Predicting the Age of Emergence of Consonants

Luis M. T. Jesus [1], **Jihen Trabelsi** [2]
[1]School of Health Sciences (ESSUA), Institute of Electronics and Informatics Engineering of Aveiro (IEETA), Intelligent Systems Associate Laboratory (LASI), University of Aveiro, Portugal
[2]School of Health Sciences (ESSUA), University of Aveiro, Portugal
lmtj@ua.pt, jihen.trabelsi@ua.pt

Abstract

Models of phonological acquisition must account for ambient language effects and the articulatory complexity (AC) of speech. There are, however, very limited assets that allow researchers of under-resourced languages to analyse the effect of predictors such as ambient frequency (AF), functional load (FL) and AC on the age of emergence (AoE) of phonemes. This paper describes the development of a new open access resource, the *Preditores da Aquisição de Consoantes* (PAC) database, which allows the exploration of these issues for European Portuguese (EP) and comparing the results with a typologically unrelated language, Tunisian Arabic (TA). Novel AC, AF, and FL values were calculated for EP and TA consonant inventories. The AoE was estimated using multiple regression models, with results showing the AC predictor had the largest effect in both languages, with AoE values within the ranges previously reported for typically developing monolingual children.

1 Introduction

The level of difficulty in producing a consonant, considering the movements and coordination of the speech articulators (articulatory complexity – AC), the importance of a consonant sound in distinguishing the meaning of words (functional load – FL) and the frequency of occurrence of a specific consonant sound in the child's surrounding language environment (ambient frequency – AF) are known to have an impact in phonological developmental patterns in various languages. In this paper "we take ease of articulation to be primarily defined by reduction of biomechanical effort" (Napoli et al., 2014, p. 426) and calculate AC based on the physics of spring-mass systems (Lindblom et al., 2011) and a taxonomy of phonemic properties (Lindblom & Maddieson, 1988). FL is estimated as the change of phoneme-level entropy in a language system resulting from the merger of a particular contrast (Cychosz, 2017; Stokes & Surendran, 2005).

This paper addresses children's phonological representations of consonants because a consonant bias emerges in children's development when there is "a sophisticated understanding" (Von Holzen & Nazzi, 2020, p. 320) of their language. Adult speakers of non-tonal languages have also been shown to have a consonant bias during lexical processing (Nazzi & Cutler, 2019), reflecting the "underlying structure of speech" (Von Holzen & Nazzi, 2020, p. 320).

The "birthplace" of the Portuguese language is Galicia, Spain, evolving over centuries from its origins in Latin. Historical, cultural, and geographical factors have influenced and driven the development of various regional varieties. European Portuguese (EP) and Brazilian Portuguese (BP) varieties' phonetics and phonology (Jesus et al., 2015) and grammar (Raposo et al., 2020) are distinct but there is a shared core vocabulary (Casteleiro, 2001) that will be the basis of the work presented in this paper (Davies & Bay, 2008). For example, regarding the phonetics and phonology of consonants that are of concern to us, we find the affricates /ʧ, ʤ/ as part of "standard" BP inventory, and not in EP.

The first Eurasian populations established on the Arabian Peninsula in West Asia situated northeast of Africa (Rodriguez-Flores et al., 2016) are the origin of what is designated as the Arabic language, including the emergence of multiple regional varieties, along the north of Africa, that coexist with Standard Arabic (SA). Tunisian Arabic (TA) is used for communication in the daily life of Tunisians, and SA in written formal documents, government communications and education (Masmoudi et al., 2014). TA is a language, widely spoken across Tunisia, impacted by centuries of colonisation, cultural interactions, and exchanges, which affect all levels of language, including phonetics, phonology, vocabulary, morphology, and syntax.

Our motivation to study typologically diverse languages (varying, for example, in consonant inventory size and type, word structure and composition) has to do with the fact that some cross-linguistic differences in the age of emergence (AoE) of consonants have not yet been fully understood based on frequency effects and universal constraints on speech production and perception (Edwards et al., 2015).

EP and TA have not been considered so far, so we have designed a study that allowed us to analyse the influence of AC, AF and FL on consonant development across these typologically distinct languages, and to better understand the challenges that children face in producing these sounds.

2 Method

2.1 Database

According to frequency dictionaries of Portuguese (Davies & Bay, 2008) and Arabic (Buckwalter & Parkinson, 2011) there are very few high frequency words after the lemma ranked 800 for both languages, as shown in Figure 1 and Figure 2.

