5 Supplementary materials

5.1 Data

Data and source code are available online.[6] Figure 1 shows the age range of each target child in each corpus in our language sample.

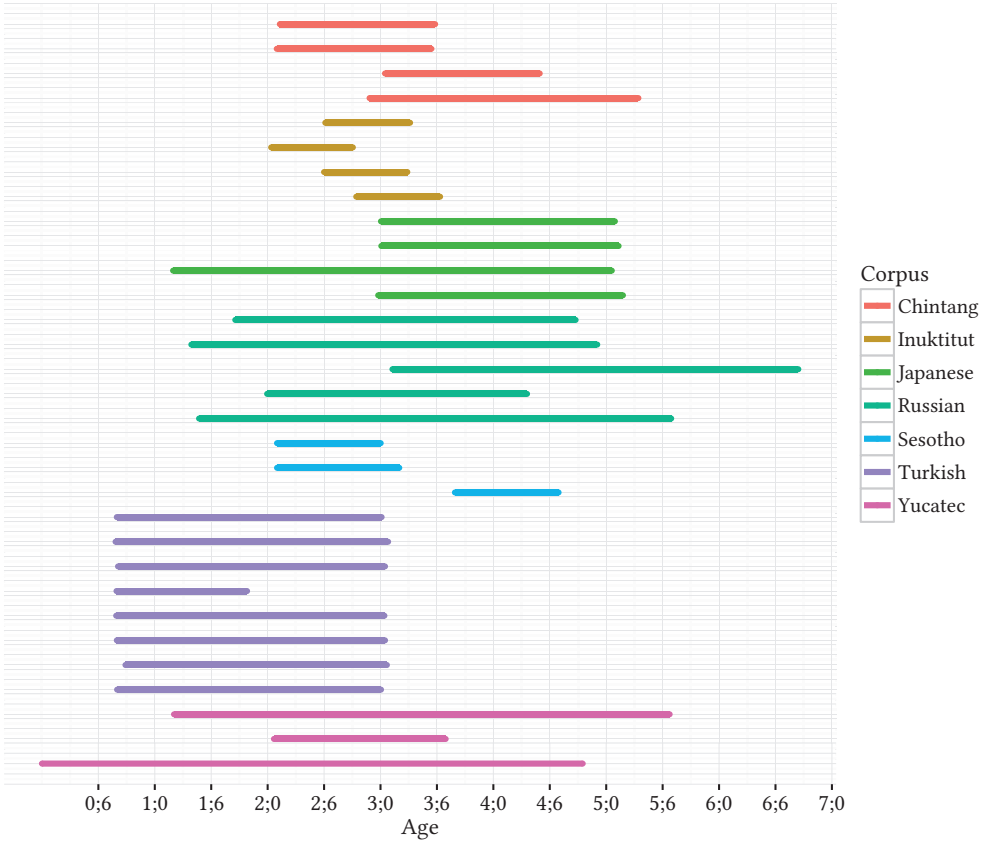

Figure 1: Age spans of children by corpus

[6] https://github.com/acqdiv/frequent-frames

Table 14 lists the grammatical category abbreviations used in the corpora. The category Particle (PTCL) includes all non-inflecting, dependent forms that have not been marked as a preposition or postposition. Morphemes also include PFX (prefix) and SFX (suffix).

Abbreviation	Category
ADJ	Adjective
ADV	Adverb
AUX	Auxiliary
CONJ	Conjunction
INTJ	Interjection
N	Noun
NUM	Numeral
POST	Postposition
PREP	Preposition
PRODEM	Pronominal / Demonstrative
PTCL	Particle
V	Verb

Table 14: Grammatical categories

5.2 Operationalization

Figure 2 shows the distribution of frequent frames across the corpora On the x-axis are relative frequency ranges. The y-axis shows the number of frames in each bin.

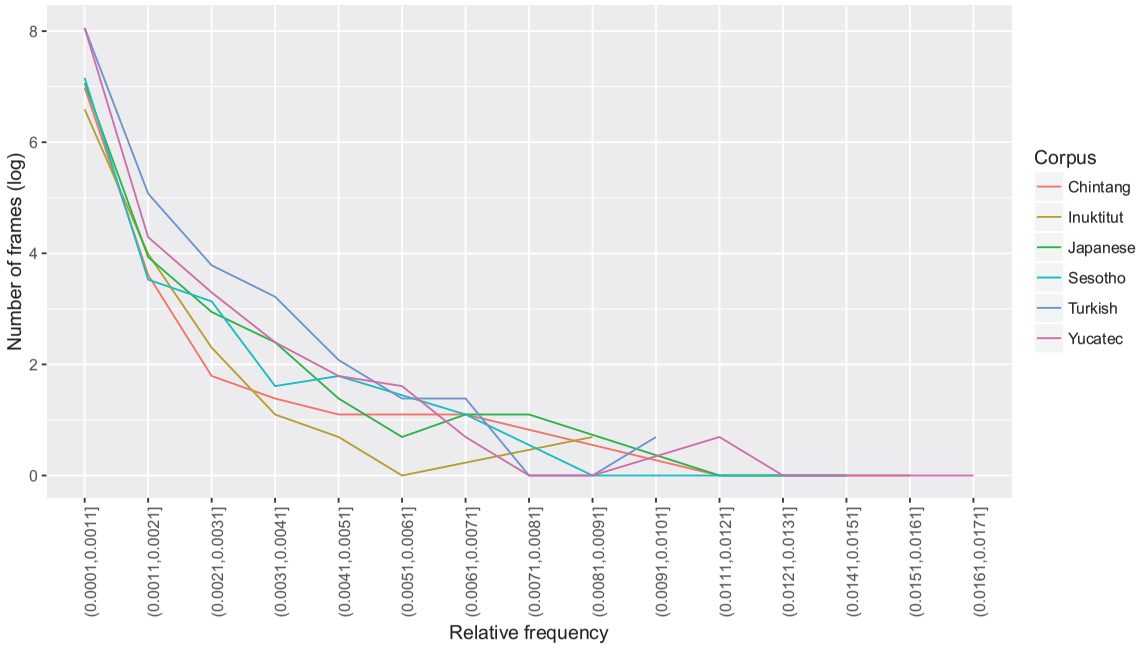

Figure 2: Distribution of most frequent morpheme frames

5.3 Analysis: individual children

Figures 3 and 4 show that frequent word frames in child-surrounding speech to individual children are not accurate or complete predictors of the intervening target word of the frame in Chintang and Russian. There are a few very accurate and often repeated frames (the large dots), but frame-based categories in general show no pattern.

