991)
- Tsova-Tush: Kadagidze and Kadagidze (1984)
- Semantic information: broad semantics categories (Male, Female, Human, Animal, Inanimate) + Concepticon (List et al., 2016)

Language	B	D	J	M	F
Chechen	320	868	1370	88	27
Tsova-Tush	462	968	836	104	30

Data

Two lexical databases

- Chechen: Erwin Komen - combined dictionaries of Matsiev (1961) and Jamalkhanov and Aliroev (1991)
- Tsova-Tush: Kadagidze and Kadagidze (1984)
- Semantic information: broad semantics categories (Male, Female, Human, Animal, Inanimate) + Concepticon (List et al., 2016)
- Form information: the first three and the last three phonemes of each noun (Dryer, 2013; Basirat et al., 2021)

Language	B	D	J	M	F
Chechen	320	868	1370	88	27
Tsova-Tush	462	968	836	104	30

Data

Noun	<i>haer</i>	<i>mar</i>
Gloss	air	husband
Gender	J	M
Concepticon_category	Person/thing	Person/thing
Concepticon_field	Physical world	Kinship
Semantic_broad	Inanimate	Male
Semantic_domain	Natural	Kinship
Borrowed_Arab	0	0
Borrowed_GE	1	0
Borrowed_Turk	0	0
Borrowed_Russian	0	0
Word length	4	3
Last first phoneme	r	r
Last second phoneme	e	a
Last third phoneme	a	m
First phoneme	h	m
Second phoneme	a	a
Third phoneme	e	r

Methods

Models

- Three computational classifiers (machine learning/ deep learning)

Methods

Models

- Three computational classifiers (machine learning/ deep learning)
- Decisions trees: one tree vs random forests (300)

Methods

Models

- Three computational classifiers (machine learning/ deep learning)
- Decisions trees: one tree vs random forests (300)
- Deep learning: neural network (feed-forward)

Methods

Models

- Three computational classifiers (machine learning/ deep learning)
- Decisions trees: one tree vs random forests (300)
- Deep learning: neural network (feed-forward)

Parameters

Methods

Models

- Three computational classifiers (machine learning/ deep learning)
- Decisions trees: one tree vs random forests (300)
- Deep learning: neural network (feed-forward)

Parameters

- 70% of the data used for training, 30% of the data used for testing

Methods

Models

- Three computational classifiers (machine learning/ deep learning)
- Decisions trees: one tree vs random forests (300)
- Deep learning: neural network (feed-forward)

Parameters

- 70% of the data used for training, 30% of the data used for testing
- **10**, 100, 1000 replications to avoid sampling bias

Methods

Models

- Three computational classifiers (machine learning/ deep learning)
- Decisions trees: one tree vs random forests (300)
- Deep learning: neural network (feed-forward)

Parameters

- 70% of the data used for training, 30% of the data used for testing
- **10**, 100, 1000 replications to avoid sampling bias
- The performance is compared with the random baseline (38.4%, 30.5%) and the majority baseline (51.3%, 39.2%)

Methods

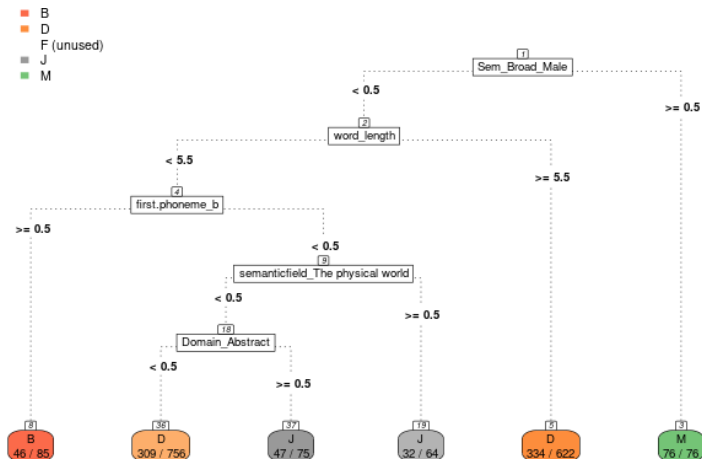
Models

- Three computational classifiers (machine learning/ deep learning)
- Decisions trees: one tree vs random forests (300)
- Deep learning: neural network (feed-forward)