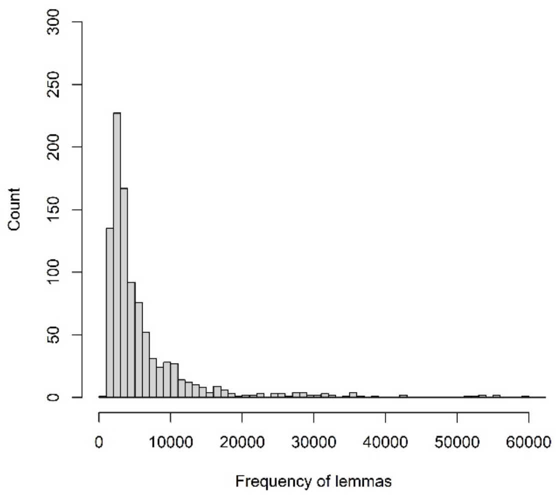

Figure 1: Portuguese frequency histogram.

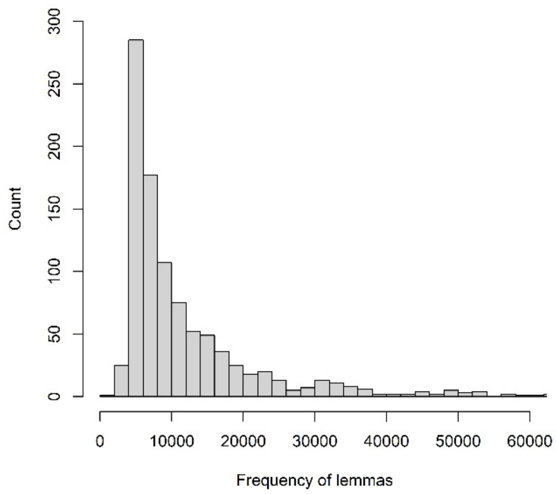

Figure 2: Arabic frequency histogram.

Therefore, a new open access resource, the *Preditores da Aquisição de Consoantes* (PAC) database[1], with the orthographic transcription, frequency, rank, parts of speech and phonemic transcriptions of the 1000 most frequently used lemmas in Portuguese and Arabic, was built in Excel Version 2308. Thirty percent (30%) of the Arabic lemmas, that are not part the Tunisian vocabulary (Bacha, 2015), were not included in the analysis presented in this paper.

The list of lemmas was compiled from a corpus of 20 (Portuguese)/ 30 (Arabic) million words (10% from spontaneous speech data; 90% from written sources). Frequency dictionaries of European (10 million words) and Brazilian (10 million words) Portuguese language varieties (Davies & Bay, 2008), and Tunisian, Egyptian, Levantine, Iraqi, Gulf and Algerian Arabic language varieties (Buckwalter & Parkinson, 2011), were used to compile the corpus.

The phonemic transcriptions were produced using the International Phonetic Alphabet (IPA), according to: An algorithm provided by FreP version 4.6.0.0 (Vigário et al., 2017) and illustrations of the IPA for EP (Jesus et al., 2015); the Convert to IPA tool (Priva et al., 2021) and phonetic descriptions of TA (Masmoudi et al., 2014). The reason why we claim that the PAC database provides data on EP and TA, is related to

[1] Available from the Advanced Communication and Swallowing Assessment (ACSA)platform.

these phonemic transcriptions, which are, to the best of our knowledge, unique as an open access resource.

2.2 Age of Emergence (AoE), Articulatory Complexity (AC), Ambient Frequency (AF) and Functional Load (FL)

A consonant inventory of 19 EP /p, b, t, d, k, g, m, n, ɲ, ɾ, f, v, s, z, ʃ, ʒ, ʁ, l, ʎ/ (Jesus et al., 2015) and 26 TA /b, t, tˤ, d, dˤ, k, q, ʔ, m, n, r, f, θ, ð, ðˤ, s, sˤ, z, ʃ, x, ɣ, ħ, ʕ, h, ʤ, l/ (Thelwall & Sa'Adeddin, 1990; Tice, 2021) phonemes, was used to compile the AoE, based on data reported by Charrua (2011) and Freitas et al. (In Press) for Portuguese, Alquattan (2015) and Elrefaie et al. (2021) and for Arabic.

