Supplementary materials

S1: Word length distributions

Table 1 reports word lengths distributions in the 12 alphabetic languages considered in this study: Dutch (labeled du), English (en), Estonian (ee), Finnish (fi), German (ge), Greek (gr), Hebrew (he), Italian (it), Norwegian (no), Russian (ru), Spanish (sp), and Turkish (tr). Word length data come from the fasttext corpus, is based on word types (unique occurrences of words) and is presented for lengths 1-16 letters as percent of total. Languages are ordered from longer to shorter average word length. Estimates lower than 1% are removed.

lang	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
fi	5.20	8.67	5.38	8.12	10.47	11.54	8.74	9.11	7.47	6.52	5.60	3.60	2.68	2.01	1.32	1.03
ee	5.61	9.13	7.13	10.94	12.12	11.12	9.84	9.26	7.34	5.13	3.62	2.48	1.81	1.34		
tr	3.35	8.15	8.95	10.09	14.79	12.23	11.45	9.09	7.17	5.18	3.27	2.35	1.44			
ru	12.93	8.07	8.24	7.21	10.43	10.84	10.24	8.30	6.71	5.38	3.85	2.73	1.72	1.10		
ge	3.77	7.39	24.50	11.41	10.43	9.24	6.93	5.74	5.04	4.02	3.26	2.28	1.67	1.15		
it	10.40	16.32	12.10	7.85	11.94	10.06	8.53	7.59	5.19	4.00	2.50	1.45				
gr	6.29	10.84	20.86	9.40	8.58	8.25	7.92	7.62	6.36	5.12	3.73	2.18	1.15			
du	4.27	18.03	19.09	12.57	8.24	9.16	6.63	5.64	4.86	3.53	2.49	1.92	1.10			
no	8.48	16.05	19.23	9.42	10.23	8.68	6.51	5.49	4.52	3.24	2.34	1.78	1.19			
en	6.68	15.91	18.72	16.47	10.94	8.60	7.80	5.59	3.79	2.49	1.42					
sp	7.58	23.46	13.42	7.95	9.92	8.55	8.47	6.72	5.02	3.52	2.29	1.28				
he	6.85	11.60	13.70	19.80	20.90	15.10	7.48	3.10	1.43							

Table 1

Word length distributions in percent (lengths 1-16), based on word tokens. Values below 1% are left blank.

S2: Partitioning data

A possible approach to establishing cross-linguistic differences in saccadic targeting is to parse existing data in a way that equated visual settings across languages. For this worked example, we created subsets of data from all languages representing unique combinations of launch site and word N+1 length. That is, each subset contained one-step saccades registered in all languages that were launched from the launch sites in

word N equally removed from word N+1 center and into the words of equal length. In each subset, we calculated the mean initial landing position for each language sample and ranked it from 1 (landing closest to the word beginning) to 12 (landing farthest into the word). If saccade planning is solely driven by the demands of its current sensory setting, we expect initial landing positions to be statistically identical across languages because they result from saccades produced in settings kept cross-linguistically constant. Conversely, if mean word length of the language exerts influence that is external to these settings, we might expect languages with longer words like Finnish to produce rankings close to 12 and those with shorter words like Hebrew or Greek close to 1. Language samples with a relatively small number of participants (e.g., Turkish) or very short words (e.g., Hebrew) contributed fewer observations to some of the matched subsets. In the estimation of ranks, we only considered the 581 subsets in which all language samples were represented by at least 50 observations and calculated the mean rank for each language. The correlation between the language's mean rank of the initial landing position and mean word length was positive and marginally significant (r=0.563, p=0.0566).

Mean rank of Finnish was 10.4 and those of Hebrew and Greek were 4.48 and 2.25, out of 12. The correlation indicates that readers of, say, Finnish tend to consistently land farther into Finnish words than, say, readers of Greek or Hebrew into words in respective languages, even when the theoretically critical visual parameters for saccade planning are kept constant across all these languages through data partitioning. This suggests – for the specific case of one-step forward saccades – that one of the premises of the discrete account is more accurate than that of the dynamic one. Namely, landing positions of initial saccades reflect global statistics of the word length distributions in specific languages, over and above ever-changing transitory conditions under which saccadic targets are selected and saccades are planned and executed.