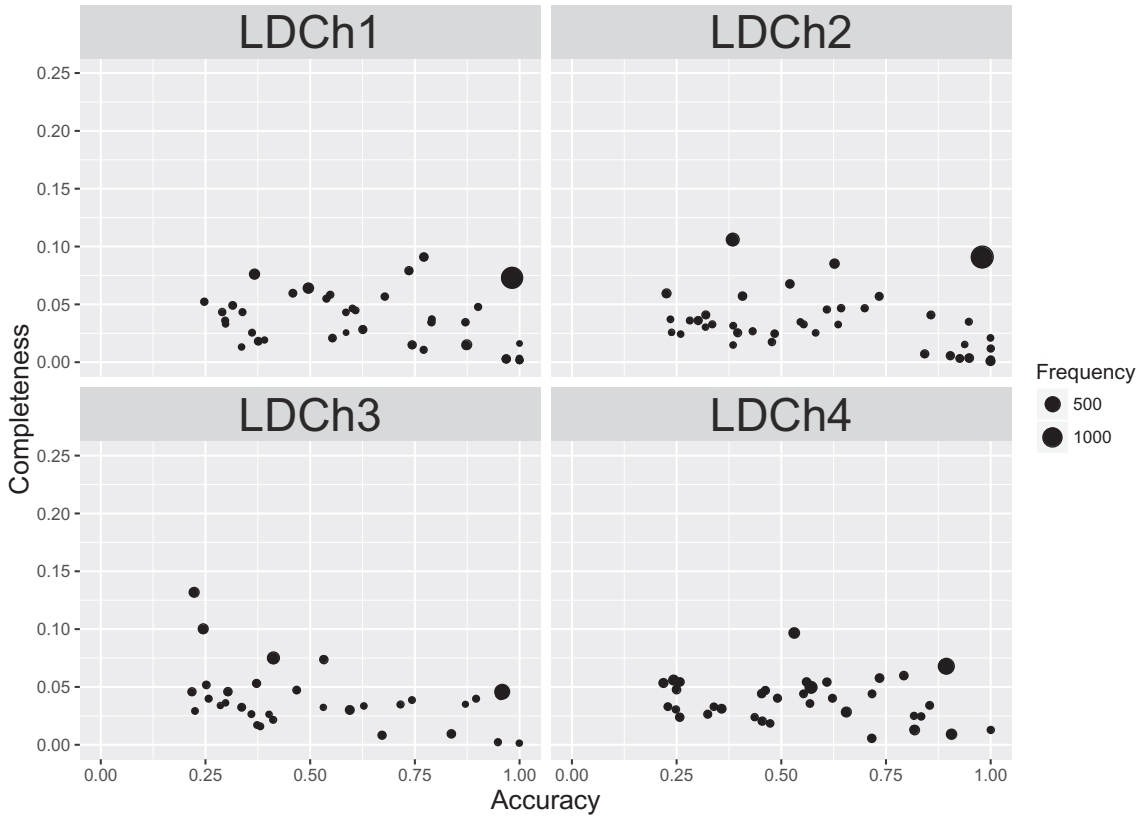

Figure 3: Accuracy and completeness: word frames to individual children in Chintang

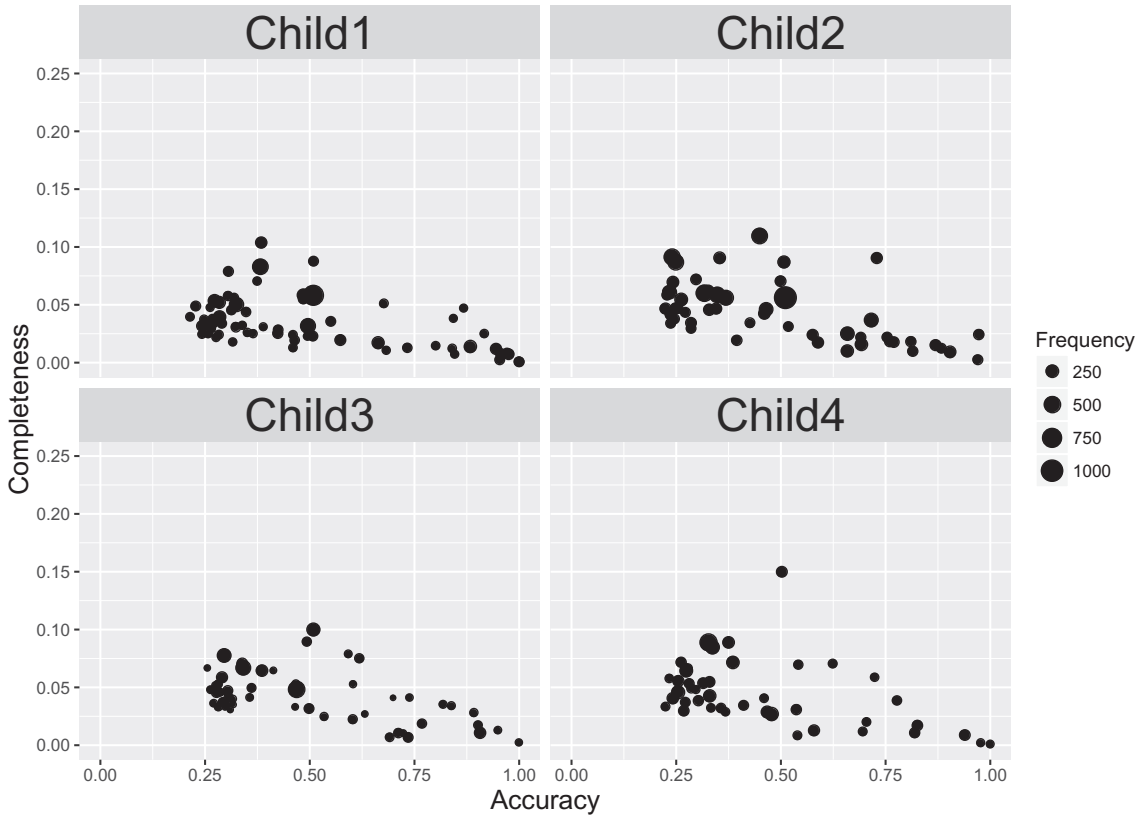

Figure 4: Accuracy and completeness: word frames to individual children in Russian

Figure 5 shows accuracy and completeness of frequent morpheme frames at the level of individual children in Chintang. There is a clear pattern shown: the majority of frames are very accurate and completeness scores are higher than in corresponding word frames.

