Appendix (Not for publication)

A The Origins of Periphrastic Future Tense

A.1 Crop Return and Periphrastic Future Tense

Table A1: Geographical Origins of Periphrastic Future Tense (OLS)

	Existence of Future Tense					
	(1)	(2)	(3)	(4)	(5)	(6)
Crop Return (pre-1500CE)	0.06**	0.06**	0.07**	0.07**	0.09***	0.11***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Absolute Latitude		-0.03	-0.03	0.03	0.05	0.15
		(0.03)	(0.04)	(0.10)	(0.10)	(0.11)
Elevation		0.04	0.03	0.05	0.05	0.03
		(0.04)	(0.04)	(0.05)	(0.04)	(0.05)
Ruggedness		-0.04	-0.03	-0.04	-0.05	-0.02
		(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Coast Length		0.08***	0.07***	0.06***	0.06***	0.06***
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Precipitation (mm/month)			0.01	0.02	0.01	0.01
			(0.07)	(0.07)	(0.07)	(0.08)
Precipitation (mm/month) (std)			0.07**	0.01	0.01	0.04
			(0.03)	(0.05)	(0.05)	(0.05)
Precipitation Volatility			-0.04	-0.03	-0.03	-0.04
			(0.08)	(0.08)	(0.08)	(0.08)
Precipitation Spatial Correlation			0.01	0.93***	0.91***	0.95***
			(0.04)	(0.26)	(0.26)	(0.30)
Temperature (Daily Mean)				0.05	0.07	0.08
				(0.07)	(0.06)	(0.08)
Temperature (Daily Mean) (std)				0.07	0.08*	0.05
				(0.05)	(0.05)	(0.04)
Temperature Volatility				-0.01	-0.06	-0.09
				(0.08)	(0.08)	(0.09)
Temperature Spatial Correlation				-0.93***	-0.91***	-0.93***
				(0.26)	(0.26)	(0.30)
Unproductive Period (pre-1500CE)					0.08**	0.09***
					(0.03)	(0.03)
Regional FE	No	No	No	No	No	Yes
Adjusted-R^2	0.01	0.02	0.02	0.06	0.07	0.11
Observations	275	275	275	275	275	275

Notes: This table establishes the positive, statistically, and economically significant association between pre-1500CE potential crop return and the existence of periphrastic future tense in the language spoken in this region, accounting for regional fixed-effects and other geographical characteristics. All specifications in this table are identical to the ones presented in Table 1, but here the coefficients of the controls are presented. All independent variables have been normalized by subtracting their mean and dividing by their standard deviation. Thus, all coefficients can be compared and show the impact of a one standard deviation in the independent variable on the probability of having a future tense in the language. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table A2: Crop Return and Periphrastic Future Tense

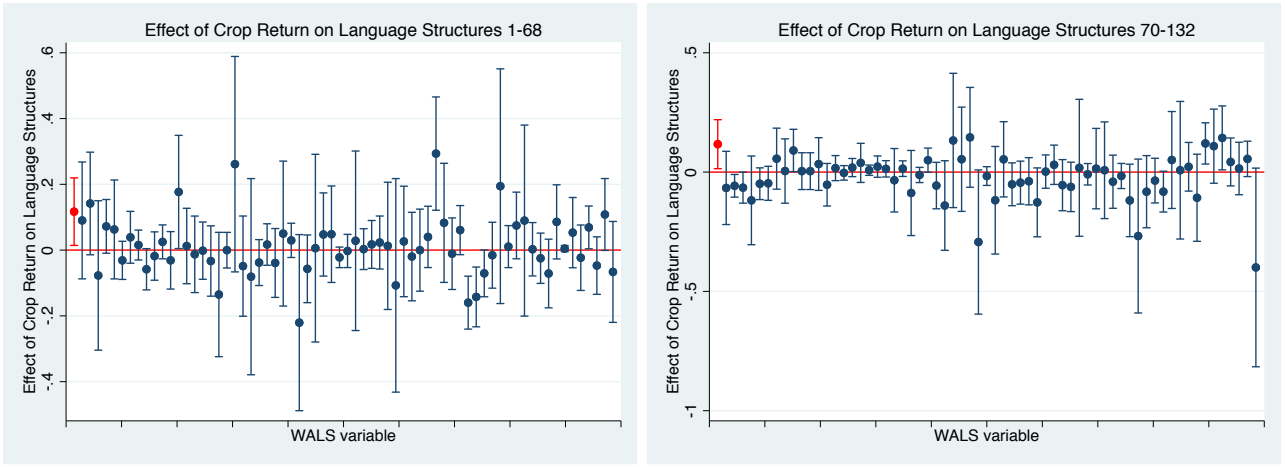
	Existence of Periphrastic Future Tense					
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Probit					
Crop Return (pre-1500CE)	0.06**	0.06**	0.06**	0.07**	0.09***	0.12***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Geographical Controls	No	Yes	Yes	Yes	Yes	Yes
Regional FE	No	No	No	No	No	Yes
Pseudo-R^2	0.01	0.03	0.04	0.08	0.10	0.14
Observations	275	275	275	275	275	275
Panel B: OLS - Spatial-Autocorrelation, Clustering and Selection On Unobservables						
Crop Return (pre-1500CE)	0.06**	0.06**	0.07**	0.07**	0.09***	0.11***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
	([0.04])	([0.04])	([0.03])	([0.03])	([0.03])	([0.03])
	[0.04]	[0.04]	[0.03]	[0.03]	[0.03]	[0.03]
	{0.03}	{0.03}	{0.03}	{0.03}	{0.03}	{0.03}
Altonji et al						-2.09
δ						-2.84
β-Oster						0.13
R^2	0.01	0.04	0.06	0.10	0.12	0.17

Notes: This table establishes the positive, statistically, and economically significant association between pre-1500CE potential crop return and the existence of periphrastic future tense in the language spoken in this region, accounting for regional fixed-effects and other geographical characteristics. All columns have the same specification and controls as Table 1. Geographical controls include absolute latitude, mean elevation, terrain ruggedness, and coast length, as well as other agriculture-related controls as precipitation and temperature means and standard deviations. All independent variables have been normalized by subtracting their mean and dividing by their standard deviation. Thus, all coefficients can be compared and show the impact of a one standard deviation in the independent variable on the probability of having a future tense in the language. Heteroskedasticity robust standard error estimates are reported in parentheses, clustered at the language genus in parenthesis and squared brackets, spatial auto-correlation corrected standard errors (Conley, 1999) in squared brackets and Cliff-Ord ML in curly brackets; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table A3: Geographic Origins of Future and Language Structures

	Language Structure						
		Temporal Structures		Non-Temporal Structures			
	Future	Past	Perfect	Possessive	Evidentiality	Consonants	Colors
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Crop Return (pre-1500CE)	0.12***	-0.06	0.05	-0.07*	0.00	0.08	0.06
	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)	(0.06)	(0.34)
All Geographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.11	0.08	0.14	0.15	0.20	0.31	-0.03
Observations	275	218	218	224	387	542	117

Notes: This table establishes the positive, statistically, and economically significant association between pre-1500CE potential crop return and the existence of periphrastic future tense in a language, and not with other language structures. The analysis accounts for regional fixed-effects and other geographical characteristics as in previous tables. Other language structures include the existence a past tense, a perfect tense, the existence of obligatory possessive inflections, semantic distinctions of evidentiality, the number of consonants, the ratio of consonants to vowels and the number of colors. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.



(a) WALS Structures 1-68

(b) WALS Structures 69-132

Figure A1: Orthogonality of Crop Return and Other Language Structures in WALS
Impact of Crop Return on Periphrastic Future Tense in Red

Table A4: Agricultural Intensity and Crop Return

	Agricultural Intensity			
	Full Sample		Future Sample	
	(1)	(2)	(3)	(4)
Crop Return (pre-1500CE)	0.19***	0.22***	0.27***	0.30***
	(0.03)	(0.02)	(0.07)	(0.06)
Regional FE	No	Yes	No	Yes
All Geographical Controls	No	Yes	No	Yes
Adjusted-R^2	0.04	0.64	0.07	0.61
Observations	1306	1306	264	264

Notes: This table establishes the positive statistically and economically significant association between a language's contemporary homeland's crop return and the level of agricultural intensity of a pre-colonial society that speaks that language. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

B Origins of Sex-Based Grammatical Gender

Table B1: Geographic Origins of Sex-Based Grammatical Gender (OLS)

	Existence of Sex-Based Grammatical Gender					
	(1)	(2)	(3)	(4)	(5)	(6)
Plow Negative Crops (Average Caloric Yield, pre-1500)	-0.13**	-0.19***	-0.25***	-0.26***	-0.29***	-0.23***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.08)
All Crops (Average Caloric Yield, pre-1500)	0.17***	0.21***	0.28***	0.29***	0.32***	0.25***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.08)
Absolute Latitude		-0.10***	-0.19***	-0.06	-0.06	-0.13
		(0.04)	(0.04)	(0.10)	(0.10)	(0.10)
Elevation		-0.01	-0.14***	-0.11**	-0.11**	-0.11**
		(0.05)	(0.04)	(0.05)	(0.05)	(0.05)
Ruggedness		-0.02	0.08	0.08	0.08	0.06
		(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Coast Length		0.08***	0.09***	0.09***	0.08***	0.08***
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Precipitation (mm/month)			-0.11*	-0.11	-0.12	-0.08
			(0.07)	(0.07)	(0.07)	(0.07)
Precipitation (mm/month) (std)			-0.02	0.03	0.04	0.03
			(0.03)	(0.05)	(0.05)	(0.05)
Precipitation Volatility			0.03	-0.01	0.00	0.01
			(0.07)	(0.08)	(0.08)	(0.08)
Precipitation Spatial Correlation			0.13***	0.14	0.15	0.12
			(0.03)	(0.28)	(0.29)	(0.28)
Temperature (Daily Mean)				0.08	0.07	0.04
				(0.07)	(0.07)	(0.07)
Temperature (Daily Mean) (std)				-0.07	-0.08*	-0.07
				(0.05)	(0.04)	(0.05)
Temperature Volatility				-0.08	-0.07	-0.10
				(0.09)	(0.09)	(0.09)
Temperature Spatial Correlation				-0.01	-0.02	-0.02
				(0.29)	(0.29)	(0.28)
Unproductive Period (pre-1500CE)					-0.06*	-0.02
					(0.03)	(0.03)
Regional FE	No	No	No	No	No	Yes
Adjusted-R^2	0.03	0.06	0.14	0.15	0.15	0.21
Observations	217	217	217	217	217	217

Notes: This table establishes the statistically and economically significant association between the historical agricultural determinants of plow usage and the existence of sex-based grammatical gender in the language spoken in this region, accounting for regional fixed-effects and other geographical characteristics. All columns have the same specification and controls as Table 5. Geographical controls include absolute latitude, mean elevation, terrain ruggedness, and coast length, as well as other agriculture-related controls as precipitation and temperature means and standard deviations. All independent variables have been normalized by subtracting their mean and dividing by their standard deviation. Thus, all coefficients can be compared and show the impact of a one standard deviation in the independent variable on the probability of having a future tense in the language. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table B2: Geographic Origins of Sex-Based Grammatical Gender

