

Errata

Publication I

Equation 2 has incorrect signs in the final publication. The correct form is $-g_i(X) + \log \sum_{j \neq i} \exp(g_j(X))$.

Publication III

The Finnish NNLM models were based on subword units and the Estonian on compound-split words, while few sentences in the published paper in Section 3.4 erroneously claim the opposite.

Publication IV

Kneser–Ney smoothing was not used when training the class-based n-gram models, because the class n-gram statistics were not suitable for the Modified Kneser–Ney implementation. While this is not said in the article, the effect of smoothing is not that significant in class-based models, because the data is not as sparse.

Publication V

A highway network layer uses a separate bias b_σ for its gate. The index σ is missing from Equation 6 in the published paper. The correct form is $g(x_t) = \sigma(W_\sigma x_t + b_\sigma)$.