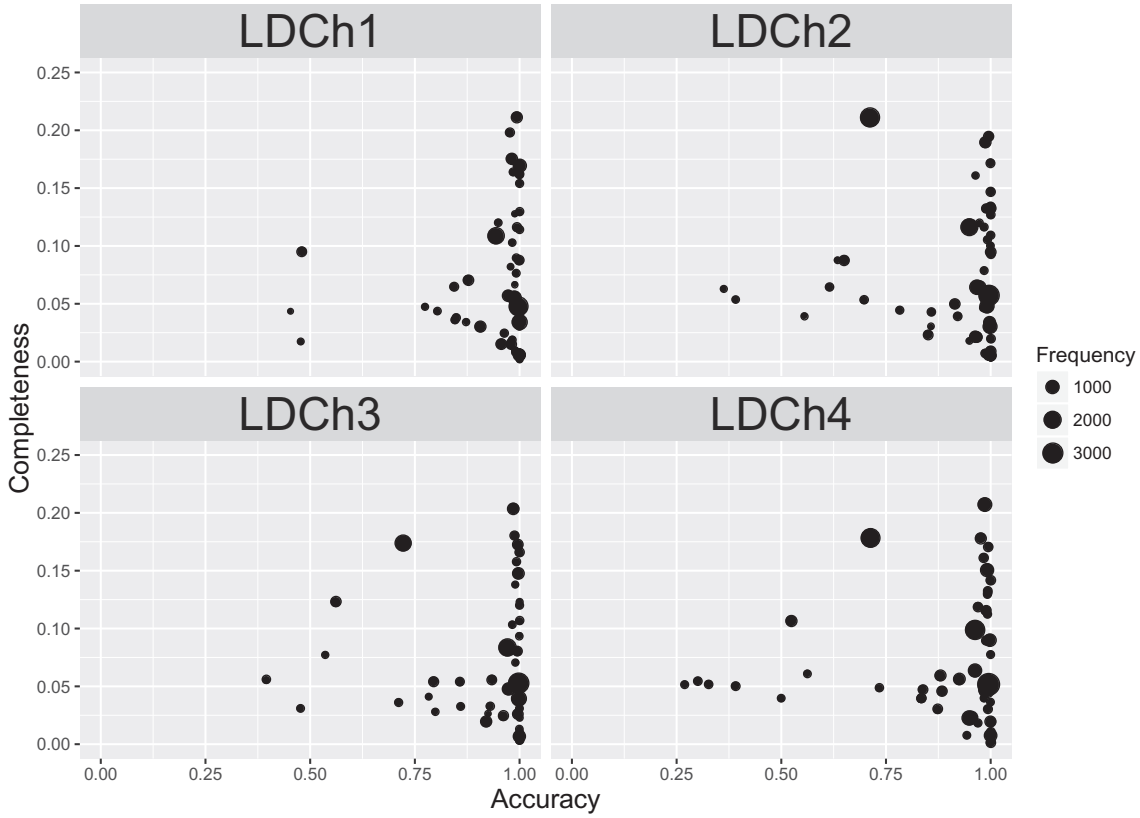

Figure 5: Accuracy and completeness: morpheme frames to individual children in Chintang

5.4 Analysis: frequent frames

Figure 6 shows the accuracy and completeness of word frames (relative frequency threshold) across pooled child-surrounding speech in each corpus. There is no discernible pattern across languages.

