Input: Source and target phrases f and e,
Source and target monolingual corpora C_f and C_e,
Phrase table pairs $T = \{(f^{(i)}, e^{(i)})\}_{i=1}^{N}$.
Output: Orientation features (p_m, p_s, p_d).

$S_f \leftarrow$ sentences containing f in C_f;
$S_e \leftarrow$ sentences containing e in C_e;
$(B_f, -, -) \leftarrow$ CollectOccurs$(f, \cup_{i=1}^{N} f^{(i)}, S_f)$;
$(B_e, A_e, D_e) \leftarrow$ CollectOccurs$(e, \cup_{i=1}^{N} e^{(i)}, S_e)$;
$c_m = c_s = c_d = 0$;
foreach *unique* f' *in* B_f **do**
    foreach *translation* e' *of* f' *in* T **do**
        $c_m = c_m + \#_{B_e}(e')$;
        $c_s = c_s + \#_{A_e}(e')$;
        $c_d = c_d + \#_{D_e}(e')$;
$c \leftarrow c_m + c_s + c_d$;
return $(\frac{c_m}{c}, \frac{c_s}{c}, \frac{c_d}{c})$

CollectOccurs(r, R, S)
    $B \leftarrow ()$; $A \leftarrow ()$; $D \leftarrow ()$;
    foreach *sentence* $s \in S$ **do**
        foreach *occurrence of phrase* r *in* s **do**
            $B \leftarrow B$ + (longest preceding r and in R);
            $A \leftarrow A$ + (longest following r and in R);
            $D \leftarrow D$ + (longest discontinuous w/ r and in R);
    return (B, A, D);