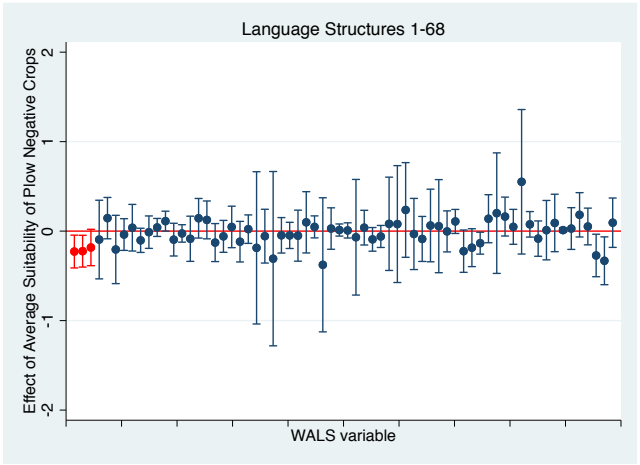
	Existence of Sex-Based Grammatical Gender					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Probit						
Plow Negative Crops (Average Caloric Yield, pre-1500)	-0.13**	-0.19***	-0.25***	-0.26***	-0.28***	-0.20**
	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.08)
All Crops (Average Caloric Yield, pre-1500)	0.16***	0.21***	0.28***	0.30***	0.32***	0.22***
	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)
Geographical Controls	No	Yes	Yes	Yes	Yes	Yes
Regional FE	No	No	No	No	No	Yes
Pseudo-R^2	0.03	0.07	0.16	0.18	0.18	0.25
Observations	216	216	216	216	216	216
Panel B: OLS - Spatial-Autocorrelation, Clustering and Selection On Unobservables						
Plow Negative Crops (Average Caloric Yield, pre-1500)	-0.13**	-0.19***	-0.25***	-0.26***	-0.29***	-0.23***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.08)
	([0.07])	([0.07])	([0.07])	([0.07])	([0.07])	([0.08])
	[0.06]	[0.07]	[0.07]	[0.07]	[0.07]	[0.07]
	{0.06}	{0.06}	{0.06}	{0.06}	{0.07}	{0.09}
Altonji et al						-2.26
δ						-1.29
β-Oster						-0.27
All Crops (Average Caloric Yield, pre-1500)	0.17***	0.21***	0.28***	0.29***	0.32***	0.25***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.08)
	([0.07])	([0.07])	([0.07])	([0.07])	([0.07])	([0.07])
	[0.06]	[0.06]	[0.06]	[0.06]	[0.06]	[0.07]
	{0.06}	{0.06}	{0.06}	{0.06}	{0.06}	{0.08}
Altonji et al						-3.02
δ						-1.56
β-Oster						0.28
R^2	0.04	0.09	0.18	0.20	0.21	0.28

Notes: This table establishes the statistically and economically significant association between the historical agricultural determinants of plow usage and the existence of sex-based grammatical gender in the language spoken in this region, accounting for regional fixed-effects and other geographical characteristics. All columns have the same specification and controls as Table A2. Geographical controls include absolute latitude, mean elevation, terrain ruggedness, and coast length, as well as other agriculture-related controls as precipitation and temperature means and standard deviations. All independent variables have been normalized by subtracting their mean and dividing by their standard deviation. Thus, all coefficients can be compared and show the impact of a one standard deviation in the independent variable on the probability of having a future tense in the language. Heteroskedasticity robust standard error estimates are reported in parentheses, clustered at the language genus in parenthesis and squared brackets, spatial auto-correlation corrected standard errors (Conley, 1999) in squared brackets and Cliff-Ord ML in curly brackets; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

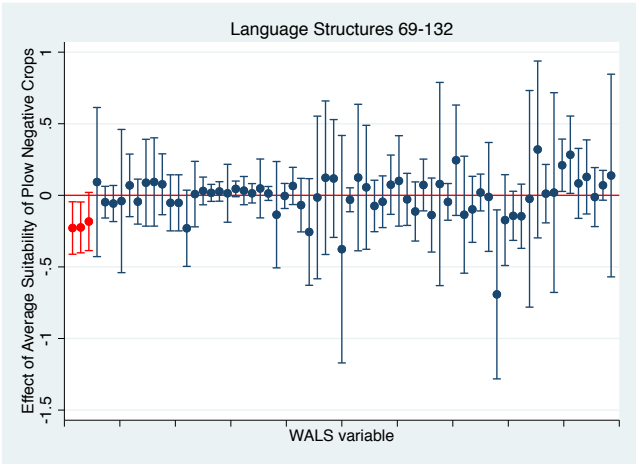
Table B3: Geographical Origins of Gender and Language Structures

	Language Structure						
		Temporal Structures		Non-Temporal Structures			
	Gender	Past	Perfect	Possessive	Evidentiality	Consonants	Colors
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Plow Negative Crops	-0.23***	-0.33***	-0.09	0.02	-0.05	-0.11	-0.76
(Average Caloric Yield, pre-1500)	(0.08)	(0.11)	(0.09)	(0.10)	(0.07)	(0.14)	(0.83)
All Crops	0.23***	0.24***	0.07	-0.06	0.02	0.08	0.73
(Average Caloric Yield, pre-1500)	(0.07)	(0.08)	(0.08)	(0.08)	(0.06)	(0.12)	(0.82)
All Geographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.21	0.12	0.14	0.14	0.20	0.30	-0.03
Observations	217	218	218	224	387	542	117

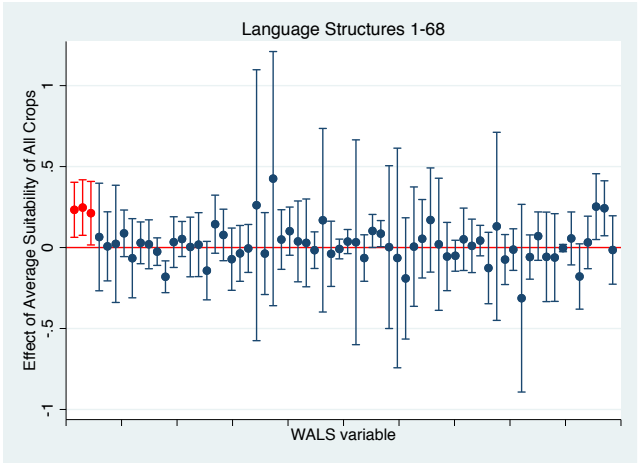
Notes: This table establishes the statistically and economically significant association between the historical agricultural determinants of plow usage and the existence of sex-based grammatical gender in a language, and compares their impact on other language structures. The analysis accounts for regional fixed-effects and other geographical characteristics as in previous tables. Other language structures include the existence a past tense, a perfect tense, the existence of obligatory possessive inflections, semantic distinctions of evidentiality, the number of consonants, the ratio of consonants to vowels and the number of colors. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.



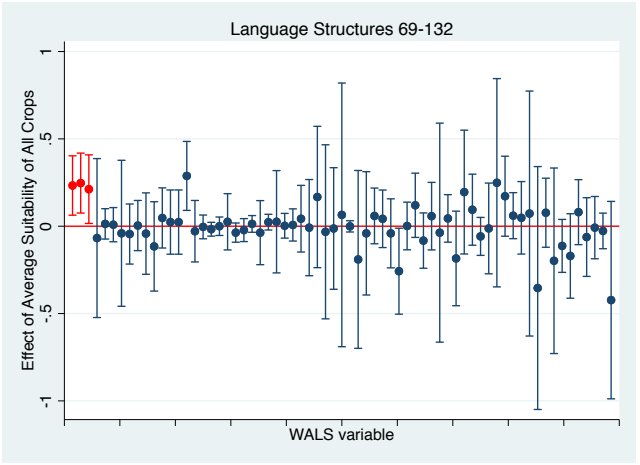
(a) Caloric Suitability of Plow Negative Crops and WALS Structures 1-68



(b) Caloric Suitability of Plow Negative Crops and WALS Structures 69-132



(c) Average Caloric Suitability and WALS Structures 1-68



(d) Average Caloric Suitability and WALS Structures 69-132

Figure B1: Orthogonality of Suitability for the Usage of the Plow and Other Language Structures in WALS
(Impact of Determinants of Plow Usage on Sex-Based Grammatical Gender in Red)

B.1 Alternative Measure for Sex-Based Grammatical Gender (Extensive Measure)

Table B4: Geographic Origins of Sex-Based Grammatical Gender (OLS)

	Existence of Sex-Based Grammatical Gender					
	(1)	(2)	(3)	(4)	(5)	(6)
Plow Negative Crops (Average Caloric Yield, pre-1500)	-0.03	-0.10	-0.16**	-0.17***	-0.23***	-0.22***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.08)
All Crops (Average Caloric Yield, pre-1500)	0.09	0.14**	0.21***	0.22***	0.27***	0.26***
	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)
Absolute Latitude		-0.10***	-0.20***	-0.09	-0.10	-0.15
		(0.04)	(0.04)	(0.10)	(0.10)	(0.09)
Elevation		0.02	-0.11**	-0.09*	-0.11**	-0.13***
		(0.05)	(0.04)	(0.05)	(0.05)	(0.04)
Ruggedness		-0.05	0.04	0.04	0.05	0.05
		(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Coast Length		0.07***	0.08***	0.08***	0.07***	0.07***
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Precipitation (mm/month)			-0.09	-0.10	-0.11	-0.06
			(0.06)	(0.07)	(0.07)	(0.07)
Precipitation (mm/month) (std)			-0.03	0.01	0.01	0.00
			(0.03)	(0.05)	(0.05)	(0.05)
Precipitation Volatility			-0.00	-0.02	-0.00	0.02
			(0.07)	(0.08)	(0.08)	(0.07)
Precipitation Spatial Correlation			0.13***	0.20	0.22	0.18
			(0.03)	(0.28)	(0.28)	(0.26)
Temperature (Daily Mean)				0.03	0.02	-0.02
				(0.07)	(0.07)	(0.07)
Temperature (Daily Mean) (std)				-0.06	-0.07	-0.06
				(0.05)	(0.04)	(0.04)
Temperature Volatility				-0.08	-0.06	-0.08
				(0.09)	(0.09)	(0.09)
Temperature Spatial Correlation				-0.07	-0.06	-0.06
				(0.29)	(0.29)	(0.26)
Unproductive Period (pre-1500CE)					-0.08***	-0.03
					(0.03)	(0.03)
Regional FE	No	No	No	No	No	Yes
Adjusted-R^2	0.01	0.04	0.13	0.13	0.15	0.24
Observations	245	245	245	245	245	245

Notes: This table establishes the statistically and economically significant association between the historical agricultural determinants of plow usage and the existence of sex-based grammatical gender in the language spoken in this region, accounting for regional fixed-effects and other geographical characteristics. All columns have the same specification and controls as Table 5. Geographical controls include absolute latitude, mean elevation, terrain ruggedness, and coast length, as well as other agriculture-related controls as precipitation and temperature means and standard deviations. All independent variables have been normalized by subtracting their mean and dividing by their standard deviation. Thus, all coefficients can be compared and show the impact of a one standard deviation in the independent variable on the probability of having a future tense in the language. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table B5: Geographic Origins of Sex-Based Grammatical Gender