Parameters

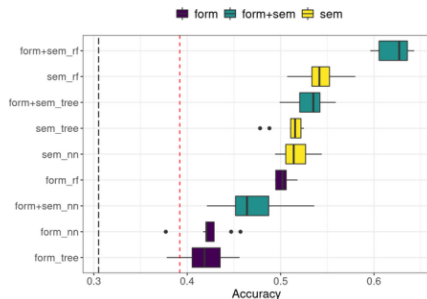
- 70% of the data used for training, 30% of the data used for testing
- **10**, 100, 1000 replications to avoid sampling bias
- The performance is compared with the random baseline (38.4%, 30.5%) and the majority baseline (51.3%, 39.2%)
- The use of precision and recall to assess the performance on each gender category (like Target-Like Usage (TLU) and Suppliance in Obligatory Context (SOC) in language acquisition)

Output

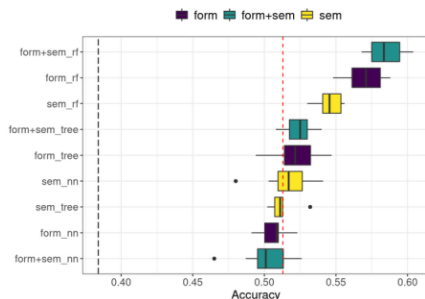


Example tree for gender assignment in Tsova-Tush

Output



Tsova-Tush



Chechen

Overall prediction accuracy across the languages of the study. The dashed lines indicate the random (black) and the majority (red) baselines.

Language-internal aspects

Tsova-Tush:

- Sustains earlier findings:

Language-internal aspects

Tsova-Tush:

- Sustains earlier findings:
 - “Male” = 100% accuracy

Language-internal aspects

Tsova-Tush:

- Sustains earlier findings:
 - “Male” = 100% accuracy
 - b- nouns → gender B, inanimate d- nouns → gender D

Language-internal aspects

Tsova-Tush:

- Sustains earlier findings:
 - “Male” = 100% accuracy
 - b- nouns → gender B, inanimate d- nouns → gender D
 - Loan word status not used as a variable

Language-internal aspects

Tsova-Tush:

- Sustains earlier findings:
 - “Male” = 100% accuracy
 - b- nouns → gender B, inanimate d- nouns → gender D
 - Loan word status not used as a variable
- New findings:

Language-internal aspects

Tsova-Tush:

- Sustains earlier findings:
 - “Male” = 100% accuracy
 - b- nouns → gender B, inanimate d- nouns → gender D
 - Loan word status not used as a variable
- New findings:
 - “word length” → gender D (high percentage of abstract nouns formed by suffixes)

Language-internal aspects

Tsova-Tush:

- Sustains earlier findings:
 - “Male” = 100% accuracy
 - b- nouns → gender B, inanimate d- nouns → gender D
 - Loan word status not used as a variable
- New findings:
 - “word length” → gender D (high percentage of abstract nouns formed by suffixes)
 - “wild plant” → gender D

Language-internal aspects

Tsova-Tush:

- Sustains earlier findings:
 - “Male” = 100% accuracy
 - b- nouns → gender B, inanimate d- nouns → gender D
 - Loan word status not used as a variable
- New findings:
 - “word length” → gender D (high percentage of abstract nouns formed by suffixes)
 - “wild plant” → gender D
 - “Inanimate” → gender J (51% accuracy)

Language-internal aspects

Chochen:

- Sustains earlier findings:

Language-internal aspects

Chochen:

- Sustains earlier findings:
 - d- nouns → gender D

Language-internal aspects

Chochen:

- Sustains earlier findings:
 - d- nouns → gender D
- New findings:

Language-internal aspects

Chochen:

- Sustains earlier findings:
 - d- nouns → gender D
- New findings:
 - “word length” → gender J (numerous suffixes deriving J gender nouns / high number of instances of J gender suffixes)

Language-internal aspects

Chechen:

- Sustains earlier findings:
 - d- nouns → gender D
- New findings:
 - “word length” → gender J (numerous suffixes deriving J gender nouns / high number of instances of J gender suffixes)
 - Russian loanwords → J gender (55/69 neuter words) (not mirrored in decision tree experiments).