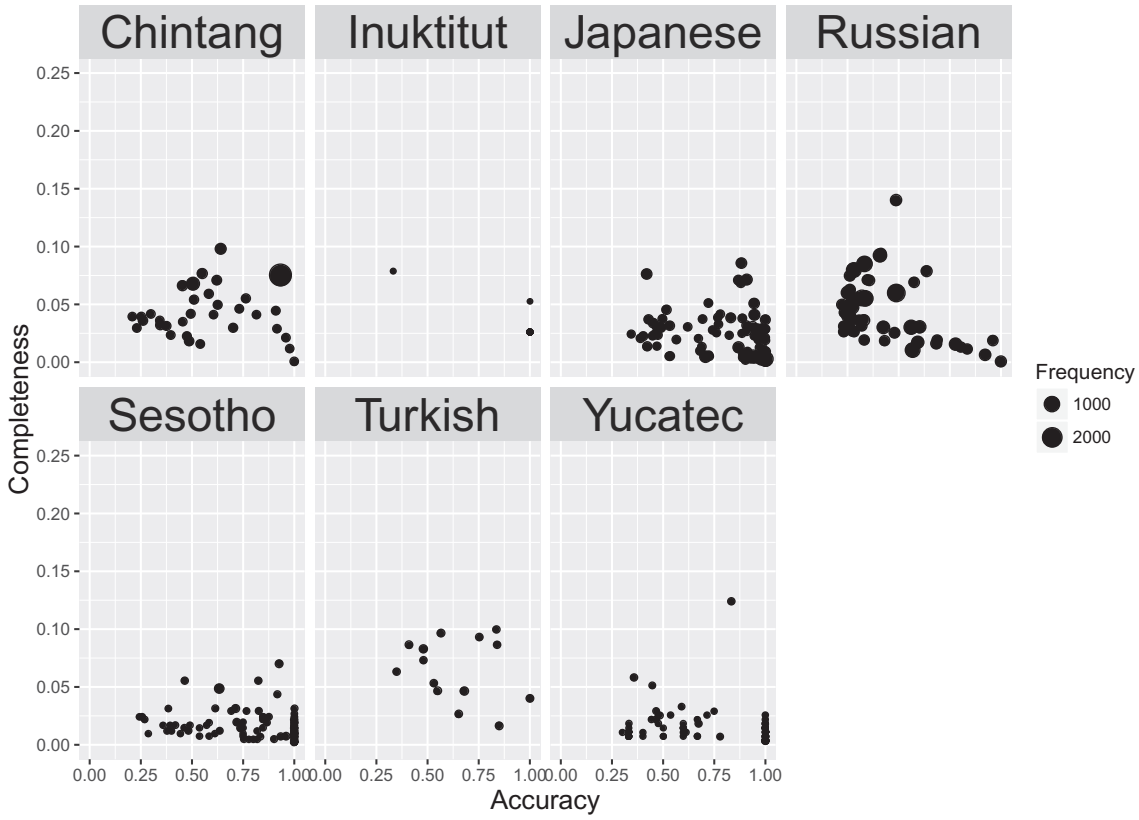

Figure 6: Frequent word frames

Figure 7 shows the accuracy and complete of frequent morpheme frames. Again at the level of morpheme frames, there is a clear pattern for frame-based categories to cluster towards higher accuracy scores and completeness scores range more than in word frames.

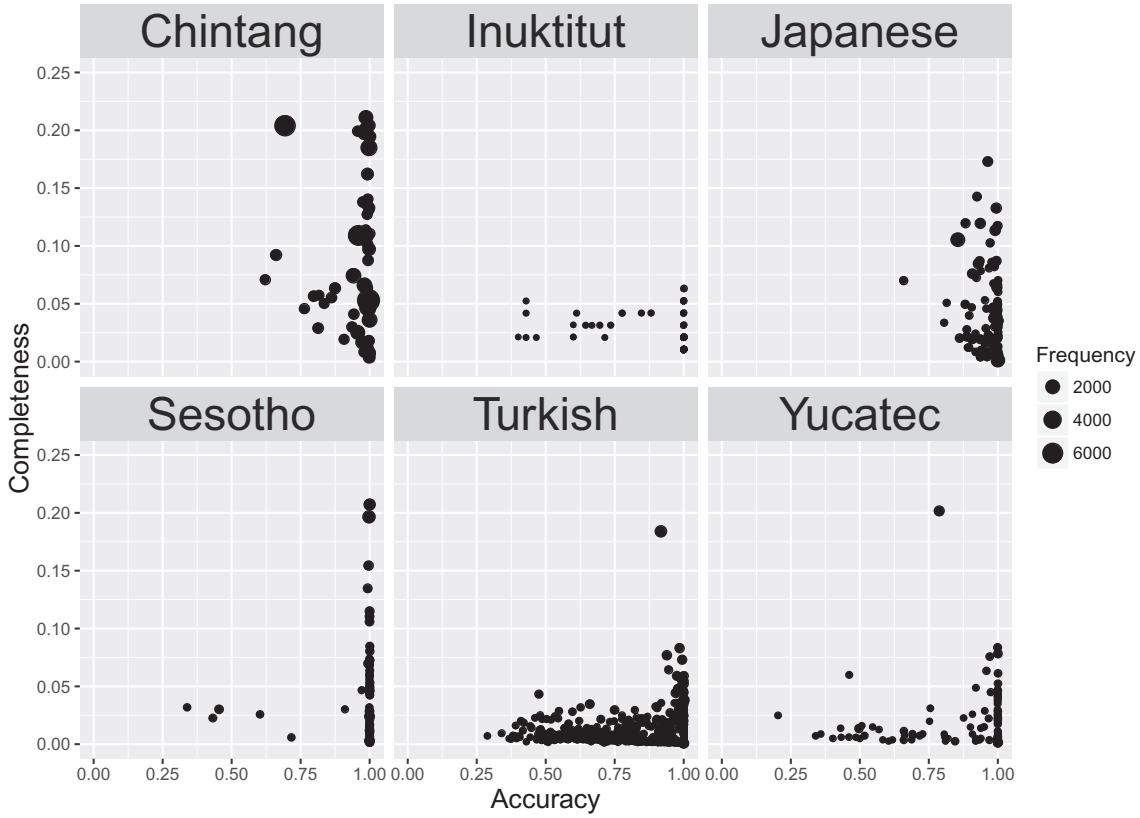

Figure 7: Frequent morpheme frames

5.5 Global accuracy and global completeness

Figure 9 shows global accuracy and completeness of all parts-of-speech at the morpheme level. There is no discernible pattern apart from the categories noun and verb.

Figure 10 plots global accuracy and completeness for nouns and verbs at the word level. Neither is particularly accurate or complete. Lastly, Figure 11 shows global accuracy and completeness for all parts-of-speech at the word level. Again, no discernible patterns appear.

Figure 8 shows global accuracy and completeness scores for nouns and verbs at the morpheme level. In nearly all cases both noun and verb frames are extremely accurate and verbs across the languages are more complete.