	Existence of Sex-Based Grammatical Gender					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Probit						
Plow Negative Crops (Average Caloric Yield, pre-1500)	-0.03	-0.10*	-0.16**	-0.17***	-0.22***	-0.21***
	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)
All Crops (Average Caloric Yield, pre-1500)	0.09	0.14**	0.21***	0.22***	0.26***	0.24***
	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)
Geographical Controls	No	Yes	Yes	Yes	Yes	Yes
Regional FE	No	No	No	No	No	Yes
Pseudo-R^2	0.01	0.05	0.14	0.15	0.17	0.26
Observations	245	245	245	245	245	245
Panel B: OLS - Spatial-Autocorrelation, Clustering and Selection On Unobservables						
Plow Negative Crops (Average Caloric Yield, pre-1500)	-0.03	-0.10	-0.16**	-0.17***	-0.23***	-0.22***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.08)
	([0.07])	([0.08])	([0.08])	([0.07])	([0.07])	([0.07])
	[0.06]	[0.07]	[0.07]	[0.07]	[0.07]	[0.07]
	{0.05}	{0.06}	{0.06}	{0.06}	{0.06}	{0.08}
Altonji et al						-1.14
δ						-0.89
β-Oster						-0.29
All Crops (Average Caloric Yield, pre-1500)	0.09	0.14**	0.21***	0.22***	0.27***	0.26***
	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)
	([0.07])	([0.07])	([0.07])	([0.07])	([0.07])	([0.07])
	[0.06]	[0.06]	[0.06]	[0.06]	[0.06]	[0.07]
	{0.05}	{0.06}	{0.06}	{0.06}	{0.06}	{0.07}
Altonji et al						-1.51
δ						-0.89
β-Oster						0.31
R^2	0.02	0.07	0.17	0.18	0.20	0.31

Notes: This table establishes the statistically and economically significant association between the historical agricultural determinants of plow usage and the existence of sex-based grammatical genderin the language spoken in this region, accounting for regional fixed-effects and other geographical characteristics. All columns have the same specification and controls as Table A2. Geographical controls include absolute latitude, mean elevation, terrain ruggedness, and coast length, as well as other agriculture-related controls as precipitation and temperature means and standard deviations. All independent variables have been normalized by subtracting their mean and dividing by their standard deviation. Thus, all coefficients can be compared and show the impact of a one standard deviation in the independent variable on the probability of having a future tense in the language. Heteroskedasticity robust standard error estimates are reported in parentheses, clustered at the language genus in parenthesis and squared brackets, spatial auto-correlation corrected standard errors (Conley, 1999) in squared brackets and Cliff-Ord ML in curly brackets; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table B6: Geographical Origins of Gender and Language Structures

	Language Structure						
		Temporal Structures		Non-Temporal Structures			
	Gender	Past	Perfect	Possessive	Evidentiality	Consonants	Colors
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Plow Negative Crops	-0.22***	-0.33***	-0.09	0.01	-0.05	-0.09	-0.83
(Average Caloric Yield, pre-1500)	(0.08)	(0.11)	(0.09)	(0.10)	(0.07)	(0.14)	(0.83)
Average Caloric Yield	0.25***	0.24***	0.07	-0.06	0.02	0.07	0.76
(All Crops, pre-1500)	(0.07)	(0.08)	(0.08)	(0.08)	(0.06)	(0.12)	(0.82)
All Geographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.24	0.12	0.14	0.14	0.20	0.31	-0.03
Observations	245	218	218	223	386	538	116

Notes: This table establishes the statistically and economically significant association between the historical agricultural determinants of plow usage and the existence of sex-based grammatical gender in a language, and compares their impact on other language structures. The analysis accounts for regional fixed-effects and other geographical characteristics as in previous tables. Other language structures include the existence a past tense, a perfect tense, the existence of obligatory possessive inflections, semantic distinctions of evidentiality, the number of consonants, the ratio of consonants to vowels and the number of colors. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table B7: Persistent Impact of Geographical Characteristics on Gender: Linguistic Homeland vs. Urheimat
(Languages Outside Urheimat)

	Existence of Sex-Based Gender System					
	Migratory Distance to Urheimat					
	Any Distance			At Least 1 Week		
	Homeland	Urheimat	Both	Homeland	Urheimat	Both
	(1)	(2)	(3)	(4)	(5)	(6)
Homeland Plow Negative Crops (Average Caloric Yield, pre-1500)	-0.17**		-0.11	-0.11		-0.10*
	(0.07)		(0.07)	(0.08)		(0.06)
Homeland All Crops (Average Caloric Yield, pre-1500)	0.16**		0.13**	0.13		0.11*
	(0.07)		(0.06)	(0.09)		(0.06)
Urheimat Plow Negative Crops (Average Caloric Yield, pre-1500)		-0.54***	-0.50***		-0.52***	-0.54**
		(0.13)	(0.13)		(0.17)	(0.20)
Urheimat All Crops (Average Caloric Yield, pre-1500)		0.67***	0.63***		0.71***	0.71***
		(0.14)	(0.13)		(0.21)	(0.24)
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes
Homeland Geographical Characteristics	Yes	No	Yes	Yes	No	Yes
Urheimat Geographical Characteristics	No	Yes	Yes	No	Yes	Yes
Adjusted-R^2	0.31	0.51	0.50	0.34	0.57	0.54
Observations	183	183	183	155	155	155
Language Families	38	38	38	30	30	30

Notes: This table explores the relative contributions of agricultural productivity in the contemporary homeland vs. the Urheimat to the presence of sex-based grammatical gender in a daughter language. Heteroskedasticity robust standard error estimates clustered at the language family level are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

B.2 Alternative Measure for Sex-Based Grammatical Gender (Intensive Measure)

Table B8: Geographic Origins of Intensity of Sex-Based Grammatical Gender System (OLS)

	Intensity of Sex-Based Grammatical Gender System					
	(1)	(2)	(3)	(4)	(5)	(6)
Plow Negative Crops (Average Caloric Yield, pre-1500)	-0.25***	-0.32***	-0.40***	-0.39***	-0.47***	-0.36**
	(0.09)	(0.10)	(0.11)	(0.11)	(0.12)	(0.14)
All Crops (Average Caloric Yield, pre-1500)	0.26**	0.32***	0.41***	0.39***	0.47***	0.40***
	(0.10)	(0.11)	(0.12)	(0.12)	(0.14)	(0.14)
Absolute Latitude		-0.04	-0.21***	0.15	0.14	0.04
		(0.04)	(0.07)	(0.16)	(0.16)	(0.15)
Elevation		0.06	-0.10	-0.02	-0.03	-0.07
		(0.08)	(0.09)	(0.08)	(0.08)	(0.08)
Ruggedness		-0.14*	-0.01	-0.04	-0.03	-0.00
		(0.07)	(0.07)	(0.09)	(0.08)	(0.08)
Coast Length		0.03	0.03	0.03	0.02	0.04
		(0.05)	(0.05)	(0.06)	(0.06)	(0.06)
Precipitation (mm/month)			-0.29**	-0.25**	-0.28**	-0.20*
			(0.11)	(0.11)	(0.11)	(0.10)
Precipitation (mm/month) (std)			0.03	-0.01	-0.00	0.01
			(0.04)	(0.08)	(0.08)	(0.08)
Precipitation Volatility			0.07	0.03	0.06	0.08
			(0.11)	(0.13)	(0.12)	(0.12)
Precipitation Spatial Correlation			0.09**	-0.09	-0.02	0.08
			(0.05)	(0.43)	(0.45)	(0.47)
Temperature (Daily Mean)				0.22*	0.20	0.08
				(0.12)	(0.14)	(0.13)
Temperature (Daily Mean) (std)				0.03	0.02	-0.02
				(0.08)	(0.08)	(0.08)
Temperature Volatility				-0.20	-0.17	-0.13
				(0.14)	(0.14)	(0.14)
Temperature Spatial Correlation				0.17	0.13	0.00
				(0.43)	(0.44)	(0.46)
Unproductive Period (pre-1500CE)					-0.14**	-0.07
					(0.07)	(0.06)
Regional FE	No	No	No	No	No	Yes
Adjusted-R^2	0.03	0.03	0.06	0.07	0.09	0.17
Observations	181	181	181	181	181	181

Notes: This table establishes the statistically and economically significant association between the historical agricultural determinants of plow usage and the existence of sex-based grammatical gender in the language spoken in this region, accounting for regional fixed-effects and other geographical characteristics. All columns have the same specification and controls as Table 5. Geographical controls include absolute latitude, mean elevation, terrain ruggedness, and coast length, as well as other agriculture-related controls as precipitation and temperature means and standard deviations. All independent variables have been normalized by subtracting their mean and dividing by their standard deviation. Thus, all coefficients can be compared and show the impact of a one standard deviation in the independent variable on the probability of having a future tense in the language. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table B9: Geographic Origins of Intensity of Sex-Based Grammatical Gender System

	Intensity of Sex-Based Grammatical Gender System					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: OLS - Spatial-Autocorrelation, Clustering and Selection On Unobservables						
Plow Negative Crops (Average Caloric Yield, pre-1500)	-0.25***	-0.32***	-0.40***	-0.39***	-0.47***	-0.36**
	(0.09)	(0.10)	(0.11)	(0.11)	(0.12)	(0.14)
	([0.11])	([0.12])	([0.14])	([0.13])	([0.15])	([0.14])
	[0.10]	[0.10]	[0.11]	[0.11]	[0.12]	[0.12]
	{0.10}	{0.11}	{0.12}	{0.12}	{0.12}	{0.15}
Altonji et al						-3.22
δ						-1.95
β-Oster						-0.40
All Crops (Average Caloric Yield, pre-1500)	0.26**	0.32***	0.41***	0.39***	0.47***	0.40***
	(0.10)	(0.11)	(0.12)	(0.12)	(0.14)	(0.14)
	([0.12])	([0.13])	([0.14])	([0.14])	([0.15])	([0.14])
	[0.11]	[0.11]	[0.12]	[0.11]	[0.12]	[0.12]
	{0.10}	{0.11}	{0.11}	{0.11}	{0.12}	{0.14}
Altonji et al						-2.85
δ						-1.36
β-Oster						0.45
R^2	0.04	0.06	0.12	0.14	0.16	0.26
Observations	181	181	181	181	181	181

Notes: This table establishes the statistically and economically significant association between the historical agricultural determinants of plow usage and the existence of sex-based grammatical gender in the language spoken in this region, accounting for regional fixed-effects and other geographical characteristics. All columns have the same specification and controls as Table A2. Geographical controls include absolute latitude, mean elevation, terrain ruggedness, and coast length, as well as other agriculture-related controls as precipitation and temperature means and standard deviations. All independent variables have been normalized by subtracting their mean and dividing by their standard deviation. Thus, all coefficients can be compared and show the impact of a one standard deviation in the independent variable on the probability of having a future tense in the language. Heteroskedasticity robust standard error estimates are reported in parentheses, clustered at the language genus in parenthesis and squared brackets, spatial auto-correlation corrected standard errors (Conley, 1999) in squared brackets and Cliff-Ord ML in curly brackets; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table B10: Geographical Origins of Gender and Language Structures

