

A Dataset

Data from grammars and other linguistic and phonological analyses of languages listed in the STRESSTYP2 database ([Goedemans et al., 2014](#)) or by [Gordon \(2002\)](#) were used to create the dataset. The STRESSTYP2 database contains information about various lects and the kinds of stress patterns exhibited by these lects. The database also has links to the sources from which the data was collected for compiling the database, and these were the sources we consulted for examples of words with stress patterns marked. All the words collected from these sources have the stress marking attested in the source – there are no cases of a given rule being used to predict the stress pattern on words. All languages used are listed in Table 1.

A.1 Olympiad data

We use the problems from [Vaduguru et al. \(2021\)](#). Language-wise results are listed in Table 2.

Language	Source	BASIC	CV	SYLLABLE	FEATURE	LSTM	Transformer
Anejom	Lynch, 2000	0	0	0.666667	0	0	0
Arabic	Wright, 1874	0.25	0.75	0.25	0.5	25	50
Araucanian	Echeverría and Contreras, 1965	0	0.666667	0.333333	0.333333	0	0
Au	Scorza, 1985	0	0.666667	0.333333	0	33.3333	0
Awtuw	Feldman, 1983	0.333333	1	1	1	0	0
Bortzerrieta Basque	Hualde, 1989	0	1	1	1	0	0
Cofan	Borman, 1962	0	0.333333	0.666667	0.666667	33.3333	0
Coreguaje	Gralow, 1985	0	1	1	1	50	50
Garawa	Furby, 1974	0	0	1	1	0	0
Hopi	Jeanne, 1978	0.4	0.4	0.8	0.6	0	20
Kaliai-Kove	Counts, 1969	0.25	1	1	1	25	0
Kara	Schlie and Schlie, 1993	0.25	0.25	0	0	0	0
Kwakiutl	Bach, 1975	0.2	0	0.2	0	0	40
Lezgian	Haspelmath, 1993	0	1	1	1	50	0
Manam	Buckley, 1998	0	0.25	1	0.25	0	0
Mathimathi	Hercus, 1986	0	0.25	0	0	0	25
Murik	Abbott, 1985	0.5	0.75	0.75	0.75	0	0
Piro	Matteson, 1965	0	0	0	0	0	0
Sarangani Monobo	Dubois, 1976	0.5	0.5	0.75	1	0	0
Seri	Marlett, 1988	0.75	0.75	0.75	0.75	50	50
Sio	Soetenga Clark, 1993	0.5	0.5	1	0.75	0	0
Tahitian	Tryon, 1976	0.666667	0.666667	0.666667	0.666667	0	66.6667
Tamazight	Abdel-Massih, 1971	0.666667	1	1	1	0	0
Tinrin	Osumi, 1995	0	1	1	1	0	0
Tol	Fleming and Dennis, 1977	0	0.25	1	1	50	25
Tumpisa	Dayley, 1989	0	0	0.333333	0.333333	33.3333	33.3333
Tzutujil	Dayley, 1985	0	0.8	0.4	0.4	20	0
Yidiny	Dixon, 1977	0	0	0	0	0	0

Table 1: Languages for which we collect data.

Problem Name	BASIC	CV	SYLLABLE	FEATURE	LSTM	Transformer
Aliutor	0.25	0	0.4375	0.4375	37.5	31.25
Chickasaw Stress	0	0	0.833333	0.166667	0	0
Creek	0.416667	0	0.5	0.083333	33.3333	0
Kabardian	0.5	1	1	1	33.3333	16.6667
Old Indic	0	0	0	0.25	25	25
Piraha	0	0	0	0	10	0

Table 2: Olympiad problems. Kabardian is an IOL problem in which stress is presented as a side phenomenon, and is easier to deduce than in other IOL problems.

