

French Nouns EDA

Exploring linguistic feature encoding in French nouns using perceptron models and word embeddings, with a focus on understanding how these features emerge in simplified representations.



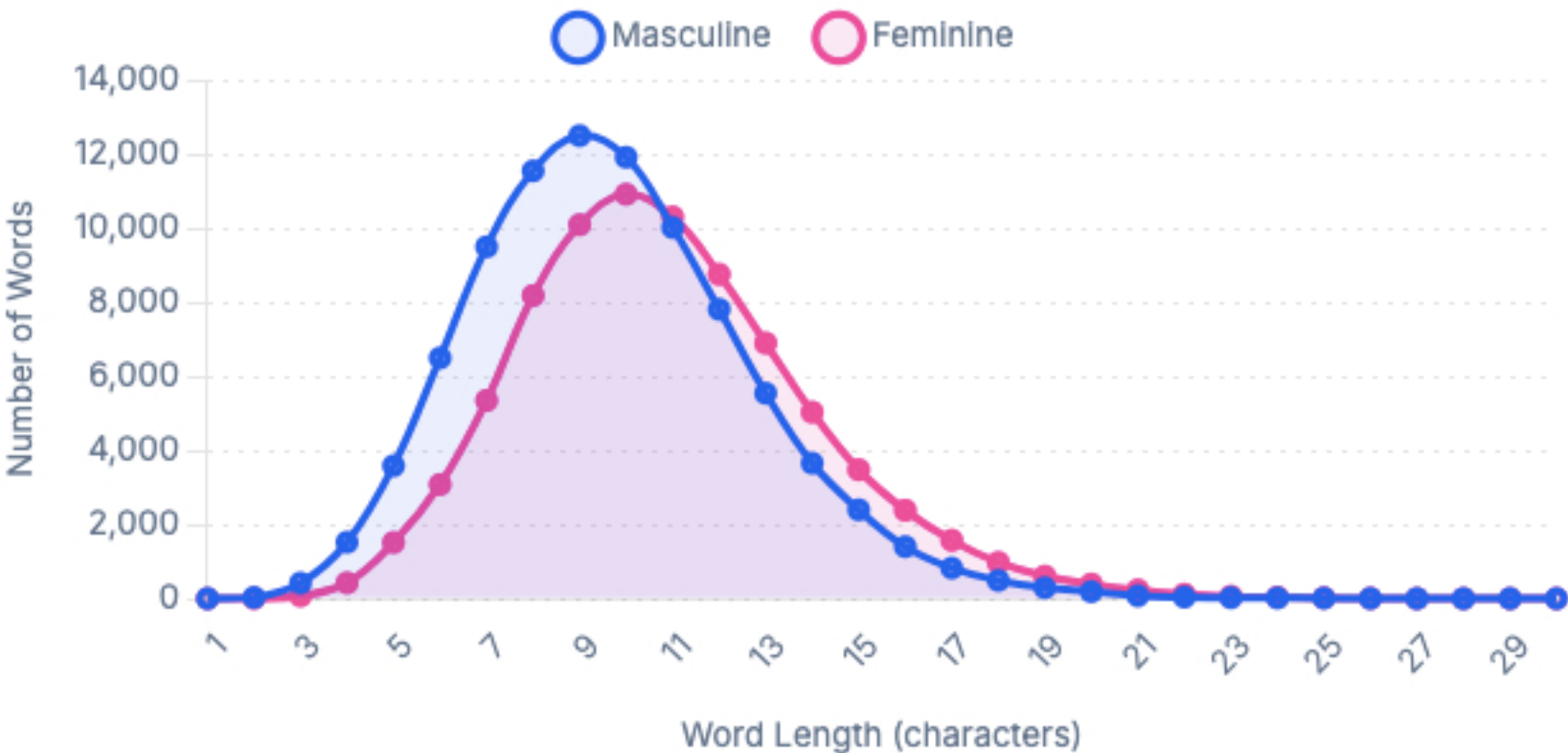
170,763
Total Words
Unique entries analyzed in dataset