Previous linguistic models of phonological emergence have included a metric of AC (Cychosz, 2017, p. 317; Stokes & Surendran, 2005, p. 582) based on Kent's (1992, pp. 74–75) original proposal, but we have developed our own classification supported by a recent interpretation (Bybee & Easterday, 2022, pp. 2–6) of Lindblom and Maddieson's (1988) framework, with an additional weight based on the articulatory cost defined by Lindblom et al. (2011, pp. 77–81), to differentiate between some of the consonants that were all originally (Lindblom & Maddieson, 1988) classified as basic.

The first step of the AC calculation involved attributing an integer score of 0 to 2 based on the classification of all consonants according to the three sets proposed by Lindblom and Maddieson (1988): 0 – basic; 1 – elaborated; 2 – complex. Then, three additional weights, were added to the initial score, regarding manner (0 to 2), place (0 to 8) and voicing (0 or 1) of the consonants, based on various literature sources (Bybee & Easterday, 2022; Lindblom et al., 2011; Lindblom & Maddieson, 1988; Napoli et al., 2014). For example, the AC value of 2 (shown in tables 2 and 3) attributed to the consonant /b/, results from the following: 0 (basic set) + 1 (manner = stop) + 0 (place = bilabial) + 1 (voicing = voiced). The score of 7 for /ʃ/ was calculated as follows: 1 (elaborated set) + 2 (manner = strident fricative) + 4 (place = postalveolar) + 0 (voicing = voiceless). The details of the AC calculations for all consonants listed in tables 1 and 2, are distributed in open access with the PAC database[1], as an Excel Version 2308 file that includes all the formulas used to produce the scores and details about the bibliography sources. This will allow as future work and with the contribution of other researchers that can download the data, to fix bugs and integrate new theoretical paradigms into the AC calculations.

The AF for the consonant inventories of both languages was derived, in Excel Version 2308, from the frequency and phonemic transcription data for the 1000 lemmas in PAC database.

The FL was calculated with the Phonological Corpus Tools 1.5.1, using the change in entropy method measured over the whole consonant inventory of each language (Cychosz, 2017, p. 314), not just from word initial consonants as in Stokes & Surendran (2005) since not all children pay more attention "to the onset of words" (Stokes & Surendran, 2005, p. 581).

2.3 Multiple regression modelling

Multiple linear regression models were developed in R version 4.3.1 running in RStudio 2023.06.1+524, with `AoE` as outcome variable, and `AC`, `AF`, and `FL` as predictors, for the two languages.

Both models satisfied the normality assumption (i.e., its residuals were approximately normally distributed) and the constant variance assumption (homoscedasticity), as assessed by the following visual diagnostics plots: Histogram of residuals; Q-Q plots of residuals; residuals plot.

We have not considered interactions between any of the predictors (`AC`, `AF`, and `FL`), because we are not aware of any previous work on EP and TA that has shown these may be theoretically motivated (Winter, 2020, p. 155).

The predictors were standardised/ z-scored (subtracting the mean and dividing the centred values by the standard deviation).

The models' marginal effects, regression lines and shading spanning the 95% confidence intervals were plotted using the sjPlot 2.8.14 package.

3 Results

3.1 European Portuguese (EP)

Table 1 presents the EP consonant inventory, AoE in years as reported in the literature, AC, AF, and FL relative values before standardisation.

Consonant	AoE	AC	AF (%)	FL (%)
/p/	1.5	1	7.17	0.96
/b/	1.5	2	1.74	2.12
/t/	2.0	3	9.54	7.09
/d/	1.5	4	13.35	21.70
/k/	2.0	8	8.58	21.59
/g/	2.5	9	1.67	0.51
/m/	1.5	2	5.28	3.86
/n/	2.5	5	3.25	2.83
/ɲ/	2.0	8	0.39	0.03
/ɾ/	4.0	4	21.14	5.21
/f/	2.5	2	1.76	0.77
/v/	3.0	4	2.77	3.33
/s/	3.0	5	9.39	20.75
/z/	4.0	7	2.01	0.53
/ʃ/	2.5	7	4.61	1.13
/ʒ/	4.0	9	1.21	0.55
/ʁ/	3.0	10	1.07	0.68
/l/	3.5	4	4.38	2.78
/ʎ/	3.5	7	0.67	3.56

Table 1: EP consonant inventory, AoE (years), AC (1 – least complex; 10 – most complex), AF (relative to the to the total number of phonemes), and FL (relative to the highest functional load).

