Supplementary material for "Why are some word orders more common than others? A uniform information density account"

Anonymous Author(s)
Affiliation
Address
email

1 Toy world probability distribution

Figure 1 shows the event distribution E for the toy world presented in Section 2 of the paper, which was used to generate the mean entropy trajectories and deviation scores shown in Figure 1.

	EAT				DRINK			
	Apple	Bread	Cake	Rice	Coffee	Cola	Juice	Water
Alice	0.05	0.00	0.03	0.02	0.07	0.03	0.00	0.00
Bob	0.02	0.00	0.04	0.04	0.02	0.04	0.02	0.02
Eve	0.00	0.01	0.00	0.9	0.03	0.01	0.00	0.06
Mallory	0.04	0.04	0.01	0.01	0.00	0.01	0.9	0.00
Trent	0.02	0.00	0.01	0.07	0.02	0.03	0.03	0.02

Figure 1: Diagrammatic representation of event distribution E for toy world. Rows represent people and columns represent items of food or drink. The leftmost block shows the probabilities of eating the various food items and the rightmost block shows the probability of drinking various items.

2 Details of Bayesian Thurstonian model

The raw data collected from our experiment is a set of integer values C_{ij}, where C_{ij} counts the number of participants who judged the event e_i to be more probable than e_j , so that $0 \leq C_{ij} \leq 3$. To transform these values into a probability distribution E, we use a simple Bayesian Thurstonian model presented in [1].

The core idea of this model is the following generative model for the values C_{ij}: we assume that each of the events e_i has a normal distribution associated with it, with mean $0 \leq \mu_i \leq 1$ and variance

σ_i^2. When a participant is asked to which of the events e_i and e_j is the most probable, they sample $s_i \sim N(\mu_i, \sigma_i^2)$ and $s_j \sim N(\mu_j, \sigma_j^2)$ and respond that event e_i is more probable if $s_i > s_j$, and that event e_j is more probable otherwise. Under this generative model, the values of C_{ij} are binomially distributed variables, with $C_{ij} \sim B(3, p_{ij})$, where the "success" probability p_{ij} is determined by μ_i, σ_i^2, μ_j and σ_j^2 according to the following equation:

$$p_{ij} = \Phi\left((\mu_i - \mu_j)/\sqrt{\sigma_i^2 + \sigma_j^2}\right), \qquad (1)$$

where Φ is the cumulative density function of the standard normal distribution $N(0, 1)$.

By placing a prior probability distribution over the set of all parameters, μ_i and σ_i^2 for $i = 1, 2, \ldots$, we can use Bayes' Law to compute a posterior probability over parameter values based on the values of C_{ij} yielded by our experiment. We place a component-wise prior on the parameters such that $P(\mu_i, \sigma_i^2) \propto \exp -\sigma_i^2$, with $P(\mu_i, \sigma_i^2) = 0$ if $\mu_i < 0$ or $\mu_i > 1$.

We perform the Bayesian inference numerically using a Metropolis Hastings algorithm to draw samples from the posterior distribution. Our proposal process for the MH algorithm selects a single parameter to change from a uniform distribution over all the parameters, and then proposes a new value for that parameter by sampling from a normal distribution centred on the parameter's current value. The normal proposal distributions are not truncated, with the requirement that $0 \leq \mu_i \leq 1$ is enforced by the prior. To obtain an estimate of the values of μ_i, we take 10 samples from each of 10 randomly initialized chains, for a total of 100 samples, with a lag of 1000 iterations between samples. Each chain is allowed to "burn in" for 15,000 iterations. This value was chosen by examining plots of the log posterior probability versus iterations and observing a plateau after around 15,000 iterations.

At the end of this process we have recovered a value of μ_i for each of the events in our experimentally defined world. These values are transformed into an event distribution E via the straightforward process of setting each event's probability to be directly proportional to its value of μ_i. This is the distribution used to produce the mean entropy trajectories and deviation scores shown in Figure 4.

References

[1] B. Miller, P. Hemmer, M. Steyvers, and M.D. Lee. The wisdom of crowds in rank ordering problems. In A. Howesa, D. Peebles, and R. Cooper, editors, *9th International Conference on Cognitive Modeling*, 2009.