	Language Structure						
		Temporal Structures		Non-Temporal Structures			
	Gender	Past	Perfect	Possessive	Evidentiality	Consonants	Colors
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Plow Negative Crops	-0.36**	-0.33***	-0.09	0.01	-0.05	-0.09	-0.83
(Average Caloric Yield, pre-1500)	(0.14)	(0.11)	(0.09)	(0.10)	(0.07)	(0.14)	(0.83)
Average Caloric Yield	0.39***	0.24***	0.07	-0.06	0.02	0.07	0.76
(All Crops, pre-1500)	(0.14)	(0.08)	(0.08)	(0.08)	(0.06)	(0.12)	(0.82)
All Geographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.17	0.12	0.14	0.14	0.20	0.31	-0.03
Observations	181	218	218	223	386	538	116

Notes: This table establishes the statistically and economically significant association between the historical agricultural determinants of plow usage and the intensity of grammatical gender in a language, and compares their impact on other language structures. The analysis accounts for regional fixed-effects and other geographical characteristics as in previous tables. Other language structures include the existence a past tense, a perfect tense, the existence of obligatory possessive inflections, semantic distinctions of evidentiality, the number of consonants, the ratio of consonants to vowels and the number of colors. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table B11: Persistent Impact of Geographical Characteristics on Intensity of Sex-Based Grammatical Gender System
Linguistic Homeland vs. Urheimat
(Languages Outside Urheimat)

	Existence of Sex-Based Gender System					
	Migratory Distance to Urheimat					
	Any Distance			At Least 1 Week		
	Homeland	Urheimat	Both	Homeland	Urheimat	Both
	(1)	(2)	(3)	(4)	(5)	(6)
Homeland Plow Negative Crops (Avg. Caloric Yield, pre-1500)	-0.32*		-0.11	-0.31		-0.39
	(0.18)		(0.16)	(0.23)		(0.26)
Homeland All Crops (Average Caloric Yield, pre-1500)	0.34*		0.20	0.30		0.22
	(0.20)		(0.16)	(0.23)		(0.19)
Urheimat Plow Negative Crops (Avg. Caloric Yield, pre-1500)		-0.46*	-0.46**		-0.61	-0.42
		(0.25)	(0.22)		(0.41)	(0.39)
Urheimat All Crops (Average Caloric Yield, pre-1500)		0.53**	0.56***		0.73*	0.60
		(0.21)	(0.20)		(0.39)	(0.42)
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes
Homeland Geographical Characteristics	Yes	No	Yes	Yes	No	Yes
Urheimat Geographical Characteristics	No	Yes	Yes	No	Yes	Yes
Adjusted-R^2	0.15	0.35	0.32	0.11	0.42	0.42
Observations	129	129	129	109	109	109
Language Families	37	37	37	30	30	30

Notes: This table explores the relative contributions of agricultural productivity in the contemporary homeland vs. the Urheimat to the presence of sex-based grammatical gender in a daughter language. Heteroskedasticity robust standard error estimates clustered at the language family level are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table B12: Geographic Origins of Usage of the Plow and Intensive Sex-Based Grammatical Genders

	Mechanism			
	Plow		Grammatical Gender	
	(1)	(2)	(3)	(4)
Plow Negative Crops (Average Caloric Yield, pre-1500)	-0.26***	-0.06**		
	(0.02)	(0.02)		
All Crops (Average Caloric Yield, pre-1500)	0.26***	0.09***		
	(0.02)	(0.02)		
Aboriginal Plow			0.65***	0.56**
			(0.20)	(0.27)
All Geographic Controls	No	Yes	No	Yes
Regional FE	No	Yes	No	Yes
Adjusted-R^2	0.19	0.47	0.09	0.21
Observations	1175	1175	133	133

Notes: This table establishes the statistically and economically significant association between the historical agricultural determinants of plow usage and the existence of sex-based grammatical gender in a language. The first two columns provide the association between the geographical determinants of plow suitability and actual usage of the plow, and columns (3) and (4) provide evidence on the association between actual usage of the plow and the emergence of sex-based grammatical gender. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

B.3 Origins of Gender Distinctions in Pronouns

Table B13: Geographic Origins of Gender Distinctions in Pronouns (OLS)

	Existence of Gender Distinctions in Pronouns					
	(1)	(2)	(3)	(4)	(5)	(6)
Plow Negative Crops (Average Caloric Yield, pre-1500)	-0.08*	-0.12**	-0.13***	-0.15***	-0.17***	-0.18***
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)
All Crops (Average Caloric Yield, pre-1500)	0.10**	0.14***	0.15***	0.18***	0.19***	0.21***
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)
Absolute Latitude		-0.04	-0.08**	0.06	0.05	0.01
		(0.03)	(0.03)	(0.07)	(0.07)	(0.08)
Elevation		0.03	-0.03	-0.01	-0.02	-0.03
		(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
Ruggedness		-0.11***	-0.05	-0.07*	-0.07*	-0.06*
		(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
Coast Length		0.07***	0.07***	0.07***	0.07***	0.08***
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Precipitation (mm/month)			0.02	-0.01	-0.02	0.01
			(0.05)	(0.06)	(0.06)	(0.06)
Precipitation (mm/month) (std)			-0.00	0.02	0.02	0.01
			(0.03)	(0.04)	(0.04)	(0.04)
Precipitation Volatility			-0.09*	-0.09*	-0.09	-0.08
			(0.05)	(0.05)	(0.05)	(0.05)
Precipitation Spatial Correlation			0.04	0.24	0.26	0.34
			(0.03)	(0.23)	(0.23)	(0.25)
Temperature (Daily Mean)				0.01	0.01	-0.03
				(0.06)	(0.06)	(0.06)
Temperature (Daily Mean) (std)				-0.04	-0.04	-0.04
				(0.04)	(0.04)	(0.04)
Temperature Volatility				-0.16***	-0.16**	-0.19***
				(0.06)	(0.06)	(0.06)
Temperature Spatial Correlation				-0.19	-0.20	-0.29
				(0.23)	(0.23)	(0.25)
Unproductive Period (pre-1500CE)					-0.03	-0.02
					(0.03)	(0.03)
Regional FE	No	No	No	No	No	Yes
Adjusted-R^2	0.01	0.06	0.08	0.10	0.10	0.13
Observations	354	354	354	354	354	354

Notes: This table establishes the statistically and economically significant association between the historical agricultural determinants of plow usage and the existence of gender distinctions in independent personal pronouns in the language spoken in this region, accounting for regional fixed-effects and other geographical characteristics. All columns have the same specification and controls as Table 5. Geographical controls include absolute latitude, mean elevation, terrain ruggedness, and coast length, as well as other agriculture-related controls as precipitation and temperature means and standard deviations. All independent variables have been normalized by subtracting their mean and dividing by their standard deviation. Thus, all coefficients can be compared and show the impact of a one standard deviation in the independent variable on the probability of having a future tense in the language. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table B14: Geographic Origins of Gender Distinctions in Pronouns (Probit)

	Existence of Gender Distinctions in Pronouns					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Probit						
Plow Negative Crops (Average Caloric Yield, pre-1500)	-0.06	-0.10**	-0.11**	-0.13***	-0.14***	-0.16**
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)
All Crops (Average Caloric Yield, pre-1500)	0.09**	0.12***	0.14***	0.16***	0.17***	0.19***
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)
Geographical Controls	No	Yes	Yes	Yes	Yes	Yes
Regional FE	No	No	No	No	No	Yes
Pseudo-R^2	0.01	0.07	0.09	0.11	0.11	0.15
Observations	350	350	350	350	350	350
Panel B: OLS - Spatial-Autocorrelation, Clustering and Selection On Unobservables						
Plow Negative Crops (Average Caloric Yield, pre-1500)	-0.08*	-0.12**	-0.13***	-0.15***	-0.17***	-0.18***
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)
	([0.05])	([0.06])	([0.05])	([0.05])	([0.05])	([0.07])
	[0.05]	[0.05]	[0.05]	[0.05]	[0.05]	[0.06]
	{0.04}	{0.05}	{0.05}	{0.05}	{0.05}	{0.07}
Altonji et al						-1.77
δ						-1.12
β-Oster						-0.21
All Crops (Average Caloric Yield, pre-1500)	0.10**	0.14***	0.15***	0.18***	0.19***	0.21***
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)
	([0.05])	([0.06])	([0.05])	([0.05])	([0.05])	([0.06])
	[0.05]	[0.05]	[0.05]	[0.04]	[0.05]	[0.06]
	{0.04}	{0.05}	{0.05}	{0.05}	{0.05}	{0.06}
Altonji et al						-1.89
δ						-0.90
β-Oster						0.24
R^2	0.01	0.08	0.11	0.13	0.14	0.18

Notes: This table establishes the statistically and economically significant association between the historical agricultural determinants of plow usage and the existence of gender distinctions in pronouns in the language spoken in this region, accounting for regional fixed-effects and other geographical characteristics. All columns have the same specification and controls as Table A2. Geographical controls include absolute latitude, mean elevation, terrain ruggedness, and coast length, as well as other agriculture-related controls as precipitation and temperature means and standard deviations. All independent variables have been normalized by subtracting their mean and dividing by their standard deviation. Thus, all coefficients can be compared and show the impact of a one standard deviation in the independent variable on the probability of having a future tense in the language. Heteroskedasticity robust standard error estimates are reported in parentheses, clustered at the language genus in parenthesis and squared brackets, spatial auto-correlation corrected standard errors (Conley, 1999) in squared brackets and Cliff-Ord ML in curly brackets; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table B15: Geographical Origins of Gender and Language Structures

	Language Structure						
		Temporal Structures		Non-Temporal Structures			
	Gender	Past	Perfect	Possessive	Evidentiality	Consonants	Colors
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Plow Negative Crops	-0.18***	-0.33***	-0.09	0.01	-0.05	-0.09	-0.83
(Average Caloric Yield, pre-1500)	(0.07)	(0.11)	(0.09)	(0.10)	(0.07)	(0.14)	(0.83)
Average Caloric Yield	0.21***	0.24***	0.07	-0.06	0.02	0.07	0.76
(All Crops, pre-1500)	(0.06)	(0.08)	(0.08)	(0.08)	(0.06)	(0.12)	(0.82)
All Geographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.13	0.12	0.14	0.14	0.20	0.31	-0.03
Observations	355	218	218	223	386	538	116

Notes: This table establishes the statistically and economically significant association between the historical agricultural determinants of plow usage and the existence of gender distinctions in pronouns in a language, and compares their impact on other language structures. The analysis accounts for regional fixed-effects and other geographical characteristics as in previous tables. Other language structures include the existence a past tense, a perfect tense, the existence of obligatory possessive inflections, semantic distinctions of evidentiality, the number of consonants, the ratio of consonants to vowels and the number of colors. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table B16: Persistent Impact of Geographical Characteristics on Gender Distinctions in Pronouns
Linguistic Homeland vs. Urheimat
(Languages Outside Urheimat)