Figure 2 shows the relationship between the EP consonant inventory, AoE (as reported in the literature), AC, AF, and FL.

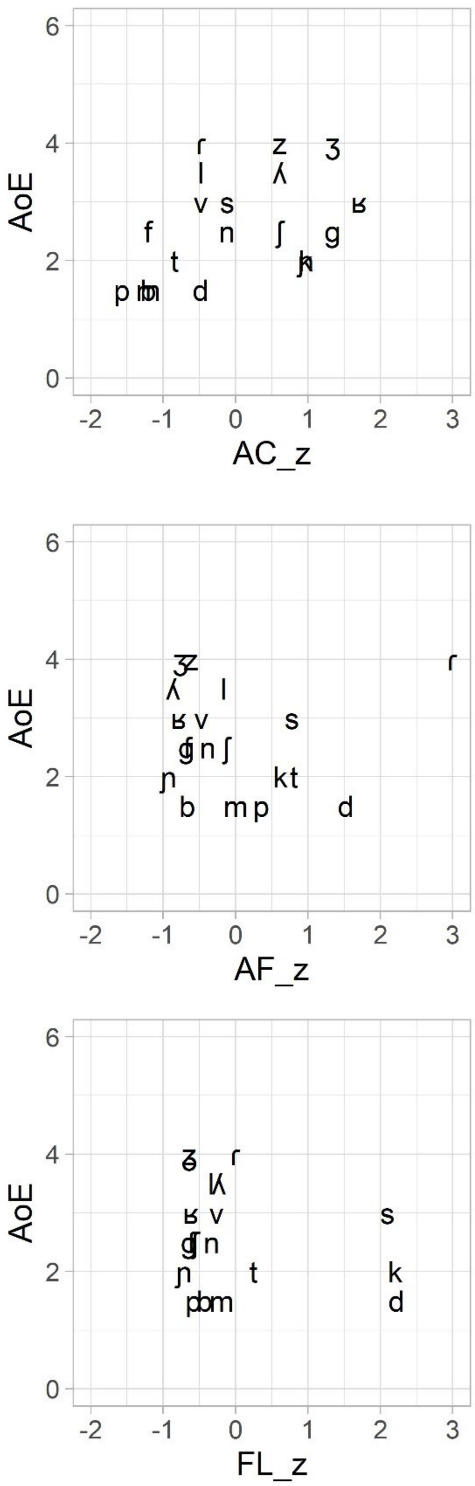

Figure 3: EP consonant inventory, AoE in years as a function of z-scored AC, AF, and FL.

A multiple linear regression model for EP with the `lm` function syntax `AoE ~ AC + AF + FL`, accounted for 20% of the variance in the `AoE` (Winter, 2020, pp. 103–116; 133–156): adjusted R^2 = 0.197. The `AC` predictor had the largest effect on the `AoE` (Winter, 2020, p. 109): For each increase in AC by one standard deviation (holding all variables constant), the AoE increased significantly by half a year (*slope* = 0.48; *SE* = 0.18; *p* = 0.030).

The slope for the AF predictor was also positive (*slope* = 0.36; *SE* = 0.24; *p* = 0.156), suggesting that the more complex (AC) and frequent (AF), a phoneme is, the later in Portuguese children's lives it is acquired. We have, however, also found the slope of the FL was negative (*slope* = -0.15; *SE* = 0.86; *p* = 0.867), meaning that when all the variables in the term increased (which is feasible since the AC and AF slopes were positive), the AoE was predicted to decrease.

Figure 3 shows the marginal effects computed for the z-scored EP model predictors (AC_z, AF_z and FL_z) at three different levels: -2, 0 and 3.

