

|                   | Monolingual training corpora |               |            | Spanish-English phrase table |           |
|-------------------|------------------------------|---------------|------------|------------------------------|-----------|
|                   | Europarl                     | Gigaword      | Wikipedia  | Phrase pairs                 | 3,093,228 |
| date range        | 4/96-10/09                   | 5/94-12/08    | n/a        | Spanish phrases              | 89,386    |
| uniq shared dates | 829                          | 5,249         | n/a        | English phrases              | 926,138   |
| Spanish articles  | n/a                          | 3,727,954     | 59,463     | Spanish unigrams             | 13,216    |
| English articles  | n/a                          | 4,862,876     | 59,463     | Avg # translations           | 98.7      |
| Spanish lines     | 1,307,339                    | 22,862,835    | 2,598,269  | Spanish bigrams              | 41,426    |
| English lines     | 1,307,339                    | 67,341,030    | 3,630,041  | Avg # translations           | 31.9      |
| Spanish words     | 28,248,930                   | 774,813,847   | 39,738,084 | Spanish trigrams             | 34,744    |
| English words     | 27,335,006                   | 1,827,065,374 | 61,656,646 | Avg # translations           | 13.5      |