

**FREQUENCY DICTIONARY  
OF  
ITALIAN WORDS**



THE ROMANCE LANGUAGES AND THEIR STRUCTURES

DIRECTED BY ALPHONSE JUILLAND

*First Series 11*

---

FREQUENCY DICTIONARY  
OF  
ITALIAN WORDS

*by*

ALPHONSE JUILLAND

*Stanford University*

VINCENZO TRAVERSA

*California State University, Hayward*

*with the assistance of*

ANTONIO BELTRAMO and SEBASTIANO DI BLASI

*University of Montana*

*San John College, Philadelphia*



---

1973

MOUTON

THE HAGUE • PARIS

© Copyright 1973 in The Netherlands  
Mouton & Co. N.V., Publishers, The Hague

*No part of this book may be translated or reproduced in any form, by print, photoprint, microfilm,  
or any other means, without written permission from the publishers.*

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 77-169997

Printed in The Netherlands by Mouton & Co., The Hague

# CONTENTS

Foreword . . . . .	
INTRODUCTION	
1. PREVIOUS ATTEMPTS . . . . .	IX
1.1 Thompson, <i>A Study in Italian Vocabulary Frequency</i> (1927) . . . . .	IX
1.2 Knease, <i>An Italian Word List from Literary Sources</i> (1931) . . . . .	X
1.3 Skinner, <i>A Comparative Study of the Vocabularies of Forty-Five Italian Textbooks</i> (1935) . . . . .	XI
1.4 Migliorini, <i>Der Grundlegende Wortschatz des Italienischen</i> (1943) . . . . .	XI
2. SAMPLE . . . . .	XII
2.1 Plays . . . . .	XII
2.2 Fiction . . . . .	XIV
2.3 Essays . . . . .	XV
2.4 Newspapers and Magazines . . . . .	XVI
2.5 Technical and Scientific Writings . . . . .	XVII
3. INTERPRETATION . . . . .	XVIII
3.1 Textual problems: Segmentation . . . . .	XVIII
3.2 Systematic problems: Identification . . . . .	XX
3.21 Heterographs . . . . .	XX
3.211 Cultural . . . . .	XX
3.212 Structural . . . . .	XX
3.2121 Syntagmatic . . . . .	XX
3.2122 Paradigmatic . . . . .	XXI
3.22 Homographs. . . . .	XXIII
3.221 Syntagmatic . . . . .	XXIII
3.222 Paradigmatic . . . . .	XXIII
3.2221 Lexical . . . . .	XXIV
3.22211 Nouns . . . . .	XXIV
3.22212 Adjectives . . . . .	XXV
3.22213 Pronouns . . . . .	XXVI
3.22214 Verbs . . . . .	XXVI
3.2222 Syntactic . . . . .	XXVI
3.22221 Nouns and Verbs . . . . .	XXVIII

## CONTENTS

3.22222	Verbs and Adjectives . . . . .	XXX
3.22223	Adjectives and Nouns . . . . .	XXXI
3.22224	Adjectives, Nouns, and Verbs . . . . .	XXXI
3.22225	Functional words . . . . .	XXXII
3.22226	Functional words and lexical words . . . . .	XXXIII
3.2223	Morphologic. . . . .	XXXIII
3.22231	Articles . . . . .	XXXIII
3.22232	Adjectives . . . . .	XXXIII
3.22233	Nouns. . . . .	XXXIV
3.22234	Pronouns . . . . .	XXXIV
3.22235	Verbs . . . . .	XXXIV
4.	COEFFICIENTS . . . . .	XXXV
4.1	Frequency . . . . .	XXXV
4.11	Frequency weighting . . . . .	XXXVI
4.12	High frequency words. . . . .	XXXVI
4.13	Low frequency words . . . . .	XXXVII
4.2	Dispersion . . . . .	XXXVII
4.21	Dispersion weighting . . . . .	XXXVII
4.22	High dispersion words . . . . .	XLI
4.23	Low dispersion words . . . . .	XLI
4.3	Usage . . . . .	XLI
4.31	Usage weighting . . . . .	XLII
4.32	High usage words . . . . .	XLIII
4.33	Low usage words . . . . .	XLIV
5.	THE DICTIONARY . . . . .	XLV
5.1	Size . . . . .	XLV
5.2	Selection . . . . .	XLVI
5.3	Structure . . . . .	XLVI

## PART ONE

### Types and Tokens: Alphabetical order

A	. . . . .	3
B	. . . . .	46
C	. . . . .	56
D	. . . . .	100
E	. . . . .	125
F	. . . . .	141
G	. . . . .	159
H	. . . . .	171
I	. . . . .	172

## CONTENTS

L	194
M	206
N	226
O	234
P	244
Q	283
R	286
S	315
T	364
U	381
V	386
Z	401

## PART TWO

### Types: Usage, Frequency, and Dispersion

First Five Hundred	405
Second Five Hundred	416
Third Five Hundred	427
Fourth Five Hundred	439
Fifth Five Hundred	450
Sixth Five Hundred	462
Seventh Five Hundred	473
Eighth Five Hundred	485
Ninth Five Hundred	496
Tenth Five Hundred	508
Last 14 words	519





# INTRODUCTION

## 1. PREVIOUS ATTEMPTS

While several statistical investigations of the French and Spanish vocabularies have been undertaken<sup>1</sup>, few have been devoted to Italian. Moreover, the limited word counts of Italian texts made in the United States have not been published, and their results have remained generally inaccessible to the scholarly public.

Essentially, there has been one small word count aiming at a basic vocabulary of 500 words drawn from a comparison of the vocabularies of Italian grammars and readers; another aiming at 2000 words extracted from literary sources; an essential Italian vocabulary of 1500 words; and still another aiming at 3000 words established on a presumably intuitive basis by comparing Italian textbooks<sup>2</sup>.

### 1.1 Thompson, *A Study in Italian Vocabulary Frequency* (1927)

The purpose of Miss Thompson's Master's thesis<sup>3</sup> was to establish an objective vocabulary of 500 words for first year Italian students. The sample underlying Miss Thompson's ranking supposedly consists of 100,000 words drawn from ten Italian reading textbooks edited in the United States. The number of words counted is actually smaller because functional words as well as some very frequent lexical words were ignored.

Words were counted on every tenth page of each book, except for articles, pronouns, demonstrative and possessive adjectives, certain functional adverbs, prepositions, conjunctions, numerals, names of days, months and seasons, cardinal points, proper and geographical names, and interjections of less than four letters. However, these words were included in the estimated size of the sample.

Verb and adjective variants, the singular and plural of nouns, etc., were not distinguished: noun variants are lumped together under the singular, adjective variants under the masculine singular, verb variants under the present infinitive. Certain syntactic homographs are distinguished by a part of speech abbreviation: *bene* adv. and *bene* n., *male* adv. and *male* n. (but not *peggio*, *meglio*, *meno*, *mezzo*, etc.); however, we are not told whether *diritto* is a noun, an adverb or an adjective, *dolce* a noun or an adjective, *parere* a noun or a verb, *verso* an adverb or a noun, etc. Nor do

<sup>1</sup> Cf. A. Juilland, D. Brodin, and C. Davidovitch, *Frequency Dictionary of French Words* (The Hague 1970), pp. XVI-LXX.

<sup>2</sup> For a discussion of the recent *Lessico di frequenza della lingua italiana* by U. Bortolini, C. Tagliavini, and A. Zampoli (Milano 1971) see our forthcoming paper *Frequency Dictionary of Italian Words and Lessico di frequenza della Lingua Italiana*.

<sup>3</sup> Mary Elizabeth Thompson, *A Study in Italian Vocabulary Frequency*, University of Iowa Master's Thesis, 1927.

## INTRODUCTION

we know whether *opera* means “work” or “opera”, *paese* “country” or “village”, *signore* “mister”, “ladies” or “Lord”, *tendere* “to tend” or “to stretch”, *nome* “name” or “noun”, etc.

Two coefficients are given for each entry, the “frequency” and the “weighted total”. The frequency coefficient (the number of times the word occurred in the sample) ranges from a high of 2,970 (the verb *essere*) to a low of 30, shared by fifteen words. The weighted total, an interesting but questionable attempt to combine frequency and dispersion, was established by multiplying the frequency by the square root of the square nearest to the number of sources in which the word occurred. For instance, *guardare* had 289 occurrences and was found in all ten sources; 289 was multiplied by 3, the square root of ten to the nearest integer, for a weighted total of 867. This procedure was borrowed from E. Horn, *A Basic Writing Vocabulary*.<sup>4</sup>

The dictionary consists of two parts. In the first, the 500 entries, followed by part of speech abbreviations only when necessary to distinguish between syntactic homonyms, are listed in decreasing order of frequency. The second part provides the same 500 entries in alphabetical order.

### 1.2 Knease, *An Italian Word List from Literary Sources* (1931)

Four years after Miss Thompson’s pioneering word count, Miss Knease defended, a more ambitious work at the same university.<sup>5</sup> While Miss Thompson was interested in establishing a vocabulary for students of beginning Italian, Knease concentrated on the needs of more advanced students, and aimed at a list of about 2,000 words.

The sample underlying *An Italian Word List* consists of 400,000 words, which is equivalent to the sample Henmon had found satisfactory for his *A French Word Book*.<sup>6</sup> However, as in Thompson’s count, the elimination of a large number of high frequency words considerably reduces the actual size of the sample drawn from the works published by forty authors from 1850 to 1930, selected from a list published in *Italica*.<sup>7</sup> Restricted to “literary sources,” the sample consists of novels, short stories and plays, some verse, a few volumes of literary criticism, and two works of history. Because of their specialized vocabulary, “no war books, no books purely scientific, religious, or philosophical” were included.

All words on every fifth page in each book were counted, except for many functional words, all foreign words, numerals, and geographical names. Miss Knease, who justifies these omissions in view of the pedagogical purpose of her word list, shows some inconsistency in explaining that such high frequency items as “*avere, buono, come, così, dopo, nuovo*, and about fifty others” were included “in order that this first extensive investigation in Italian may present data for all words which might be of interest to any teacher or student.” The author’s somewhat arbitrary decision concerning what might be of interest is unfortunate.

Inflected forms of nouns, verbs, and adjectives were counted together under the invariant form. Some variants such as *leggero* and *leggiero*, *danaro* and *denaro*, *domani*, *dimani* and *dimane*, etc., were listed as sub-entries but not counted separately. Feminine nouns with masculine equivalents, such as *ragazza*, *zia*, etc., were counted under the masculine *ragazzo*, *zio*, etc. Syntactic homonyms

<sup>4</sup> Ernest Horn, “A Basic Writing Vocabulary”, University of Iowa Monographs in Education. First Series, No. 4, 1926.

<sup>5</sup> Tacie Mary Knease, *An Italian Word List From Literary Sources*, Iowa State University Doctoral Dissertation, 1931.

<sup>6</sup> V.A.C. Henmon, *A French Word Book, Based on a Count of 400,000 Running Words*, Bureau of Educational Research Bulletin, No. 3, September, 1924. University of Wisconsin.