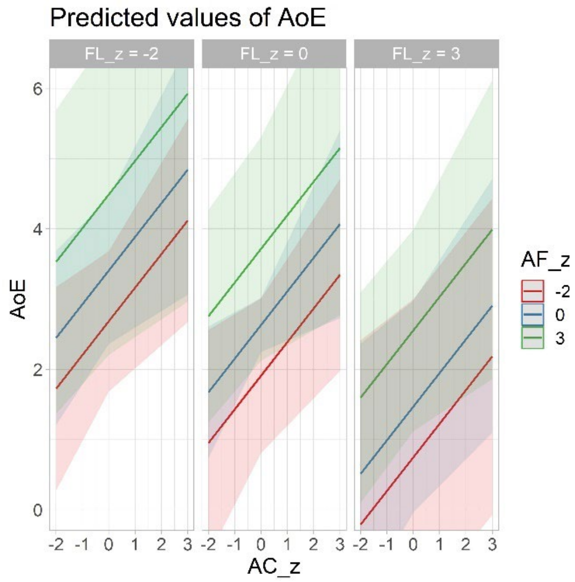

Figure 4: Marginal effects of the EP predictors.

3.2 Tunisian Arabic (TA)

Table 2 presents the TA consonant inventory, AoE in years as reported in the literature, AC, AF, and FL values before standardisation. We estimated an FL value of zero for /ðˤ/ because only ten different lemmas in the PAC database included this phoneme.

Consonant	AoE	AC	AF (%)	FL (%)
/b/	1.5	2	5.14	18.68
/t/	1.5	3	2.07	1.13
/tˤ/	5.0	4	1.02	0.52
/d/	1.5	4	3.23	4.56
/dˤ/	5.0	5	0.71	1.30
/k/	3.0	8	2.93	1.75
/q/	5.5	10	2.62	6.11
/ʔ/	1.5	1	15.89	1.29
/m/	1.5	2	8.37	9.08
/n/	1.5	4	6.97	2.01
/r/	5.0	4	4.85	3.46
/f/	3.5	2	3.72	1.28
/θ/	5.5	2	0.48	0.50
/ð/	5.5	4	0.92	1.16
/ðˤ/	5.5	5	0.18	0.00
/s/	4.5	5	2.61	2.85
/sˤ/	4.5	6	0.92	0.54
/z/	5.0	7	0.43	1.71
/ʃ/	4.0	7	1.29	1.57
/x/	4.0	7	1.10	2.26
/ɣ/	4.5	8	0.33	1.31
/ħ/	2.5	1	2.32	3.12
/ʕ/	2.5	3	5.66	4.99
/h/	1.5	0	2.66	4.67
/ʤ/	3.5	9	1.40	3.31
/l/	2.0	4	22.18	20.82

Table 2: TA consonant inventory, and non-standardised AoE, AC (0 – least complex; 10 – most complex), AF, and FL.

Figure 3 shows the relationship between the TA consonant inventory, AoE (as reported in the literature), AC, AF, and FL.

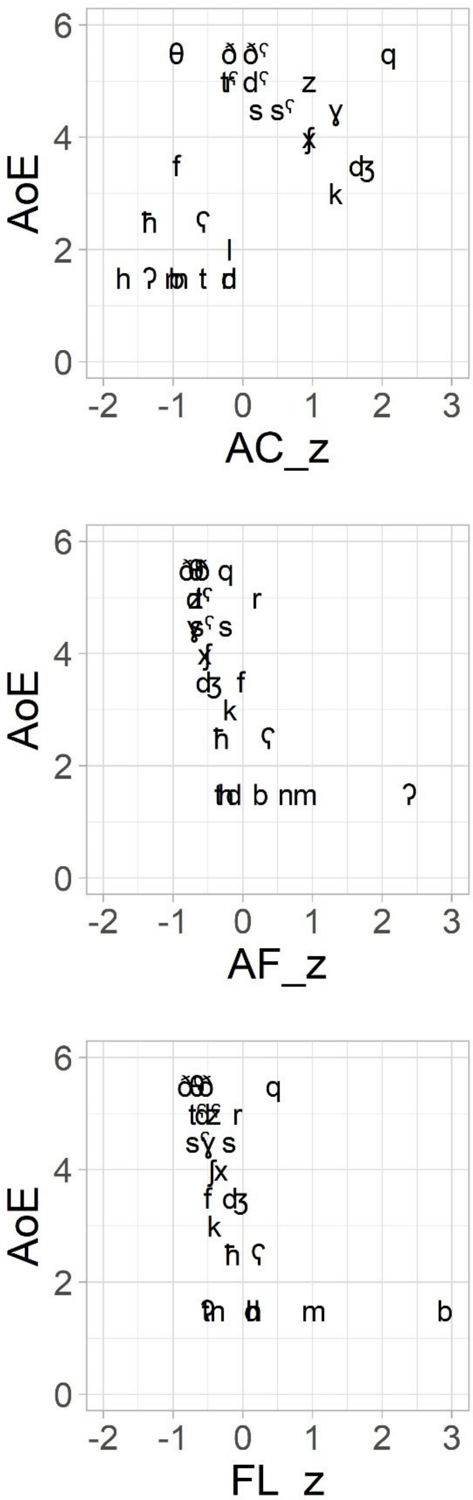

Figure 5: TA consonant inventory, AoE as a function of z-scored AC, AF, and FL.

