<u>Supplementary materials</u>

1. Additional controls for simulation 1 using artificial words

1a. In additional simulations using the coding scheme of Simulation 1, we examined whether other manipulations which favor relying upon identity information relative to position information produce similar results to the ones reported in the main text. Specifically, we either: a) multiplied the number of identity units in the input layer by 5 (such that the identity information is repeated 5 times compared to the single copy of the position information, corresponding to the scaling of the position information by 0.2 used in the main text); b) multiplied the number of hidden units in the representation layer of the identity pathway by 5, thus increasing its influence on the second hidden layer compared to the representations of position; c) directly injected random Gaussian noise to the activation of each unit in the positional input layer, under the constraint that activation could not go below 0. We have found that all of these manipulations yield similar results to those reported in the paper.

1b. Omitting the two internal representations of position and identity prior to connecting to the integration layer (i.e., the simplest possible three-layer network, similar to that used in Simulation 2), we observed qualitatively similar results to those reported in the main text. But note that without those additional levels of representation, it is not possible to conduct the separate analysis of the internal representations of position and identity that are integral to understanding the operation of the model.

1c. To rule out the possibility that our results stem from differences in orthographic density in the artificial English and Hebrew vocabularies (given that the presence of anagrams reduces the variability of letters across the stimuli and thus increases the average orthographic density), we ran additional simulations in which the artificial English alphabet only contained 13 letters instead of 26. This simulation produced TL priming that was almost identical in magnitude to the one observed when using the English vocabulary with the original 26 letter alphabet. This indicates that it was the presence of anagrams in the Hebrew vocabulary that led to the differences in TL priming rather than a reduction in orthographic density.

2. Additional analyses and modeling of the effects of word frequency, noise, and sample size on transposed letter priming in the model

2a. As reported in the main text, examination of the descriptive statistics of the training samples in Simulation 1b, as well as their corresponding corpora, showed that the English and Hebrew stimuli differed in terms of word frequency. We therefore examined whether word frequency differences contributed to the observed cross-linguistic differences (the training of the network to the same homeostatic criterion in both languages notwithstanding). We correlated word frequency with the magnitude of TL facilitation using words within the typical frequency ranges used in human experiments (10-80 words per million). Given the wide range of frequency values and the paucity of data at the extreme ends, to increase the reliability of these analyses, multivariate outliers—the 5% of items with the highest Mahalanobis distance statistics—were first screened before calculating the correlations. These analyses revealed no significant relationship between frequency and TL effect magnitude either in English or in Hebrew (all correlation values close to 0). This result is consistent with findings of previous empirical investigations studying the relationship between TL effect magnitude and word frequency (Forster et al., 1987). Similar results were also obtained on the log-transformed frequency data with or without the outlier screening.

Interestingly, in exploratory analyses below the standard range of word frequencies used in psychological experiments, significant correlations were detected; that is, word frequency for relatively low frequency words (< 10 words per million) was positively correlated with the magnitude of the TL effect. No significant effects were detected in the analyses of the relatively high frequency (> 100 words per million) words, although there was a numeric trend towards a negative correlation. These results provide another interesting novel prediction for future empirical research to investigate and against which the model's performance can be evaluated, again highlighting the value of coordinated computational and empirical research.

2b. To examine and evaluate the effect of noise on our findings, we examined a modification of Simulation 1 that did not include any noise (i.e., that did not include scaling the output of the position input layer), and modifications of Simulation 2 that employed different amounts of noise (i.e., smaller standard deviations, corresponding to values equal to ¼ and ½ of the range of the Gaussian). The TL priming effects and Euclidian distances for the extensions of Simulation 1 are presented in Figure 1 and Figure 2, respectively. The analogous plots for the extensions of Simulation 2 are presented in Figure 3 and Figure 4. For convenience, the original results appearing in the main text are presented as well.