Summary

Corroborates earlier findings in East Caucasian

- Broad semantic categories such as animacy, humanness and abstractness show high correlation with certain gender classes (Carling et al., 2021).

Summary

Corroborates earlier findings in East Caucasian

- Broad semantic categories such as animacy, humanness and abstractness show high correlation with certain gender classes (Carling et al., 2021).
- 1st segment /b/ or /d/ found to correlate highly with corresponding genders B and D

Summary

Corroborates earlier findings in East Caucasian

- Broad semantic categories such as animacy, humanness and abstractness show high correlation with certain gender classes (Carling et al., 2021).
- 1st segment /b/ or /d/ found to correlate highly with corresponding genders B and D

Summary

Corroborates earlier findings in East Caucasian

- Broad semantic categories such as animacy, humanness and abstractness show high correlation with certain gender classes (Carling et al., 2021).
- 1st segment /b/ or /d/ found to correlate highly with corresponding genders B and D

Broader linguistic implications

- Descriptive analyses of gender systems can be assisted by computational methods to identify the relative weights of semantics and form in gender assignment

Summary

Corroborates earlier findings in East Caucasian

- Broad semantic categories such as animacy, humanness and abstractness show high correlation with certain gender classes (Carling et al., 2021).
- 1st segment /b/ or /d/ found to correlate highly with corresponding genders B and D

Broader linguistic implications

- Descriptive analyses of gender systems can be assisted by computational methods to identify the relative weights of semantics and form in gender assignment
- However, more data is always needed... (lexicon and semantic features)



Thank you!

References

- Allasonnière-Tang, Marc & Brown, Dunstan & Fedden, Sebastian. 2021. Testing semantic dominance in Mian gender: Three machine learning models. *Oceanic Linguistics* 60(2). 302-334. doi.org/10.1353/ol.2021.0018.
- Balam, Osmer. 2016. Semantic Categories and Gender Assignment in Contact Spanish: Type of Code-Switching and its Relevance to Linguistic Outcomes. *Journal of Language Contact* 9. 405-435. doi.org/10.1163/19552629-00903001.
- Basirat, Ali & Allasonnière-Tang, Marc & Berdicevskis, Aleksandrs. 2021. An empirical study on the contribution of formal and semantic features to the grammatical gender of nouns. *Linguistics Vanguard* 7(1). 20200048. doi.org/10.1515/lingvan-2020-0048.
- Bellamy, Kate & Jesse Wichers Schreur. 2021. When semantics and phonology collide: Gender assignment in mixed Tsova-Tush-Georgian nominal constructions. *The International Journal of Bilingualism* 0(0). 1-29. doi.org/10.1177/13670069211039559.
- Breiman, Leo & Jerome H. Friedman & Richard A. Olshen & Charles J. Stone. 1984. *Classification and Regression Trees*. Boca Raton: Routledge.
- Brown, Dunstan. 1998. Defining 'sub-gender': Virile and devirilized nouns in Polish. *Lingua* 104. 187-233.
- Carling, Gerd, Kate Bellamy & Jesse Wichers Schreur. 2021. Gender stability in Nakh-Daghestanian, paper presented at Languages, Dialects and Isoglosses of Anatolia, the Caucasus and Iran, March 2021, Paris.
- Contini-Morava, Ellen & Marcin Kilarski. 2013. Functions of nominal classification. *Language Sciences* 40. 263-299. doi.org/10.1016/j.langsci.2013.03.002.
- Corbett, Greville. 1982. Gender in Russian: an Account of Gender Specification and its Relationship to Declension. *Russian Linguistics* 6. 197-232.
- Corbett, Greville. 1991. *Gender*. Cambridge: Cambridge University Press.
- Corbett, Greville. 2013. Systems of Gender Assignment. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/32>, Accessed on 2022-02-02.)
- Corbett, Greville G. & Norman M. Fraser. 1993. Network Morphology: A DATR account of Russian nominal inflection. *Journal of Linguistics* 29(1). 113-142. doi:10.1017/S0022226700000074.
- Corbett, Greville & Norman M. Fraser. 2000. Default genders. In Barbara Unterbeck & Matti Rissanen (eds.), *Gender in Grammar and Cognition I: Approaches to Gender*. 55-98. Berlin: Mouton de Gruyter.
- Cruz, Abel. 2021. A syntactic approach to gender assignment in Spanish-English bilingual speech. *Glossa: a journal of general linguistics* 6(1). 1-40. doi.org/10.16995/glossa.5878.