A multiple linear regression model for AT, with the same predictors as EP, accounted for 37% of the variance in the AoE: Adjusted R^2 = 0.365. The AC predictor had the largest effect on the AoE: For each increase in AC by one standard deviation, the AoE increased significantly by seven months (*slope* = 0.58; *SE* = 0.26; *p* = 0.038). The slopes for the AF and FL predictors were also negative (AF – *slope* = -0.47; *SE* = 0.34; *p* = 0.182; FL – *slope* = -0.30; *SE* = 0.33; *p* = 0.360), suggesting that the more frequent a phoneme is (higher AF values) and the more meaningful a contrast is (higher FL values), the earlier in Tunisian children's lives it is acquired. Figure 5 shows the TA marginal effects.

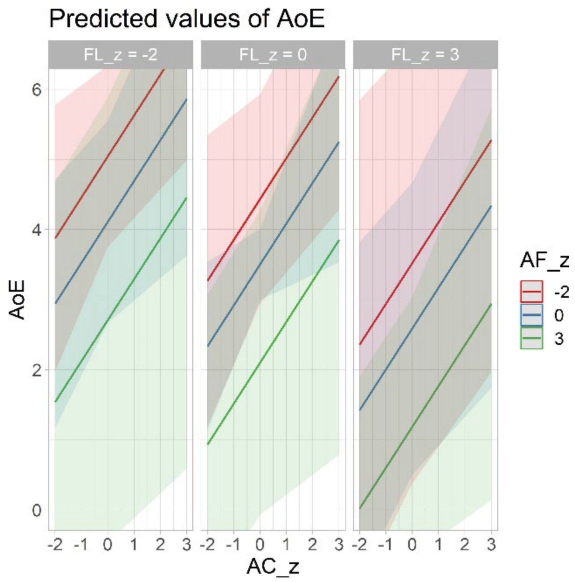

Figure 6: Marginal effects of the TA predictors.

4 Discussion

This paper discusses predictors of the AoE of consonants estimated from adult Portuguese (Davies & Bay, 2008) and Arabic (Buckwalter & Parkinson, 2011) languages' rank-ordered listings of the top one thousand lemmas, starting with the most frequent word.

These words constitute the core vocabulary that people encounter in regular conversation since they are the most used in the language. Frequency dictionaries have long been recognised, by language teachers and learners, as the most effective way of acquiring a vocabulary (Davies & Bay, 2008). Also, speech and language therapists can ensure that their work addresses the most relevant linguistic features, focusing their intervention on these high-frequency lemmas; "the more words containing a sound that a child has learned to say, the more practiced the child becomes at recognizing and reproducing the sound abstracted away from the phonological contexts of a few specific words" (Edwards et al., 2015, p. 307). Therefore, calculating the frequency of phonemes based on high-frequency words accounts for the fact that "type rather than token frequency" (Edwards et al., 2015, p. 307) should be the predictor for AF.

The fact that the morphological structure of the two languages is different should be considered. For example, in EP, the regular plural is formed by adding /ʃ/ to the end of the word, which means that there are many instances of the voiceless postalveolar fricative that are often present in spontaneous speech, and which are omitted here.

According to the literature (Alqattan, 2015; Charrua, 2011; Elrefaie et al., 2021; Freitas et al., In Press), most EP and TA consonants are acquired after a year and half and before children are six years old.

Comparing the AoE for the eleven consonants /b, t, d, k, m, n, f, s, z, ʃ, l/ that can be observed in both languages, voiced bilabial /b, m/ and dental /d/ stops emerge at the same early stage, but different acquisition ages are reported for the other phonemes.

Reverse acquisition orders were observed for /ʃ/ and /l/ in the two languages. For example, /l/ is acquired at an early stage in TA and only emerges very late in some contexts of Portuguese children's phonology (Freitas et al., In Press).