10.25
Average Word Length
Characters per word (spaces removed)

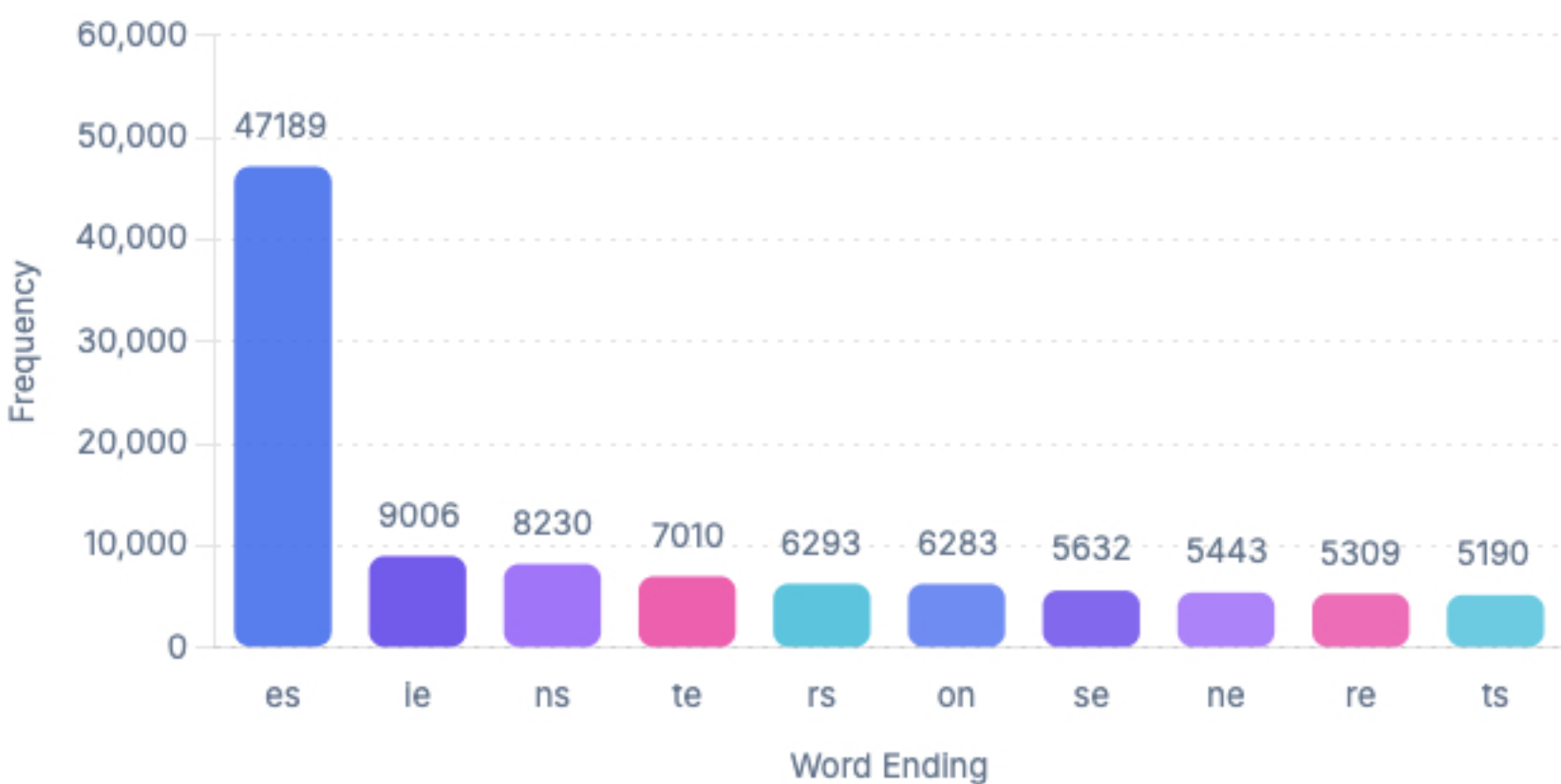
52.90
Masculine Ratio
Of total word corpus