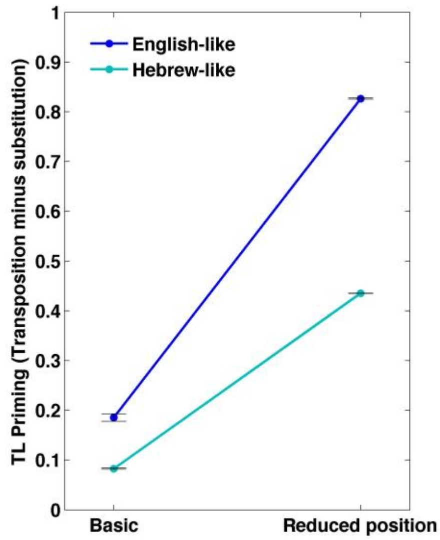

Figure 1. *TL priming effects in Simulation 1 trained on artificial words, for the basic ("no noise") and reduced-position ("noisy") conditions.*

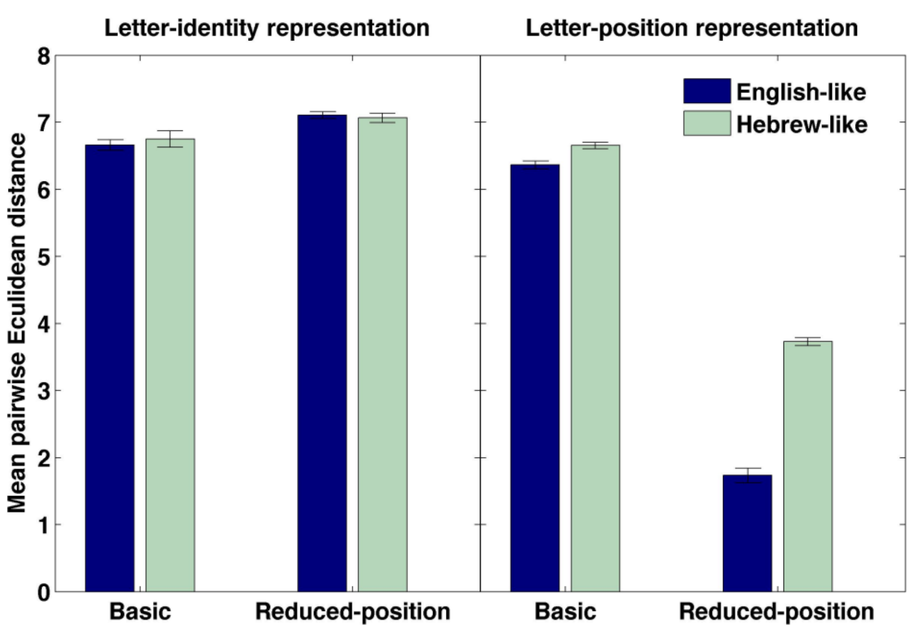

Figure 2. *Mean pairwise Euclidean distances between the representations formed for identity and position. Results for the basic ("no noise") and reduced-position ("noisy") network are presented as separate bars.*

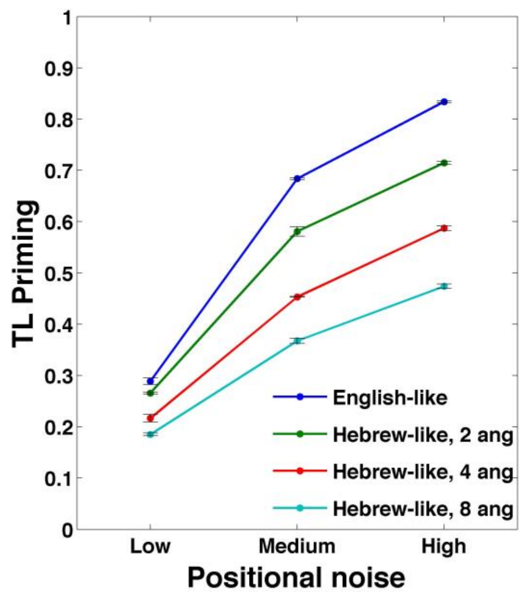

Figure 3. *TL priming effects in Simulation 2 trained on artificial words with various noise levels (reflecting the variance of the Gaussian activation function). 'ang': number of anagrams per letter string in the Hebrew-like condition*