	Existence of Gender Distinctions in Independent Personal Pronouns					
	Migratory Distance to Urheimat					
	Any Distance			At Least 1 Week		
	Homeland	Urheimat	Both	Homeland	Urheimat	Both
	(1)	(2)	(3)	(4)	(5)	(6)
Homeland Plow Negative Crops (Avg. Caloric Yield, pre-1500)	-0.14*		-0.06	-0.11		-0.07
	(0.08)		(0.07)	(0.10)		(0.08)
Homeland All Crops (Average Caloric Yield, pre-1500)	0.16**		0.11**	0.12		0.08
	(0.07)		(0.05)	(0.09)		(0.07)
Urheimat Plow Negative Crops (Avg. Caloric Yield, pre-1500)		-0.31***	-0.32***		-0.21**	-0.20**
		(0.08)	(0.09)		(0.08)	(0.08)
Urheimat All Crops (Average Caloric Yield, pre-1500)		0.33***	0.33***		0.23***	0.22***
		(0.07)	(0.07)		(0.08)	(0.07)
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes
Homeland Geographical Characteristics	Yes	No	Yes	Yes	No	Yes
Urheimat Geographical Characteristics	No	Yes	Yes	No	Yes	Yes
Adjusted-R^2	0.17	0.30	0.34	0.17	0.34	0.37
Observations	245	245	245	207	207	207
Language Families	51	51	51	36	36	36

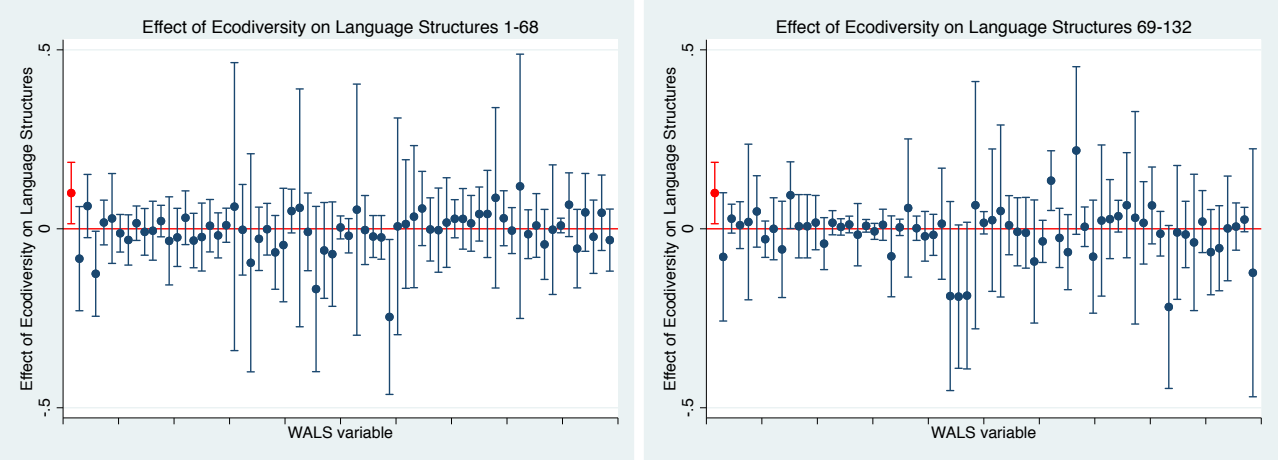
Notes: This table explores the relative contributions of agricultural productivity in the contemporary homeland vs. the Urheimat to the presence of sex-based grammatical gender in a daughter language. Heteroskedasticity robust standard error estimates clustered at the language family level are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table B17: Geographic Origins of Usage of the Plow and Gender Distinctions in Pronouns

	Mechanism			
	Plow		Grammatical Gender	
	(1)	(2)	(3)	(4)
Plow Negative Crops (Average Caloric Yield, pre-1500)	-0.26***	-0.06**		
	(0.02)	(0.02)		
All Crops (Average Caloric Yield, pre-1500)	0.26***	0.09***		
	(0.02)	(0.02)		
Aboriginal Plow			0.24***	0.22**
			(0.08)	(0.11)
All Geographic Controls	No	Yes	No	Yes
Regional FE	No	Yes	No	Yes
Adjusted-R^2	0.19	0.47	0.04	0.10
Observations	1175	1175	220	220

Notes: This table establishes the statistically and economically significant association between the historical agricultural determinants of plow usage and the existence of gender distinctions in independent personal pronouns in a language. The first two columns provide the association between the geographical determinants of plow suitability and actual usage of the plow, and columns (3) and (4) provide evidence on the association between actual usage of the plow and the emergence of sex-based grammatical gender. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

C Origins of Politeness Distinctions



(a) WALS Structures 1-68

(b) WALS Structures 69-132

Figure C1: Orthogonality of Ecological Diversity and Other Language Structures (WALS) (Impact of Ecological Diversity on Politeness Distinctions in Red)

Table C1: Geographic Origins of Politeness Distinctions (OLS)

	Existence of Politeness Distinctions					
	(1)	(2)	(3)	(4)	(5)	(6)
Ecological Diversity	0.15***	0.10***	0.10***	0.10**	0.10**	0.10**
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Average Caloric Yield (All Crops, pre-1500)		0.11***	0.10***	0.11***	0.11***	0.12***
		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Absolute Latitude		0.07**	0.09**	0.23**	0.24**	0.07
		(0.03)	(0.04)	(0.11)	(0.11)	(0.11)
Elevation		0.01	0.05	0.09*	0.09*	0.10**
		(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Ruggedness		-0.04	-0.06*	-0.07*	-0.06*	-0.09**
		(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Coast Length		0.07*	0.06	0.06	0.06	0.04
		(0.04)	(0.04)	(0.05)	(0.05)	(0.05)
Precipitation (mm/month)			0.03	0.03	0.04	-0.02
			(0.06)	(0.06)	(0.06)	(0.06)
Precipitation (mm/month) (std)			0.04	0.05	0.06	0.07
			(0.04)	(0.07)	(0.07)	(0.07)
Precipitation Volatility			-0.06	-0.07	-0.08	-0.04
			(0.07)	(0.07)	(0.06)	(0.07)
Precipitation Spatial Correlation			-0.08**	0.31	0.31	0.35
			(0.04)	(0.28)	(0.29)	(0.27)
Temperature (Daily Mean)				0.07	0.07	0.04
				(0.08)	(0.08)	(0.07)
Temperature (Daily Mean) (std)				-0.01	-0.03	-0.07
				(0.07)	(0.07)	(0.06)
Temperature Volatility				-0.09	-0.07	0.03
				(0.09)	(0.08)	(0.08)
Temperature Spatial Correlation				-0.39	-0.39	-0.42
				(0.29)	(0.30)	(0.28)
Unproductive Period (pre-1500CE)					-0.06*	-0.07*
					(0.04)	(0.04)
Regional FE	No	No	No	No	No	Yes
Adjusted-R^2	0.09	0.18	0.20	0.20	0.21	0.32
Observations	198	198	198	198	198	198

Notes: This table establishes the positive, statistically, and economically significant association between ecological diversity and the existence of politeness distinctions in the language spoken in this region, accounting for regional fixed-effects and other geographical characteristics. All columns have the same specification and controls as Table A1. Geographical controls include absolute latitude, mean elevation, terrain ruggedness, and coast length, as well as other agriculture-related controls as precipitation and temperature means and standard deviations. All independent variables have been normalized by subtracting their mean and dividing by their standard deviation. Thus, all coefficients can be compared and show the impact of a one standard deviation in the independent variable on the probability of having a future tense in the language. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table C2: Geographic Origins of Politeness Distinctions

	Existence of Politeness Distinctions					
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Probit					
Ecological Diversity	0.14***	0.09***	0.09***	0.10***	0.10***	0.10***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Geographical Controls	No	Yes	Yes	Yes	Yes	Yes
Regional FE	No	No	No	No	No	Yes
Pseudo-R^2	0.07	0.21	0.23	0.24	0.25	0.36
Observations	180	180	180	180	180	180
Panel B: OLS - Spatial-Autocorrelation, Clustering and Selection On Unobservables						
Ecological Diversity	0.15***	0.10***	0.10***	0.10**	0.10**	0.10**
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
	([0.04])	([0.04])	([0.04])	([0.04])	([0.04])	([0.03])
	[0.03]	[0.04]	[0.04]	[0.04]	[0.04]	[0.04]
	{0.03}	{0.03}	{0.03}	{0.03}	{0.03}	{0.03}
Altonji et al						3.10
δ						5.02
β-Oster						0.08
R^2	0.10	0.21	0.24	0.26	0.27	0.39

Notes: This table establishes the positive, statistically, and economically significant association between ecological diversity and the existence of politeness distinctions in the language spoken in this region, accounting for regional fixed-effects and other geographical characteristics. All columns have the same specification and controls as Table 1. Geographical controls include absolute latitude, mean elevation, terrain ruggedness, and coast length, as well as other agriculture-related controls as precipitation and temperature means and standard deviations. All independent variables have been normalized by subtracting their mean and dividing by their standard deviation. Thus, all coefficients can be compared and show the impact of a one standard deviation in the independent variable on the probability of having a future tense in the language. Heteroskedasticity robust standard error estimates are reported in parentheses, clustered at the language genus in parenthesis and squared brackets, spatial auto-correlation corrected standard errors (Conley, 1999) in squared brackets and Cliff-Ord ML in curly brackets; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table C3: Geographical Origins of Politeness and Language Structures

	Language Structure						
		Temporal Structures		Non-Temporal Structures			
	Politeness	Past	Perfect	Possessive	Evidentiality	Consonants	Colors
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ecological Diversity	0.10**	0.04	0.05	-0.03	0.01	-0.10*	-0.49
	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.05)	(0.37)
All Geographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.32	0.07	0.14	0.14	0.20	0.31	-0.02
Observations	198	218	218	224	387	542	117

Notes: This table establishes the positive, statistically, and economically significant association between ecological diversity and the existence of politeness distinctions in a language, and compares their impact on other language structures. The analysis accounts for regional fixed-effects and other geographical characteristics as in previous tables. Other language structures include the existence a past tense, a perfect tense, the existence of obligatory possessive inflections, semantic distinctions of evidentiality, the number of consonants, the ratio of consonants to vowels and the number of colors. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

C.1 Alternative Measure of Politeness Distinctions (Extensive Margin)

Table C4: Geographic Origins of Politeness Distinctions (Alternative measure - OLS)