<sup>7</sup> “A Five Foot Shelf of Modern Italian Literature”, *Italica*, August 1925, Vol. 2, No. 3.

such as *ora* n. and *ora* adv. are distinguished by part of speech abbreviations; lexical homonyms are not distinguished.

Coefficients of range (the number of sources in which the word occurred) and frequency (the number of occurrences) are given for each entry.

The final list contains 2,097 entries, presumably all words with a range of 9 or more. The entries are presented in decreasing order of range ("since it is generally conceded that a word appearing in a wide range of sources is more useful than one of higher frequency but of lower range") and in alphabetical order. Part of speech is provided only when necessary to distinguish between certain syntactic homonyms.

### 1.3 Skinner, *A Comparative Study of the Vocabularies of Forty-Five Italian Textbooks* (1935)

Mr. Skinner's purpose was to establish a list of about 3,000 words, based on the vocabulary of Italian grammars and readers used in American high schools and colleges.<sup>8</sup>

The sample was drawn from "all Italian texts supplied with vocabularies issued by standard American publishers," and it included twenty grammars, composition books, and readers, as well as twenty-five literary reading texts. Instead of counting all textual occurrences, Mr. Skinner took note only of words included in the vocabulary section of each book, a method which detracts from the validity of the results: the verb *essere*, for instance, which obviously must appear in all 45 sources, has a range of 41, apparently because text editors differ on what words should be included in vocabulary sections.

Inflected forms, identical cognates, and a large number of functional words were excluded from the count.

Lexical homonyms are distinguished by an English translation, and syntactic homonyms by a part of speech abbreviation. Occasional variants are indicated, for example: *fra* and *tra*, *entro* and *dentro*, *comprare* and *comperare*, *movere* and *muovere*, etc.

Because of the method of counting, only range coefficients could be given. The list includes 2,962 words with a range of 10 or more. English translations and part of speech abbreviations accompany the entries only when required by the distinction of homonyms.

### 1.4 Migliorini, *Der Grundlegende Wortschatz des Italienischen* (1943)

Although Professor Migliorini provides, in his brief introduction, no indication of how he arrived at the 1,500 most common words ("die 1500 wesentlichsten Wörter") contained in his dictionary,<sup>9</sup> we assume that his choices were guided by partial word counts.

Each entry is accompanied by a phonetic transcription, a gender abbreviation in the case of nouns, principal parts of verbs, and a German translation. 500 words are printed in bold-face type, and are recommended by Professor Migliorini for first year students.

\* Laurence Harvey Skinner, "A Comparative Study of the Vocabularies of Forty-Five Italian Textbooks," *Modern Language Journal*, XX, 1935.

\* Bruno Migliorini, *Der Grundlegende Wortschatz des Italienischen* (Marburg 1943).

## INTRODUCTION

### 2. SAMPLE

In establishing the lexical universe from which sentences were selected at random, we followed the principles outlined in the Introduction to the *Frequency Dictionary of Spanish Words*<sup>10</sup> and further elaborated in the Introduction to the *Frequency Dictionary of French Words*.<sup>11</sup>

The purpose of the selection was to strike an appropriate balance between *unity* (homogeneity) and *diversity* (heterogeneity): the sample had to be homogeneous enough to represent more than a mechanical agglomeration of data borrowed from different periods or from different levels of style; and sufficiently diversified to be truly representative of the standard language, not only of a particular genre or style.

To reach these objectives, we restricted our samples to publications issued mainly between the years 1920 and 1940, as listed in the Union Catalogue of the Library of Congress; we discarded papers as well as writings that were too dialectal or popular in character.

For reasons discussed elsewhere<sup>12</sup>, we chose to operate with a lexical universe of about 500,000 words divided into five equal subsamples of 100,000 words as follows:

- 1) Dramatic literature: plays;
- 2) Fictional literature: novels and short stories;
- 3) Essayistic literature: essays, memoirs and correspondence;
- 4) Periodical literature: editorials, news, advertizing, etc., in dailies, weeklies, and monthlies;
- 5) Technical literature: medicine, botany, engineering, agronomy, etc.

The method of selection in the three "literary" or "creative" genres, plays, fiction, and essays, differed from the method followed in the "nonliterary" genres of periodical and technical literature. For the creative genres, we first established a list of about 100 representative authors, on the basis of information contained in the indices and tables of contents of a number of histories of Italian literature.<sup>13</sup> This first list served as a basis for a second one, consisting of nearly 500 titles by selected authors and listed in the Union Catalogue of the Library of Congress. After assigning each title to the appropriate genre, we selected at random three lists, one for dramatic literature, another for fictional literature, still another for essayistic literature.

A different procedure was followed for the "non-literary" genres: for periodical literature, we aimed for as wide a selection as possible of news, editorials, advertizing, etc.; for technical literature, we tried to secure texts from a great variety of scientific disciplines: aeronautics, agriculture, archeology, architecture, astronomy, anthropology, botany, chemistry, economics, fishery, genetics, geography, geology, home economics, hygiene, linguistics, law, mathematics, medicine, peletonology, physics, political science, sociology, zoology, etc.

#### 2.1 Plays

Sentences were sampled from the following 60 plays by 27 different authors:

<sup>10</sup> The Hague 1964, sections 1.1211 and 1.1212.

<sup>11</sup> The Hague 1970.

<sup>12</sup> Cf. *FDSW*

<sup>13</sup> Pietro Azzarone, *Storia della Letteratura Italiana*, Napoli 1965. Francesco Flora, *Storia della Letteratura Italiana*, Firenze 1957. Attilio Momigliano, *Storia della Letteratura Italiana*, Milano 1963. *Orientamenti Culturali: Letteratura Italiana*, Milano 1963. Francesco Pedrina, *Poesia e Critica*, Milano 1960. Arturo Pompeati, *Storia della Letteratura Italiana*, Torino 1962. Natalino Sapegno, *Compendio di Storia della Letteratura Italiana*, Firenze 1965. Raffaele Spongano, *Schemi di Storia della Letteratura Italiana*, Bologna 1965. Rosario Tosto, *Storia della Letteratura Italiana*, Firenze 1957.

L. Antonelli	<i>La fiaba dei tre maghi</i> (1920)
L. Antonelli	<i>Bernardo l'Eremita</i> (1927)
L. Antonelli	<i>Il dramma, la commedia, la farsa</i> (1927)
S. Benelli	<i>Ali</i> (1921)
S. Benelli	<i>La santa primavera</i> (1923)
S. Benelli	<i>Adamo ed Eva</i> (1932)
S. Benelli	<i>Il vezzo di perle</i> (1934)
S. Benelli	<i>L'elefante</i> (1937)
S. Benedetti	<i>La vendetta del chiaro di luna</i> (1927)
M. Bontempelli	<i>Nostra Dea</i> (1925)
M. Bontempelli	<i>Minnie, la candida</i> (1929)
M. Bontempelli	<i>Bassano, padre geloso</i> (1934)
A. Campanile	<i>Centocinquanta la gallina canta</i> (1925)
A. Campanile	<i>L'inventore del cavallo</i> (1925)
A. Campanile	<i>L'amore fa fare questo e altro</i> (1930)
G. Cantini	<i>È tornato carnevale</i> (1930)
L. Chiarelli	<i>Chimere</i> (1930)
L. Chiarelli	<i>Pulcinella</i> (1939)
G. Civinini	<i>Rottami</i> (1929)
E. Corradini	<i>Giulio Cesare</i> (1926)
E. Corradini	<i>L'apologo delle due sorelle</i> (1929)
G. Forzano	<i>Lorenzino</i> (1923)
G. Forzano	<i>I fiordalisi d'oro</i> (1924)
G. Forzano	<i>Gutlibi</i> (1926)
G. Forzano	<i>Campo di Maggio</i> (1931)
G. Forzano	<i>Napoleone e le donne</i> (1930)
G. Forzano e F. Paolieri	<i>Maestro Landi</i> (1925)
A. Fraccaroli	<i>Biraghin</i> (1924)
A. Fraccaroli	<i>Peccato biondo</i> (1929)
S. Gotta	<i>Peccato di poesia</i> (1924)
S. Gotta	<i>Il primo peccato</i> (1939)
A. Guglielminetti	<i>Nei e cicisbei</i> (1926)
F. Lanza	<i>Fiordispina</i> (1928)
C. V. Lodovici	<i>La donna di nessuno</i> (1926)
L. Macina	<i>Il dormiglione risvegliato</i> (1930)
F. M. Martini	<i>L'altra Nanetta</i> (1923)
F. M. Martini	<i>La sera del trenta</i> (1926)
E. L. Morselli	<i>Glauco</i> (1922)
E. L. Morselli	<i>Belfagor</i> (1930)
D. Niccodemi	<i>L'alba, il giorno, la notte</i> (1925)
D. Niccodemi	<i>Scena vuota</i> (1922)
L. Pirandello	<i>Sei personaggi in cerca d'autore</i> (1923)
L. Pirandello	<i>La vita che ti diedi</i> (1924)
L. Pirandello	<i>La signora Morli una e due</i> (1925)
L. Pirandello	<i>Diana e la tuda</i> (1927)
L. Pirandello	<i>O di uno o di nessuno</i> (1929)

## INTRODUCTION

F. V. Ratti	<i>Il solco quadrato</i> (1921)
F. V. Ratti	<i>Bruto</i> (1925)
F. V. Ratti	<i>Socrate</i> (1927)
F. V. Ratti	<i>Termiti</i> (1928)
L. Repaci	<i>La barba alla morte</i> (1926)
L. Repaci	<i>Favola di Martini</i> (1927)
L. Repaci	<i>Luciano</i> (1930)
G. Rocca	<i>Le pecorelle</i> (1925)
G. Rocca	<i>Il re povero</i> (1939)
R. di San Secondo	<i>Il minuetto dell'anima nostra</i> (1923)
R. di San Secondo	<i>Lazzarina tra i coltelli</i> (1925)
O. Vergani	<i>Il cammino sulle acque</i> (1927)
C. G. Viola	<i>Il cuore in due</i> (1926)
C. G. Viola	<i>Fine del protagonista</i> (1931)
G. Zorzi	<i>Le due metà</i> (1922)
G. Zorzi	<i>La vita degli altri</i> (1926)

## 2.2 Fiction

Sentences were sampled from the following 55 novels or short stories by 40 different authors:

