In numbers,

$$\hat{\beta}_j = (0.3, 0.6, 0.04, 0.06)^\top. \tag{10}$$

The posterior probability that the native encountered by the linguist is a northerner, for example, is 34%. Wagner's notation is completely different and never specifies or provides the joint probabilities, but I hope the reader appreciates both the analogy to (2) underlined by this notation as well as its efficiency in delivering a correct PME solution for us. The solution that Wagner attributes to PME is misleading because of Wagner's Dempsterian setup which does not take into account that proponents of PME are likely to be proponents of the classical Bayesian position that type II prior probabilities are specified and determinate once the agent attends to the events in question. Some Bayesians in the current discussion explicitly disavow this requirement for (possibly retrospective) determinacy (especially James Joyce in [33] and other papers). Proponents of PME (a proper subset of Bayesians), however, are unlikely to follow Joyce—if they did, they would indeed have to address Wagner's example to show that their allegiances to PME and to indeterminacy are compatible.

That (9) follows from JUP is well-documented in Wagner's paper. For the PME solution for this problem, I will not use (9) or JUP, but maximize the entropy for the joint probability matrix M and then minimize the cross-entropy between the prior probability matrix M and the posterior probability matrix $\hat{M}$. The PME solution, despite its seemingly different ancestry in principle, formal method, and assumptions, agrees with (9). This completes our argument.

What follows may only be accessible to PME cognoscenti, since it involves the Lagrange multiplier method (see [12] (p.327ff) and [34] (p.244)). Others may read the conclusion and find a sketch for an easier, but much less rigorous proof in the appendix. To maximize the Shannon entropy of M and minimize the Kullback-Leibler divergence between $\hat{M}$ and M, consider the Lagrangian functions:

$$\Lambda(\mu_{ij}, \xi) = \sum_{\kappa_{ij}=1} \mu_{ij} \log \mu_{ij} + \sum_{j=1}^{n} \xi_j \left(\beta_j - \sum_{\kappa_{kj}=1} \mu_{kj} \right) + \lambda_m \left(x - \sum_{j=1}^{n} \mu_{mj} \right) \tag{11}$$

and

$$\hat{\Lambda}(\hat{\mu}_{ij}, \hat{\lambda}) = \sum_{\hat{\kappa}_{ij}=1} \hat{\mu}_{ij} \log \frac{\hat{\mu}_{ij}}{\mu_{ij}} + \sum_{i=1}^{m} \hat{\lambda}_i \left(\hat{\alpha}_i - \sum_{\hat{\kappa}_{il}=1} \hat{\mu}_{il} \right). \tag{12}$$

For the optimization, we set the partial derivatives to 0, which results in

$$\begin{aligned} M &= rs^\top \circ \kappa && (13)\\ \hat{M} &= \hat{r}s^\top \circ \hat{\kappa} && (14)\\ \beta &= S\kappa^\top r && (15)\\ \hat{\alpha} &= \hat{R}\kappa s && (16) \end{aligned}$$

where $r_i = e^{\zeta_i \lambda_m}, s_j = e^{-1-\xi_j}, \hat{r}_i = e^{-1-\hat{\lambda}_i}$ represent factors arising from the Lagrange multiplier method (ζ was defined in (5)). The operator $\circ$ is the entry-wise Hadamard product in linear algebra. $r, s, \hat{r}$ are the vectors containing the $r_i, s_j, \hat{r}_i$, respectively. $R, S, \hat{R}$ are the diagonal matrices with $R_{il} = r_i \delta_{il}, S_{kj} = s_j \delta_{kj}, \hat{R}_{il} = \hat{r}_i \delta_{il}$ (δ is Kronecker delta).