References

- Stan Abbott. 1985. *A Tentative Multilevel Multiunit Phonological Analysis of the Murik Language*, volume 63 of *Pacific Linguistics, Series A*. Australian National University, Canberra.
- E.T. Abdel-Massih. 1971. *A Reference Grammar of Tamazight: A Comparative Study of the Berber Dialects of Ayt Ayache and Ayt Seghrouchen*. Number v. 1 in Publications of the Center for Near Eastern and North African Studies, University of Michigan. Center for Near Eastern and North African Studies, University of Michigan.
- E. Bach. 1975. Long vowels and stress in kwakiutl. *Texas Linguistic Forum*, 2:9–19.
- M. B. Borman. 1962. Cofan phonemes. In Benjamin F. Elson, editor, *Studies in Ecuadorian Indian languages 1*, volume 7 of *Linguistic Series*, pages 45–59. Summer Institute of Linguistics of the University of Oklahoma, Norman.
- Eugene Buckley. 1998. [Alignment in manam stress](#). *Linguistic Inquiry*, 29(3):475–496.
- David R. Counts. 1969. [A grammar of kaliai-kove](#). *Oceanic Linguistics Special Publications*, (6):i–170.
- Jon P. Dayley. 1985. *Tzutujil Grammar*, volume 107 of *University of California Publications in Linguistics*. University of California Press, Berkeley and Los Angeles. Bibliography: p. 409–412.
- Jon P. Dayley. 1989. *Tümpisa (Panamint) Shoshone Grammar*, volume 115 of *University of California Publications in Linguistics*. University of California Press, Berkeley.
- Robert M. W. Dixon. 1977. Some phonological rules in yidin. *Linguistic Inquiry*, 8:1–34.
- Carl D. Dubois. 1976. *Sarangani Manobo: An Introductory Guide*, volume 6 of *Special monograph issue*. Linguistic Society of the Philippines, Manila.
- Max S. Echeverría and Heles Contreras. 1965. [Araucanian phonemics](#). *International Journal of American Linguistics*, 31(2):132–135.
- H. Feldman. 1983. *A Grammar of Awtuw*. Australian National University.
- Ilah Fleming and Ronald K. Dennis. 1977. [Tol \(jicaque\): Phonology](#). *International Journal of American Linguistics*, 43(2):121–127.
- Christine E. Furby. 1974. *Garawa Phonology*, volume 37 of *Pacific Linguistics, Series A*, pages 1–11. Australian National University, Canberra.
- Rob Goedemans, Jeffrey Heinz, and Harry Van der Hulst. 2014. *Stresstyp2*. *University of Connecticut, University of Delaware, Leiden University, and the US National Science Foundation*.
- Matthew Gordon. 2002. [A factorial typology of quantity-insensitive stress](#). *Natural Language & Linguistic Theory*, 20(3):491–552.
- Frances L. Gralow. 1985. [Coreguaje: Tone, stress and intonation](#). In Ruth M. Brend, editor, *From phonology to discourse: Studies in six Colombian languages*, volume 9 of *Language Data, Amerindian Series*, pages 3–11. Summer Institute of Linguistics, Dallas.
- Martin Haspelmath. 1993. *A Grammar of Lezgian*, volume 9 of *Mouton Grammar Library*. Mouton de Gruyter, Berlin.
- Luise A. Hercus. 1986. Outline of the madimadi language. In L. A. Hercus, editor, *Victorian Languages: A Late Survey*, volume 77 of *Pacific Linguistics, Series B*, pages 101–151. Australian National University, Canberra.
- José Ignacio Hualde. 1989. A lexical phonology of basque.
- L. Jeanne. 1978. Aspects of hopi grammar.
- John Lynch. 2000. *A Grammar of Anejom*, volume 507 of *Pacific Linguistics*. Australian National University, Canberra.
- Stephen A. Marlett. 1988. [The syllable structure of seri](#). *International Journal of American Linguistics*, 54(3):245–278.
- Esther Matteson. 1965. *The Piro (Arawakan) Language*. University of California Press, Berkeley.
- Midori Osumi. 1995. [Tinrin grammar](#). *Oceanic Linguistics Special Publications*, (25):i–304.
- P. Schlie and G. Schlie. 1993. A kara phonology. In J. Clifton, editor, *Phonologies of Austronesian Languages 2*, pages 99–130. Summer Institute of Linguistics, Ukarumpa.
- David Scorza. 1985. *A Sketch of Au Morphology and Syntax*, volume 63 of *Pacific Linguistics, Series A*. Australian National University, Canberra.
- D. Soetenga Clark. 1993. The phonology of the sio language. In J. Clifton, editor, *no booktitle*, volume 2 of *Phonologies of Austronesian Languages*, pages 25–70. Summer Institute of Linguistics, Ukarumpa.
- Darrell T. Tryon. 1976. *Conversational Tahitian: an introduction to the Tahitian Language of French Polynesia*. Australian National University Press, Canberra.
- Saujas Vaduguru, Aalok Sathe, Monojit Choudhury, and Dipti Sharma. 2021. [Sample-efficient linguistic generalizations through program synthesis: Experiments with phonology problems](#). In *Proceedings of the 18th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 60–71, Online. Association for Computational Linguistics.

William Wright. 1874. *A Grammar of the Arabic Language*, 3 edition. Librairie du Liban, Beirut.