

LanCorp Metadata

Each utterance in LanCorp is linked to file-, utterance-, and participant-level metadata, which can be used for (socio)linguistic analyses. There are 13 file- and utterance-level variables and 40 participant-level sociolinguistic variable (groups). The complete list of available metadata is provided in Table 1 and Table 2, where I also note the type of metadata/variable and provide examples for each of them.

Table 1. File- and utterance-level metadata linked to the Lannang Corpus

variable	variable name in corpus	type	example
tag	tag	categorical	<CLIN-18-68:1>
unique utterance number	no	continuous	1, 2, 3...
ELAN file	filename	categorical	PC0001-CLIN18.eaf
year collected	year	continuous	18 (2018), 19 (2019)
region collected	region	categorical	MNL (Manila)
audio file time alignment (begin/end) in minutes, seconds, milliseconds	begin, end	time	00:03.0
duration of utterance in minutes, seconds, milliseconds	duration	time	00:04.3
interlocutor identification	interlocutor.no	categorical	1, 2, 3...
language of utterance	lg.clause	categorical	l (Lánang-uè), x (unknown), t (Tagalog), g (English with Tagalog borrowings), etc.
context of utterance	context	categorical	CLIN (interview), PROT (unstructured conversation), FRST (narrative)
transcriber identification	transcriber.id	categorical	1, 2, 3...
file index	index.file	categorical	1, 2, 3...
relative point in audio file - in percentage	filepercent	categorical	0.002298851

Table 2. Informant-level sociolinguistic metadata linked to the Lannang Corpus

variable	variable name in corpus	type	example
individual identification number	idno	continuous	1,2,3
date of collection	date.of.collection	continuous	2019, 2020
occupation	occupation	categorical	student, pastor
Chinese ancestral origin of mother (by province)/ (by city)	origin.ancestral.province.maternal	categorical	Fujian, Shishi

Chinese ancestral origin of father (by province)/ (by city)	origin.ancestral.province.paternal	categorical	Guangdong, Taishan
age	age	continuous	25
sex	sex	categorical	M, F
distance from Binondo (average)/ (place of current residence)/ (place of longest residence)	distance.average, distance.curren.binondo, distance.longres.binondo	continuous	2.7
residence in Manila (at all)/ (current)/ (longest)	bin.manila, bin.curren.manila, bin.longres.manila	categorical	1,-1
religion (protestant or not)	bin.christian	categorical	1,-1
religion	religion	categorical	Protestant, Catholic
institution of worship	institution.worship	categorical	UECP
identity as Filipino/ Filipino-Chinese/ Chinese Filipino/ Huakiau/ Chinese immigrant/ Chinese	identity.fil, identity.filchi, identity.chifil, identity.huakiau, identity.chineseimmigrant, identity.chinese	categorical	1,0
identity (main)	main.identity	categorical	Lannang, Chinese Filipino
generation	generation	numerical	2,3,4
language used with mother	language.with.mother	categorical	L (Lánnang-uè), H (Hokkien)
language used with father	language.with.father	categorical	T (Tagalog), CD (Taishanese)
native language is Tagalog/Philippine Hokkien/ Lánnang-uè/ English/Standard Cantonese/ Taishanese/ Standard Hokkien/ Mandarin	native.language.Tagalog, native.language.PHHokkien, native.language.Lannangue, native.language.English, native.language. StandardCantonese, native.language.Taishanese, native.language.StandardHokkien, native.language.Mandarin	categorical	1,0
first language	language.L1	categorical	L (Lánnang-uè), H (Hokkien)
second language	language.L2	categorical	T (Tagalog), CD (Taishanese)
language used at home	language.home	categorical	L (Lánnang-uè), H (Hokkien)
highest level of education	education.highest	categorical	Doctorate, Masters

place of education (elementary)/ (high school)/ (college)	school.elementary, school.highschool, school.college	categorical	Hope Christian High School
number of languages used in education (elementary)/ (high school)/ (college)	number.moi. elementary, number.moi.highschool, number.moi.college	continuous	3
medium of instruction in Chinese education (elementary)/ (high school)/ (college)	moi.elementary.chinese, moi.highschool.chinese, moi.college.chinese	categorical	L (Lánnang-uè), H (Hokkien)
place of current residence (city)/(district)	residence.current.city, residence.current.district	categorical	Manila
place of longest residence (city)/(district)	residence.longest.city, residence.longest.district	categorical	Manila
latitude/longitude of current city of residence	residence.current.city.latitude, residence.current.city.longitude	continuous	14.599512, 120.984222
latitude/longitude of longest city of residence	residence.longest.city.latitude, residence.longest.city.longitude	continuous	14.599512, 120.984222
length of stay in the Philippines in years	years.in.philippines	continuous	15, 49
percentage of life in the Philippines	length.stay.philippines.proportion	continuous	0.6
length of stay in Manila in years	years.in.manila	continuous	40, 49
percentage of life in Manila	length.stay.manila.proportion	continuous	0.2
self-reported proficiency in Tagalog/ Hokkien/ Lánnang-uè/ English/ Cantonese/ Taishanese/ Mandarin	prof.t, prof.h, prof.l, prof.e, prof.cs, prof.cd, prof.m	continuous	1 to 7
self-reported comfort in languages (see list above)	comf.t, comf.h, comf.l, comf.e, comf.cs, comf.cd, comf.m	continuous	1 to 7
self-reported confidence in languages (see list above)	conf.t, conf.h, conf.l, conf.e, conf.cs, conf.cd, conf.m	continuous	1 to 7
self-reported frequency of language use -	freq.t, freq.h, freq.l, freq.e, freq.cs, freq.cd, freq.m	continuous	1 to 7

languages (see list above)			
self-reported pride in languages (see list above)	pride.t, pride.h., pride.l, pride.e, pride.cs, pride.cd, pride.m	continuous	1 to 7
self-reported importance of language in society - languages (see list above)	imp.t, imp.h, imp.l, imp.e., imp.cs, imp.cd, imp.m	continuous	1 to 7
attitudes - Lánnang-uè is barok/ konyo/ comical/ natural/ bad-sounding/ bastardized/ prestigious/ reflective of the Lannang identity	barok.l, conyo.l, comical.l, natural.l, badsounding.l, bast.l, prestige.l, ref.l	continuous	1 to 7