	Existence of politeness Distinctions (Alt. Measure)					
	(1)	(2)	(3)	(4)	(5)	(6)
Ecological Diversity	0.08***	0.09***	0.07**	0.09***	0.09***	0.08***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Average Caloric Yield (All Crops, pre-1500)		0.03	0.04	0.04	0.03	0.06**
		(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Absolute Latitude		-0.03	-0.04	-0.05	-0.05	-0.10
		(0.02)	(0.03)	(0.06)	(0.06)	(0.07)
Elevation		-0.06**	-0.08***	-0.07*	-0.07*	-0.06*
		(0.02)	(0.03)	(0.04)	(0.04)	(0.03)
Ruggedness		0.01	0.02	0.03	0.04	-0.00
		(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Coast Length		0.01	0.01	0.01	0.02	0.02
		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Precipitation (mm/month)			-0.04	-0.05	-0.05	-0.07
			(0.03)	(0.05)	(0.04)	(0.04)
Precipitation (mm/month) (std)			0.06*	0.12**	0.12**	0.12**
			(0.03)	(0.05)	(0.06)	(0.05)
Precipitation Volatility			0.01	-0.00	-0.01	-0.01
			(0.04)	(0.04)	(0.04)	(0.04)
Precipitation Spatial Correlation			0.02	0.05	0.05	0.00
			(0.02)	(0.19)	(0.19)	(0.19)
Temperature (Daily Mean)				-0.03	-0.03	-0.06
				(0.07)	(0.06)	(0.06)
Temperature (Daily Mean) (std)				-0.08	-0.09*	-0.11**
				(0.05)	(0.05)	(0.05)
Temperature Volatility				-0.01	0.00	0.06
				(0.05)	(0.05)	(0.05)
Temperature Spatial Correlation				-0.03	-0.04	0.01
				(0.19)	(0.19)	(0.19)
Unproductive Period (pre-1500CE)					-0.03*	-0.04*
					(0.02)	(0.02)
Regional FE	No	No	No	No	No	Yes
Adjusted-R^2	0.06	0.09	0.09	0.09	0.10	0.25
Observations	198	198	198	198	198	198

Notes: This table establishes the positive, statistically, and economically significant association between ecological diversity and the existence of politeness distinctions in the language spoken in this region, accounting for regional fixed-effects and other geographical characteristics. All columns have the same specification and controls as Table A1. Geographical controls include absolute latitude, mean elevation, terrain ruggedness, and coast length, as well as other agriculture-related controls as precipitation and temperature means and standard deviations. All independent variables have been normalized by subtracting their mean and dividing by their standard deviation. Thus, all coefficients can be compared and show the impact of a one standard deviation in the independent variable on the probability of having a future tense in the language. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table C5: Geographic Origins of Politeness Distinctions (Alternative measure - Probit)

	Existence of Politeness Distinctions (Alternative Measure)					
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Probit			
Ecological Diversity	0.09***	0.11***	0.08***	0.08***	0.08***	0.07***
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Geographical Controls	No	Yes	Yes	Yes	Yes	Yes
Regional FE	No	No	No	No	No	Yes
Pseudo-R^2	0.09	0.21	0.26	0.27	0.29	0.51
Observations	140	140	140	140	140	140
Panel B: OLS - Spatial-Autocorrelation, Clustering and Selection On Unobservables						
Ecological Diversity	0.08***	0.09***	0.07**	0.09***	0.09***	0.08***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
	([0.04])	([0.05])	([0.04])	([0.05])	([0.05])	([0.03])
	[0.03]	[0.04]	[0.04]	[0.04]	[0.04]	[0.03]
	{0.02}	{0.02}	{0.02}	{0.02}	{0.02}	{0.02}
Altonji et al						-7.30
δ						-12.31
β-Oster						0.09
R^2	0.07	0.11	0.14	0.16	0.17	0.33
Observations	198	198	198	198	198	198

Notes: This table establishes the positive, statistically, and economically significant association between ecological diversity and the existence of politeness distinctions in the language spoken in this region, accounting for regional fixed-effects and other geographical characteristics. All columns have the same specification and controls as Table 1. Geographical controls include absolute latitude, mean elevation, terrain ruggedness, and coast length, as well as other agriculture-related controls as precipitation and temperature means and standard deviations. All independent variables have been normalized by subtracting their mean and dividing by their standard deviation. Thus, all coefficients can be compared and show the impact of a one standard deviation in the independent variable on the probability of having a future tense in the language. Heteroskedasticity robust standard error estimates are reported in parentheses, clustered at the language genus in parenthesis and squared brackets, spatial auto-correlation corrected standard errors (Conley, 1999) in squared brackets and Cliff-Ord ML in curly brackets; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table C6: Geographical Origins of Politeness and Language Structures

	Language Structure						
		Temporal Structures		Non-Temporal Structures			
	Politeness	Past	Perfect	Possessive	Evidentiality	Consonants	Colors
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ecological Diversity	0.09***	0.04	0.05	-0.04	0.01	-0.08	-0.49
	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)	(0.05)	(0.38)
All Geographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.25	0.07	0.14	0.15	0.20	0.31	-0.02
Observations	198	218	218	223	386	538	116

Notes: This table establishes the positive, statistically, and economically significant association between ecological diversity and the existence of strong politeness distinctions in a language, and compares their impact on other language structures. The analysis accounts for regional fixed-effects and other geographical characteristics as in previous tables. Other language structures include the existence a past tense, a perfect tense, the existence of obligatory possessive inflections, semantic distinctions of evidentiality, the number of consonants, the ratio of consonants to vowels and the number of colors. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table C7: Persistent Impact of Geographical Characteristics on Politeness (Alternative Measure)
Linguistic Homeland vs. Urheimat
(Languages Outside Urheimat)

	Existence of Politeness Distinctions					
	Migratory Distance to Urheimat					
	Any Distance			At Least 1 Week		
	Homeland	Urheimat	Both	Homeland	Urheimat	Both
	(1)	(2)	(3)	(4)	(5)	(6)
Homeland Ecological Diversity	0.11**		0.11***	0.10**		0.08**
	(0.04)		(0.04)	(0.04)		(0.03)
Urheimat Ecological Diversity		-0.01	-0.06*		0.05	-0.01
		(0.04)	(0.03)		(0.05)	(0.04)
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes
Homeland Geographical Characteristics	Yes	No	Yes	Yes	No	Yes
Urheimat Geographical Characteristics	No	Yes	Yes	No	Yes	Yes
Adjusted-R^2	0.29	0.13	0.33	0.34	0.23	0.41
Observations	146	146	146	126	126	126
Language Families	36	36	36	28	28	28

Notes: This table explores the relative contributions of ecological diversity in the contemporary homeland vs. the Urheimat to the presence of politeness distinctions in a daughter language. Heteroskedasticity robust standard error estimates clustered at the language family level are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table C8: Geographic Origins of Jurisdictional Hierarchy and Politeness Distinctions

	Mechanism			
	Jurisdictional Hierarchy		Politeness	
	(1)	(2)	(3)	(4)
Ecological Diversity	0.23***	0.12***		
	(0.04)	(0.03)		
Average Caloric Yield (All Crops, pre-1500)	0.17***	0.23***		
	(0.03)	(0.03)		
Jurisdictional Hierarchy Beyond Local Community			0.11***	0.10***
			(0.02)	(0.02)
All Geographic Controls	No	Yes	No	Yes
Regional FE	No	Yes	No	Yes
Adjusted-R^2	0.07	0.32	0.18	0.41
Observations	1154	1154	139	139

Notes: This table establishes the positive statistically and economically significant association between the geographical determinants of statehood, as measured by jurisdictional hierarchy beyond the local level, and politeness distinctions in a language. The first two columns provide the evidence on the association between ecological diversity and statehood, and columns (3) and (4) show the association between statehood and the emergence of politeness distinctions. The table shows the estimated coefficients in an OLS regression as the dependent variable in columns (1) and (2) is not binary. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table C9: Hierarchy and Politeness (Alternate Measure)

	Hierarchy Index			
	All		Old World	
	(1)	(2)	(3)	(4)
Politeness Distinctions	0.31***	0.37***	0.36***	0.43***
	(0.07)	(0.11)	(0.09)	(0.12)
Ecological Diversity		0.18*		0.08
		(0.10)		(0.08)
Regional FE	Yes	Yes	Yes	Yes
Main Geographical Controls	No	Yes	No	Yes
Adjusted-R^2	0.42	0.49	0.42	0.57
Observations	53	53	50	50

Notes: This table establishes the statistically and economically positive association between of politeness distinctions and preferences for hierarchy in society. The analysis accounts for the geographical origins of politeness distinctions and other geographical characteristics of the contemporary homeland of the language. Heteroskedasticity robust standard error estimates clustered at the language family level are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

C.2 Alternative Measure of Politeness Distinctions (Intensive Margin)

Table C10: Geographic Origins of Politeness Distinctions (Intensive measure - OLS)

	Intensity of Politeness Distinctions					
	(1)	(2)	(3)	(4)	(5)	(6)
Ecological Diversity	0.26***	0.21***	0.20***	0.21***	0.21***	0.22***
	(0.06)	(0.07)	(0.07)	(0.08)	(0.08)	(0.07)
Average Caloric Yield (All Crops, pre-1500)		0.14***	0.14***	0.14***	0.14***	0.19***
		(0.05)	(0.05)	(0.05)	(0.05)	(0.06)
Absolute Latitude		0.02	0.02	0.16	0.17	-0.01
		(0.05)	(0.06)	(0.16)	(0.15)	(0.15)
Elevation		-0.08	-0.06	-0.02	-0.01	0.01
		(0.07)	(0.07)	(0.09)	(0.08)	(0.08)
Ruggedness		-0.01	-0.02	-0.01	-0.00	-0.06
		(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
Coast Length		0.12	0.11	0.12	0.12	0.10
		(0.09)	(0.09)	(0.10)	(0.10)	(0.09)
Precipitation (mm/month)			-0.01	-0.02	-0.01	-0.10
			(0.08)	(0.10)	(0.09)	(0.10)
Precipitation (mm/month) (std)			0.12	0.22	0.23	0.26**
			(0.08)	(0.14)	(0.14)	(0.13)
Precipitation Volatility			-0.06	-0.10	-0.10	-0.06
			(0.09)	(0.10)	(0.09)	(0.10)
Precipitation Spatial Correlation			-0.03	0.45	0.46	0.41
			(0.06)	(0.44)	(0.45)	(0.45)
Temperature (Daily Mean)				0.05	0.04	-0.03
				(0.14)	(0.13)	(0.13)
Temperature (Daily Mean) (std)				-0.13	-0.15	-0.23**
				(0.12)	(0.12)	(0.11)
Temperature Volatility				-0.10	-0.06	0.11
				(0.12)	(0.12)	(0.12)
Temperature Spatial Correlation				-0.51	-0.52	-0.44
				(0.45)	(0.46)	(0.46)
Unproductive Period (pre-1500CE)					-0.10**	-0.12**
					(0.05)	(0.05)
Regional FE	No	No	No	No	No	Yes
Adjusted-R^2	0.11	0.17	0.17	0.17	0.18	0.32
Observations	198	198	198	198	198	198

Notes: This table establishes the positive, statistically, and economically significant association between ecological diversity and an intensive measure of politeness distinctions in the language spoken in this region, accounting for regional fixed-effects and other geographical characteristics. All columns have the same specification and controls as Table A1. Geographical controls include absolute latitude, mean elevation, terrain ruggedness, and coast length, as well as other agriculture-related controls as precipitation and temperature means and standard deviations. All independent variables have been normalized by subtracting their mean and dividing by their standard deviation. Thus, all coefficients can be compared and show the impact of a one standard deviation in the independent variable on the probability of having a future tense in the language. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table C11: Geographical Origins of and Politeness Distinctions (Intensive Measure - OLS)
Robustness to Spatial-Autocorrelation, Clustering and Selection