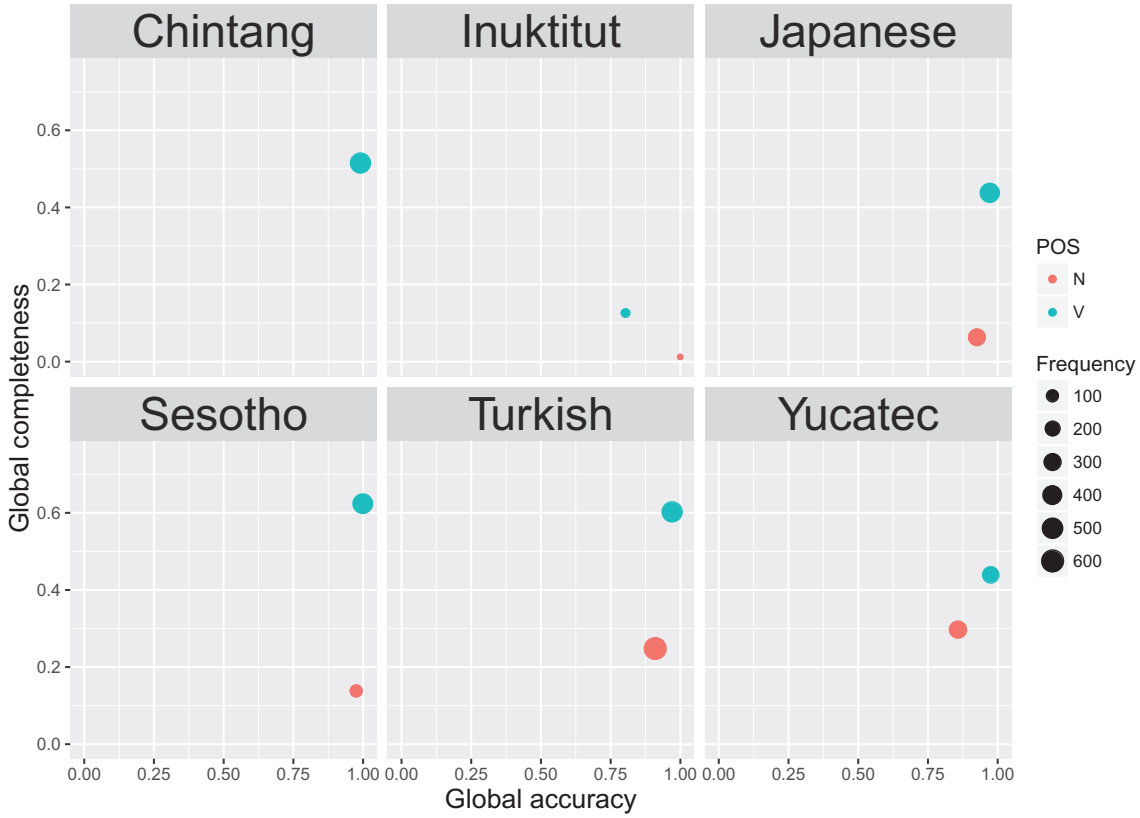

Figure 8: Nouns and verbs by morpheme frames

Figure 9 shows global accuracy and completeness for all parts-of-speech at the morpheme level. Except for nouns and verbs, there is no discernible pattern.

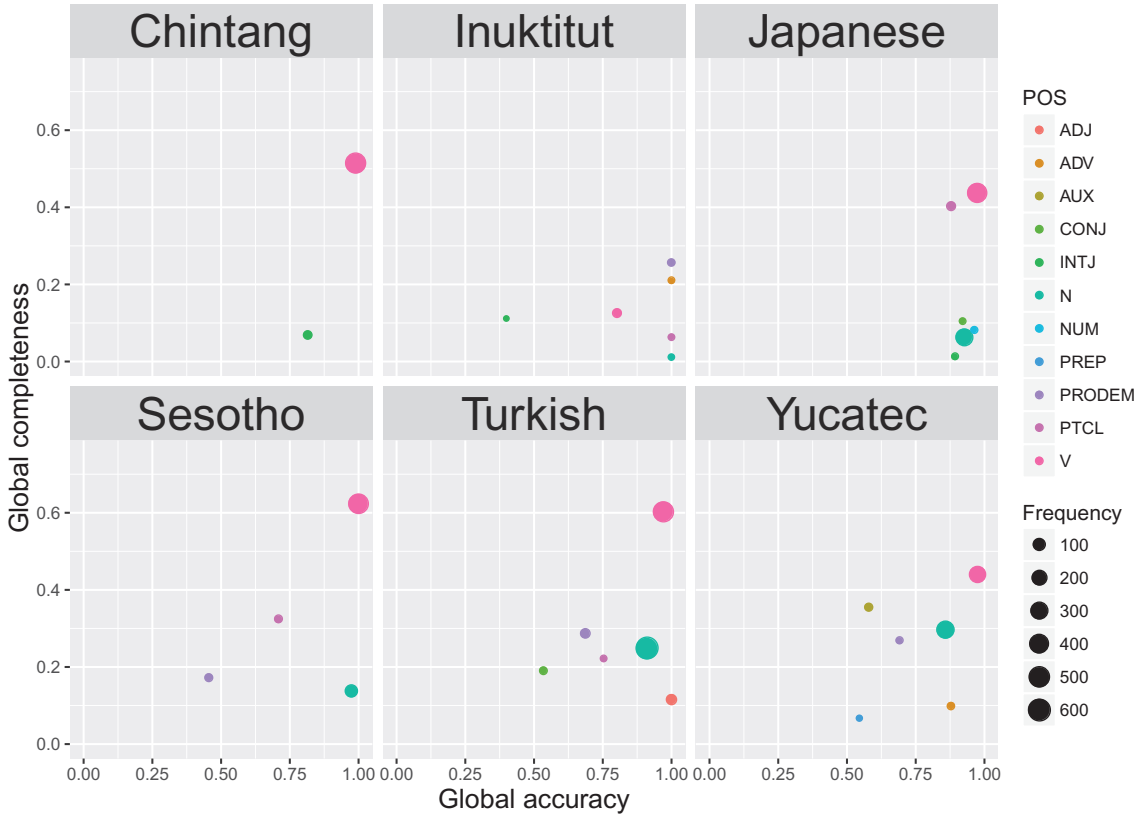

Figure 9: POS categories in morpheme frames

Figure 10 shows global accuracy and completeness scores for noun and verb word frames. There is no discernible pattern and neither is particularly accurate, except perhaps Japanese, which also happens to be the phonologically most impoverished language in our sample.

