

# Measuring morphological similarities for low-resource languages

**ALFONSO MEDINA URREA**  
**El Colegio de México**

**Linguistic Data Consortium Workshop on Languages of the Americas**  
in the City of Mexico, May 24<sup>th</sup> and 25<sup>th</sup>, 2018

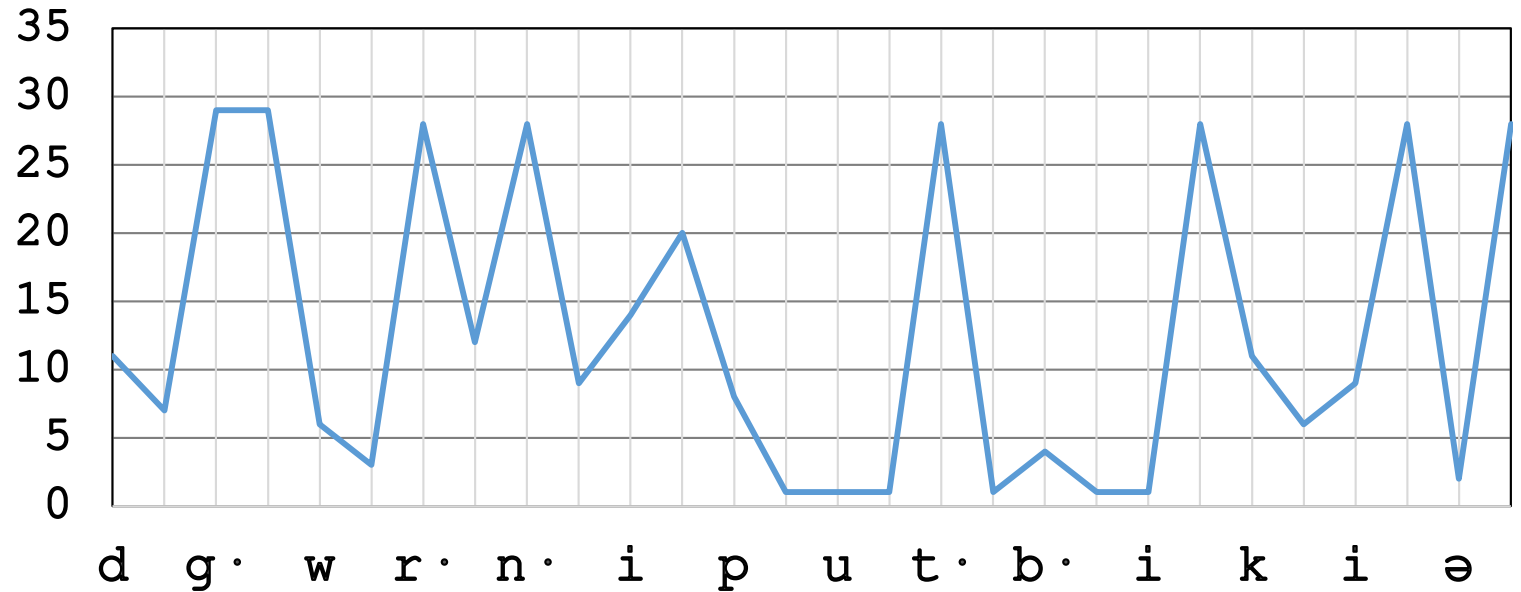
# El Colegio de México

- Founded in 1940
- Hispanic Literature & Linguistics PhD Program opened in 1960
- Home of the Dictionary of Mexican Spanish (1974-present)
  - Inspired on the Noah Webster's Dictionary (circa 1840)
  - <http://www.colmex.mx>
  - Vocabulary selected by means of corpus statistical methods



# Discovering Affixes in Corpora

- Unsupervised word segmentation
- Some well known methods
  - Harris, phoneme counts
  - Goldsmith, *Linguistica*
  - Creuz, *Morfessor*
  - ...
- In the following experiment, affix catalogs for some Mayan languages (Chuj, Tojolabal, Yucateco and Huasteco) are compiled



# Unsupervised Morphological Segmentation of graphical Words

- Automatic separation of stems and affixes
- *Affixality* estimation:
  - Measurements of structure (economy and square counts) and information (entropy)
- To build affix catalogs (suffixes, prefixes and sets of affixes)
  - Each affix is inserted in a catalog along with its values
  - Affix entries are ordered from most to least *affixal*
  - These sets of values characterize languages like “fingerprints”

# Example of suffix catalog

Table 1. Sufijos del *CHEM* s. xvi en orden de *afijalidad*

rango	sufijo	frec.	cuadros	economía	entropía	afijalidad		
1.	~a	967	0.6145	0.9741	0.9464	0.4027	0.4361	0.845
2.	~s	1321	1.001	1	0.4452	0.4206	0.501	0.8153
3.	~ó	331	0.5993	0.9653	0.8512	0.8173	0.898	0.8053
4.	~o	955	0.6421	0.9795	0.7905	0.3674	0.5968	0.804
5.	~as	413	0.3256	0.9657	0.9341	0.4693	0.5869	0.7418
6.	~e	548	0.4668	0.9728	0.7851	0.3229	0.191	0.7415
7.	~ar	275	0.2859	0.9616	0.9513	0.618	0.7585	0.7329
8.	~ado	260	0.2627	0.9744	0.9572	0.6147	0.7781	0.7314
9.	~os	514	0.3665	0.9706	0.8437	0.3909	0.5382	0.7269
10.	~en	255	0.3439	0.9674	0.7996	0.5403	0.2047	0.7036
11.	~ase	103	0.1742	0.9693	0.9666	0.741	0.8354	0.7034
12.	~aron	129	0.2173	0.9778	0.9138	0.7633	0.9289	0.703
13.	~ados	118	0.1588	0.9742	0.9554	0.599	0.6371	0.6961
14.	~an	303	0.2261	0.952	0.8944	0.4742	0.4324	0.6908
15.	~asen	70	0.1326	0.9672	0.9684	0.7527	0.8239	0.6894
16.	~ando	121	0.1516	0.9566	0.9456	0.6954	0.8559	0.6846
17.	~ava	95	0.1453	0.96	0.9241	0.7197	0.7986	0.6765
18.	~ará	39	0.08687	0.9441	0.9837	0.75	0.5596	0.6715
19.	~ido	127	0.1609	0.956	0.8957	0.6546	0.6423	0.6709
20.	~ada	95	0.1159	0.9543	0.9419	0.5901	0.488	0.6707
21.	~aba	97	0.162	0.9608	0.8788	0.7405	0.8569	0.6672
22.	~er	121	0.147	0.9191	0.9152	0.6685	0.7295	0.6605
23.	~n	719	0.4461	0.9886	0.5279	0.394	0.2299	0.6542
24.	~i	57	0.06176	0.8982	1	0.3904	0.02974	0.6533
25.	~adas	51	0.08442	0.9468	0.9287	0.5795	0.7404	0.6533
26.	~avan	64	0.1061	0.955	0.896	0.7901	0.8612	0.6524
27.	~fa	165	0.235	0.9486	0.7609	0.8333	0.8456	0.6482
28.	~ara	28	0.06945	0.9251	0.9444	0.6829	0.02649	0.6463
29.	~amente	34	0.05795	0.9603	0.9112	0.6415	0.7545	0.6431
30.	~iendo	78	0.1273	0.9485	0.8436	0.7959	0.7042	0.6398

- Average of normalized measurements in last column (affixality)
- Items ordered from more to less affixality
- Greater frequency does not necessarily implies greater affixality

