Notation:

- Denote the set of vocabulary by V and the target word by $t \in V$.
- Denote the number of occurrences of a word $w \in V$ in the context of t (i.e., in all sentences containing t) by n_w.
- Denote the number of occurrences of w in the entire corpus, i.e., the raw term frequency of w, by N_w.
- Let $T = \left(n_w + \frac{1}{2}\right)_{w \in V}$ be called the target vector and let $B = \left(N_w + \frac{1}{2}\right)_{w \in V}$ be called the background vector.
- Let $T' = \frac{1}{\|T\|} T$ and $B' = \frac{1}{\|B\|} B$, where $\|\cdot\|$ is the ℓ_1-norm, i.e., $\|(x_1, \ldots, x_n)\| = x_1 + \cdots + x_n$.
- Let C be the size of the corpus and let $p_w = \frac{N_w}{C}$.

Assume that the proportion p_w of word w is constant with respect to the size of the corpus, i.e., $\frac{\partial p_w}{\partial C} = 0$. Since it is empirically the case that $|V|$ is linear in C, i.e., $|V| = kC$, we have

$$\|B\| = \sum_{w \in V} \left(N_W + \frac{1}{2}\right) = \sum_{w \in V} \left(p_w C + \frac{1}{2}\right) = C \sum_{w \in V} p_w + \frac{1}{2}|V| = \left(1 + \frac{1}{2}k\right) C$$

$$\|T\| = \sum_{u \in V} \left(n_u + \frac{1}{2}\right) = \left(\sum_{u \in V} n_u\right) + \frac{1}{2}|V| = \tau + \frac{1}{2}kC,$$

where $\tau = \sum_{u \in V} n_u$ is the number of words (counting multiplicity) appearing in all sentences containing the target t. It is likely that $\frac{\partial \tau}{\partial C} > 0$ since $\frac{\partial n_w}{\partial C} > 0$ probably holds for almost all w. Then

$$\begin{aligned}
KL &= \sum_{w \in V} \left(T'_w \log \frac{T'_w}{B'_w} + B'_w \log \frac{B'_w}{T'_w} \right) \\
&= \sum_{w \in V} \left(\left(\frac{n_w + \frac{1}{2}}{\|T\|}\right) \log \frac{\left(\frac{n_w + \frac{1}{2}}{\|T\|}\right)}{\left(\frac{N_w + \frac{1}{2}}{\|B\|}\right)} + \left(\frac{N_w + \frac{1}{2}}{\|B\|}\right) \log \frac{\left(\frac{N_w + \frac{1}{2}}{\|B\|}\right)}{\left(\frac{n_w + \frac{1}{2}}{\|T\|}\right)} \right) \\
&= \sum_{w \in V} \left(\left(\frac{n_w + \frac{1}{2}}{\tau + \frac{1}{2}kC}\right) \log \frac{\left(\frac{n_w + \frac{1}{2}}{\tau + \frac{1}{2}kC}\right)}{\left(\frac{p_w C + \frac{1}{2}}{\left(1 + \frac{1}{2}k\right)C}\right)} + \left(\frac{p_w C + \frac{1}{2}}{\left(1 + \frac{1}{2}k\right)C}\right) \log \frac{\left(\frac{p_w C + \frac{1}{2}}{\left(1 + \frac{1}{2}k\right)C}\right)}{\left(\frac{n_w + \frac{1}{2}}{\tau + \frac{1}{2}kC}\right)} \right)
\end{aligned}$$

The problem is, how do n_w and τ vary (stochastically) with C?