A. Albertini	<i>Cresco</i> (1937)
R. Bacchelli	<i>Mal d'Africa</i> (1935)
R. Bacchelli	<i>Il raddomante</i> (1936)
A. Beltramelli	<i>Tre bimbe a vendere!</i> (1920)
M. Bontempelli	<i>"522", Racconto di una giornata</i> (1932)
M. Bontempelli	<i>Gente nel tempo</i> (1937)
G. A. Borgese	<i>Le belle</i> (1927)
G. Brocchi	<i>La rocca sull'onda</i> (1926)
V. Brocchi	<i>Il destino in pugno</i> (1924)
V. Brocchi	<i>L'arcolaio</i> (1924)
A. Campanile	<i>Se la luna mi porta fortuna</i> (1928)
G. Caprin	<i>Quirina e Floriana</i> (1931)
L. Capuana	<i>La sfinge</i> (1921)
F. Chiesa	<i>Sant'Amarillide</i> (1938)
B. Cicognani	<i>Il museo delle figure viventi</i> (1928)
D. Cinelli	<i>Cinquemila lire</i> (1930)
G. Comisso	<i>Avventure terrene</i> (1935)
G. Comisso	<i>Gente di mare</i> (1929)
M. Dandolo	<i>Il vento nella foresta</i> (1922)
M. Dandolo	<i>Il dolore degli altri</i> (1928)
M. Dandolo	<i>Le stelle nel mare</i> (1924)
M. Dandolo	<i>Come agnelli tra i lupi</i> (1931)
M. T. Dazzi	<i>Città</i> (1936)
G. Deledda	<i>La madre</i> (1928)
G. Deledda	<i>La casa del poeta</i> (1930)

G. Deledda	<i>Il paese del vento</i> (1931)
A. Gatti	<i>Ilia e Alberto</i> (1930)
S. Gotta	<i>La donna mia</i> (1924)
S. Gotta	<i>Portofino</i> (1936)
A. Guglielminetti	<i>Quando avevo un amante</i> (1933)
C. Linati	<i>Sulle orme di Renzo</i> (1919)
C. Linati	<i>La principessa delle stelle</i> (1924)
A. Marpicati	<i>Piccolo romanzo di una vela</i> (1922)
M. Messina	<i>Alla deriva</i> (1920)
P. Mignosi	<i>Gioia d'Agave</i> (1934)
G. Milanesi	<i>La sera di Santa Barbara</i> (1938)
G. Milanesi	<i>La sperduta di Allah</i> (1927)
I. Montanelli	<i>Giorno di festa</i> (1939)
A. Moravia	<i>La bella vita</i> (1935)
M. Moretti	<i>La voce di Dio</i> (1921)
M. Moretti	<i>La vera grandezza</i> (1925)
A. Negri	<i>Stella mattutina</i> (1927)
A. Panzini	<i>Il mondo è rotondo</i> (1920)
A. Panzini	<i>Novelline divertenti per bambini intelligenti</i> (1934)
L. Pietravalle	<i>I racconti della terra</i> (1924)
M. Puccini	<i>Dove è il peccato è Dio</i> (1922)
L. Repaci	<i>L'ultimo cireneo</i> (1923)
G. Rocca	<i>L'uragano</i> (1929)
B. Sanminiati	<i>Le pecore pazze</i> (1920)
R. di San Secondo	<i>La festa delle rose</i> (1921)
M. Saponaro	<i>La giovinezza</i> (1926)
N. Savarese	<i>Ricordi di strada</i> (1922)
G. da Verona	<i>Mata Hari</i> (1926)
C. G. Viola	<i>Pricò</i> (1924)
L. Zuccoli	<i>Kif Tebbi</i> (1923)

### 2.3 Essays

Sentences were sampled from the following 50 essays by 42 different authors:

C. Alvaro	<i>Calabria</i> (1931)
C. Angelini	<i>Testimonianze cattoliche</i> (1928)
A. Baldini	<i>Amici allo spiedo</i> (1932)
B. Barilli	<i>Il sorcio nel violino</i> (1926)
A. G. Bragaglia	<i>Sotto palco</i> (1937)
A. Cajumi	<i>Galleria</i> (1930)
R. Calzini	<i>Spagna</i> (1930)
A. Capasso	<i>Saper distinguere</i> (1934)
G. Caprin	<i>Sommario storico della guerra universale</i> (1921)
E. Cecchi	<i>Scrittori inglesi e americani</i> (1935)
F. Chiesa	<i>Svizzera italiana</i> (1931)

## INTRODUCTION

E. Corradini	<i>L'unità è la potenza delle nazioni</i> (1926)
B. Croce	<i>Nuove curiosità storiche</i> (1922)
B. Croce	<i>La poesia di Dante</i> (1921)
B. Croce	<i>L'Italia dal 1914-1918</i> (1919)
B. Croce	<i>Cultura e vita morale</i> (1926)
A. Fraccaroli	<i>Hollywood paese d'avventura</i> (1929)
R. Franchi	<i>L'europeo sedentario</i> (1929)
G. Gentile	<i>Giordano Bruno e il pensiero del rinascimento</i> (1925)
G. Gentile	<i>Che cosa è il fascismo</i> (1925)
S. Di Giacomo	<i>I quattro antichi conservatori musicali di Napoli</i> (1924)
U. E. Imperatori	<i>Italia prodiga (Gli italiani all'estero)</i> (1925)
C. Linati	<i>Irlanda</i> (1940)
C. Malaparte	<i>Italia barbara</i> (1926)
C. Marchesi	<i>Seneca</i> (1944)
F. T. Marinetti	<i>Il fascino dell'Egitto</i> (1933)
A. Marpicati	<i>Nella vita del mio tempo</i> (1934)
A. Marpicati	<i>Il dramma politico di U. Foscolo</i> (1928)
F. M. Martini	<i>Cronache teatrali</i> (1923)
F. M. Martini	<i>Cronache drammatiche</i> (1922, 1923)
P. Monelli	<i>Viaggio alle isole freddazzurre</i> (1926)
B. Mussolini	<i>L'Italia nel mondo</i> (1928)
D. Niccodemi	<i>Tempo passato</i> (1929)
U. Ojetti	<i>Scrittori che si confessano</i> (1926)
A. Panzini	<i>Diario sentimentale della guerra</i> (1923)
G. Papini	<i>Storia di Cristo</i> (1923)
G. Papini	<i>Gog</i> (1931)
P. Rebora	<i>Jonathan Swift</i> (1922)
L. Repaci	<i>Ribalte a lumi spenti</i> (1941)
G. De Ruggiero	<i>L'Impero britannico dopo la guerra</i> (1921)
L. Russo	<i>Giovanni Verga</i> (1934)
E. Ruta	<i>Politica e ideologia</i> (1929)
N. Savarese	<i>Cose d'Italia</i> (1940)
G. Stuparich	<i>Colloqui con mio fratello</i> (1925)
G. Stuparich	<i>Guerra del '15</i> (1931)
A. Tilgher	<i>Homo faber</i> (1929)
G. Toffanin	<i>La fine dell'Umanesimo</i> (1920)
D. Valeri	<i>Fantasie veneziane</i> (1934)
L. Viani	<i>Storie di umili titani</i> (1934)
E. Vittorini e G. Ferrata	<i>La tragica vicenda di Carlo III</i> (1939)

## 2.4 Newspapers and Magazines

Sentences were sampled from the following 32 newspapers and magazines:

*Affari Esteri* (1935)

*Corriere della Sera* (1921, 1929, 1936, 1937)



*Corriere Mercantile* (1920)  
*La Critica Politica* (1938)  
*La Cultura* (1934, 1935)  
*Casabella* (1939)  
*Domus* (1928, 1930, 1936)  
*Dramma* (1925, 1940)  
*Il Giornale d'Italia* (1929, 1931)  
*L'Illustrazione Italiana* (1920, 1924, 1931, 1932, 1933)  
*L'Italiano* (1931, 1932, 1933)  
*Il Lavoro d'Italia* (1926)  
*Il Mattino* (1924, 1926)  
*Il Mondo Classico* (1931)  
*L'Osservatore Romano* (1922, 1925, 1939)  
*Panorama* (1939)  
*Il Popolo d'Italia* (1925, 1930, 1933, 1936, 1940)  
*Il Popolo di Roma* (1925, 1926)  
*Rassegna Monetaria* (1937)  
*Rassegna Numismatica* (1931)  
*Relazioni Internazionali* (1935, 1939, 1940)  
*Rivista del Lavoro Commerciale* (1933)  
*Rivista Geografica Italiana* (1930, 1936)  
*Sapere* (1937, 1938)  
*Il Secolo* (1921, 1927)  
*Il Secolo, La Sera* (1928, 1939)  
*Secolo Nostro* (1939)  
*Tempo* (1939, 1940)  
*Le Tre Venezie* (1925)  
*La Tribuna* (1923, 1939)  
*Umanità nova* (1920)  
*Viaggi e Soste in Italia* (1939)

## 2.5 Technical and Scientific Writings

Sentences were sampled from the following 33 publications:

*Annali dell'Ufficio Presagi, Ministero dell'Aeronautica, Direzione Generale dei Servizi del Materiale e degli Aeroporti* (1932)  
*Annali di Botanica* (1933, 1934)  
*Annali di Matematica Pura ed Applicata* (1924)  
*Archivio Botanico* (1936)  
*Archivio Glottologico Italiano* (1929)  
*Archivio per l'Antropologia e l'Etnologia* (1927)  
*Archivio Zoologico Italiano* (1926)  
*Atti della Reale Accademia d'Italia* (1932)  
*Atti della Reale Accademia Nazionale dei Lincei* (1930, 1931)  
*Atti della Reale Accademia delle Scienze Fisiche e Matematiche* (1927)

## INTRODUCTION

*Atti della Accademia Reale delle Scienze di Torino* (1927, 1928)  
*Bollettino del Comitato per la Geodesia e la Geofisica del Consiglio Nazionale delle Ricerche* (1934, 1935, 1936, 1937)  
*Bollettino del Reale Ufficio Geologico d'Italia* (1922, 1923)  
*Bollettino d'Informazioni del Consiglio Nazionale delle Ricerche* (1930, 1931)  
*Bollettino di Paleontologia Italiana* (1930, 1935)  
*Circolo Matematico di Palermo* (1931)  
*Consiglio Nazionale delle Ricerche - Istituti e Laboratori Scientifici Italiani* (1931)  
*Consiglio Nazionale delle Ricerche - La Ricerca Scientifica ed il Progresso Tecnico nell'Economia Nazionale* (1932, 1934, 1935, 1936, 1938, 1940)  
*L'Elettrotecnica* (1936)  
*Minerva Medico-Legale* (1923)  
*Ministero dell'Aeronautica - Aviazione Civile e Traffico Aereo* (1926)  
*Nuovo Bollettino di Archeologia Cristiana* (1920)  
*Nuovo Cimento* (1924)  
*Ricerche Astronomiche* (1939)  
*Rivista di Antropologia* (1928, 1929)  
*Rivista di Archeologia Cristiana della Pontificia Commissione di Archeologia Sacra* (1930)  
*Rivista di Parasitologia* (1937)  
*Rivista Geografica Italiana* (1921, 1922, 1934, 1937)  
*Rivista Internazionale di Scienze Sociali e Discipline Ausiliarie* (1920)  
*Rivista Italiana di Sociologia* (1921)  
*Scientia Genetica* (1939)  
*Servizio Geologico d'Italia* (1924)  
*Unione Matematica Italiana - Bollettino* (1935)

From these sources, sentences were sampled in such a way as to approximate as closely as possible the desired amount of 100,000 words per genre, by taking into consideration the number of volumes retained in each genre and the estimated average number of pages per volume, sentences per page, and words per sentence.<sup>14</sup>

## 3. INTERPRETATION

A word count must determine which words occur in the sample, how many times each word occurs, and how evenly the occurrences are spread through the sample. Such information enables the linguist to make predictions about the occurrence of words, parts of words, classes of words, or categories of words in larger samples and, under certain conditions, in the language as a whole. To this end, one must determine forms correctly by resolving ambiguous instances of segmentation in the text, of identification in the system.