<b>prefijo</b>	<b>frec.</b>	<b>cuadros</b>	<b>econ.</b>	<b>entrop.</b>	<b>afijalidad</b>
a~	2074	0.05189	0.9511	0.963	0.6553
re~	1866	0.07725	0.9639	0.9108	0.6507
semi~	63	0.004754	0.943	0.9617	0.6365
des~	1041	0.05772	0.97	0.847	0.6249
auto~	82	0.01445	0.9076	0.9144	0.6122
pro~	481	0.02093	0.9166	0.8892	0.6089
kontra~	92	0.01568	0.8883	0.9219	0.6086
anti~	66	0.004184	0.9108	0.8976	0.6042
in~	1066	0.03447	0.9551	0.8176	0.6024
sub~	198	0.01012	0.8532	0.9334	0.5989
radio~	31	0.02116	0.8803	0.8728	0.5914
porta~	16	0.04622	0.9062	0.8216	0.5913
intra~	21	0.001767	0.9113	0.8444	0.5858
sobre~	126	0.01693	0.9298	0.8063	0.5844
idro~	32	0.002454	0.9122	0.8157	0.5768
pre~	425	0.02103	0.9115	0.7917	0.5748

# What is the axis of comparison among languages?

- Cognates – words with a common origin
  - **among languages:** ‘posible’/‘possible’; ‘starve’/‘sterben’; ‘exquisite’/‘exquisito’ (Portuguese)
  - **within a language:** ‘delicado’/‘delgado’; ‘shirt’/‘skirt’
- Morphological cognates
  - Verbal clitics – me, te, se, nos, lo, los, la, las, etc. (Romance Languages)
  - Dervative suffixes ‘-idad’, ‘-ité’, ‘-ity’, ‘-ität’
  - Mayan shared suffixes

# Language samples for this experiment

- Chuj: 15,900 tokens  
stories compiled by Cristina Buenrostro
- Tojolabal: 22,590 tokens  
stories compiled by Antonio Gómez et al (*Palabras de nuestro corazón*)
- Yucateco: 21,740 tokens  
stories in *Yucatán, identidad y cultura*, UAY
- Huasteco: 3,120 tokens  
stories compiled by Lucero Meléndez



# Important remarks

- Are these small samples representative of these languages?
- Writing normalization is necessary
  - An affix in one language may be a clitic in other
  - Phoneme representation varies: “tz” vs. “ts”, “x” vs. “sh”
  - Etc.
- Short and frequent affixes exhibit great polysemy or homography (-a, -o, -e, -as, -es, -os, -an, -en, etc.; -∀n)
- In short, error is inevitable

# Chuj affixes

1 in<	94	0.5278	0.8278	1.0000	0.9895	0.6548	0.9139
2 ix<	180	1.0000	0.8000	0.9198	0.9836	0.4050	0.8599
3 s<	186	0.3871	0.6517	0.9774	0.9588	0.9086	0.8146
4 ko<	71	0.3064	0.6710	0.8742	0.9221	0.8512	0.7726
5 xschi'<	1	0.0429	1.0000	0.5202	0.2500	0.0033	0.7601
6 tz<	358	0.7286	0.6143	0.8976	0.9862	0.9608	0.7560
7 tzin<	47	0.1708	0.4278	0.9512	0.9216	0.8312	0.6895
8 a<	165	0.1687	0.4051	0.9594	0.8967	0.2630	0.6823
9 olin<	26	0.1775	0.4572	0.8982	0.8387	0.7679	0.6777
10 ol<	190	0.4844	0.5276	0.8163	0.9694	0.9245	0.6719
11 ach<	11	0.0571	0.4247	0.8534	0.6111	0.1310	0.6390
12 olach<	26	0.1747	0.4944	0.7805	0.8125	0.7447	0.6375
13 w<	74	0.4338	0.7204	0.5346	0.9136	0.9441	0.6275
14 tzs<	49	0.0776	0.4267	0.8252	1.0000	1.0000	0.6259
15 ay<	31	0.2198	0.5414	0.7003	0.7381	0.2094	0.6208
16 ixin<	28	0.1199	0.3443	0.8335	0.8235	0.7188	0.5889
17 y<	128	0.3488	0.5739	0.5830	0.8767	0.4703	0.5784
18 k'a<	11	0.0156	0.3636	0.7591	0.9167	0.8800	0.5613
19 ak'<	12	0.0321	0.3172	0.8027	0.6667	0.4545	0.5599
20 tzonh<	17	0.0782	0.2754	0.8365	0.8947	0.8966	0.5559
21 ma<	31	0.0115	0.2202	0.8884	0.8158	0.5124	0.5543
22 al<	15	0.0352	0.2123	0.8902	0.7143	0.6667	0.5513
23 e<	63	0.0304	0.1663	0.9240	0.8514	0.2207	0.5452
24 x<	44	0.0601	0.2883	0.7708	0.7857	0.6803	0.5295
25 b'ati<	1	0.0286	0.7500	0.2601	0.5000	0.3333	0.5051
26 ixs<	24	0.0274	0.1688	0.8306	0.9600	0.9744	0.4997
27 olonh<	5	0.0229	0.4250	0.5715	0.5556	0.5000	0.4982
28 k'e<	3	0.0476	0.2500	0.7153	0.3750	0.0400	0.4826
29 tzach<	11	0.0701	0.3610	0.6039	0.5500	0.4615	0.4825
30 yo<	14	0.0163	0.1786	0.7741	0.6364	0.2222	0.4764

1 >i'	573	1.0000	1.0000	0.8769	0.9879	0.8176	0.9384
2 >ja	76	0.1814	0.9243	0.9171	0.9744	0.0425	0.9207
3 >b'i	342	0.5806	0.9241	0.8571	1.0000	1.0000	0.8906
4 >e'i'	45	0.1449	0.9453	0.8129	1.0000	1.0000	0.8791
5 >uk	133	0.3073	0.7699	0.8952	0.9925	0.9805	0.8326
6 >xa	63	0.1846	0.6705	0.9236	0.9844	0.9966	0.7971
7 >e'	220	0.2067	0.6816	0.8959	0.9735	0.9400	0.7888
8 >kan	32	0.0746	0.7443	0.8326	0.9697	0.3247	0.7885
9 >ma	57	0.1622	0.6798	0.8840	0.9661	0.8871	0.7819
10 >b'a	49	0.1357	0.7420	0.8146	0.9800	0.3152	0.7783
11 >a	980	0.6003	0.6247	0.9269	0.9722	0.4035	0.7758
12 >k'a	42	0.1239	0.7013	0.8087	1.0000	1.0000	0.7550
13 >tik	105	0.1577	0.6039	0.8973	0.9906	0.9962	0.7506
14 >'e'	38	0.0536	0.6291	0.8606	1.0000	1.0000	0.7449
15 >oj	26	0.0829	0.6441	0.8408	0.8966	0.1052	0.7424
16 >jel	47	0.1446	0.6284	0.8238	0.9792	0.4152	0.7261
17 >tiki'	22	0.0371	0.6572	0.7766	1.0000	1.0000	0.7169
18 >ye'	38	0.1794	0.7012	0.6833	0.9744	0.9368	0.6923
19 >'a	281	0.1520	0.5768	0.7914	0.9929	0.6803	0.6841
20 >ta	112	0.1236	0.4384	0.9114	0.9739	0.5683	0.6749
21 >an	192	0.0894	0.3283	0.9884	0.9505	0.4967	0.6584
22 >xta	22	0.0320	0.6543	0.6064	1.0000	1.0000	0.6303
23 >ya'	17	0.0621	0.3115	0.9283	0.9444	0.3247	0.6199
24 >e'a	52	0.0360	0.3928	0.8422	1.0000	1.0000	0.6175
25 >ali'	36	0.0137	0.4220	0.8118	1.0000	1.0000	0.6169
26 >xtab'i	16	0.0387	0.6465	0.5847	1.0000	1.0000	0.6156
27 >on	142	0.0910	0.5447	0.6857	0.9530	0.8272	0.6152
28 >ito	7	0.0101	0.6272	0.5997	1.0000	1.0000	0.6134
29 >e	83	0.0430	0.2734	0.9491	0.8737	0.5257	0.6113
30 >a'	134	0.0273	0.2184	1.0000	0.9371	0.5909	0.6092