When based on 12 observations only (one mean value per language), the Pearson's correlation has an 80% power to detect correlations with $r \geq 0.718$ at the nominal 0.05 significance level. The planned addition of language samples to the MECO database will

increase the power and verify whether the observed correlation is statistically significant.

S3: Refixation probabilities for skipping saccades

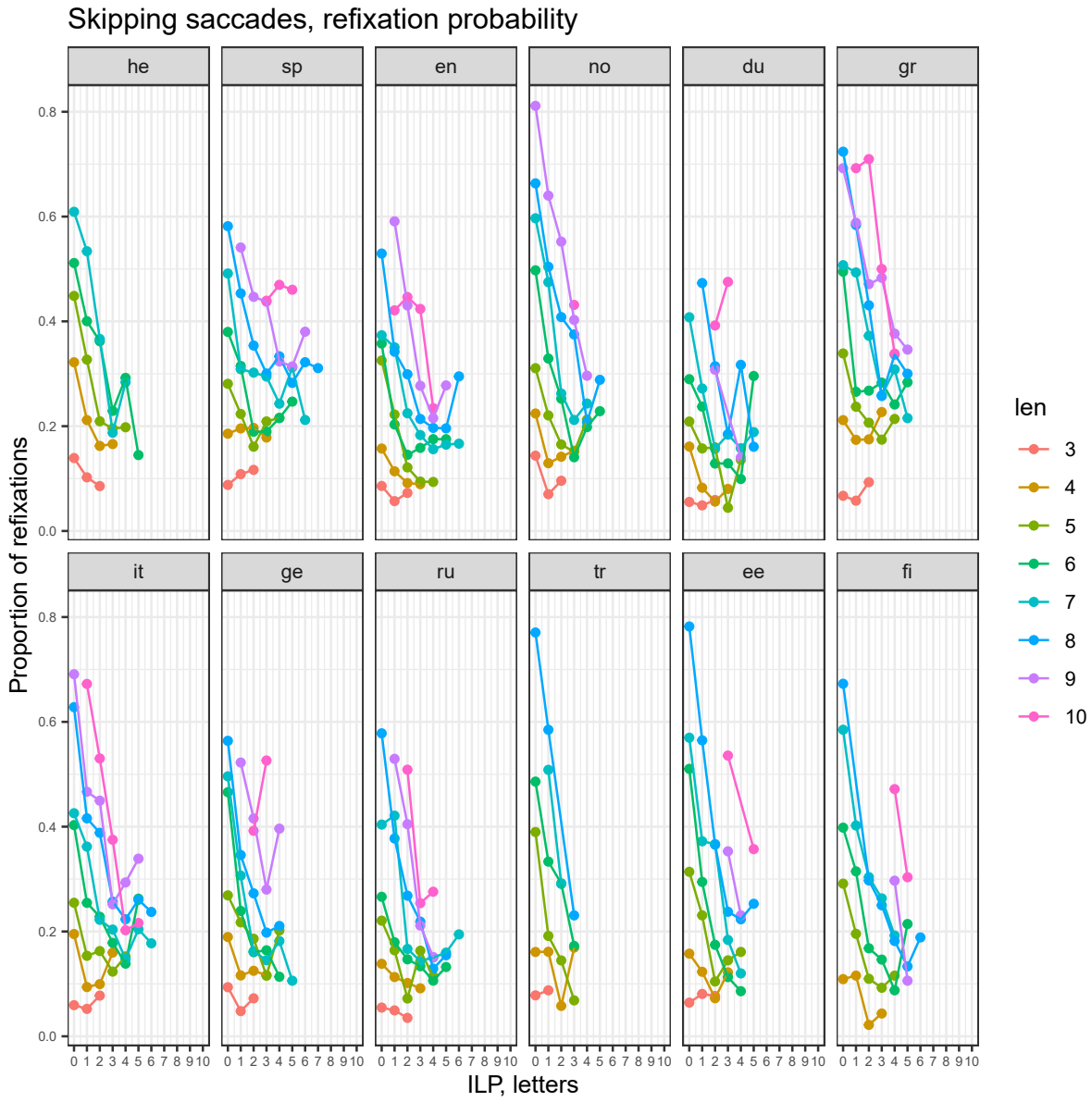

Figure 1. Refixation probability as a function of initial landing position of skipping saccades in 12 alphabetic languages for 3-10 letter words.