### 3.1 Textual problems: Segmentation

In the text, the question is whether certain constructions are to be interpreted as one word or two

<sup>14</sup> Cf. *FDSW*.

or more words, whether adjacent morphemes are to be interpreted conjunctively, as parts of the same word, or disjunctively, as parts of different words.

Problems of segmentation arise in languages in which words are not well individualized: phonologically, their autonomy is not marked by segmental or suprasegmental features; grammatically, they seldom occur as minimum utterances, have narrow distributional or substitutional ranges, show restricted privileges of occurrence relative to larger syntactic units, are not normally permutable in the sentence, cannot be disjoined by means of insertion, etc. In such languages the autonomy of the word cannot be verified by means of distributional, substitutional, permutational, transformational, or insertional testing.

Fortunately, Italian words are well defined. An inflectional language, Italian expresses grammatical relations by predominantly morphological means: functional morphemes, be they free like particles and clitics, or bound like affixes and flexives, mark the boundaries of lexical words, which are syntactically free, i.e., have wide distributional and substitutional ranges and are easily permutable or disjoined by insertion. Physically, the independence of Italian words is marked by a relatively free stress which correlates highly with lexical words: unlike French, where phenomena of *liaison* and *enchaînement* often place word boundaries at the very heart of a syllable, Italian word boundaries are strengthened by syllabic ictus, with every word boundary a syllable boundary. The autonomy of Italian words is further reinforced by distributional rules which restrict the occurrence of final consonants.

The unambiguous status of Italian words is illustrated by the scarcity of constructions spelt sometimes as one word (with no space between the constituents,) sometimes as two (with a blank space in between), and sometimes linked by a hyphen. In inflectional languages, ambiguous segmentation generally involves compounds and clitics. Italian compounds are spelt as a rule in one word: *benchè, chissà, finora, giacchè, iersera, poichè, stamane, stasera*, etc. Functional words, however, are sometimes agglutinated to an adjacent word: prepositions to a following article, e.g., *al, all', agli, alla, alle*, etc., pronouns to a preceding verb, e.g., *venderlo, vendergli, vendermelo*, etc.

Conjoined prepositions and articles are:

**a:** agl', agli, all', alla, alle, allo  
**con:** cogli, coi, col, coll', colla, colle  
**da:** dagl', dagli, dai, dal, dall', dalla, dalle, dallo  
**di:** degl', degli, dei, del, dell', della, delle, dello  
**in:** negl', negli, nei, nel, nell', nella, nelle, nello  
**su:** sugli, sui, sul, sull', sulla, sulle, sullo

Object pronouns conjoined with the preceding verb are: *-ci, -gli, -la, -le, -li, -lo, -mi, -si, -ti*, and *-vi*. When two pronominal clitics are agglutinated to the same verb, the indirect, *-ce-, -glie-, -me-, -te-*, or *-ve-*, precede the direct, *-la, -le, -li, -lo*, or the particle *-ne*. Since distributional limitations restrict the combination of ten pronominal forms and 25 indirect + direct pronominal sequences with verbal forms, a typical verb such as *mandare* may exhibit no less than 269 enclitically agglutinated variants.

We have interpreted both types of construction disjunctively, as two or three words, rather than follow the conjunctive interpretation embodied in spelling. Occurrences of such graphic units have been counted two or three times: *allo* etc. under the preposition **a** and under the article **lo**, *venderlo* etc. under the verb **vendere** and under the direct pronoun **lo**, *vendermelo* etc. under the verb **vendere**, under the indirect pronoun **mi**, and under the direct pronoun **lo**. Constructions conjoining a preposition with an article have been entered under both: *allo* appears under the preposition **a**

## INTRODUCTION

and under the article **lo** because of the relatively limited number of such constructions and the very high frequency of their occurrence. However, constructions conjoining pronominal objects with a preceding verb (*vendiamolo*, *vendiamogli*, *vendiamogliela*, etc.) are not entered separately under **vendere**, but computed under *vendiamo* in order to provide accurate totals for the occurrences of verbal heterographs.

### 3.2 Systematic problems: Identification

In the system, problems of identification involve *homographs*, which are identical occurrences of different words; and *heterographs*, which are different occurrences of the same word.

#### 3.21 Heterographs

Heterographs are “structural” or “cultural” depending on whether they are conditioned by linguistic factors in the text or in the system, or by interference of a different level of style, an earlier stage of the language, or another dialect.

3.211 *Cultural*. — The variants we call “cultural” are sometimes restricted to spelling, but there are instances in which they reflect also differences in pronunciation.

A. *Spelling*. With its strong literary tradition, Italian offers relatively few instances of alternative spelling. Most graphic variants involve geminate or affricate consonants: *immaginare* or *immaginare*, *immaginare* or *immaginare*, *sodisfatto* or *soddisfatto*, *sopratutto* or *soprattutto*; *bilance* or *bilancie*, *facce* or *faccie*, *guance* or *guancie*, *minacce* or *minaccie*, *rocce* or *roccie*, *tracce* or *traccie*, *sufficente* or *sufficiente*, as well as *piogge* or *pioggie*. We may add rare instances of hesitation between a vowel and a semivowel such as *aiuta* or *ajuta*, *aiuti* or *ajuti*.

B. *Pronunciation*. Alternative spellings which imply differences in pronunciation are stylistic, archaic, or dialectal. They involve words containing affricates such as *annunciare* or *annunziare*, *rinunciare* or *rinunziare*, as well as *servizio* or *servigio*. Other words show hesitation between a simple vowel and a diphthong: *bono* or *buono*, *core* or *cuore*, *figliolo* or *figliuolo*, *giocare* or *giuocare*, *risonare* or *risuonare*, *sonare* or *suonare*. Still others are plurals which alternate between a simple and double vowel: *anniversari* and *anniversarii*, *domini* or *dominii*, *finanziari* or *finanziarii*, *necessari* or *necessarii*, *principi* or *principii*, *propri* or *proprii*, *vari* or *varii*. One must add isolated heterographs such as *denaro* or *danaro*, *stamane* or *stamani*, *stromento* or *strumento*.

Such variants have been kept apart, their coefficients being totalized under the same entry word.

3.212 *Structural*. — The variants we call “structural” are syntagmatic, determined by and predictable from their textual environment, and paradigmatic, expressing different grammatical relations or categories.

3.2121 *Syntagmatic*. — Among variants conditioned by the phonological environment, Italian distinguishes elisions, which suppress a final vowel before a word beginning with a vowel (*la madre* but *l'artista*), from truncations, which suppress a final vowel or consonant-vowel before words beginning with a consonant (*ancora ieri* but *ancor più*, *grande affare* but *gran donna*).

A. *Elision*. This phenomenon affects words ending in =a, =e, =i, or =o. Forms with elided variants which appear in our dictionary are:

- =a: cosa, la *art.*, la *pron.*, mezza, nessuna, nulla, ora, quaranta, senza, sopra, tanta, terza, trenta, tutta, ultima, una.
- =e: che, come, diciassette, dove, ne, pure, se, sette, tale, vale
- =i: ci, mi, si, ti, venti, vi
- =o: cinquecento, cento, lo *art.*, lo *pron.*, molto, poco, povero, presso, quanto, quattro, quello, questo, santo, solo, sotto, uno.

B. *Truncation*. This phenomenon affects words ending in *-a*, *-e*, *-i*, and *-o*. Forms with truncated variants which appear in our dictionary are:

=*a*: allora, ancora, suora

=*e*: affare, altare, amore, ardore, bicchiere, bene, canale, cane, carbone, cavaliere, colore, commendatore, correzione, cuore, direttore, dottore, ebbene, emigrazione, essere, fare, legione, maggiore, male, mare, migliore, minore, mortale, neppure, niente, occidentale, odore, padrone, pane, particolare, professore, quale, ragione, resurrezione, rigore, sale, salone, sapore, sebbene, signore, singolare, sole, sostanziale, speciale, spessore, splendore, superiore, umore, valore, vile, and many infinitives in *-are*, such as abbandonare.

=*i*: cavalieri, fuori

=*o*: alcuno, anziano, cammino, capitano, castello, ciascuno, filo, fino, loro, meno, mio, nemmeno, nessuno, ognuno, perfino, piano, pieno, seno, sino, tenero, uomo, vano, vero, and many verbal forms in *-ano*, such as abbandonano.

3.2122 *Paradigmatic*. — These are word variants which express differences in number for nouns; in gender and number for adjectives and articles; in gender, number, person, and case for pronouns; in mode, tense, number, person, and gender for verbs.

A. *Nouns*. Singular and plural forms, masculines such as *abate* and *abati* or *libro* and *libri*, feminines such as *donna* and *donne* or *luce* and *luci*, have been grouped and computed under the same heading, a procedure also followed in the case of irregulars such as *dio* and *dei* or *uomo* and *uomini*. Dual plurals, in *-a* and *-i*, were grouped with the appropriate singulars under **braccio**, **dito**, **grido**, **membro**, **muro**, **urlo**, and so was the irregular *films*, entered along with *filmi* under **film**. Syncretic plurals such as *orecchie* were grouped with the appropriate singulars, *orecchia* or *orecchio*. The same distinction was made between the singular and plural occurrences of *piè*, grouped with the regular *pie* under **pie**.

We have departed from Italian lexicographic practice in separating masculines and feminines such as *imperatore* and *imperatrice* or *principe* and *principessa*, which were treated as independent entries.

Again departing from the approach of most Italian dictionaries, we have treated augmentatives and diminutives as separate words: *portone* or *tavolina* are considered words by themselves rather than variants of *porta* or *tavola*.

B. *Adjectives*. We have grouped under the same heading the masculine and feminine singular and plural of adjectives, including the superlative forms, so that adjectives in *-e* may have as many as six variants: *verde*, *verdi*, *verdissima*, *verdissime*, *verdissimi*, and *verdissimo*; adjectives in *-o*, as many as eight: *bianca*, *bianche*, *bianchi*, *bianchissima*, *bianchissime*, *bianchissimi*, *bianchissimo*, and *bianco*. It goes without saying that adjectives with syntagmatically conditioned variants may exhibit even more: *begli*, *bei*, *bel*, *bell'*, *bella*, *belle*, *belli*, *bellissima*, *bellissime*, *bellissimi*, *bellissimo*, and *bello*, or *gran*, *grand'*, *grandi*, *grandissima*, *grandissime*, *grandissimi*, and *grandissimo*.