# Tojolabal affixes

1 masan<	1	0.1471	0.9900	0.6117	0.2000	0.0909	0.8009
2 yajnib'i<	5	0.0735	0.7272	0.7181	0.5556	0.0327	0.7226
3 oj<	43	0.1867	0.4314	1.0000	0.8958	0.1005	0.7157
4 ja<	246	0.4907	0.4538	0.9498	0.9572	0.2441	0.7018
5 lek<	11	0.2072	0.4426	0.9009	0.6111	0.0694	0.6717
6 yuj<	15	0.1941	0.4593	0.8779	0.7143	0.2035	0.6686
7 tan<	15	0.0980	0.5638	0.7713	0.8333	0.0749	0.6675
8 ay<	26	0.1476	0.3705	0.9450	0.7222	0.1667	0.6578
9 jun<	25	0.0847	0.3828	0.8951	0.7812	0.1407	0.6390
10 la'<	6	0.0343	0.5194	0.7496	0.5455	0.2059	0.6345
11 sjoko<	1	0.1618	1.0000	0.2635	0.5000	0.6667	0.6317
12 ko'<	4	0.1213	0.3936	0.8567	0.3333	0.1538	0.6252
13 och<	21	0.2339	0.3160	0.9242	0.7000	0.1799	0.6201
14 sjipa<	1	0.0441	0.7333	0.5064	0.2000	0.2143	0.6198
15 syama<	2	0.0368	0.6417	0.5932	0.3333	0.1176	0.6174
16 s<	650	0.2855	0.2564	0.9754	0.9774	0.8837	0.6159
17 a'ji<	2	1.0000	0.5460	0.6810	0.3333	0.1515	0.6135
18 el<	28	0.2237	0.4041	0.8189	0.7368	0.2621	0.6115
19 k'ot<	15	0.2157	0.4211	0.7885	0.6818	0.3265	0.6048
20 cham<	12	0.1961	0.3597	0.8382	0.6316	0.4348	0.5989
21 wan<	44	0.1143	0.3006	0.8954	0.8462	0.4346	0.5980
22 waj<	50	0.2185	0.3411	0.8533	0.8475	0.3547	0.5972
23 b'a'y<	2	0.0441	0.4583	0.7245	0.2500	0.0909	0.5914
24 jas<	48	0.0705	0.3311	0.8498	0.8421	0.4488	0.5904
25 a<	331	0.1421	0.2253	0.9525	0.9430	0.6996	0.5889
26 mach'<	1	0.0294	0.5500	0.6117	0.2000	0.1429	0.5809
27 x<	240	0.1956	0.2808	0.8796	0.9836	0.9908	0.5802
28 kan<	29	0.1962	0.2706	0.8766	0.7838	0.2988	0.5736
29 k'ak'<	1	0.1176	0.5500	0.5932	0.1667	0.0667	0.5716
30 ek'<	13	0.1052	0.2746	0.8466	0.6190	0.2041	0.5606

1 >i'	573	1.0000	1.0000	0.8769	0.9879	0.8176	0.9384
2 >ja	76	0.1814	0.9243	0.9171	0.9744	0.0425	0.9207
3 >b'i	342	0.5806	0.9241	0.8571	1.0000	1.0000	0.8906
4 >e'i'	45	0.1449	0.9453	0.8129	1.0000	1.0000	0.8791
5 >uk	133	0.3073	0.7699	0.8952	0.9925	0.9805	0.8326
6 >xa	63	0.1846	0.6705	0.9236	0.9844	0.9966	0.7971
7 >e'	220	0.2067	0.6816	0.8959	0.9735	0.9400	0.7888
8 >kan	32	0.0746	0.7443	0.8326	0.9697	0.3247	0.7885
9 >ma	57	0.1622	0.6798	0.8840	0.9661	0.8871	0.7819
10 >b'a	49	0.1357	0.7420	0.8146	0.9800	0.3152	0.7783
11 >a	980	0.6003	0.6247	0.9269	0.9722	0.4035	0.7758
12 >k'a	42	0.1239	0.7013	0.8087	1.0000	1.0000	0.7550
13 >tik	105	0.1577	0.6039	0.8973	0.9906	0.9962	0.7506
14 >'e'	38	0.0536	0.6291	0.8606	1.0000	1.0000	0.7449
15 >oj	26	0.0829	0.6441	0.8408	0.8966	0.1052	0.7424
16 >jel	47	0.1446	0.6284	0.8238	0.9792	0.4152	0.7261
17 >tiki'	22	0.0371	0.6572	0.7766	1.0000	1.0000	0.7169
18 >ye'	38	0.1794	0.7012	0.6833	0.9744	0.9368	0.6923
19 >'a	281	0.1520	0.5768	0.7914	0.9929	0.6803	0.6841
20 >ta	112	0.1236	0.4384	0.9114	0.9739	0.5683	0.6749
21 >an	192	0.0894	0.3283	0.9884	0.9505	0.4967	0.6584
22 >xta	22	0.0320	0.6543	0.6064	1.0000	1.0000	0.6303
23 >ya'	17	0.0621	0.3115	0.9283	0.9444	0.3247	0.6199
24 >e'a	52	0.0360	0.3928	0.8422	1.0000	1.0000	0.6175
25 >ali'	36	0.0137	0.4220	0.8118	1.0000	1.0000	0.6169
26 >xtab'i	16	0.0387	0.6465	0.5847	1.0000	1.0000	0.6156
27 >on	142	0.0910	0.5447	0.6857	0.9530	0.8272	0.6152
28 >ito	7	0.0101	0.6272	0.5997	1.0000	1.0000	0.6134
29 >e	83	0.0430	0.2734	0.9491	0.8737	0.5257	0.6113
30 >a'	134	0.0273	0.2184	1.0000	0.9371	0.5909	0.6092

# Yucateco affixes

1 aaj<	1	0.6552	1.0000	0.7738	0.1111	0.0526	0.8869
2 al<	32	0.9397	0.5378	0.8783	0.6531	0.2475	0.7081
3 a<	329	0.8418	0.3540	0.9990	0.8545	0.3454	0.6765
4 j<	303	1.0000	0.3595	0.9722	0.8938	0.6850	0.6658
5 paa<	36	0.5307	0.3953	0.9255	0.9231	0.8658	0.6604
6 aa<	94	0.6049	0.3594	0.9481	0.8868	0.4795	0.6538
7 aj<	14	0.1823	0.4594	0.8068	0.6364	0.0574	0.6331
8 nuuka<	1	0.1379	0.7917	0.4643	0.3333	0.0526	0.6280
9 i<	123	0.4965	0.2485	1.0000	0.7834	0.3328	0.6243
10 k'uch<	1	0.3103	0.5864	0.6595	0.1667	0.0625	0.6230
11 x<	129	0.3066	0.2268	0.9980	0.8487	0.5437	0.6124
12 u<	132	0.3166	0.2275	0.9906	0.7543	0.1696	0.6091
13 kaa<	37	0.1286	0.3030	0.9098	0.9024	0.7905	0.6064
14 yilaj<	2	0.3103	0.4691	0.7386	0.2857	0.0807	0.6039
15 o<	138	0.5682	0.2878	0.9112	0.7500	0.2489	0.5995
16 an<	25	0.4455	0.2573	0.9225	0.6098	0.1748	0.5899
17 jun<	29	0.1415	0.3203	0.8582	0.9355	0.7442	0.5893
18 ti<	20	0.1638	0.2642	0.9076	0.6452	0.0661	0.5859
19 oo<	49	0.2906	0.2699	0.8916	0.8596	0.6131	0.5807
20 uu<	37	0.1752	0.2575	0.8849	0.8409	0.7655	0.5712
21 aat<	4	0.3793	0.4318	0.6886	0.1905	0.0420	0.5602
22 wa<	39	0.2529	0.3540	0.7643	0.7647	0.3067	0.5592
23 oke<	1	0.0690	0.5278	0.5858	0.2500	0.1250	0.5568
24 chuu<	10	0.0621	0.3782	0.7331	0.7143	0.5652	0.5557
25 e<	68	0.2358	0.1795	0.9266	0.6415	0.1698	0.5531
26 aak<	10	0.1310	0.2264	0.8741	0.5882	0.2435	0.5503
27 w<	130	0.7801	0.4838	0.6165	0.8966	0.5751	0.5501
28 ax<	5	0.3379	0.3008	0.7965	0.3846	0.0373	0.5486
29 ala<	3	0.1264	0.3167	0.7802	0.2727	0.0262	0.5484
30 le<	20	0.1534	0.3928	0.6936	0.8696	0.0827	0.5432