	Intensity of Politeness Distinctions					
	(1)	(2)	(3)	(4)	(5)	(6)
Ecological Diversity	0.26***	0.21***	0.20***	0.21***	0.21***	0.22***
	(0.06)	(0.07)	(0.07)	(0.08)	(0.08)	(0.07)
	([0.07])	([0.09])	([0.09])	([0.09])	([0.09])	([0.07])
	[0.07]	[0.09]	[0.09]	[0.09]	[0.09]	[0.08]
	{0.05}	{0.05}	{0.05}	{0.06}	{0.06}	{0.05}
Geographical Controls	No	Yes	Yes	Yes	Yes	Yes
Regional FE	No	No	No	No	No	Yes
Altonji et al						28.16
δ						49.64
β-Oster						0.22
R^2	0.11	0.19	0.21	0.23	0.25	0.39
Observations	198	198	198	198	198	198

Notes: This table establishes the positive, statistically, and economically significant association between ecological diversity and the existence of politeness distinctions in the language spoken in this region, accounting for regional fixed-effects and other geographical characteristics. All columns have the same specification and controls as Table 1. Geographical controls include absolute latitude, mean elevation, terrain ruggedness, and coast length, as well as other agriculture-related controls as precipitation and temperature means and standard deviations. All independent variables have been normalized by subtracting their mean and dividing by their standard deviation. Thus, all coefficients can be compared and show the impact of a one standard deviation in the independent variable on the probability of having a future tense in the language. Heteroskedasticity robust standard error estimates are reported in parentheses, clustered at the language genus in parenthesis and squared brackets, spatial auto-correlation corrected standard errors (Conley, 1999) in squared brackets and Cliff-Ord ML in curly brackets; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table C12: Geographical Origins of Politeness and Language Structures

	Language Structure						
		Temporal Structures		Non-Temporal Structures			
	Politeness	Past	Perfect	Possessive	Evidentiality	Consonants	Colors
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ecological Diversity	0.23***	0.04	0.05	-0.04	0.01	-0.08	-0.49
	(0.07)	(0.04)	(0.04)	(0.04)	(0.03)	(0.05)	(0.38)
All Geographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.32	0.07	0.14	0.15	0.20	0.31	-0.02
Observations	198	218	218	223	386	538	116

Notes: This table establishes the positive, statistically, and economically significant association between ecological diversity and the existence of politeness distinctions in a language, and compares their impact on other language structures. The analysis accounts for regional fixed-effects and other geographical characteristics as in previous tables. Other language structures include the existence a past tense, a perfect tense, the existence of obligatory possessive inflections, semantic distinctions of evidentiality, the number of consonants, the ratio of consonants to vowels and the number of colors. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table C13: Persistent Impact of Geographical Characteristics on Politeness (Intensive Measure)
Linguistic Homeland vs. Urheimat
(Languages Outside Urheimat)

	Intensity of Politeness Distinctions					
	Migratory Distance to Urheimat					
	Any Distance			At Least 1 Week		
	Homeland	Urheimat	Both	Homeland	Urheimat	Both
	(1)	(2)	(3)	(4)	(5)	(6)
Homeland Ecological Diversity	0.28***		0.27***	0.28***		0.26***
	(0.07)		(0.06)	(0.08)		(0.07)
Urheimat Ecological Diversity		0.03	-0.07		0.17	-0.03
		(0.08)	(0.08)		(0.11)	(0.09)
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes
Homeland Geographical Characteristics	Yes	No	Yes	Yes	No	Yes
Urheimat Geographical Characteristics	No	Yes	Yes	No	Yes	Yes
Adjusted-R^2	0.40	0.34	0.52	0.39	0.39	0.57
Observations	146	146	146	126	126	126
Language Families	36	36	36	28	28	28

Notes: This table explores the relative contributions of ecological diversity in the contemporary homeland vs. the Urheimat to the presence of politeness distinctions in a daughter language. Heteroskedasticity robust standard error estimates clustered at the language family level are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table C14: Geographic Origins of Juridictional Hierarchy and Politeness Distinctions (Intensive Measure)

	Mechanism			
	Jurisdictional Hierarchy		Politeness	
	(1)	(2)	(3)	(4)
Ecological Diversity	0.23***	0.12***		
	(0.04)	(0.03)		
Average Caloric Yield (All Crops, pre-1500)	0.17***	0.23***		
	(0.03)	(0.03)		
Jurisdictional Hierarchy Beyond Local Community			0.38***	0.33***
			(0.05)	(0.05)
All Geographic Controls	No	Yes	No	Yes
Regional FE	No	Yes	No	Yes
Adjusted-R^2	0.07	0.32	0.35	0.54
Observations	1154	1154	139	139

Notes: This table establishes the positive statistically and economically significant association between the geographical determinants of statehood, as measured by jurisdictional hierarchy beyond the local level, and politeness distinctions in a language. The first two columns provide the evidence on the association between ecological diversity and statehood, and columns (3) and (4) show the association between statehood and the emergence of politeness distinctions. The table shows the estimated coefficients in an OLS regression as the dependent variable in columns (1) and (2) is not binary. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table C15: Hierarchy and Politeness

	Hierarchy Index					
	All			Old World		
	(1)	(2)	(3)	(4)	(5)	(6)
Politeness Distinctions	0.48***	0.37**	0.37**	0.55***	0.56***	0.56***
	(0.15)	(0.14)	(0.16)	(0.16)	(0.16)	(0.10)
Ecological Diversity			0.18*			0.07
			(0.09)			(0.09)
Regional FE	No	Yes	Yes	No	Yes	Yes
Main Geographical Controls	No	No	Yes	No	No	Yes
Adjusted-R^2	0.22	0.43	0.49	0.28	0.47	0.60
Observations	53	53	53	50	50	50

Notes: This table establishes the statistically and economically positive association between of politeness distinctions and preferences for hierarchy in society. The analysis accounts for the geographical origins of politeness distinctions and other geographical characteristics of the contemporary homeland of the language. Heteroskedasticity robust standard error estimates clustered at the language family level are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

D Variable Definitions, Sources and Summary Statistics

D.1 Variable Definition and Sources

- **Absolute latitude**: The absolute value of the latitude of a homeland's approximate geodesic centroid. Author's computations.
- **Mean Elevation**: The mean elevation of a homeland in km above sea level, calculated using geospatial elevation data taken from GLOBE Task Team and others (1999). Author's computations.
- **Terrain Ruggedness**: The mean change in elevation across cells in a homeland in km, calculated following the methodology of Riley et al. (1999), using geospatial elevation data taken from GLOBE Task Team and others (1999). Author's computations.
- **Caloric Suitability**: Pre-1500CE Caloric suitability is the potential caloric output in a region as reported in Galor and Özak (2016).
- **Coast length**: Length, in thousands of km, of a country's coastline. Author's computations.
- **Ecological Diversity**: Herfindahl index of share's of a country's area in various ecologies. Author's computations following the method of Fenske (2014) and Depetris-Chauvin and Özak (2016).
- **Volatility (temperature and precipitation)**: Volatility of temperature and precipitation constructed using v3.2 of the Climatic Research Unit (CRU) database following the method of Durante (2010).
- **Diversification (temperature and precipitation)**: Spatial Correlation of temperature and precipitation shocks constructed using v3.2 of the Climatic Research Unit (CRU) database following the method of Durante (2010).

D.2 Summary Statistics

Table D1: Summary Statistics of the Existence of Periphrastic Future Tense by Region

Region	Observations	Mean	Std. Dev.
Sub-Saharan Africa	66	0.53	0.503
Middle East and North Africa	8	0.5	0.53
Europe and Central Asia	56	0.48	0.50
South Asia	21	0.19	0.40
East Asia and Pacific	71	0.55	0.50
North America	22	0.41	0.50
Latin America	31	0.55	0.50
Total	275	0.49	0.50

Table D2: Summary Statistics of the Existence of Sex-Based Grammatical Gender Systems by Region

Region	Observations	Mean	Std. Dev.
Sub-Saharan Africa	27	0.63	0.49
Middle East and North Africa	7	0.71	0.49
Europe and Central Asia	40	0.48	0.51
South Asia	16	0.63	0.50
East Asia and Pacific	70	0.27	0.45
North America	25	0.08	0.28
Latin America	32	0.28	0.46
Total	227	0.37	0.48

Table D3: Summary Statistics of the Existence of Politeness Distinctions by Region

Region	Observations	Mean	Std. Dev.
Sub-Saharan Africa	36	0.14	0.35
Middle East and North Africa	4	0.25	0.50
Europe and Central Asia	34	0.71	0.46
South Asia	19	0.63	0.50
East Asia and Pacific	59	0.32	0.47
North America	18	0.00	0.00
Latin America	28	0.18	0.39
Total	207	0.34	0.48

Table D4: Summary statistics

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
Absolute Latitude	0.096	1.025	-1.302	2.613	275
Elevation	0.027	1.026	-0.92	4.827	275
Ruggedness	-0.014	0.979	-0.877	6.162	275
Coast Length	0.024	1.154	-0.302	11.692	275
Precipitation	-0.078	0.928	-1.3	4.4	275
Precipitation (std)	-0.02	0.911	-0.667	8.314	275
Precipitation Volatility	-0.064	0.926	-1.531	4.665	275
Precipitation Spatial Correlation	0.064	0.939	-2.133	0.810	275
Temperature (Daily Mean)	-0.054	0.977	-2.996	1.176	275
Temperature (Daily Mean) (std)	-0.017	0.929	-0.877	4.876	275
Temperature Volatility	0.079	0.991	-1.641	3.504	275
Temperature Spatial Correlation	0.068	0.939	-2.161	0.683	275

Table D5: Summary statistics

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
Intensity of Agriculture	8.890	3.061	2	12	264

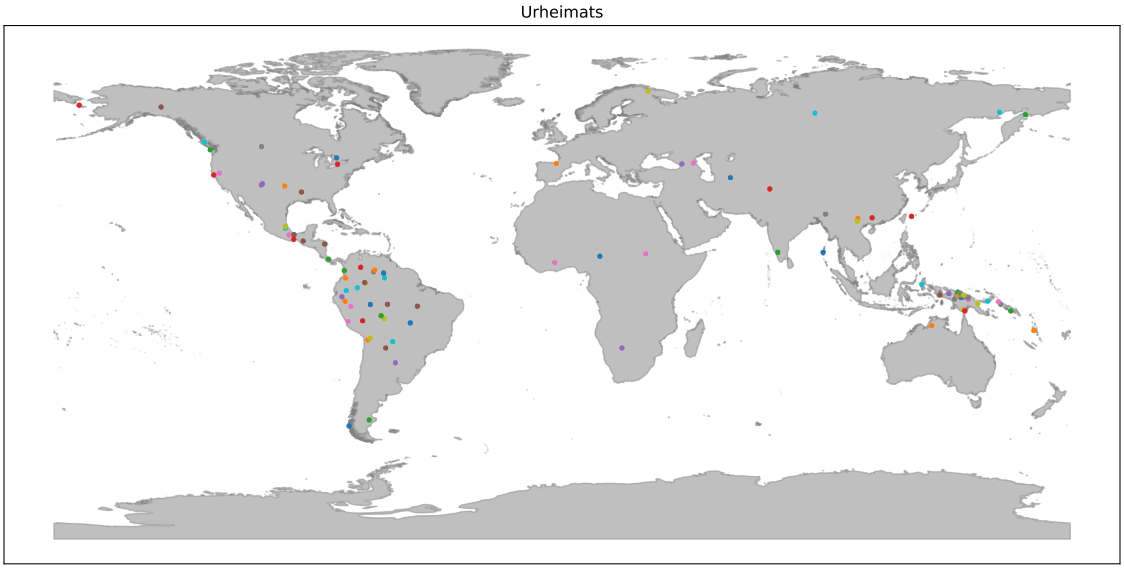


Figure E1: Location of Urheimats for 147 Language Families in Ethnologue

E Identifying the Location of Urheimats of Language Families

The location of language's Urheimat (i.e., the geographical region where the language family emerged) is identified based on the frontier methodology proposed by linguists (Wichmann et al., 2010), according to which the geographical area that maximized linguistic diversity within a language family corresponds to the Urheimat.[52]

In particular, using data on the linguistic diversity of 4169 languages from version 12 of the ASJP (Wichmann et al., 2009), Wichmann et al. (2010) identify the Urheimat of 82 language families based on the comparison of the level of linguistic diversity of each language within each language family. Following their methodology the analysis identifies the location of Urheimats for 147 language families based on the updated version 18 of the ASJP (Wichmann et al., 2018). Figure E1 depicts the location of these Urheimats, while Figures E2 and E3 depict the extent of the Indo-European and Niger-Congo language families and the location of their Urheimats.