C. *Articles*. The number of article variants depends on the number of articles recognized in the analysis. Inasmuch as Italian articles vary in *gender* (masculine/feminine), *number* (singular/plural), and *determination* (definite/indefinite), one may reckon with as few as one article (with gender, number, and determination variants) or as many as nine (if *lo* is considered a separate article). Intermediary solutions would postulate two, four, six, or eight articles, depending on which grammatical distinctions are considered relevant and on the way in which they are coordinated or subordinated. For reasons partly parallel to those which justify the interpretation of pronouns, we have adopted an intermediary solution, with three definite articles, **il**, **la**, and **lo**, and one indefinite, **uno**, the masculine and feminine being treated as variants on the basis of the parallelism with the adjective. Because definite articles combine with preceding prepositions, the following variants account for the occurrence of the four Italian articles:

## INTRODUCTION

**il:** al, coi, col, dai, dal, dei, del, i, il, nei, nel, pei, pel, sui, sul  
**la:** all', all' *p.*, alla, alle, coll' *m.*, colla, colle, dall', dalla, dalle, dell', dell' *p.*, della, delle, l', l' *p.*, la, le, nell', nell' *p.*, nella, nelle, sull', sulla, sulle.  
**lo:** agl', agli, all', allo, cogli, coll', dagl', dagli, dal, dall', dallo, degl', dell', dello, gl', gli, l', lo, negl', negli, nell', nello, sugli, sull', sullo  
**uno:** un, un' *f.*, una, uno.

**D. Pronouns.** Italian has personal, impersonal, reflexive, reciprocal, disjunctive, and demonstrative pronouns. Difficulties are encountered in the interpretation of personal pronouns, which differ according to person (1st/2nd/3rd), number (singular/plural), gender (masculine/feminine), case (subject/object, direct/indirect), and emphasis (conjunctive/disjunctive). As in the case of the articles, the number of variants grouped under each personal pronoun is conditioned by the number of pronouns recognized. Depending on which grammatical categories are considered relevant, and how these categories are coordinated or subordinated, one may reckon with as few as two personal pronouns (one masculine, the other feminine, one singular, the other plural, one subject, the other object, etc.) or as many, say, as 24, if all distinctions are considered relevant. Intermediary solutions would postulate 3 pronouns (one for each person), or 4 (by distinguishing masculine and feminine pronouns in the 3rd person), or 6 (by distinguishing singular and plural pronouns for each person while grouping the masculine and feminine under the 3rd), or 8 (by considering person, gender, and number relevant), or 16 (by distinguishing subject and object pronouns for each person, number, and gender), or 24, by further distinguishing direct and indirect object pronouns, etc.

We have opted for one of the more complex solutions in which person, number, gender, case, and emphasis are relevant in determining independent pronouns, as follows:

1) Singular first	subject	:	<b>io</b>
2) Singular first	object conjunctive:		<b>mi</b>
3) Singular first	object disjunctive :		<b>me</b>
4) Singular second	subject	:	<b>tu</b>
5) Singular second	object conjunctive:		<b>ti</b>
6) Singular second	object disjunctive :		<b>te</b>
7) Singular third masculine	subject	:	<b>egli</b>
8) Singular third masculine	object conjunctive:		<b>lo</b>
9) Singular third masculine	object disjunctive :		<b>lui</b>
10) Singular third feminine	subject	:	<b>ella</b>
11) Singular third feminine	object conjunctive:		<b>la</b>
12) Singular third feminine	object disjunctive :		<b>lei</b>
13) Plural first	subject	:	<b>noi</b>
14) Plural first	object conjunctive:		<b>ci</b>
15) Plural first	object disjunctive :		<b>noi</b>
16) Plural second	subject	:	<b>voi</b>
17) Plural second	object conjunctive:		<b>vi</b>
18) Plural second	object disjunctive :		<b>voi</b>
19) Plural third masculine	subject	:	<b>essi</b>
20) Plural third masculine	object conjunctive:		<b>li</b>
21) Plural third masculine	object disjunctive :		<b>loro</b>
22) Plural third feminine	subject	:	<b>esse</b>
23) Plural third feminine	object conjunctive:		<b>le</b>
24) Plural third feminine	object disjunctive :		<b>loro</b>

On the basis of this interpretation, paradigmatic heterographs are *me* (vede me, me lo metto) and *mi* (mi chiama, mi scrive, mi lavo) of **mi** (2); *te* (vede te, te lo metti) and *ti* (ti chiama, ti scrive, ti lavi) of **ti** (5); *gli* (glielo scrivo), *le* (le scrivo), *li* (li chiamo) and *lo* (lo chiamo) of **lo** (8); *gli* (gli scrivo) and *le* (le chiamo, le scrivo, le parlo) of **la** (11); *ce* (ce lo dice, ce lo mettiamo), and *ci* (ci vede, ci scrive, ci mettiamo) of **ci** (14).

Other pronouns are treated like nouns, with singular and plural variants grouped under the masculine or the feminine: *questo* and *questi* versus *questa* and *queste*, *quello* and *quelli* versus *quella* and *quelle*, *alcuno* and *alcuni* versus *alcuna* and *alcune*, etc.

E. *Verbs*. We have grouped under the same entry the *voice* (personal/impersonal, transitive/reflexive, active/passive), *mode* (indicative/imperative/conditional/subjunctive/participle/infinitive/gerund), *tense* (present/descriptive past/absolute past/present perfect/past descriptive perfect/future perfect/past absolute perfect), *number* (singular/plural) and *person* (1st/2nd/3rd), so that there may be more than 200 variants per verb. Including homographs, there are 78 different forms under *essere* and 48 under the auxiliary, 83 under *avere* and 50 under the auxiliary.

è, era, eran, erano, eravam, eravamo, eravate, eri, ero, essendo, esser, essere, fora, fosse, fossero, fossi, fossimo, foste, fosti, fu, fui, fummo, furon, furono, sarà, sarai, saran, saranno, sarebbe, sarebbero, sarei, saremmo, saremo, sareste, saresti, sarete, sarò, sei, sia, siam, siamo, sian, siano, siate, sien, sieno, siete, sii, son, sono.

à, abbi, abbia, abbiam, abbiamo, abbian, abbiano, abbiate, ài, avea, avemmo, avendo, avente, aventi, aver, avere, avesse, avessero, avessi, avessimo, aveste, avete, aveva, avevamo, avevan, avevano, avevate, avevi, avevo, avrà, avrai, avranno, avrebbe, avrebbero, avrei, avremmo, avremo, avreste, avresti, avrete, avrò, avuta, avute, avuti, avuto, ebbe, ebbero, ebbi, ha, hai, han, hanno, ho, ò.

### 3.22 Homographs

Homographs are *syntagmatic*, resulting from the reduction of different words, or *paradigmatic*, resulting from the phonological convergence of different words or from the functional or semantic differentiation of the same word.

3.221 *Syntagmatic*. — Syntagmatic homographs are created by the elision of a final vowel before words beginning with a vowel. They involve articles, pronouns, and adjectives, e.g., the article *l'*, which may be an occurrence of *la* (l'anima), of *lo* (l'orologio), or of *le* (l'erbe); the demonstrative *quell'* which may be an occurrence of *quello* (quell'albero), of *quella* (quell'isola) or of *quelle* (quell'erbe), just as *quest'* may represent *questo* (quest'uomo) or *questa* (quest'amica), *questi* (quest'istinti) or *queste* (quest'erbe). Add adjectival forms such as *molt'* for *molto* (molt'onore) or *molta* (molt'acqua), *tant'* for *tanto* (tant'affetto) or *tanta* (tant'aria), etc.

3.222 *Paradigmatic*. — Paradigmatic homographs are *historical* or *functional*.

Historical homographs are different forms whose shapes have merged as a result of phonetic change, such as *sei* in *sei giorni* and *tu sei regina*, *fine* in *persona fine* and *fine del mondo*, *riso* in *riso bollito* and *pianto e riso*, *sapere* in *sapere la regola* and *sapere di menta*, etc.

More complex, functional homographs are identical forms which have acquired different meanings by semantic or syntactic differentiation. Compare the meaning of *senso* in *senso tattile*, in *senso peggiorativo*, in *buon senso*, *senso morale*, or in *senso unico*. Even more diversified, a form like *piano* has not only different meanings in *gli eserciti si sono scontrati nel piano*, in *piano di lavoro*, in *piano geometrico*, or in *quarto piano*, it also functions as an adjective in *pavimento piano*, as an adverb in *camminare piano*. The latter example suggests forms with an even greater diversity of syntactic functions: *tanto* may modify a noun in *tanto lavoro*, an adjective in *tanto buono*, a verb in *scrive tanto*, an adverb in *tanto rapidamente*, just as it may function as a subject in *il tanto dell'affare*, as an object in *ne ha speso tanto*. A particle like *ci* may be a direct object in *ci vedono* an indirect object in *ci scrivono*, an impersonal in *ci si vende*. Somewhat apart are homographs with the same syntactic function which express different grammatical categories: indicative (*egli canta*) or an imperative (*canta*). indicative (*tu parli*) or subjunctive (*che tu parli*), a 1st person (*che io parli*) or a third (*che lui parli*), a masculine (*il fine*), or a feminine (*la fine*), a singular (*la virtù*) or a plural (*le virtù*), etc. Are these variants of the same invariant, or is it preferable to take

## INTRODUCTION

them as occurrences of the same word? Since criteria of expression are inconclusive, their identification will rest heavily on considerations of content based on meaning and use.

If the solution is obvious in the case of syncretic homographs which refer to clearly different objects, notions, or actions such as “six” and “are” (*sei*), “fine” and “end” (*fine*), “rice” and “laughter” (*riso*), “know” and “taste” (*sapere*), identification becomes difficult when an unbroken chain of semantic derivation can still be perceived between homographs such as “plan”, “plain”, “plane”, and “floor” (*piano*), or “feeling”, “sense”, “common sense”, and “direction” (*senso*). Identification is even more involved in the case of homographs with the same lexical meaning which assume different functions in the sentence: subject or predicate (*potere*), predicate or noun modifier (*amato*), noun modifier or verb modifier (*tanto*), verb modifier or clause governor (*abbastanza*), etc; or which express different grammatical categories: indicative or imperative (*canta*), first person or second person (*parli*), singular or plural (*virtù*), masculine or feminine (*fine*), subject or object (*noi*), direct object or indirect object (*mi*), personal or impersonal (*ci*), etc. Are differences in syntactic function or grammatical category sufficient to justify the interpretation of functional homographs as occurrences of different words, to be entered under different headings and computed separately? Or is it preferable to consider them variants of the same invariants, to be entered and totalized under the same heading?

Since arguments can be found to support either interpretation, the answer will depend on the purpose of the investigation, e.g., the grammatical exploitation of word counts. If coefficients of word frequency are to be structurally revealing, then surely they must provide a sound basis for predictions about the share of classes and categories expressed by Italian words in representative texts. To this end, homographs are a problem, as no sound inferences about word classes can be drawn from coefficients which fail to distinguish between noun and verb, adverb and adjective, preposition and conjunction, etc., just as no accurate predictions can be made about grammatical categories from data which fail to discriminate between singular and plural, masculine and feminine, indicative and imperative, subject and object, 1st person and 2nd person, etc.