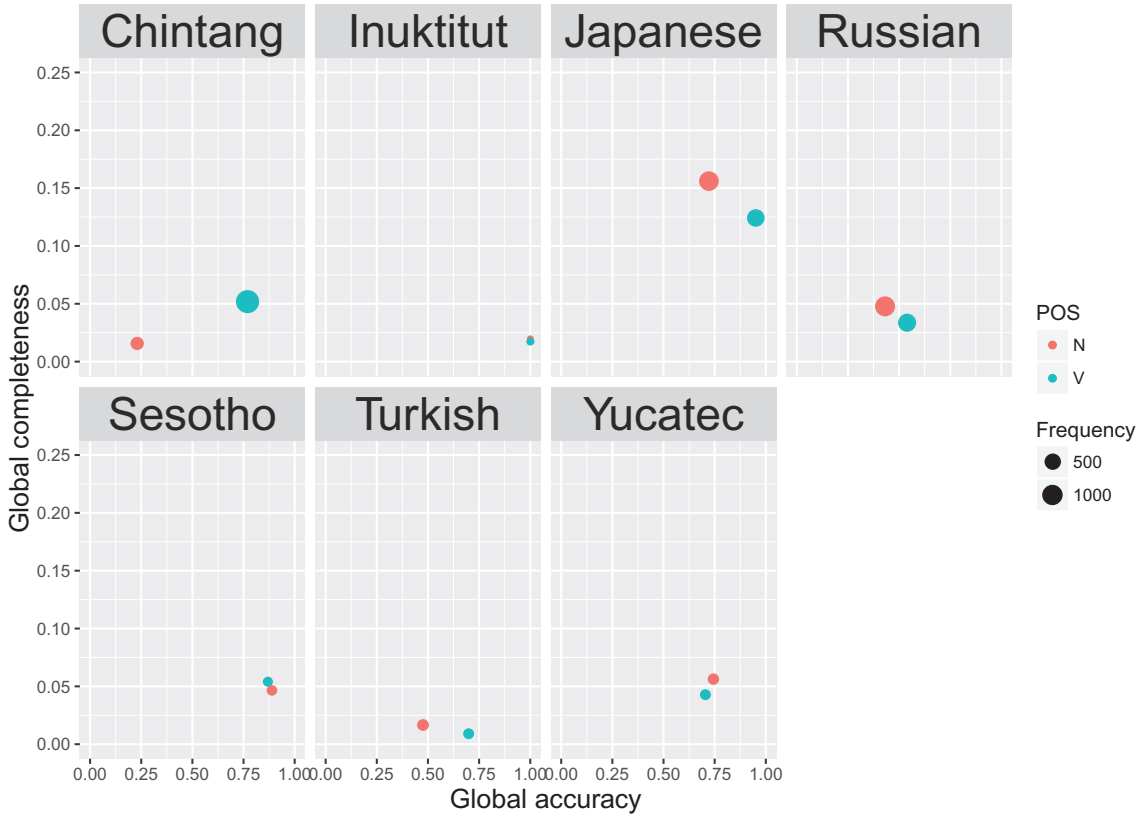

Figure 10: Nouns and verbs by word frames

Figure 11 shows global accuracy and completeness scores for parts-of-speech in word level frames (relative frequency threshold). Parts-of-speech frames classified by modal categories are similar to those with high precision, i.e. Inuktitut, Japanese, Sesotho and Yucatec. Frequent frames are most accurate and frequent in Japanese, which may be expected because its phonology contains few phonemes and syllable types (at least much less than the rest of the languages in this language sample).

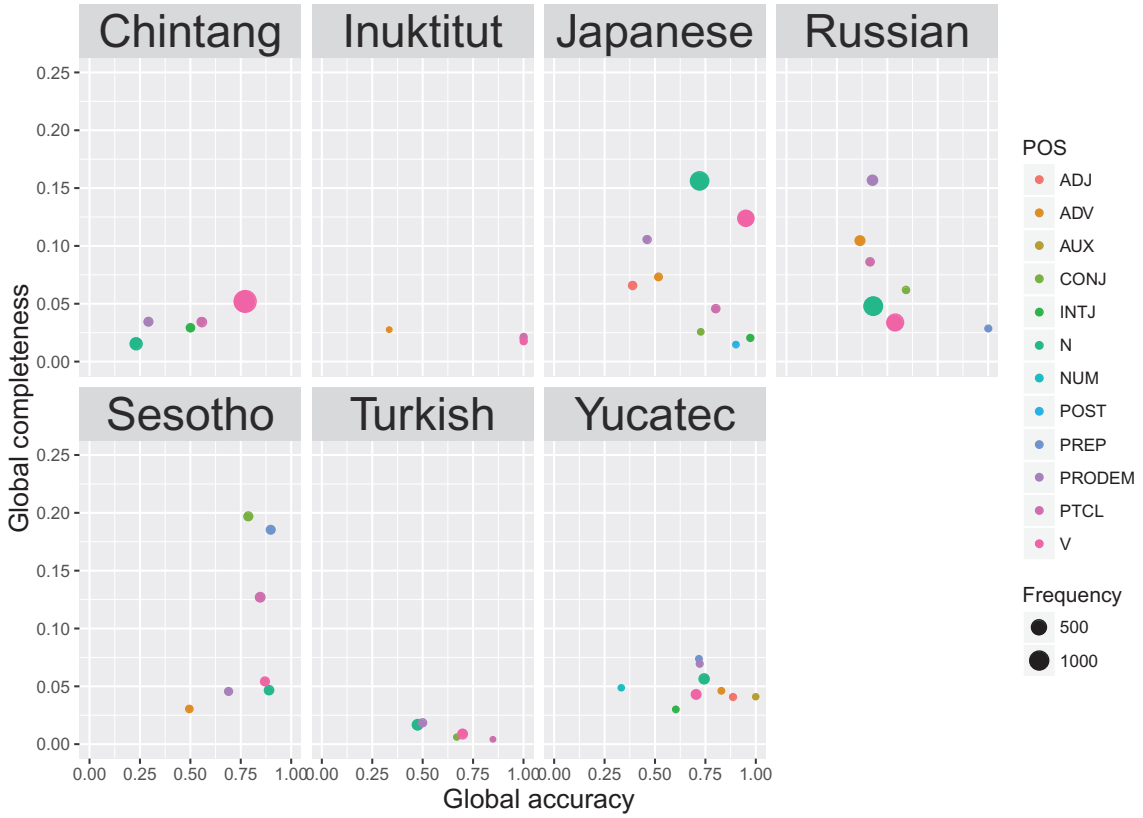

Figure 11: POS categories in word frames

Tables 15 and 16 provide the raw global accuracy and completeness scores per grammatical category in each language in our sample.