1 >o	403	1.0000	0.9747	0.9385	0.8916	0.5169	0.9566
2 >e'	84	0.3027	1.0000	0.8774	0.9767	0.9737	0.9387
3 >e	370	0.9977	0.9901	0.8844	0.8959	0.3228	0.9372
4 >ik	190	0.4371	0.9182	0.8590	0.9548	0.8424	0.8886
5 >i	158	0.4026	0.8425	0.8731	0.8360	0.3425	0.8578
6 >il	186	0.4022	0.7839	0.9213	0.9789	0.9256	0.8526
7 >aak	16	0.1350	0.7212	0.9251	0.9412	0.6814	0.8231
8 >a	225	0.3167	0.6338	1.0000	0.8065	0.1543	0.8169
9 >ak	71	0.1713	0.6443	0.9750	0.8554	0.6154	0.8096
10 >o'	57	0.1506	0.7666	0.8497	1.0000	1.0000	0.8081
11 >aan	24	0.1138	0.6456	0.9274	0.9600	0.8437	0.7865
12 >ile	33	0.1333	0.6082	0.9447	0.9706	0.9273	0.7764
13 >en	111	0.2424	0.7110	0.8140	0.9569	0.8360	0.7625
14 >taj	56	0.1754	0.5596	0.9619	0.9825	0.9752	0.7608
15 >o'ob	30	0.0748	0.6531	0.8620	1.0000	1.0000	0.7575
16 >tal	31	0.1695	0.5928	0.9207	0.9688	0.9833	0.7567
17 >ej	38	0.2541	0.6493	0.8088	0.9048	0.8060	0.7291
18 >al	150	0.1802	0.5316	0.9105	0.9494	0.7015	0.7211
19 >aj	256	0.2177	0.5164	0.9226	0.9517	0.7456	0.7195
20 >tik	71	0.1671	0.5115	0.8984	0.9861	0.9481	0.7049
21 >ij	33	0.1567	0.4542	0.9372	0.9429	0.8667	0.6957
22 >ta	50	0.0957	0.4668	0.9168	0.9434	0.5487	0.6918
23 >sik	22	0.1391	0.4503	0.9231	0.9565	0.9355	0.6867
24 >ike	32	0.1888	0.5284	0.8244	0.9697	0.9744	0.6764
25 >ilo	40	0.0961	0.4037	0.9445	0.9756	0.9634	0.6741
26 >ch	130	0.1287	0.6015	0.7100	0.9420	0.6881	0.6557
27 >an	51	0.0603	0.5356	0.7679	0.8500	0.6929	0.6518
28 >iko	43	0.1723	0.5792	0.7203	0.9773	0.9178	0.6497
29 >el	47	0.0462	0.3780	0.9214	0.9038	0.7137	0.6497
30 >la	30	0.0899	0.5237	0.7685	0.8333	0.5023	0.6461

# Huasteco affixes

1 aaj<	1	0.6552	1.0000	0.7738	0.1111	0.0526	0.8869
2 al<	32	0.9397	0.5378	0.8783	0.6531	0.2475	0.7081
3 a<	329	0.8418	0.3540	0.9990	0.8545	0.3454	0.6765
4 j<	303	1.0000	0.3595	0.9722	0.8938	0.6850	0.6658
5 paa<	36	0.5307	0.3953	0.9255	0.9231	0.8658	0.6604
6 aa<	94	0.6049	0.3594	0.9481	0.8868	0.4795	0.6538
7 aj<	14	0.1823	0.4594	0.8068	0.6364	0.0574	0.6331
8 nuuka<	1	0.1379	0.7917	0.4643	0.3333	0.0526	0.6280
9 i<	123	0.4965	0.2485	1.0000	0.7834	0.3328	0.6243
10 k'uch<	1	0.3103	0.5864	0.6595	0.1667	0.0625	0.6230
11 x<	129	0.3066	0.2268	0.9980	0.8487	0.5437	0.6124
12 u<	132	0.3166	0.2275	0.9906	0.7543	0.1696	0.6091
13 kaa<	37	0.1286	0.3030	0.9098	0.9024	0.7905	0.6064
14 yilaj<	2	0.3103	0.4691	0.7386	0.2857	0.0807	0.6039
15 o<	138	0.5682	0.2878	0.9112	0.7500	0.2489	0.5995
16 an<	25	0.4455	0.2573	0.9225	0.6098	0.1748	0.5899
17 jun<	29	0.1415	0.3203	0.8582	0.9355	0.7442	0.5893
18 ti<	20	0.1638	0.2642	0.9076	0.6452	0.0661	0.5859
19 oo<	49	0.2906	0.2699	0.8916	0.8596	0.6131	0.5807
20 uu<	37	0.1752	0.2575	0.8849	0.8409	0.7655	0.5712
21 aat<	4	0.3793	0.4318	0.6886	0.1905	0.0420	0.5602
22 wa<	39	0.2529	0.3540	0.7643	0.7647	0.3067	0.5592
23 oke<	1	0.0690	0.5278	0.5858	0.2500	0.1250	0.5568
24 chuu<	10	0.0621	0.3782	0.7331	0.7143	0.5652	0.5557
25 e<	68	0.2358	0.1795	0.9266	0.6415	0.1698	0.5531
26 aak<	10	0.1310	0.2264	0.8741	0.5882	0.2435	0.5503
27 w<	130	0.7801	0.4838	0.6165	0.8966	0.5751	0.5501
28 ax<	5	0.3379	0.3008	0.7965	0.3846	0.0373	0.5486
29 ala<	3	0.1264	0.3167	0.7802	0.2727	0.0262	0.5484
30 le<	20	0.1534	0.3928	0.6936	0.8696	0.0827	0.5432

1 >o	403	1.0000	0.9747	0.9385	0.8916	0.5169	0.9566
2 >e'	84	0.3027	1.0000	0.8774	0.9767	0.9737	0.9387
3 >e	370	0.9977	0.9901	0.8844	0.8959	0.3228	0.9372
4 >ik	190	0.4371	0.9182	0.8590	0.9548	0.8424	0.8886
5 >i	158	0.4026	0.8425	0.8731	0.8360	0.3425	0.8578
6 >il	186	0.4022	0.7839	0.9213	0.9789	0.9256	0.8526
7 >aak	16	0.1350	0.7212	0.9251	0.9412	0.6814	0.8231
8 >a	225	0.3167	0.6338	1.0000	0.8065	0.1543	0.8169
9 >ak	71	0.1713	0.6443	0.9750	0.8554	0.6154	0.8096
10 >o'	57	0.1506	0.7666	0.8497	1.0000	1.0000	0.8081
11 >aan	24	0.1138	0.6456	0.9274	0.9600	0.8437	0.7865
12 >ile	33	0.1333	0.6082	0.9447	0.9706	0.9273	0.7764
13 >en	111	0.2424	0.7110	0.8140	0.9569	0.8360	0.7625
14 >taj	56	0.1754	0.5596	0.9619	0.9825	0.9752	0.7608
15 >o'ob	30	0.0748	0.6531	0.8620	1.0000	1.0000	0.7575
16 >tal	31	0.1695	0.5928	0.9207	0.9688	0.9833	0.7567
17 >ej	38	0.2541	0.6493	0.8088	0.9048	0.8060	0.7291
18 >al	150	0.1802	0.5316	0.9105	0.9494	0.7015	0.7211
19 >aj	256	0.2177	0.5164	0.9226	0.9517	0.7456	0.7195
20 >tik	71	0.1671	0.5115	0.8984	0.9861	0.9481	0.7049
21 >ij	33	0.1567	0.4542	0.9372	0.9429	0.8667	0.6957
22 >ta	50	0.0957	0.4668	0.9168	0.9434	0.5487	0.6918
23 >sik	22	0.1391	0.4503	0.9231	0.9565	0.9355	0.6867
24 >ike	32	0.1888	0.5284	0.8244	0.9697	0.9744	0.6764
25 >ilo	40	0.0961	0.4037	0.9445	0.9756	0.9634	0.6741
26 >ch	130	0.1287	0.6015	0.7100	0.9420	0.6881	0.6557
27 >an	51	0.0603	0.5356	0.7679	0.8500	0.6929	0.6518
28 >iko	43	0.1723	0.5792	0.7203	0.9773	0.9178	0.6497
29 >el	47	0.0462	0.3780	0.9214	0.9038	0.7137	0.6497
30 >la	30	0.0899	0.5237	0.7685	0.8333	0.5023	0.6461