The solutions we have adopted in treating such ambiguities differ according to the types of homographs, lexical, syntactic, or morphologic. LEXICAL homographs are identical occurrences of different members of the same class: *fine* “end” and *fine* “objective”, *aria* “air” and *aria* “aria”, *natura* “nature” and *natura* “character”, *provare* “try” and *provare* “demonstrate”, *povero* “poor” and *povero* “wretched”, etc. SYNTACTIC homographs are identical occurrences of members of different classes, such as *critica* “critique” and *critica* “critical”, *invitato* “guest” and *invitato* “invited”, *letto* “bed” and *letto* “read”, *misero* “miserable” and *misero* “put”, *marcia* “walk” and *marcia* “walks”, etc. MORPHOLOGIC homographs are identical occurrences of the same word which express different grammatical categories such as *virtù* “virtue” and *virtù* “virtues”, *canta* “he sings” and *canta* “sing!”, *parli* “that I speak” and *parli* “that he speak”, *mi* “me” and *mi* “to me”, *noi* “we” and *noi* “to us”, etc.

3.2221 *Lexical homographs.* — Lexical homographs are identical occurrences of different members of the same class. They involve mostly nouns, although there are also instances of lexical homography among verbs, adjectives, and pronouns. Certain lexical homographs may be distinguished with reference to grammatical characteristics such as gender (*il fine* “the purpose” and *la fine* “the end”, *il nipote* “nephew” and *la nipote* “niece”, *il braccio* and *le braccia* versus *il braccio* and *i bracci*), morphological class (*presento* “present” from *presentare* and *presento* “forebode” from *presentire*), or syntactic characteristics (*povero amico* “wretched friend” and *amico povero* “poor friend”, etc.).

3.22211 *Nouns.* — Nominal homography involves two masculines, or two feminines, or one



masculine and one feminine noun.

A. *Masculines*. — Homographic masculines which appear in our dictionary are: *caffè* “coffee” and “coffee house”; *maestro* “master” and “teacher”; *mezzo* “means”, “middle” and “half”; *pezzo* “piece” and “musical composition”; *piano* “plan”, “plane”, “plain” and “floor”, *paese* “country” and “town” (also *paesi*); *principio* “principle” and “beginning”; *senso* “sense”, “common sense”, “moral sense”, “feeling”, and “direction”; *stato* “state” and “condition”; *studio* “study” and “study room”; *tempo* “time” and “weather”.

Masculine homographs in our corpus at least one of which failed to qualify are: *accordo* “agreement” and “harmony”; *affare* “business” and “affair”; *albero* “tree” and “mast”; *angolo* “angle” and “corner”; *bagno* “bath” and “bathroom”; *bordo* “edge” and “(on) board”; *brivido* “shiver” and “thrill”; *campione* “champion” and “sample”; *carbone* “coal” and “carbon”; *ciglio* “eyelash” and “edge”; *cinese* “Chinese man” and “Chinese language”; *corso* “street”, “course”, and “circulation”; *dente* “tooth” and “cog”; *ferro* “iron” and “horseshoe”; *fondo* “bottom” and “fund”; *genere* “kind”, “gender” and “genre”; *giudizio* “judgment” and “mind”; *giuoco* “sport” and “gambling”; *impasto* “mixture” and “impasto”; *indice* “index” and “index finger”; *indirizzo* “address” and “direction”; *legno* “wood” and “ship”; *libretto* “small book” and “libretto”; *membro* “limb” and “member”; *nodo* “nautical measure” and “knot”; *nome* “name” and “noun”; *sacco* “sack” and “bag”, “pillage”; *scurio* “darkness” and “shutter”; *stadio* “stage”, “period” and “stadium”; *tipo* “type” and “person”.

B. *Feminines*. — Homographic feminines in the basic vocabulary are: *aria* “air” and “aria”; *nature* “character” and “nature”; *prova* “proof” and “rehearsal”; *maestà* “majesty” and “Majesty”; *signora* “lady” and “Madam”; *società* “association” and “society”; *storia* “history” and “story”; *voce* “voice” and “word”.

Feminine homographs in our corpus at least one of which failed to qualify are: *altezza* “height” and “Highness”; *arte* “art” and “craft trade”; *azione* “action” and “stock”; *banda* “side”, “band”, and “gang”; *borsa* “purse”, “bag”, and “exchange”; *camera* “room” and “camera (oscura)”; *campagna* “country” and “campaign”; *carica* “position, rank,” and “charge”; *carta* “paper” and “playing card”; *cateratta* “waterfall” and “cataract”; *causa* “cause” and “court cause”; *cittadina* “female citizen of a nation”, “female citizen of a city”, and “small city”; *dama* “woman” and “checkers”; *divisione* “a dividing” and “military and arithmetic operation”; *eccellenza* “excellence” and “Excellency”; *facoltà* “(mental) power” and “school”; *forma* “form” and “mold”; *gravità* “importance” and “gravity”; *gru*, “bird” and “machine”; *influenza* “influence” and “influenza”; *lira*, “currency”, and “lyre”; *lingua* “tongue” and “language”; *macchia* “spot” and “trees”; *nota*, written and musical, and “bill” (of costs); *pena* “pain” and “punishment”; *penna* “pen” and “feather”; *poesia* “poem” and “poetry”; *rivista* “magazine” and “parade”; *stampa* “press” and “print”; *tenda* “tent” and “curtain”; *usura* “wear and tear” and “usury”; *voce* “voice”, “rumor”, and “word”.

C. *Masculines and Feminines*. — The only homographic nouns of different genders which qualified for our dictionary are *amante* “lover” and “mistress”, *fronte* “front line” and “forehead”, *morti* “deaths” and “dead ones”, *signore* “ladies” and “Sir!”

However, *capitale* “capital (funds)” and “capital (city)”, *fine* “purpose” and “end”, *fronte* “forehead” and “front line”, *morale* “spirit” and “ethics”, *nipote* “nephew” and “niece”, and *signore* “ladies” and “gentleman” have occurred in our corpus.

3.22212 *Adjectives*. — Only one instance of homography between adjectives appears in our dictionary: *povero* “poor” (*amico povero*) and “wretched” (*povero amico*). Others occurred in our corpus, but failed to qualify: *capace* “capable” and “capacious”; *capitale* “fund, capital”, “death

## INTRODUCTION

(sentence),” and “chief, main”; *caro* “expensive” and “dear”; *familiare* “familiar” and “familial, (pertaining to) the family”; *grave* “grave” and “heavy”; *occupato* “busy” and “filled”; *romantico* “Romantic” and “romantic”; *secolare* “secular, worldly” and “secular, old”; *sinistro* “left” and “sinister”.

3.22213 *Pronouns*. — Instances of lexical homography between occurrences of different pronouns are: *me* (me ne manda, me lo metto) of *mi* (2) and *me* (chiamano me, lavorano per me) of *me* (3), *te* (te ne manda, te lo metti) of *ti* (5) and *te* (chiamano te, lavorano per te) of *te* (6), *gli* (gli parlo) of *lo* (8) and *gli* (gliene manda) of *la* (11), *lei* (chiamano lei) of *lei* (8) and *lei* (polite), *c’* (c’invita) of *ci* (14) and *ce* (ce ne sono) of *ci* (impers.), *ci* (ci conosce, ci manda, ci laviamo) of *ci* (14) and *ci* (ci sono) of *ci* (imp.), *v’* (v’invita) of *vi* (17) and *vi* (vi sono), of *vi* (impers.), *vi* (vi chiama, vi parla, vi lavate) of *vi* (17) and *vi* (impers.), *loro* (lavora per loro, i ragazzi) of *loro* (21) and *loro* (lavora per loro, le ragazze) of *loro* (24).

3.22214 *Verbs*. — Instances of verbal homography in our dictionary are: *provare* “try” and “demonstrate”, *presento* (*presentare* and *presentire*), *sentire* “feel” and “hear”. To these we must add: *avere* and *essere* as regular verbs and auxiliary verbs; and *dovere*, *potere*, *volere*, *sapere* as regular and modal verbs.

Verbal homographies which have failed to qualify are *accusare* “accuse” and “acknowledge”; *alterarsi* “be altered” and “get angry”; *andare* “go” and “fit”; *assistere* “attend” and “help”; *caricare* “charge” and “load”; *comprendere* “understand” and “include”; *contare* “tell”, “count on”, “depend”, and “to be of account” (intrans.); *dare* “give” and “open, face”; *durare* “last” and “endure”; *edificare* “build” and “edify”; *esperimentare* “experiment” and “experience”; *girare* “turn” and “shoot (a movie)”; *importare* “matter” and “import”; *intendere* “intend”, “mean”, and “understand”; *sentire* “hear”, “feel”, and “taste”; *passare* “pass”, “spend time”; *portare* “carry” and “wear”; *riflettere* “think” and “reflect”; *scoprire* “discover” and “uncover”; *spiegare* “explain” and “unfold”; *tentare* “try” and “tempt”.

3.2222 *Syntactic*. — Syntactic homographs are identical occurrences of members of different word classes: the same shapes, often with the same lexical meaning, are distinguished by sentence function, e.g.: *abbandono*, *accordi*, *arma*, *chiese*, *combattente*, *componenti*, *coperte*, *date*, *durata*, *emigrato*, *fabbricati*, *fossi*, *legge*, *passante*, *proposte*, *sale*, etc., may be nouns or verbs; *agitato*, *brillante*, *calmi*, *definiti*, *misero*, *perduta*, *popolare*, *secca*, *unite*, *vivente*, etc., may be adjectives or verbs; *amico*, *borghese*, *critica*, *diritti*, *fedeli*, *fine*, *morte*, etc., may be nouns or adjectives; *amante*, *dato*, *precedenti*, *vedute*, etc., may be nouns, adjectives, or verbs. Among functional words, *che* may be a conjunction or a pronoun, *ne* a preposition or a pronoun, *nonostante* a conjunction or a preposition, *ove* an adverb or a pronoun, *la* an article or a pronoun, *benché* an adverb or a conjunction, etc.

A predominantly substantial interpretation would consider these forms as occurrences of one and the same word with nominal functions in certain constructions, verbal functions in certain others, etc.; and a predominantly formal interpretation would consider them identical occurrences of different words, one a noun, the other a verb, etc. The first interpretation, which can be defended on non-formal grounds, either substantial (phonetic, semantic) or historical, compromises the grammatical exploitation of the data: no accurate predictions can be made about the relative share of word classes if the subject occurrences of *abbandono*, for instance are not distinguished from its predicate occurrences. Instead of allowing the same word to assume sometimes nominal and sometimes verbal functions, the formal interpretation postulates different words which occur under the same shape. A method of interpretation which prohibits words from belonging to more than one class—every word belongs to one class and no word belongs to more than one—illustrates the

paradoxical situation of units whose identification depends on their classification. The formal solution has the additional advantage of being easily convertible into its substantial counterpart, whereas the contrary is not true: the coefficients of words capable of assuming different roles in the sentence can be easily determined by computing the homographic occurrences of the noun and of the verb, while the occurrences of the noun, for instance, cannot be extracted from coefficients established through a method which does not distinguish between the subject and predicate functions of the same shape.