	Corpus	POS	Accuracy	Completeness	Frames
1	Japanese	ADJ	0.39	0.07	41
2	Yucatec	ADJ	0.89	0.04	13
3	Inuktitut	ADV	0.33	0.03	2
4	Japanese	ADV	0.52	0.07	26
5	Russian	ADV	0.36	0.10	98
6	Sesotho	ADV	0.50	0.03	21
7	Yucatec	ADV	0.83	0.05	8
8	Yucatec	AUX	1.00	0.04	5
9	Japanese	CONJ	0.73	0.03	7
10	Russian	CONJ	0.59	0.06	17
11	Sesotho	CONJ	0.79	0.20	62
12	Turkish	CONJ	0.67	0.01	4
13	Chintang	INTJ	0.50	0.03	45
14	Japanese	INTJ	0.97	0.02	14
15	Yucatec	INTJ	0.60	0.03	7
16	Chintang	N	0.23	0.02	234
17	Inuktitut	N	1.00	0.02	6
18	Japanese	N	0.72	0.16	893
19	Russian	N	0.43	0.05	937
20	Sesotho	N	0.89	0.05	81
21	Turkish	N	0.48	0.02	139
22	Yucatec	N	0.75	0.06	120
23	Yucatec	NUM	0.33	0.05	5
24	Japanese	POST	0.90	0.01	6
25	Russian	PREP	1.00	0.03	11
26	Sesotho	PREP	0.90	0.19	59
27	Yucatec	PREP	0.72	0.07	8
28	Chintang	PRODEM	0.29	0.03	61
29	Inuktitut	PRODEM	1.00	0.02	6
30	Japanese	PRODEM	0.46	0.11	35
31	Russian	PRODEM	0.43	0.16	124
32	Sesotho	PRODEM	0.69	0.05	33
33	Turkish	PRODEM	0.50	0.02	37
34	Yucatec	PRODEM	0.72	0.07	11
35	Chintang	PTCL	0.56	0.03	81
36	Inuktitut	PTCL	1.00	0.02	11
37	Japanese	PTCL	0.80	0.05	41
38	Russian	PTCL	0.41	0.09	44
39	Sesotho	PTCL	0.85	0.13	91
40	Turkish	PTCL	0.85	0.00	2
41	Chintang	V	0.77	0.05	1447
42	Inuktitut	V	1.00	0.02	13
43	Japanese	V	0.95	0.12	628
44	Russian	V	0.54	0.03	690
45	Sesotho	V	0.87	0.05	62
46	Turkish	V	0.70	0.01	95
47	Yucatec	V	0.70	0.04	96

Table 15: Global accuracy and completeness of word frames

	Corpus	POS	Accuracy	Completeness	Frames
1	Turkish	ADJ	1.00	0.12	51
2	Inuktitut	ADV	1.00	0.21	4
3	Yucatec	ADV	0.88	0.10	9
4	Yucatec	AUX	0.58	0.36	16
5	Japanese	CONJ	0.92	0.11	4
6	Turkish	CONJ	0.53	0.19	12
7	Chintang	INTJ	0.82	0.07	22
8	Inuktitut	INTJ	0.40	0.11	1
9	Japanese	INTJ	0.89	0.01	5
10	Inuktitut	N	1.00	0.01	3
11	Japanese	N	0.93	0.06	291
12	Sesotho	N	0.97	0.14	105
13	Turkish	N	0.91	0.25	601
14	Yucatec	N	0.86	0.30	312
15	Japanese	NUM	0.96	0.08	6
16	Yucatec	PREP	0.55	0.07	2
17	Inuktitut	PRODEM	1.00	0.26	10
18	Sesotho	PRODEM	0.45	0.17	14
19	Turkish	PRODEM	0.69	0.29	38
20	Yucatec	PRODEM	0.69	0.27	7
21	Inuktitut	PTCL	1.00	0.06	4
22	Japanese	PTCL	0.88	0.40	27
23	Sesotho	PTCL	0.71	0.33	13
24	Turkish	PTCL	0.75	0.22	4
25	Chintang	V	0.99	0.51	479
26	Inuktitut	V	0.80	0.13	26
27	Japanese	V	0.97	0.44	422
28	Sesotho	V	1.00	0.62	448
29	Turkish	V	0.97	0.60	479
30	Yucatec	V	0.98	0.44	262

Table 16: Global accuracy and completeness of morpheme frames

5.6 Utterance boundaries

Utterance boundaries are a salient feature in the input to the language learner. Figures 12 and 13 show the frequency of parts-of-speech in utterance initial and final positions in the seven languages in our sample. The typological features of these languages differ and they do not necessarily correlate with frequency of part-of-speech, e.g. word order (see Table 5 in the main text), which raises the question: what roles do frequency and position salience play in language acquisition? Consider the study by Longobardi et al. (2015), who show that early vocabulary development of nouns by Italian-speaking children is facilitated by the mother's use of nouns in utterance final position – even as verbs are more frequent in their child-directed speech.