Consonants that are acquired before two years of age have low AF (less than 9%), except for the EP /d/ and the TA /ʔ/. The consonants with the highest FL (/d/ in EP and /b/ in TA), are acquired at an early developmental stage (around a year and half). Language specific AF and FL values were observed for EP and AP.

Even if more words were used to compute the FL of /ðˤ/ its value is likely to be low, because for some languages' emphatics "appearing in a small set of words" exercise "a limited functional load" (Anonby, 2020, p. 292).

The `AC` predictor had the largest (significant) effect on the `AoE` for both languages with a very similar increase (6-7 months) in `AoE` for an increase in `AC` of one standard deviation.

The slope of the `AF` predictor was positive for Portuguese which is a "counterintuitive" result that apparently contradicts most literature on the effect of frequency in phonological acquisition (Edwards et al., 2015). One must bear in mind that these are the results of a multiple linear regression model, so the effect of the other predictors is factored in. If we were to run a simple regression model with the `lm` function syntax `AoE ~ AC` the slope for the `AF` predictor would be negative (*slope* = -0.02; *SE* = 0.21; *p* = 0.933), suggesting the opposite effect. However, this model only accounts for 6% of the variance in the `AoE` and the slope is close to zero (flat slope/ near no effect of the `AC` predictor).

5 Conclusions

The AoE, AC, AF and FL estimates discussed in this paper can provide guidance in establishing intervention strategies to facilitate the acquisition of consonants for effective communication in Portugal and Tunisia. The proposed reference values of AC constitute a departure from previously ill-defined values of AC that reflected what we knew about the growth in motor control more than thirty years ago. This skewed previous models of `AoE` that used `AC` predictor values based on a biased hierarchy.

The multiple regression models presented in this paper can be used by researchers, educators, and clinicians to estimate a typical range for the `AoE` of consonants. Current results showed that `AC` was the only significant predictor.

We plan to expand the size of the PAC database to more than 2000 lemmas, to explore the orthographic transcription and parts of speech as future work.

Acknowledgments

This work was supported by National Funds through the FCT – Foundation for Science and Technology, in the context of the project UIDB/00127/2020.

References

Alqattan, S. (2015). *Early Phonological Acquisition by Kuwaiti Arabic Children* [Ph.D. Thesis]. Newcastle University, UK.

Anonby, E. (2020). Emphatic consonants beyond Arabic: The emergence and proliferation of uvular-pharyngeal emphasis in Kumzari. *Linguistics*, *58*(1), 275–328. https://doi.org/10.1515/ling-2019-0039

Bacha, M. (2015). *Tunisian Arabic - English Dictionary*. CreateSpace.

Buckwalter, T., & Parkinson, D. (2011). *A Frequency Dictionary of Arabic Core Vocabulary for Learners*. Routledge.

Bybee, J., & Easterday, S. (2022). Primal consonants and the evolution of consonant inventories. *Language Dynamics and Change*, *13*(1), 1–33. https://doi.org/10.1163/22105832-bja10020

Casteleiro, J. M. (2001). *Dicionário da língua portuguesa contemporânea da Academia das Ciências de Lisboa*. Academia das Ciências de Lisboa e Editorial Verbo.

Charrua, C. (2011). *Aquisição Fonética-Fonológica do Português Europeu dos 18 aos 36 meses [Phonetic-Phonological Acquisition of European Portuguese from 18 to 36 months]* [M.Sc. Thesis]. Instituto Politécnico de Setúbal (IPS), Portugal.

Cychosz, M. (2017). Functional load and frequency predict consonant emergence across five languages. In *UC Berkeley Phonetics and Phonology Lab Annual Report* (pp. 312–320).

Davies, M., & Bay, M. (2008). *Frequency Dictionary of Portuguese*. Routledge.

Edwards, J., Beckman, M., & Munson, B. (2015). Frequency effects in phonological acquisition. *Journal of Child Language*, *42*(2), 306–311. https://doi.org/10.1017/S0305000914000634

Elrefaie, D. A., Hegazi, M. A. E.-F., El-Mahallawi, M. M., & Khodeir, M. S. (2021). Descriptive analysis of the development of the Arabic speech sounds among typically developing colloquial Egyptian Arabic-speaking children. *The Egyptian Journal of Otolaryngology*, *37*. https://doi.org/10.1186/s43163-021-00094-w

Freitas, M., Lousada, M., & Ramalho, A. (In Press). Portuguese (European) speech development. In S. McLeod (Ed.), *The Oxford handbook of speech development in languages of the world*. Oxford University Press.