Most word counts have bypassed the difficulties involved in defining word classes by computing coefficients for shapes rather than forms: parts-of-speech abbreviations were added after the word count was completed, when the function of homographs taken out of context could no longer be determined. Unresolved by multiple class markers such as “*n.* and *v.*”, syntactic ambiguities are “resolved” by retaining the function assumed to be the more frequent. The question of word classes is subject to a variety of solutions, the appropriateness of classification depending on intended use. We have discussed elsewhere the principles underlying the major approaches to word classes in a statistical context,<sup>15</sup> the morphologic solutions in 3.2221,<sup>16</sup> the syntactic solutions in 3.2222,<sup>17</sup> and the morphosyntactic solutions in 3.2223.<sup>18</sup> We may now outline the system of criteria integrated in a hierarchy which provides the definitions we have used to resolve instances of syntactic homography. Basically we assume that investigations whose results are to provide sound insights into the structural properties of representative texts must: 1) subordinate substantial (phonetic, semantic) to formal (grammatical) criteria, and 2) among the latter, syntactic relations take precedence over considerations of morphologic constituency.

The FUNCTION of words in the sentence is the most fundamental criterion for assigning words to classes; the basic *constituents*, subject and predicate, as well as their objects and their modifiers; and the *function words* which govern, introduce, or link them. Words which function as *subjects* or *objects* are likely to be nouns or pronouns; *predicates* are verbs or adjectives; *modifiers of nouns* are articles or adjectives; *modifiers of verbs* are adverbs; *noun substitutes* are pronouns; functional words which *introduce nouns or nominal phrases* are prepositions, those which *govern verbs, clauses, or sentences* are conjunctions. Next are the grammatical CATEGORIES expressed by words: *gender* characterizes nouns, articles, adjectives, pronouns, exceptionally verbs; *number* is manifested by nouns, articles, adjectives, pronouns, and verbs; *person* characterizes pronouns and verbs; *case* is associated with nouns and pronouns; *mode* and *tense* characterize exclusively verbs, just as a category such as *superlative* is associated exclusively with adjectives. Finally, there is a variety of morphologic criteria based on morphemes associated with particular classes: among suffixes *=zione* characterizes nouns, *=mente* characterizes adverbs, *=issim=* adjectives, etc.; among inflexions plural *=i* characterizes nouns and adjectives, *=are* verbs, etc.

Structural definitions of parts of speech can be derived from this hierarchy of criteria. *Nouns*, exhibit gender and number, function as subjects or objects and are modified by adjectives, determined by articles, governed by prepositions, replaced by pronouns, extended by certain suffixes (*=zion=*), characterized by certain flexives (*=i*). *Verbs*, express mode, tense, number, and person, function as predicates and are modified by adverbs, introduced by conjunctions, extended by certain suffixes (*=iz=*), characterized by certain inflexions (*=are*, etc.). *Articles*, exhibit gender and number, determine nouns. *Adjectives*, exhibit gender and number, modify nouns and may function

<sup>15</sup> FDRW, 3.222 Syntactic homographs, pp. xxix-xxxviii.

<sup>16</sup> pp. xxx-xxxi

<sup>17</sup> pp. xxxi-xxxiii

<sup>18</sup> pp. xxxiii-xxxvi

## INTRODUCTION

as predicates, are extended by the suffix *-issim-* and characterized by certain inflexions (*=o*, *=i*, *=a*, *=e*). *Pronouns*, exhibit gender, number, person, and case, function as subjects or objects and replace nouns. *Prepositions* govern nouns and nominal phrases. *Conjunctions* introduce verbs, clauses, sentences. *Interjections* and numerals form an easily identifiable (peripheral) class.

These criteria and definitions may be used to resolve instances of syntactic ambiguity by assigning homographic shapes to different classes. Note that certain criteria are exclusive, while others are not. On the one hand, mode and tense characterize verbs unambiguously; only conjunctions introduce verbs, clauses, sentences; pronouns are the only words which replace nouns, and they are the only words which can express case; adverbs are exclusively characterized by the suffix *-mente*, adjectives by the suffix *-issim-*, etc. On the other hand, both nouns and pronouns may function as subjects or objects, both verbs and adjectives may be the predicate of a sentence, both articles and adjectives may modify nouns; verbs as well as pronouns exhibit person; nouns, adjectives, articles, and certain pronouns and verbs exhibit gender and number; adjectives and nouns share certain inflexions in common (plural *=i*); etc. If the usefulness of exclusive criteria is obvious, the others may be used in combination, to swing the balance one way or the other.

Note that certain definitions are circular: an adjective is defined as a noun modifier, a noun as a form modified by an adjective; prepositions are forms that govern nouns, while nouns are defined as forms governed by prepositions, etc. Reciprocal definitions based on modifier/modified relations can be used to determine the status of an ambiguous term with reference to its unambiguous counterpart, either by an exclusive criterion or by a combination of non-exclusive criteria: to decide whether a participial form is a verb or a noun, determination by an article or modification by an unambiguous adjective (a superlative) may decide in favor of the latter; conversely, when intrinsic considerations do not allow us to decide whether a modifier is an adjective or an adverb, the fact that it modifies an unambiguous noun (the subject of the sentence determined by an article) may decide in favor of the former.

The preceding system of criteria and definitions underlies a classification which resolves instances of syntactic homography involving mostly lexical words, nouns, verbs, and (lexical) adjectives. Functional words (including functional adjectives and adverbs) are also involved, generally among themselves, but also with lexical words (mostly nouns).

3.22221 *Nouns and Verbs*. — Instances of homography between nouns and verbs may be classified relative to the grammatical category they express, indicative (subjunctive, imperative), infinitive, present participle, and past absolute for verbs, masculine or feminine, singular or plural for nouns. The most frequent verbal forms are indicatives and past participles, but there are also infinitives, present participles, and past absolutes.

### A. Indicative (Subjunctive, Imperative)

#### 1. Masculine Singular:

*=are:* abbandono, abito, affetto, acquisto, aiuto, arrivo, ascolto, aspetto, bacio, ballo, calcolo, cambio, cammino, canto, consiglio, disprezzo, disturbo, dono, esito, fondo, fumo, giro, grido, guadagno, inganno, inizio, invito, lamento, lavoro, medico, merito, mostro, passo, porto, racconto, regalo, respiro, resto, richiamo, ricordo, rifiuto, riguardo, rimprovero, rischio, risparmio, risveglio, ritorno, salto, saluto, sbaglio, scambio, scoppio, sogno, sospetto, sposo, studio, taglio, tiro, tratto, viaggio, \*perdono.

*=ire:* sale, servo

#### 2. Feminine Singular:

*=are:* arma, bestemmia, caccia, calma, causa, condanna, consegna, corona, costa, cura, disputa, domanda, fabbrica, faccia, figura, fila, firma, forma, guida, insegna, lega, leva, marcia, meraviglia, minaccia, mira, misura, mostra, nota, opera, paga, pianta, piega, porta, posa, predica, protesta, prova, rassegna, regola,

## INTRODUCTION

ricerca, riforma, riserva, rovina, scusa, serra, sosta, sposa, squadra, stampa, traccia, turba, vergogna, visita, volta.

=*ere*: era, legge, tenda (subj)

=*ire*: veste

### 3. Masculine Plural:

=*are*: abiti, accenni, accordi, affetti, aiuti, appoggi, appunti, arrivi, aspetti, attacchi, baci, cambi, canti, cerchi, consigli, conti, depositi, disturbi, gridi, impegni, inizi, lavori, meriti, mostri, ordini, passi, porti, registri, resti, ricordi, rimproveri, riposi, risvegli, ritorni, scherzi, scoppi, segni, sfoghi, sogni, sospiri, spaventi, sposi, suoni, tagli, tiri, tormenti, tratti, usi, versi, voli\*

=*ere*: fossi

### 4. Feminine Plural:

=*are*: \*basi, date, urla

=*ere*: leggi, tende

=*ire*: vesti

## B. Past Participle

### 1. Masculine Singular:

=*are*: emigrato, fabbricato, impiegato, peccato, risultato, trattato

=*ere*: letto, messo, percorso, permesso, pianto, posto, presupposto, prodotto, ricorso, scelto, sorriso, stato\*, successo, volto.

=*ire*: aperto, partito

### 2. Feminine Singular:

=*are*: chiamata, durata, entrata, ritirata.

=*ere*: battuta, caduta, condotta, difesa, discesa, messa, posta, presa, promessa, proposta, resa, richiesta, ripresa, scossa, sorpresa, spinta, tenuta

=*ire*: aperta, salita

### 3. Masculine Plural:

=*are*: fabbricati, impiegati, risultati, trattati

=*ere*: caduti, dipinti, messi, posti, risultati, tratti

=*ire*: partiti

### 4. Feminine Plural:

=*are*: chiamate, date

=*ere*: battute, cadute, messe, poste, proposte, raccolte, sorte, spinte, tenute

=*ire*: aperte, coperte

## C. Present Participle

### 1. Singular:

=*are*: abitante, passante

=*ere*: combattente, corrente

### 2. Plural:

=*are*: abitanti, passanti

=*ere*: componenti, concorrenti

## D. Past Absolute

=*ere*: chiese, pose, riprese\*, scosse\*, sorprese\*, spese

## E. Infinitive

=*are*: fare

=*ere*: dovere, esser(e), potere, volere

=*ire*: avvenire, morire

## INTRODUCTION

In a few instances, homography involves one nominal and two verbal forms: *ricorsi* and *sorrisi*, as well as *difese*, *pose*, *riprese*, *scesse*, *sorprese*, and *spese* may be a noun, a past participle, and a past absolute, whereas a form such as *date* occurs as a noun, as a present indicative and as a past participle.

3.22222 *Verbs and Adjectives*. — Verbal forms involved in syntactic homography with adjectives are mostly past participles, but indicatives (subjunctives, imperatives), past absolutes, infinitives, and present participles have been also registered. Adjective forms are masculine or feminine, singular or plural.

### A. Past Participle

#### 1. Masculine Singular

**=are:** abbassato, accennato, agitato, alzato, amato, armato, avanzato, bagnato, conservato, considerato, desiderato, destinato, determinato, dimenticato, elevato, illuminato, indicato, isolato, limitato, mancato, mutato, obbligato, preparato, rassegnato, rialzato, rilevato, rinnovato, trasformato, turbato, usato

**=ere:** acceso, avvolto, chiuso, commosso, composto, compreso, confuso, convinto, deciso, diffuso, diretto, disposto, distinto, distrutto, diviso, dovuto, esteso, nato, nascosto, perduto, preso, raccolto, retto, ridotto, ripetuto, rotto, sciolto, seduto, sospeso, sparso, spento, stretto, vinto, visto.