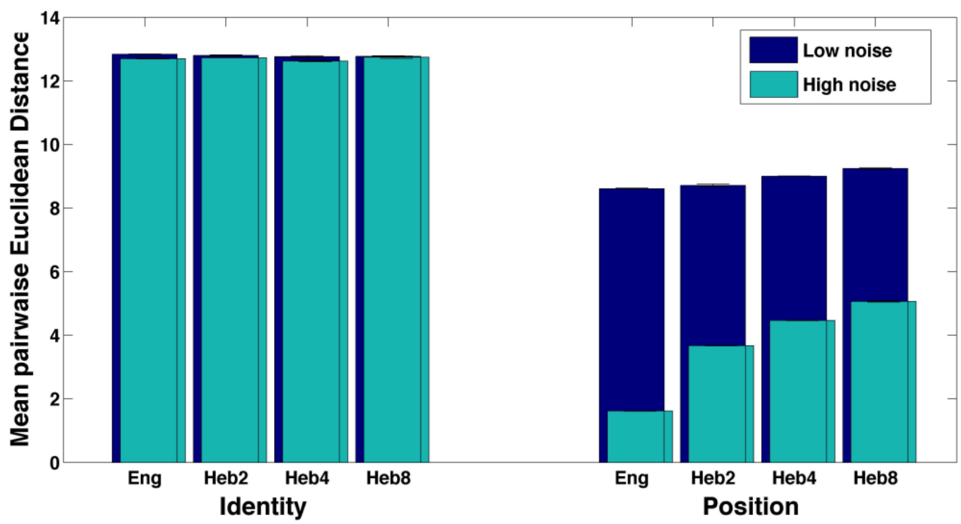

Figure 4. *Mean Euclidean distances between pairs of hidden representations of non-words differing from each other by letter position or letter identity in Simulation 2 using artificial words. Results are presented for the high noise simulation (i.e., the noise level used in the paper) alongside the lowest noise value we simulated here. The number besides the label 'Heb' refers to the number of anagrams per letter string in the Hebrew-like condition.*

As can be seen above, all the effects presented in the paper are still apparent, but are reduced in strength. Specifically, the difference in the network performance between the two languages is reduced mostly due to the English condition behaving similarly to the Hebrew condition and relying more on position information. In other words, when positional noise is minimized, both languages become sensitive to the position of letters to a more similar degree, although a slightly higher reliance on position is still apparent in Hebrew. This highlights the importance of the assumption of neurobiological noise, which is widely accepted in models of orthographic processing, in generating TL effects when combined with a learning model.

2c. As reported in the main text, the corpora from which our training samples were taken included a significantly larger number of items in Hebrew than in English (61,383 compared to 17,530, respectively). Although in all simulations using real words we chose an equal number of items from each corpus to equate conditions between the two linguistic conditions, an interesting question that arises is whether our results would be affected had the number of items chosen for training for each language was proportional to their corpus sizes; and how, in general, are TL effects influenced by sample size, all else being equal. To explore this issue, we re-run Simulation 2b using smaller sample sizes for each language. Specifically, we ran the simulation with either 3,000 or 5,000 words instead of the 10,000 reported in the main text (choosing a sample size smaller than 3,000 – which is equal to the number of units in the hidden layer – does not sufficiently restrict the mapping task that the network needs to perform, thus leading to transposed letters having no effect on recognition in both languages and consequently to strong ceiling effects when computing TL priming; conversely, a sample size larger than 10,000 does not allow the network to learn the mapping in reasonable time). Results of these simulations, alongside the original ones with 10,000 items, are presented in Table 1 and Figure 5.

		3,000 words	**5,000 words**	**10,000 words**
English	**Letter transposition**	0.99	0.97	0.9
	1-Letter substituaion	0.21	0.15	0.07
Hebrew	**Letter transposition**	0.87	0.82	0.68
	1-Letter substitution	0.13	0.09	0.05

Table 1 *Effects of letter transposition and 1-letter substitution on word recognition for English and Hebrew as a function of training-sample size (based on the network presented in Simulation 2b)*

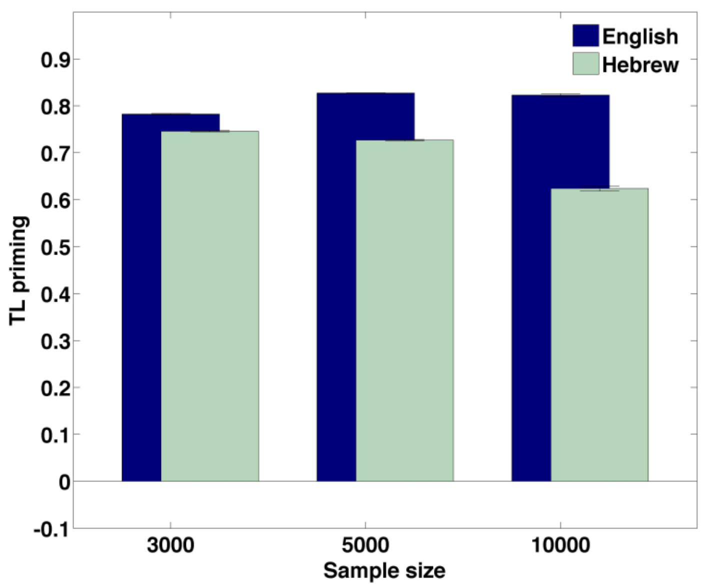

Figure 5 *TL priming effects as a function of sample size (using the network presented in Simulation 2b)*