# Chuj pre- and postforms

to<	378	8.55	49.45	40.90	58	40.90	1.0000
t'a<	660	31.70	43.51	11.80	75.21	11.81	0.8799
winh<	1278	37.86	40.14	2.28	78.01	2.28	0.8117
ay<	187	13.74	27.82	14.08	41.56	14.08	0.5626
xo<	145	5.43	20.89	15.46	26.32	15.46	0.4224
yik<	47	5.95	18.40	12.44	24.35	12.45	0.3721
yo<	112	4.83	16.64	11.82	21.47	11.81	0.3365
malaj<	87	4.88	15.94	11.07	20.82	11.07	0.3223
pero<	181	1.51	15.91	14.40	17.43	14.40	0.3217
wal<	53	8.42	13.93	5.51	22.35	5.51	0.2817
axo<	281	7.23	13.46	6.23	20.69	6.23	0.2722
wan<	26	3.09	12.72	9.63	15.8	9.63	0.2572
max<	84	8.12	12.12	4.01	20.24	4.00	0.2451
tekan<	49	1.77	11.21	9.44	12.98	9.44	0.2267
wach'<	42	7.50	10.65	3.15	18.15	3.15	0.2154
mach<	30	2.11	9.28	7.17	11.39	7.17	0.1877
y<	19	1.14	8.13	6.99	9.272	6.99	0.1644
tom<	52	0.25	7.79	7.54	8.047	7.54	0.1576
maj<	57	6.17	7.01	0.84	13.18	0.84	0.1418
yuj<	161	4.53	6.75	2.22	11.28	2.22	0.1364
yol<	25	0.61	6.57	5.96	7.173	5.96	0.1328
anh<	27	5.30	6.22	0.92	11.52	0.92	0.1258
a<	350	5.81	5.99	0.18	11.8	0.18	0.1211
b'aj<	97	3.89	5.86	1.97	9.749	1.97	0.1185
pax<	22	0.76	5.22	4.46	5.983	4.46	0.1056
tzyak'<	14	2.66	5.07	2.41	7.728	2.41	0.1025
ach<	67	0.40	4.82	4.42	5.225	4.42	0.0975
yaj<	11	3.10	4.37	1.28	7.467	1.28	0.0884
a'<	29	3.80	4.22	0.42	8.022	0.42	0.0853
ingana<	27	2.36	3.96	1.60	6.323	1.60	0.0801

>eb'	345	51.66	0.55	-51.12	52.21	51.11	1.0000
>nok'	247	69.40	23.06	-46.34	92.45	46.34	0.9066
>ix	516	52.89	18.10	-34.79	70.98	34.79	0.6806
>laj	184	46.61	21.90	-24.71	68.51	24.71	0.4834
>chi'	656	53.43	30.84	-22.59	84.28	22.59	0.4419
>no'	64	17.21	3.23	-13.98	20.44	13.98	0.2735
>t'ay	163	13.57	0.29	-13.28	13.86	13.28	0.2599
>jun	522	41.50	31.09	-10.41	72.60	10.41	0.2037
>yet'	83	11.75	3.54	-8.21	15.29	8.21	0.1606
>icha	124	7.81	0.30	-7.50	8.11	7.50	0.1467
>ke	34	8.00	0.56	-7.45	8.56	7.45	0.1457
>tik	433	26.22	18.83	-7.40	45.05	7.39	0.1446
>b'ian	93	7.71	1.04	-6.67	8.74	6.67	0.1304
>te	27	10.83	4.21	-6.63	15.04	6.62	0.1296
>anima'	26	7.75	2.49	-5.26	10.23	5.26	0.1029
>kasar	19	5.62	0.77	-4.85	6.39	4.85	0.0948
>ixkot	12	5.84	1.06	-4.77	6.90	4.77	0.0934
>yowal	20	3.88	0.37	-3.51	4.26	3.51	0.0686
>ma'ay	88	5.15	1.84	-3.30	6.99	3.30	0.0646
>komam	11	4.37	1.22	-3.16	5.59	3.16	0.0617
>ixk'och	13	3.69	0.97	-2.72	4.67	2.72	0.0532
>k'en	77	17.61	15.02	-2.59	32.63	2.59	0.0507
>man	83	4.66	2.10	-2.56	6.76	2.57	0.0502
>tzb'at	21	3.91	1.74	-2.17	5.65	2.16	0.0423
>bueno	65	4.84	2.71	-2.14	7.55	2.14	0.0418
>sb'a	21	4.20	2.08	-2.12	6.28	2.12	0.0415
>tzw'al	12	2.86	0.79	-2.06	3.65	2.06	0.0404
>tzaj	9	2.83	0.95	-1.88	3.77	1.88	0.0367
>winhaj	17	3.80	1.97	-1.83	5.77	1.83	0.0359
>pensar	9	2.73	0.95	-1.78	3.67	1.78	0.0348

# Tojolabal pre- and postforms

oj<	676	50.02	192.80	142.80	242.80	142.78	1.0000
ja<	2724	349.90	417.40	67.42	767.30	67.50	0.4728
jun<	192	24.14	66.80	42.66	90.95	42.66	0.2988
b'a<	289	12.62	52.72	40.10	65.34	40.10	0.2809
cha<	145	4.93	44.75	39.83	49.68	39.82	0.2789
yajni<	139	6.33	29.07	22.74	35.40	22.74	0.1593
jel<	131	20.45	39.47	19.02	59.92	19.02	0.1332
wax<	246	53.02	71.89	18.87	124.90	18.87	0.1322
tan<	295	0.41	17.56	17.15	17.97	17.15	0.1201
yal<	132	4.69	20.00	15.31	24.69	15.31	0.1073
chichi'<	101	0.48	13.25	12.77	13.73	12.77	0.0894
waj<	189	9.76	21.77	12.00	31.53	12.01	0.0841
och<	74	6.49	18.22	11.73	24.71	11.73	0.0821
ixuki'<	58	1.41	11.52	10.11	12.93	10.11	0.0708
ka'<	42	0.22	10.02	9.80	10.25	9.80	0.0686
wan<	65	12.81	22.29	9.48	35.10	9.48	0.0664
lom<	44	0.26	8.92	8.66	9.18	8.66	0.0606
mib'i<	53	5.48	13.21	7.74	18.69	7.73	0.0542
jelxa<	30	3.97	11.43	7.46	15.41	7.46	0.0522
-<	29	2.59	9.94	7.36	12.53	7.36	0.0515
ja'xa<	66	2.33	9.54	7.21	11.87	7.21	0.0505
yala<	91	1.19	7.84	6.65	9.03	6.65	0.0466
winiki'<	58	3.05	9.43	6.38	12.47	6.38	0.0447
la'<	20	1.16	7.37	6.21	8.53	6.21	0.0435
ja'<	84	6.96	13.09	6.14	20.05	6.13	0.0430
wajkon<	29	0.36	6.38	6.02	6.74	6.02	0.0421
ma'<	36	2.45	7.92	5.47	10.37	5.47	0.0383
yajnib'i<	145	21.92	27.32	5.41	49.24	5.40	0.0378
ch'ak<	66	6.62	12.00	5.38	18.62	5.38	0.0377
kal<	18	1.32	5.88	4.56	7.20	4.56	0.0320