[52]This method builds upon a similar methodology employed by geneticists to determine the location in which domestication of crops was originated.

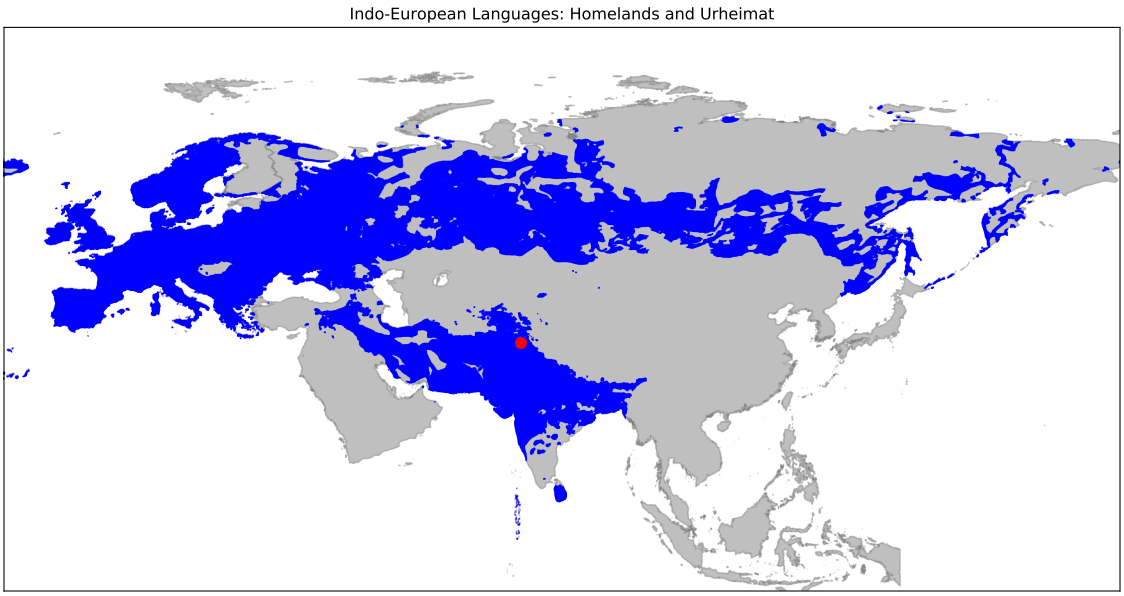


Figure E2: Indo-European Language Family: Homelands and Urheimat

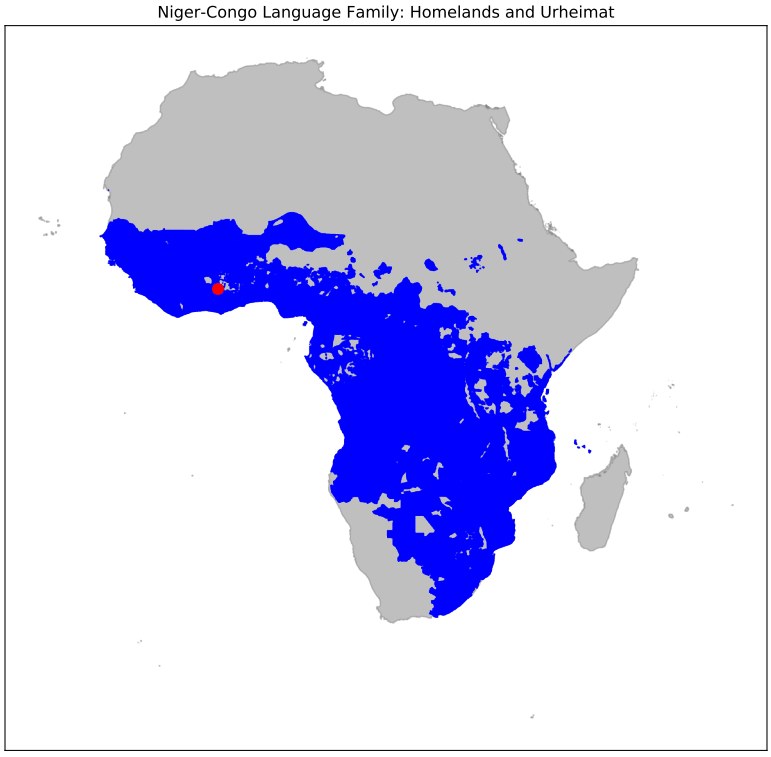


Figure E3: Niger-Congo Language Family: Homelands and Urheimat

F The Timing of the Appearance of Ancient Plows

Figure F1: Ancient Egyptian ard. Burial chamber of Sennedjem (ca. 1200 BCE).

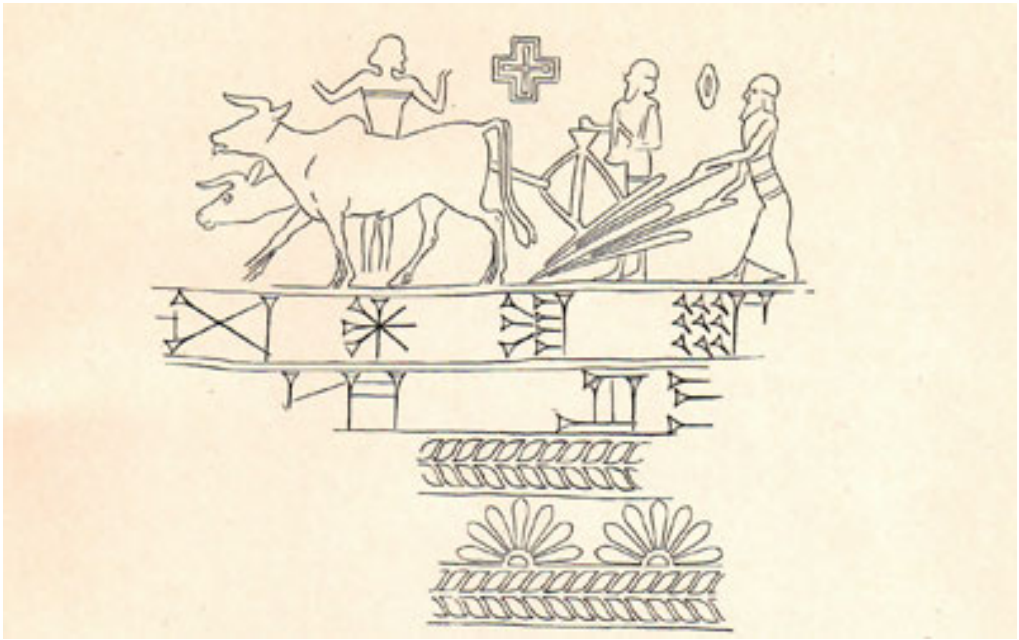


Figure F2: Gang of men plowing and sowing. Babylonian Seal Impression (1500 BCE)

The hypothesized coevolution of gender biases and sex-based grammatical gender based on their common geographical origin necessitates the appearance of the plow prior to the emergence of significant fraction of daughter languages within each language family. Indeed, paleo-linguistic and archeological evidence suggest that the plow appears prior to the formation of languages with a sex-

Figure F3: Gang of men plowing and sowing. Sumerian (ca.4500-2000BCE).

Figure F4: Picture of a field plowed along the Indus river (ca. 2000 BCE)

based grammatical gender. Evidence indicates that the plow has originated during the emergence of agriculture. In particualr, evidence suggests that sporadic use of the plow is present as early as 10-13 thousands years ago in the fertile crescent, and in particualr along the Tigris, Euphrates, Nile, as well as in Indus and Yangtze River valleys (Lal et al., 2007). Primitive prototypes of plows that were sufficently demanding physically to provide a comprative advantage in their use for male (e.g, the *ard*) were more prevalent in the fertile crescent and in particualr in Mesopotamia and along the Nile river 5000-8000 years ago (Lal et al., 2007; Lal, 2009). Moreover, nearly 4000 years ago, the plow was adopted by the Greeks, was further developed by the Indus Valley civilizations (i.e., the forefathers of the Dravidian language family), and by civilizations that resided along the Yangtze river (Lal, 2009). In additon, the plow was prevelant in both Northern Africa and Ethiopia, where the ancestors of the Afro-Asiatic language family resided, at least 2500 years ago (Blench, 2013).

Linguistic evidence suggests that early Indo-Europeans used the plow during the late Neolithic period, before their proto-language separated into its different daughter languages around 3000 BCE. In particualr,German emerged around the 1st century BCE, English and Spanish in about the 5th century CE and Italian around the 14th century CE. The proto-language spoken around 4500 BCE consists of a word for the plow prior to the demic diffusion of the Indo-European languages to Asia and Europe (Piggott, 1983). Reconstructions of the proto-Indo-European language suggest that the verb, plow, was *ar(e)*, which is the root of *ear* (English), *erian* (Old English), *erja* (Old Norse), *arjan* (Gothic), *arare* (Latin), *arar* (Spanish), as well as *ard* (English) and *arado* (Spanish) (Pokorny, 1948).

In contrast to temperate zones where plow positive crops were suitable for cultivation, tropical zones as well as the Americas have not adopted the earlier versions of the plow. Instead, in order to till the soil these areas adopted a hoe (i.e., a pointy stick used by hunter-gatherers) that was more suitable for their climatic environment. In these areas, the plow was introduced by Europeans during the colonial period.[53]

[53]Figures F1-F3 depict ancient plow agriculture in Egypt, Babylon and Sumer around 2000BCE. Additionally, Figure F4 depicts a plowed field in the Indus Valley around 2000BCE. Literary evidence for the early adoption of the plow in these regions is "Debate between the pickaxe and the plough" (*The Sumerian Disputation/Debate*; 3rd millennium BCE), the Old and the New Testements, and the Babylonian *Code of Hammurabi* (ca. 1750BCE) where references to the plow or plowing is present in laws 43, 44, 242, and 260.