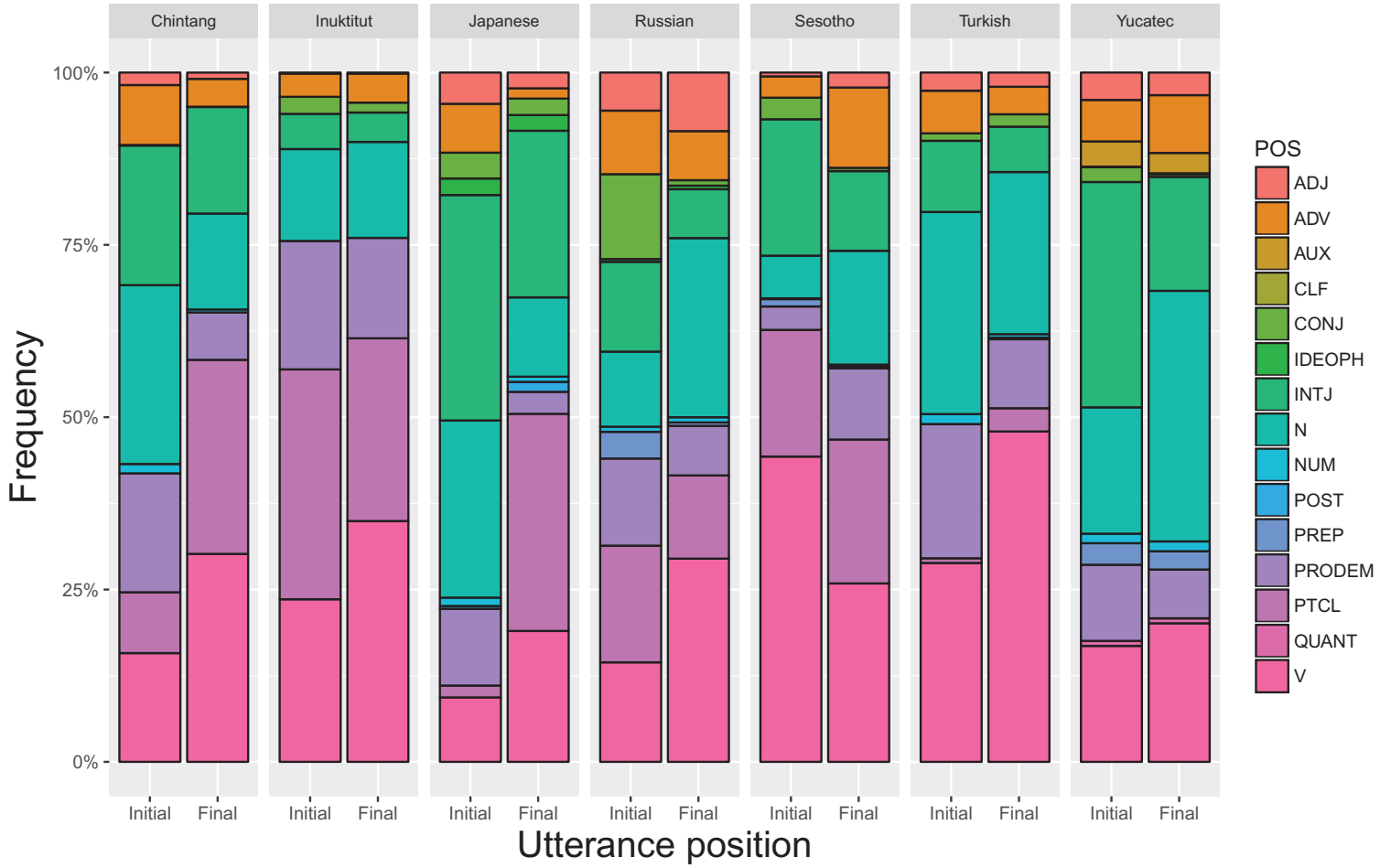

Figure 12: Utterance boundaries: all parts-of-speech

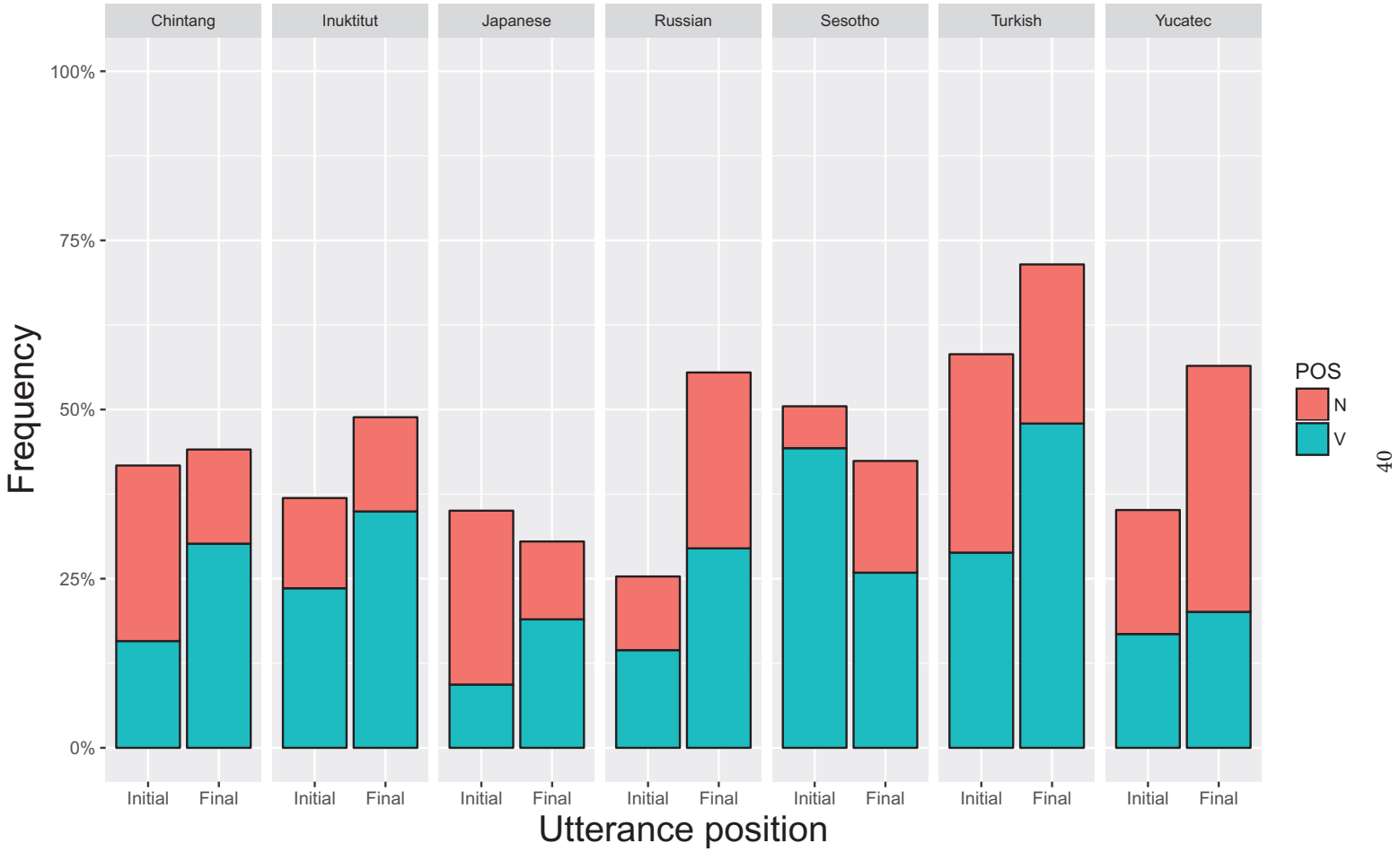

Figure 13: Utterance boundaries: nouns and verbs