>sok	130	42.74	0.54	-42.21	43.28	42.20	1.0000
>yi'	168	35.58	4.81	-30.76	40.39	30.77	0.7290
>kan	131	33.76	7.64	-26.12	41.40	26.12	0.6189
>jawi'	172	44.85	20.20	-24.65	65.04	24.65	0.5841
>jan	80	18.84	1.39	-17.46	20.23	17.45	0.4135
>el	95	28.96	11.63	-17.33	40.59	17.33	0.4106
>yuj	174	19.89	2.65	-17.24	22.54	17.24	0.4085
>lek	250	39.97	23.03	-16.94	63.00	16.94	0.4014
>sb'aj	40	14.77	3.15	-11.62	17.92	11.62	0.2753
>ek'	41	13.52	1.98	-11.54	15.50	11.54	0.2734
>99	97	25.55	16.93	-8.62	42.47	8.62	0.2043
>k'ot	46	10.61	2.21	-8.40	12.82	8.40	0.1990
>yi'oj	18	7.87	0.40	-7.47	8.27	7.47	0.1769
>wewo	17	7.23	0.68	-6.56	7.91	6.56	0.1554
>iti'	38	11.85	5.83	-6.02	17.68	6.02	0.1426
>ki'	55	11.52	6.34	-5.19	17.86	5.19	0.1229
>ta	154	24.54	19.37	-5.17	43.90	5.17	0.1225
>pwes	68	12.57	7.55	-5.03	20.12	5.03	0.1191
>ke	74	19.86	14.88	-4.98	34.74	4.98	0.1180
>yab'	73	5.43	0.92	-4.51	6.35	4.51	0.1068
>ajyi	20	5.68	1.43	-4.25	7.10	4.25	0.1007
>k'a'uj	21	5.94	1.96	-3.98	7.90	3.98	0.0944
>wajye'	15	4.27	0.35	-3.92	4.63	3.92	0.0929
>mas	43	9.78	6.07	-3.70	15.85	3.70	0.0878
>jastal	102	7.53	3.83	-3.70	11.36	3.70	0.0877
>jasa	75	7.24	3.57	-3.67	10.81	3.67	0.0870
>junb'i	27	7.77	4.23	-3.54	12.00	3.54	0.0839
>b'ej	11	4.37	0.95	-3.42	5.32	3.42	0.0811
>lado	13	5.11	1.95	-3.15	7.06	3.15	0.0747
>ganar	11	3.43	0.48	-2.94	3.91	2.94	0.0698

# Yucateco pre- and postforms

u<	1731	196.90	323.10	126.20	520.10	126.20	1.0000
a<	521	63.02	135.80	72.74	198.80	72.78	0.5767
tu<	1043	56.85	93.47	36.62	150.30	36.62	0.2902
in<	484	69.80	93.74	23.94	163.50	23.94	0.1897
k<	616	34.11	57.08	22.97	91.20	22.97	0.1820
chan<	116	4.52	25.59	21.07	30.11	21.07	0.1670
jach<	128	17.49	37.69	20.20	55.19	20.20	0.1601
eel<	105	0.98	19.70	18.73	20.68	18.72	0.1484
tun<	80	1.54	19.77	18.23	21.31	18.23	0.1445
juntuul<	129	21.39	38.15	16.76	59.55	16.76	0.1328
p<	126	11.70	27.52	15.82	39.22	15.82	0.1254
beetik<	61	2.42	17.99	15.57	20.40	15.57	0.1234
j<	50	0.54	15.62	15.09	16.16	15.08	0.1195
ta<	78	15.86	30.67	14.81	46.53	14.81	0.1174
aj<	219	2.13	15.31	13.17	17.44	13.18	0.1044
mix<	57	2.54	13.41	10.87	15.95	10.87	0.0861
tene<	51	2.79	12.43	9.63	15.22	9.64	0.0764
ol<	87	2.93	11.53	8.60	14.46	8.60	0.0682
el<	44	0.98	9.46	8.48	10.44	8.48	0.0672
jook<	65	4.26	12.60	8.34	16.86	8.34	0.0661
maak<	45	2.35	10.53	8.19	12.88	8.19	0.0649
aak<	72	5.43	13.39	7.96	18.82	7.96	0.0631
bin<	272	10.69	18.36	7.67	29.05	7.67	0.0608
uch<	24	0.27	7.93	7.65	8.20	7.65	0.0606
yool<	50	1.38	8.75	7.37	10.12	7.37	0.0584
e<	199	9.97	17.31	7.34	27.28	7.34	0.0582
maan<	64	5.25	11.83	6.59	17.08	6.59	0.0522
yan<	40	4.34	10.61	6.26	14.95	6.27	0.0497
kin<	52	8.65	14.62	5.98	23.27	5.97	0.0473
ba'ax<	23	2.39	8.02	5.63	10.41	5.63	0.0446

>ka	1215	133.80	41.82	-91.93	175.60	91.98	1.0000
>ob	331	103.10	16.74	-86.36	119.80	86.36	0.9389
>le	1175	173.70	98.47	-75.21	272.10	75.23	0.8179
>yeetel	190	62.98	15.10	-47.88	78.08	47.88	0.5205
>ti	390	51.32	18.95	-32.37	70.27	32.37	0.3519
>ts	440	35.82	14.56	-21.26	50.39	21.26	0.2311
>ma	343	34.38	19.20	-15.18	53.58	15.18	0.1650
>ba	324	13.83	0.86	-12.97	14.69	12.97	0.1410
>ik	62	17.06	4.97	-12.10	22.03	12.09	0.1315
>obe	69	18.98	7.76	-11.22	26.74	11.22	0.1220
>beyo	31	13.51	3.45	-10.06	16.96	10.06	0.1093
>taan	167	10.31	0.43	-9.87	10.74	9.88	0.1074
>ja	107	14.10	4.76	-9.34	18.85	9.34	0.1016
>tech	37	18.52	10.10	-8.42	28.63	8.42	0.0915
>an	91	13.92	6.46	-7.46	20.38	7.46	0.0811
>tulaakal	43	7.95	1.59	-6.35	9.54	6.35	0.0690
>tuun	131	19.82	13.82	-5.99	33.64	6.00	0.0652
>beya	30	7.36	1.59	-5.77	8.96	5.77	0.0627
>yaalkab	48	13.58	8.00	-5.58	21.59	5.58	0.0606
>bix	65	13.13	7.81	-5.32	20.94	5.32	0.0578
>ti'	29	8.31	3.10	-5.21	11.41	5.20	0.0566
>i	44	7.83	2.65	-5.19	10.48	5.19	0.0564
>che	32	6.68	1.51	-5.17	8.19	5.17	0.0562
>nuxi	86	13.94	8.81	-5.13	22.75	5.13	0.0557
>ten	41	11.12	6.07	-5.06	17.19	5.05	0.0549
>obo	96	10.75	5.73	-5.02	16.48	5.02	0.0546
>yaan	132	14.26	9.47	-4.79	23.73	4.79	0.0521
>yuum	54	7.17	3.01	-4.16	10.17	4.16	0.0452
>teech	23	4.68	0.56	-4.12	5.23	4.12	0.0448
>jwaane	17	4.90	0.86	-4.04	5.75	4.04	0.0439