As can be seen, whereas TL priming effects in English show a small increase, TL effects in Hebrew display the opposite pattern. The decrease in the TL priming effect for Hebrew is expected since larger sample sizes increase the competition between representations, forcing the network to accommodate its mapping based on more features of the input, including letter position; thus, transposing letters in a word is more likely to distort its correct mapping for larger sample sizes. This effect is also apparent for letter substitution, but to a lesser extent since letter substitution almost completely abolish the mapping for any sample size, leading to a net decrease in TL priming. While the same effects are also prominent in English, notice that in English the letter transposition effect is close to ceiling levels for the smaller sample sizes whereas letter substitution is not; therefore, the net effect on TL priming is a small increase. To summarize, the cross-linguistic TL effects are evident in the model regardless of sample size, but, as can be expected, are higher for large sample sizes where competition between representations is stronger and no ceiling effects distort the effects. Note also that the same is true when comparing the small sample size in English to the large sample size in Hebrew, mimicking the original differences in corpora sizes (which is about three-fold in favor of Hebrew).

3. Detailed analyses of the weights formed during training of the network in Simulation 1, and other similarity metrics

In an attempt to further understand the network behavior, we examined the strengths of the connection weights formed during training between the input layer and the representation layer in Simulation 1. We observed that the mean magnitude of these weights, as well as the outgoing weights connecting the representation layer to the second hidden layer, showed similar trends to those observed when comparing the distance between the internal representations (this was true for stress as well; see Plaut, 1997). In other words, the magnitude of the weights connected to the identity representation layer were roughly equal for both languages, whereas the weights connected to the position representation layer were considerably smaller in English compared to Hebrew. Collectively, this suggests that relative to the network trained on English, the network trained on Hebrew tries from the outset of processing a word to receive more information from the position input layer and uses this information to drive hidden unit activations to more extreme values. This leads to the greater separation between the internal representations reported in detail in the main text, and, finally, causes these differences to propagate forward more strongly, all else being equal. Consequently, our findings are true not only for the specific analyses we report based on unit activations, but are also true at the various other steps that mediate between a network's input and its ultimate output, as well.

4. Statistical analysis of the results

As noted in footnote 6 in the main text, standard errors of all of the reported effects in the current work where miniscule compared to the effects sizes, obviating the need for detailed statistical analysis. This situation is typical to network models of human behavior since the question of whether a particular effect reaches significance or not can be controlled by changing the noise parameters and/or the number of items that are averaged across, and these are indeed regularly chosen to reduce unwarranted variability (something which cannot be done as easily in human experiments). This is clearly demonstrated by the tiny error bars of each of the effects presented in the Figures throughout the current work. One exception to this rule are the interaction effects between word-length and language in Simulations 1b and 2b (Figures 5 and 9), whose significance may not be as evident to the reader. To confirm the validity of these results, the TL effects were subject to a mixed-effect two-way ANOVA, with word-length (ranging from 4 to 10) as a repeated-measures factor and language (English, Hebrew) as a between-group factor. For Simulation 1b, this analysis has yielded highly significant main effects for word-length ($F(6,12) = 66.67$, $p<0.001$) and language ($F(1,2) = 5594.79$, $p<0.001$), as well as a highly significant interaction ($F(6,12) = 13.92$, $p<0.001$). Likewise, these effects were highly significant in the analysis performed for the TL effects of Simulation 2b ($F(6,12) = 718.95$, $p<0.001$; $F(1,2) = 3419.48$, $p<0.001$; $F(6,12) = 346.37$, $p< 0.001$, for the main effects of word-length and language, and the interaction, respectively). Therefore, all effects reported in the paper, including the aforementioned interactions, are highly significant and represent faithful outcomes of the model.