92,636
Unique Lemmas
Base word forms identified

Gender Distribution by Word Length



Most Common Word Endings



Dataset Preview

| WORD | GENDER | NUMBER | LEMMA | LENGTH |
|-----------|-----------|----------|----------|--------|
| aabam | Masculine | singular | aabam | 5 |
| aalénien | Masculine | singular | aalénien | 8 |
| aaléniens | Masculine | plural | aalénien | 9 |
| abadie | Feminine | singular | abadie | 6 |
| abadies | Feminine | plural | abadie | 7 |
| abacas | Masculine | plural | abaca | 6 |
| abacule | Masculine | singular | abacule | 7 |
| abacules | Masculine | plural | abacule | 8 |
| abaddir | Masculine | singular | abaddir | 7 |
| abaddirs | Masculine | plural | abaddir | 8 |

Character-wise Analysis

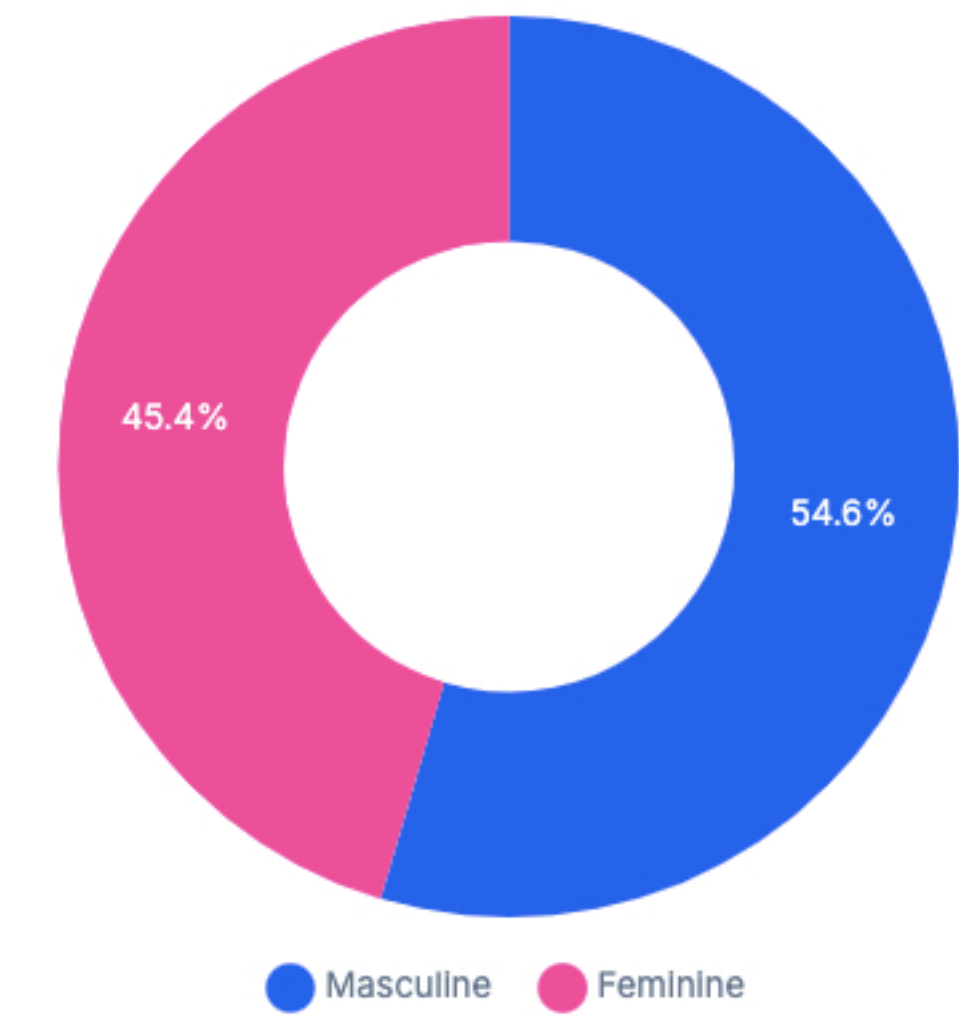


A

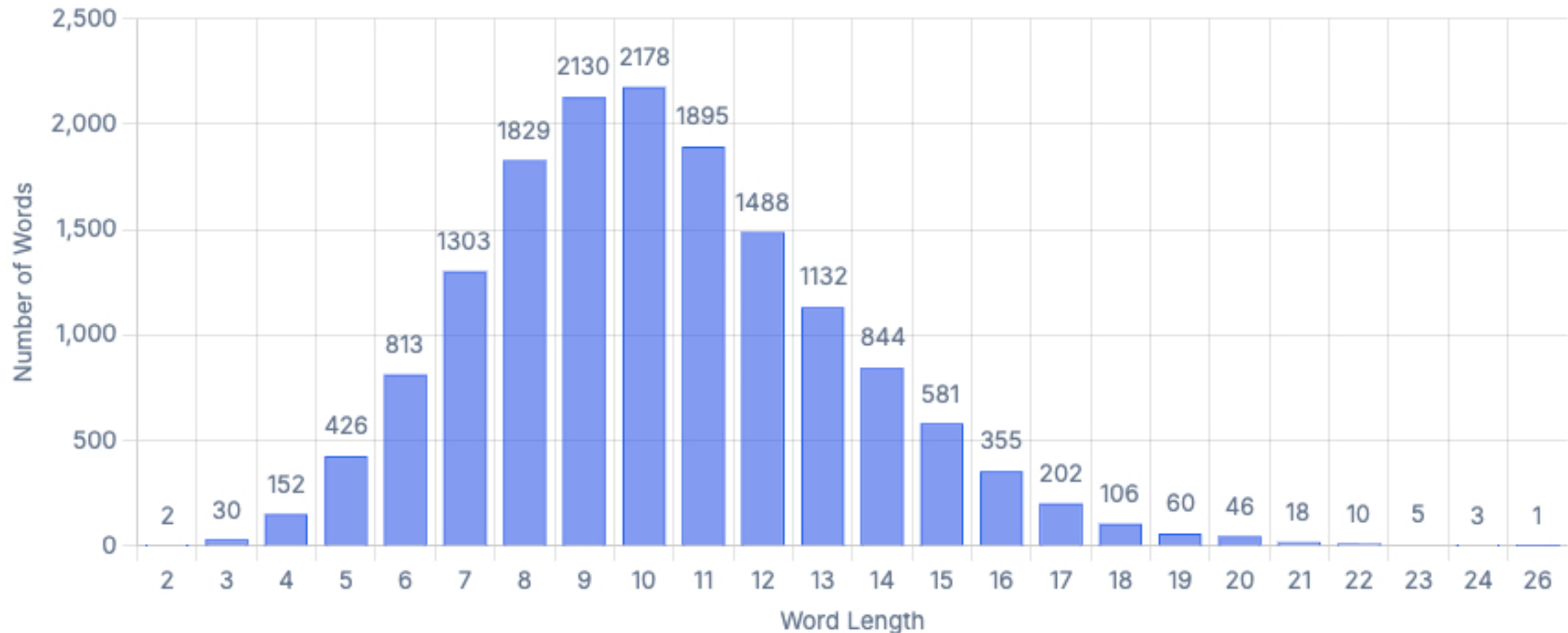
15,609 Words Total
8,524 Masculine (54.6%)
7,085 Feminine (45.4%)
Average length: 10.27 characters



Distribution



Length Analysis



Some Examples

| WORD | GENDER | NUMBER | LENGTH |
|-----------|-----------|----------|--------|
| aabam | Masculine | singular | 5 |
| aalénien | Masculine | singular | 8 |
| aaléniens | Masculine | plural | 9 |
| aba | Masculine | singular | 3 |
| abaca | Masculine | singular | 5 |
| abacas | Masculine | plural | 6 |
| abacule | Masculine | singular | 7 |
| abacules | Masculine | plural | 8 |