# Huasteco pre- and postforms

i<	87	0.39	22.93	22.53	23.32	22.54	1.0000
ti<	103	6.68	26.88	20.19	33.56	20.20	0.8962
an<	171	23.74	38.08	14.34	61.83	14.34	0.6363
ka<	67	12.78	25.27	12.49	38.05	12.49	0.5542
ko'ol<	24	1.02	12.42	11.40	13.44	11.40	0.5059
yab<	29	0.48	6.45	5.97	6.93	5.97	0.2648
abal<	34	5.34	10.94	5.60	16.28	5.60	0.2484
kin<	57	11.11	14.97	3.86	26.09	3.86	0.1713
cju'u<	17	1.99	5.27	3.28	7.26	3.28	0.1454
k'ale<	13	1.11	3.89	2.78	5.00	2.78	0.1234
ku<	16	0.31	2.86	2.55	3.18	2.55	0.1131
akal<	17	0.91	2.95	2.05	3.86	2.05	0.0908
usnal<	12	0.28	2.28	2.01	2.56	2.01	0.0890
ani<	28	2.50	4.28	1.78	6.78	1.78	0.0790
ki<	31	3.43	5.16	1.73	8.59	1.73	0.0766
tu<	25	2.83	4.48	1.66	7.31	1.66	0.0734
t'aja<	24	4.10	5.56	1.46	9.66	1.46	0.0648
utzal<	11	0.72	2.12	1.40	2.84	1.40	0.0621
tataa'<	13	1.33	2.55	1.22	3.87	1.22	0.0540
tin<	33	4.82	6.01	1.19	10.83	1.19	0.0527
ulicj<	8	1.38	2.50	1.13	3.88	1.13	0.0499
ojontzi<	8	0.35	1.25	0.90	1.60	0.90	0.0399
tzuthey<	8	0.41	1.25	0.84	1.66	0.84	0.0375
le<	18	0.80	1.49	0.70	2.29	0.70	0.0309
ban<	11	1.01	1.64	0.63	2.65	0.63	0.0280
kw'a<	8	0.80	1.38	0.57	2.18	0.57	0.0254
cjo'ob<	7	0.71	1.26	0.55	1.97	0.55	0.0245
na<	21	2.04	2.55	0.51	4.58	0.51	0.0227
tey<	5	0.46	0.96	0.50	1.42	0.50	0.0224
pithan<	5	0.76	0.96	0.20	1.72	0.20	0.0089

>jun	69	19.96	0.22	-19.74	20.18	19.74	1.0000
>jacjan	49	13.43	0.35	-13.09	13.78	13.08	0.6627
>ta	49	13.56	4.75	-8.82	18.31	8.81	0.4464
>tam	52	10.91	2.56	-8.35	13.47	8.35	0.4228
>kw'ajat	27	8.38	1.54	-6.84	9.92	6.84	0.3466
>in	140	41.60	36.51	-5.09	78.10	5.09	0.2578
>ne'ecj	22	6.14	1.49	-4.66	7.63	4.66	0.2360
>a	26	7.62	3.00	-4.62	10.62	4.62	0.2342
>jey	22	5.11	1.66	-3.44	6.77	3.44	0.1744
>u	46	14.78	11.58	-3.20	26.36	3.20	0.1621
>eela	8	4.00	1.08	-2.92	5.08	2.92	0.1480
>ne	30	2.90	1.04	-1.86	3.95	1.86	0.0941
>xo	18	3.84	2.23	-1.60	6.07	1.60	0.0813
>k'al	30	4.77	3.19	-1.59	7.96	1.59	0.0804
>janti	12	2.28	0.79	-1.49	3.08	1.49	0.0756
>ke	12	4.63	3.25	-1.38	7.88	1.38	0.0697
>cju'ucj	4	1.00	0.27	-0.73	1.27	0.73	0.0370
>xata	5	1.52	0.96	-0.56	2.48	0.56	0.0284
>kw'ajay	11	2.04	1.70	-0.34	3.74	0.34	0.0172
>juni	25	4.68	4.47	-0.21	9.16	0.21	0.0108
>i	87	0.39	22.93	22.53	23.32	0.00	0.0000
>ti	103	6.68	26.88	20.19	33.56	0.00	0.0000
>an	171	23.74	38.08	14.34	61.83	0.00	0.0000
>ka	67	12.78	25.27	12.49	38.05	0.00	0.0000
>ko'ol	24	1.02	12.42	11.40	13.44	0.00	0.0000
>yab	29	0.48	6.45	5.97	6.93	0.00	0.0000
>abal	34	5.34	10.94	5.60	16.28	0.00	0.0000
>kin	57	11.11	14.97	3.86	26.09	0.00	0.0000
>cju'u	17	1.99	5.27	3.28	7.26	0.00	0.0000
>k'ale	13	1.11	3.89	2.78	5.00	0.00	0.0000

# Simple affix counts

	Chuj	Tojolabal	Yucateco	Huasteco
Chuj	1241	448	300	145
Tojolabal		2283	420	180
Yucateco			1592	187
Huasteco				477

- Of the 1241 Chuj affixes, 448 were also extracted from Tojolabal and only 300 were found in Yucateco
- Tojolabal and Yucateco share 420 affixes
- Huasteco had considerably less affixes than the other languages; it shares less morphology with them

# Distancias euclidianas

$$D(chuj, tojolabal) = \sqrt{\frac{\sum_{i=1}^n (af_{chuj}(i) - af_{tojolabal}(i))^2}{n}}$$

	<b>Chuj</b>	<b>Tojolabal</b>	<b>Yucateco</b>
<b>-a</b>	0.6433	0.7758	0.8169
<b>-al</b>	0.7554	0.5907	0.7211
<b>-in</b>	0.6010	0.4273	0.5420
<b>-il</b>	0.7050	0.5869	0.8526
<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>

# Distancias morfológicas

	Chuj	Tojolabal	Yucateco	Huasteco
Chuj	0.0000	<b>0.1779</b>	0.2284	0.2065
Tojolabal		0.0000	0.1997	0.2067
Yucateco			0.0000	0.2594
Huasteco				0.0000

- The smallest distance is between **Chuj and Tojolabal**; the greatest is between *Yucateco and Huasteco*
- Yucateco is the most distant from Chuj
- Huasteco similarly distant to Chuj and to Tojolabal