References

- Desheriev, Y. D. 1953. *Bacbijskij jazyk: fonetika, morfologija, sintaksis, leksika* [The Tsova-Tush language: phonetics, morphology, syntax, lexicon]. Moscow: Izdatel'stvo AN SSSR. Demsar, Janez & Zupan, Blaz & Leban, Gregor & Curk, Tomaž. 2004. Orange: From Experimental Machine Learning to Interactive Data Mining, White Paper. European Conference of Machine Learning: 2004; Pisa, Italy 3202. 537-539.
- Dryer, Matthew S. 2013. Prefixing vs. suffixing in inflectional morphology. In: Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/33>, Accessed on 2022-02-02.).
- Evans, Nicholas & Brown, Dunstan & Corbett, Greville. 2002. The semantics of gender in Mayali: partially parallel systems and formal implementation. *Language* 78. 109-153. Evans, Roger & Gazdar, Gerald. 1989a. Inference in DATR. Proceedings of the fourth conference of the European Chapter of the Association for Computational Linguistics, Manchester, England. 66-71.
- Evans, Roger & Gazdar, Gerald. 1989b. The semantics of DATR. In A. G. Cohn (ed.), *Proceedings of the seventh conference of the Society for the Study of Artificial Intelligence and Simulation of Behaviour*, 79-87. London: Pitman/Morgan Kaufmann.
- Evans, Roger & Gazdar, Gerald. 1996. DATR: A language for lexical knowledge representation. *Computational Linguistics* 22. 167-216.
- Fedden, Sebastian. 2011. *A Grammar of Mian*. Berlin / Boston: De Gruyter Mouton.
- Fraser, Norman M. & Corbett, Greville G. 1995. Gender, animacy, and declensional class assignment: A unified account for Russian. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1994*, 123-150. Amsterdam: Kluwer Academic Publishers.
- Fraser, Norman M. & Corbett, Greville G. 1997. Defaults in Arapesh. *Lingua* 103. 25-57.
- Gagliardi, Annie & Jeffrey Lidz. 2014. Statistical insensitivity in the acquisition of Tsez noun classes. *Language* 90(1). 58-89. doi:10.1353/lan.2014.0013.
- Haykin, S. 1998. *Neural Networks: A Comprehensive Foundation*. Prentice-Hall: Englewood Cliffs.
- Hockett, Charles F. 1958. *A course in modern linguistics*. New York: MacMillan.
- Jamalkhanov, Z. D. & Aliroev, I. Y. []. 1991. *Slovar' pravopisanija literaturnogo čečenskogo jazyka* [Orthographical dictionary of literary Chechen]. Grozny: Kniga.
- Kadagidze, E. 2009. *c'ova-tušuri t'ekst'ebi* [Tsova-Tush texts]. Tbilisi: TSU gamomcemloba.
- Kadagidze, D. & Kadagidze, N. []. 1984. *c'ova-tušur-kartul-rusul leksik'oni* [Tsova-Tush-Georgian-Russian dictionary]. Tbilisi: Mecniereba.