Jesus, L. M. T., Valente, A. R. S., & Hall, A. (2015). Is the Portuguese version of the passage 'The North Wind and the Sun' phonetically balanced? *Journal of the International Phonetic Association*, *45*(01), 1–11. https://doi.org/10.1017/S0025100314000255

Kent, R. (1992). The biology of phonological development. In C. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 65–90). York Press.

Lindblom, B., Diehl, R., Park, S.-H., & Salvi, G. (2011). Sound systems are shaped by their users. In G. Clements & R. Ridouane (Eds.), *Where Do Phonological Features Come From? Cognitive, physical and developmental bases of distinctive speech categories* (pp. 65–98). John Benjamins. https://doi.org/10.1075/lfab.6.04lin

Lindblom, B., & Maddieson, I. (1988). Phonetic Universals in Consonant Systems. In L. Hyman & C. Li (Eds.), *Language, Speech and Mind: Studies in Honour of Victoria A. Fromkin* (pp. 62–78). Routledge.

Masmoudi, A., Esteve, Y., Khmekhem, M., Bougares, F., & Belguith, L. (2014). Phonetic Tool for the Tunisian Arabic. *Proceedings of the 4th Workshop on Spoken Language Technologies for Under-Resourced Languages (SLTU 2014)*, 252–256.

Napoli, D., Sanders, N., & Wright, R. (2014). On the Linguistic Effect of Articulatory Ease, with a focus on Sign Languages. *Language*, *90*(2), 424–456.

Nazzi, T., & Cutler, A. (2019). How Consonants and Vowels Shape Spoken-Language Recognition. *Annual Review of Linguistics*, *5*(1), 25–47. https://doi.org/10.1146/annurev-linguistics-011718-011919

Priva, U. C., Strand, E., Yang, S., Mizgerd, W., Creighton, A., Bai, J., Mathew, R., Shao, A., Schuster, J., & Wiepert, D. (2021). *The Cross-linguistic Phonological Frequencies (XPF) Corpus*. Brown University, Providence, Rhode Island, USA.

Raposo, E., Nascimento, M., Mota, M., Segura, L., Mendes, A., Andrade, A., Vicente, G., & Veloso, R. (Eds.). (2020). *Gramática do Português*. Fundação Calouste Gulbenkian.

Rodriguez-Flores, J. L., Fakhro, K., Agosto-Perez, F., Ramstetter, M. D., Arbiza, L., Vincent, T. L., Robay, A., Malek, J. A., Suhre, K., Chouchane, L., Badii, R., Al-Nabet Al-Marri, A., Abi Khalil, C., Zirie, M., Jayyousi, A., Salit, J., Keinan, A., Clark, A. G., Crystal, R. G., & Mezey, J. G. (2016). Indigenous Arabs are descendants of the earliest split from ancient Eurasian populations. *Genome Research*, *26*(2), 151–162. https://doi.org/10.1101/gr.191478.115

Stokes, S. F., & Surendran, D. (2005). Articulatory Complexity, Ambient Frequency, and Functional Load as Predictors of Consonant Development in Children. *Journal of Speech, Language, and Hearing Research*, *48*(3), 577–591.

Thelwall, R., & Sa'Adeddin, M. A. (1990). Arabic. *Journal of the International Phonetic Association*, *20*(2), 37–39. https://doi.org/10.1017/S0025100300004266

Tice, P. (2021). *Language and Performance in Post-revolution Tunisia* [Ph.D. Thesis]. The Ohio State University.

Vigário, M., Martins, F., Cruz, M., Paulino, N., & Frota, S. (2017). Basic research in phonology, resources and applications–the case of frequency. *Cadernos de Estudos Lingüísticos*, *59*(3), 599–616. https://doi.org/10.20396/cel.v59i3.8651000

Von Holzen, K., & Nazzi, T. (2020). Emergence of a consonant bias during the first year of life: New evidence from own-name recognition. *Infancy*, *25*(3), 319–346. https://doi.org/10.1111/infa.12331

Winter, B. (2020). *Statistics for Linguists: An Introduction Using R*. Routledge.