# Morphological Similarity (1 - distance)

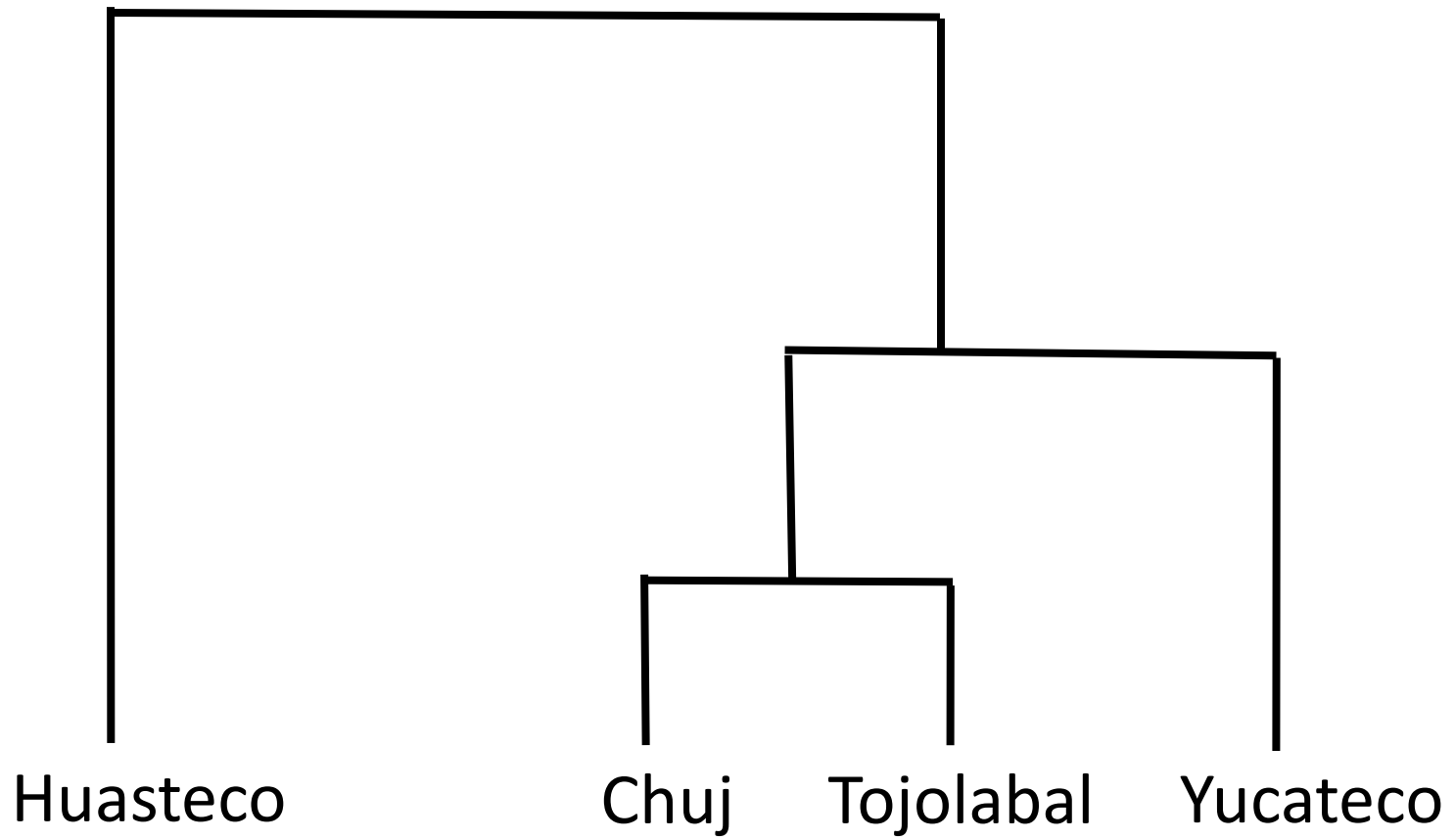
	Chuj	Tojolabal	Yucateco	Huasteco
Chuj	1.0000	<b>0.8221</b>	0.7716	0.7935
Tojolabal		1.0000	0.8003	0.7933
Yucateco			1.0000	<i>0.7406</i>
Huasteco				1.0000

- Greatest similarity is among **Chuj and Tojolabal**; the smallest is among *Yucateco and Huasteco*
- Yucateco is more similar to Tojolabal than to Chuj

# Cosine Similarity among Suffixes

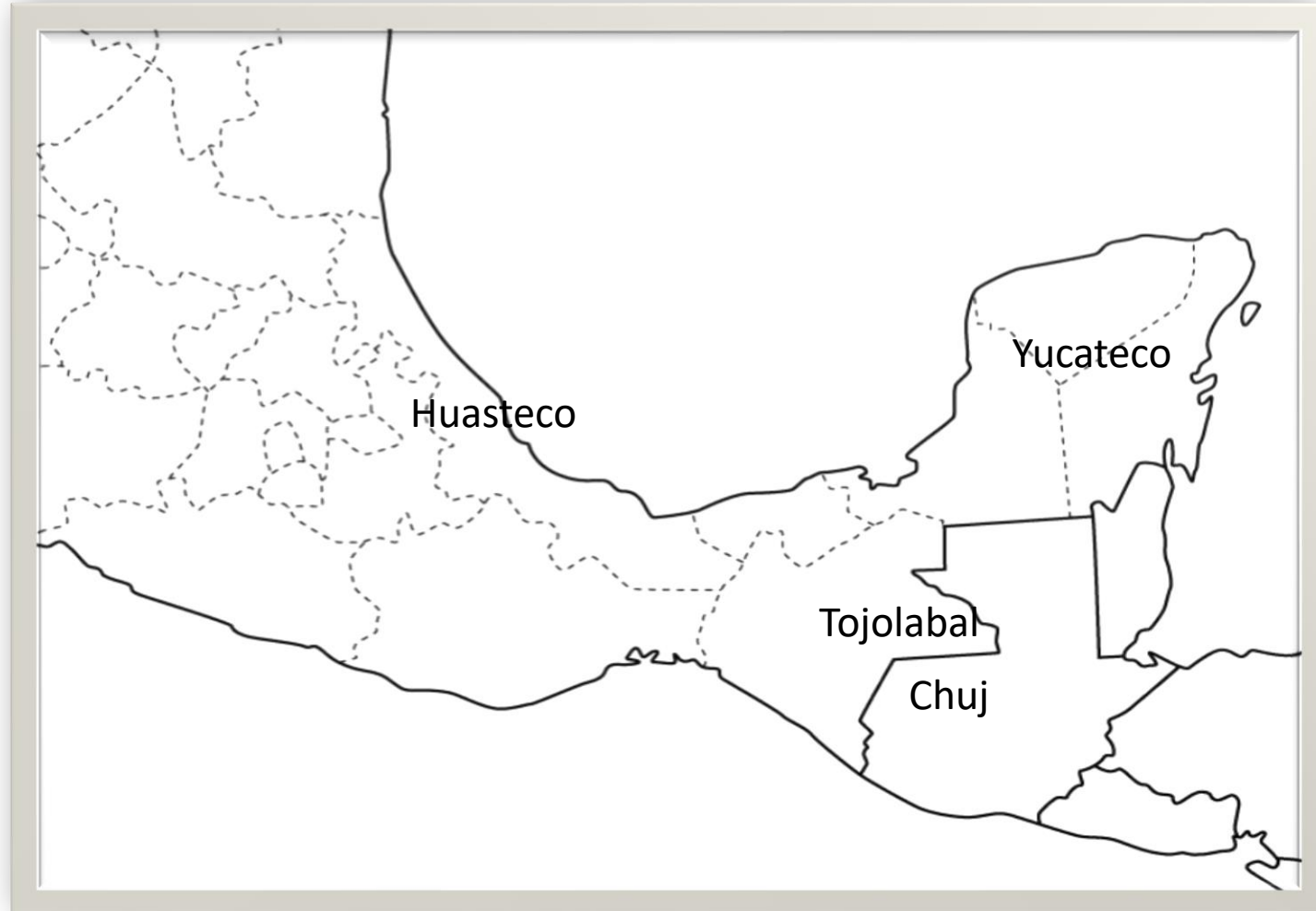
	Chuj	Tojolabal	Yucateco	Huasteco
Chuj	1.00000	<b>0.86919</b>	0.84433	0.75634
Tojolabal		1.00000	0.86820	0.74634
Yucateco			1.00000	0.74103
Huasteco				1.00000

- Again, greatest similarity is among **Chuj and Tojolabal**; the smallest is among *Yucateco and Huasteco*
- Yucateco is similar to Tojolabal than to Chuj



- Necessary to include Kanjobal in future experiments

# Geographical Distances





# Final Remarks

- Geographical and genetic relations among Mayan languages can be examined by measuring distances among sets of cognate affixes.
- Each cognate can be characterized for each language by a value describing its affixality, i. e. its quality of being an affix.
- The set of affixality values for each language can be seen as a fingerprint of that language
- These values can then be used to measure Euclidean distances, Cosine similarities, etc.
- Interestingly, these measurements seem to correspond to genetic and geographical relations among these languages

# Measuring morphological similarities for low-resource languages

[amedinau@colmex.mx](mailto:amedinau@colmex.mx)

**Linguistic Data Consortium Workshop on Languages of the Americas**  
in the City of Mexico, May 24<sup>th</sup> and 25<sup>th</sup>, 2018