References

- Karmiloff-Smith, Annette. 1979. *A Functional Approach to Child Language*. New York / London: Cambridge University Press.
- Khalilov, M. S. 1999. *Cezsko-russkij slovar' [Tsez-Russian dictionary]*. Moscow: Academia
- Kilarski, Marcin. & Allasonnière-Tang, Marc. 2021. Classifiers in Morphology. In M. Aronoff (ed.), *Oxford Research Encyclopedia of Linguistics*, 1-28. Oxford: Oxford University Press. doi:10.1093/acrefore/9780199384655.013.546.
- Lemus-Serrano, Magdalena & Allasonnière-Tang, Marc & Dediu, Dan. 2021. What conditions tone paradigms in Yukuna: Phonological and machine learning approaches. *Glossa: a journal of general linguistics* 6(1). 60. doi.org/10.5334/gjgl.1276.
- List, Johann-Mattis, Michael Cysouw & Robert Forkel. 2016. Concepticon: A Resource for the Linking of Concept Lists. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. 2393-2400.
- Matsiev, A. G. 1961. *Slovar' čečenskogo jazyka [Chechen dictionary]*. Moscow: Gosudarstvennoe izdatel'stvo inostrannyx i nacional'nyx sloverej.
- Nichols, Johanna. 1989. The Nakh evidence for the history of gender in Nakh-Daghestanian. In Howard I. Aronson (ed.) *The non-Slavic languages of the USSR: linguistic studies*. University of Chicago. pp. 158-175.
- Nichols, Johanna. 1994. Chechen. In Riks Smeets (ed.) *The North East Caucasian languages, part 2, 1-78*. Delmar: Caravan.
- Nichols, Johanna. 2003. The Nakh-Daghestanian consonant correspondences. In Dee Ann Holisky & Kevin Tuite (eds.) *Current trends in Caucasian, East European and Inner Asian linguistics: papers in honor of Howard I. Aronson*. Amsterdam: John Benjamins. pp. 207-264.
- Nichols, Johanna. 2007. Chechen morphology with notes on Ingush. In Alan S. Kaye (ed.) *Morphologies of Africa and Asia*. Penn University Press. pp. 1188-1207. doi.org/10.1515/9781575065663-044.
- Nichols, Johanna. 2011. *Ingush grammar*. Berkeley, Los Angeles: University of California Press.
- Her, One-Soon & Tang, Marc. 2020. A statistical explanation of the distribution of sortal classifiers in languages of the world via computational classifiers. *Journal of Quantitative Linguistics* 27(2). 93-113. doi.org/10.1080/09296174.2018.1523777.
- Parks, Randolph, Daniel S. Levine & Debra L. Long (eds.). 1998. *Fundamentals of Neural Network Modeling: Neuropsychology and Cognitive Neuroscience*. Boston: MIT Press.
- Plaster, Keith & Polinsky, Maria & Harizanov, Boris. 2013. Noun classes grow on trees: noun classification in the North-East Caucasus. In Balthazar Bickel, Lore A. Grenoble, David A. Peterson & Alan Timberlake (eds.), *Language typology and historical contingency: In honor of Johanna Nichols*, 153-170. Amsterdam: John Benjamins.
- Polinsky, Maria & Ezra Van Everbroeck. 2003. Development of gender classifications: Modeling the historical change from Latin to French. *Language* 79. 356-390.

References

- Quinlan, J. Ross. 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers.
- Quinlan, J. Ross. 1996. Improved use of continuous attributes in c4.5. *Journal of Artificial Intelligence Research* 4. 77-90.
- Rajabov, Ramazan. Undated. *Tsez Dictionary*. Ms. Los Angeles: University of Southern California.
- Senft, Gunter (ed.). 2000. *Systems of nominal classification*. Cambridge: Cambridge University Press.
- Sokolik, M. E. & Michael E. Smith. 1992. Assignment of gender to French nouns in primary and secondary language: a connectionist model. *Second Language Research* 8(1). 39-58.
- Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, forests and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change* 24(2). 135-178. doi:10.1017/S0954394512000129.
- Ting, Kai Ming. 2010. Precision and Recall. In Claude Sammut & Geoffrey I. Webb (eds.), *Encyclopedia of Machine Learning*, 781-781. Boston: Springer. doi:10.1007/978-0-387-30164-8_652.
- Ulrich, Natalja & Allasonnière-Tang, Marc & Pellegrino, François & Dediu, Dan. 2021. Identifying the Russian voiceless non-palatalized fricatives /f/, /s/ and /ʃ/ from acoustic cues using Machine Learning. *Journal of the Acoustical Society of America* 150(3). 1806-1820. doi.org/10.1121/10.0005950.
- Wichers Schreur, Jesse. 2021. Nominal borrowings in Tsova-Tush (Nakh-Daghestanian, Georgia) and their gender assignment. In Diana Forker & Lore A. Grenoble (eds.), *Language contact in the territory of the former Soviet Union*, 15-33. Amsterdam: John Benjamins.
- Wurm, S. A., Heyward, I., & Unesco. 2001. *Atlas of the world's languages in danger of disappearing*. Paris: Unesco Pub. Website <http://www.unesco.org/languages-atlas/> consulted on 7-12-2021.