**=ire:** definito, detto, distribuito, finito, fiorito, maledetto, smarrito, stupito, unito, scuito

#### 2. Feminine Singular:

**=are:** abituata, accennata, agitata,alzata, amata, armata, avanzata, conservata, considerata, destinata, determinata, dimenticata, elevata, fatta, illuminata, indicata, limitata, mutata, obbligata, passata, preparata, rilevata, rinnovata, serrata, sod(d)isfatta, trasformata, turbata

**=ere:** accesa, attesa, chiusa, composta, compresa, confusa, contenuta, diffusa, dipinta, diretta, disposta, distrutta, divisa, dovuta, nata, opposta, perduta, presa, ridotta, rotta, sciolta, scritta, sospesa, spenta, veduta, vinta, vista

**=ire:** definita, detta, distribuita, finita, fiorita, smarrita

#### 3. Masculine Plural:

**=are:** abituati, amati, armati, avanzati, conservati, considerati, destinati, determinati, dimenticati, elevati, indicati, isolati, obbligati, passati, preparati, usati

**=ere:** avvolti, composti, convinti, diretti, distinti, distrutti, dovuti, nati, nascosti, opposti, perduti, raccolti, ripetuti, riuniti, rotti, spenti

**=ire:** detti, finiti

#### 4. Feminine Plural:

**=are:** alzate, amate, armate, bagnate, destinate, determinate, elevate, fatte, indicate, interessate, isolate, limitate, passate, serrate, usate.

**=ere:** attese, contenute, dipinte, dirette, dovute, nate, opposte, perdute, ridotte, sciolte, scritte, spente, sperdute, strette, viste

**=ire:** definite, finite, riunite, unite

### B. Indicative (Subjunctive, Imperative)

#### 1. Masculine Singular:

**=are:** carico, contento, continuo, fisso, muto, noto, \*pagano, secco

#### 2. Feminine Singular:

continua, dura, fissa, improvvisa, libera, matura, muta, oscura, presta, pubblica, secca, sveglia, varia

### C. Present Participle

**=are:** abbondante, brillante, dominante, rilevante

**=ere:** crescente, vivente

## D. Infinitive

leggere, popolare

## E. Past Absolute

misero

A few homographs involve an adjective with a past participle and a past absolute: masculines such as *accesi*, *chiusi*, *compresi*, *divisi*, *presi*, and *sparsi*, as well as feminines such as *accese*, *attese*, *chiuse*, *comprese*, *decise*, *diffuse*, *divise*, *sospese*, and *sparse*.

3.22223 *Adjectives and Nouns*. — Homographs which involve adjectives and nouns are masculine or feminine, singular or plural.

## 1. Masculine Singular:

- =e: borghese, breve, capitale, centrale, colpevole, comune, folle, francese, generale, giovane, giovine, grande, grave, ideale, industriale, inglese, mal(e), materiale, militare, morale, nobile, onorevole, particolare, personale, presente, principale, provinciale, seguente, simile, superiore, ufficiale, verde
- =o: alto, americano, amico, antico, atto, azzurro, basso, bello, bianco, britannico, buio, buono, caldo, caro, chiaro, cieco, cittadino, comodo, complesso, contrario, critico, deserto, diritto, disperato, esterno, estero, estraneo, fisico, franco, freddo, fresco, futuro, giusto, grasso, greco, grosso, ignoto, improvviso, infinito, innamorato, intento, italiano, largo, latino, legato, maestro, maschio, massimo, mezzo, minuto, moderno, morto, necessario, nemico, nero, nord, nuovo, opportuno, pazzo, piccolo, pieno, povero, profondo, prossimo, quarto, ricco, rosso, russo, saggio, santo, segreto, serio, soggetto, sol(o), sommo, straniero, tecnico, tedesco, terzo, tessuto, turchino, vecchio, vero, vestito, vivo, vuoto
- =i: pari

## 2. Feminine Singular:

- =e: fine
- =a: carica, critica, curva, destra, ferita, fiera, legata, libera, logica, matematica, novella, politica, retta, rosa, sant(a), sinistra, somma, stanca, tecnica, vecchia

## 3. Masculine Plural:

- =e: borghesi, capitali, colpevoli, comuni, fedeli, francesi, generali, giovani, grandi, gravi, ideali, industriali, inglesi, mali, materiali, militari, morali, morti, nobili, principali, provinciali, seguenti, simili, superiori, ufficiali, verdi
- =o: americani, amici, antichi, atti, bassi, belli, bianchi, britannici, buoni, caldi, cari, ciechi, comodi, complessi, critici, deserti, diritti, dubbi, equivoci, esterni, esteri, estranei, fini, freddi, freschi, giusti, grassi, greci, grossi, ignoti, infiniti, intenti, italiani, latini, legati, liberi, lontani, lustri, maschi, massimi, medici, mezzi, minuti, moderni, necessari, nemici, nuovi, piccoli, pieni, poveri<sup>2</sup>, profondi, pubblici, ricchi, rossi, russi, saggi, santi, segreti, seri, soggetti, stranieri, tecnici, tedeschi, terzi, tessuti, turchi, vecchi, vestiti, vicini, vivi, vuoti

## 4. Feminine Plural:

- =o: corte, critiche, fiere, legate, matematiche, morte, note, novelle, politiche, pratiche, scoperte, sedute, serie, sole, vecchie

3.22224 *Adjectives, Nouns, and Verbs*. — Some 25 forms appear three times in our dictionary, as adjectives, as nouns, and as verbs. The adjectives and nouns are masculine or feminine, singular or plural, whereas the verbs are indicatives (*calmo*, *pubblica*), present participles (*amante*, *precedente*), or past participles.

## 1. Masculine Singular:

- =e: amante, precedente, presente
- =o: calmo, contenuto, dato, fermo, dipinto, fatto, ferito, opposto, passato, \*san, scritto

## INTRODUCTION

### 2. Feminine Singular:

data, nota, distesa, pratica, pubblica, raccolta, scelta, seduta, stretta, uscita

### 3. Masculine Plural:

=e: precedenti, presenti

=o: dati, fatti

### 4. Feminine Plural

distese, ferite, note

3.2225 *Functional Words*. — Analytically, words of inflexional languages can be divided into three: constituents of the sentence (nouns, verbs), their modifiers (adjectives, adverbs), and the functional words which govern or introduce them (prepositions, conjunctions). Synthetically, the same words are divided into lexical and functional: lexical words consist of a stem carrying lexical meaning, and an ending expressing grammatical functions or categories, whereas functional words, which cannot be normally analyzed into such constituents, express relational meaning. In this scheme, modifiers occupy a borderline position and must be subdivided accordingly. *Lexical adjectives* consist of a stem expressing a quality or a property, and an ending (suffixes like *-issim-*; flexives like *=o*, *=a*, etc.), whereas *functional adjectives* (modifiers of nouns) cannot be so analyzed and express relational, quantitative, deictic, etc. meanings. *Lexical adverbs* consist of a stem expressing the property of an action and the suffix *-mente*, whereas *functional adverbs*, which cannot be so analyzed, express relational, quantitative, deictic, etc. meanings. Notice that functional adjectives (and sometimes adverbs) can also modify modifiers, which is not the case for lexical adjectives. More significant in this context, lexical modifiers are involved in homography with lexical words (see preceding sections), whereas functional modifiers occur in shapes that are homographic with functional words, prepositions, conjunctions, pronouns. Adjectives and adverbs considered in this section are of the functional kind.

#### A. Adjective and Adverb

abbastanza, alquanto, altrettanto, appena, assai, forte, molto, proprio, pure, quasi, spesso, tardi, tutti

#### B. Adjective and Pronoun

altro etc., ciascuna, ciascuno, cui, loro, miei, nessuno, niente, nostro etc., parecchi, parecchio, etc., pochi, pochissimo, primi, primo, quali, qualunque, quanta, quante, quanti, quella, quelle, quello, questo etc., stesso, sua, sue, suo, tale, tali, taluni, taluno, tanta, tante, tanti, terza, troppi, tutti, ultimo, etc.

#### C. Adverb and Conjunction

benchè, comunque, dunque, quando, intanto, mentre, poichè, sebbene, tuttavia

#### D. Adverb and Preposition

accanto, addosso, attorno, davanti, dentro, dietro, dinanzi, dopo, fino, innanzi, invece, perfino, persino, presso, senza, sino

#### E. Adverb and Pronoun

v', ve, vi

#### F. Article and Pronoun

l', la, le, lo, una, uno

#### G. Conjunction and Pronoun

che



## H. Conjunction and Preposition

nonostante

## I. Conjunction and Interjection

o, se

There are instances of three-way homography among functional words: *come* may be an adverb, a preposition, or a conjunction; *oltre* may be an adjective, an adverb, or a preposition; *più* may be an adjective (or a modifier of an adjective), an adverb (or a modifier of an adverb), or a pronoun; *quale* may be an adjective, a conjunction, or a pronoun.

3.22226 *Functional words and Lexical words.* — There are several instances of homography between functional words and lexical words, mostly nouns.

a) adverbs: *domani, giù, ieri, no, oggi, \*onde, ora, poi, qua, specie, via*

b) numerals: *due, cento, cinque, tre, dieci, sette, venti*, to which designations of centuries such as *cinquecento* must be added.

c) prepositions: *colle, confronto, verso, salvo*

d) pronouns: *io, si*

An example of homography between a functional word and a verb is *sei* (numeral).

There are examples of three-way homography between two functional words and one lexical word: *appunti, avanti, fuori, intorno, sotto, sopra*, and *su* may be an adjective, a preposition, or a noun; *bene, certo, lontano, male, piano*<sup>4</sup>, and *vicino* may be an adjective, an adverb, or a noun; *incontro* may be a preposition, a noun, or a verb; *presto* may be an adjective, an adverb, or a verb.

Four-way homographies are *lungo*, which may be an adjective, an adverb, a preposition, or a noun; *o* may be a conjunction, an interjection, a preposition, or a verb; *prima*, a preposition, an adverb, a noun, or a pronoun; *tanto*, an adjective, an adverb, a conjunction, or a noun; etc.

3.2223 *Morphologic.* — Morphologic homographs are identical occurrences of different variants of the same invariant: the same shapes with the same lexical meaning express different grammatical categories such as masculine/feminine, singular/plural, present/past, subject/object, 1st/3rd person, direct/indirect object, indicative/imperative, etc. For masculine and feminine compare *l'uomo* and *l'uscita* with *il padre* and *la madre*, *uomo vivace* and *donna vivace* with *bel bambino* and *bella bambina*; for singular and plural *la specie* and *le specie* with *la sedia* and *le sedie*; for direct and indirect object *Giorgio ti conosce* and *Giorgio ti scrive* with *Giorgio lo conosce* and *Giorgio gli scrive*; for indicative and subjunctive, *tu studi* and *che tu studi* with *tu scrivi* and *che tu scriva*; for 1st and 3rd person, compare *che io finisca*, *che tu finisca*, *che egli finisca* with *io finisco*, *tu finisci*, *egli finisce*; etc.

Morphologic homographs occur in the paradigms of articles, adjectives, nouns, pronouns, and verbs.

3.22231 *Articles.* — The masculine and feminine, singular and plural are homographic in the elided form *l'* of the definite articles, for which compare *l'uomo* and *l'uscita* with *il padre* and *la madre*. In the indefinite forms, the homophonous masculine and feminine singular are kept graphically apart by the apostrophe, *un uomo* but *un'ora*.

3.22232 *Adjectives.* — The masculine and feminine variants of adjectives in *-e* are homographic: compare *allievo intelligente* and *allieva intelligente* with *allievo pigro* and *allieva pigra*. The singular and plural of such adjectives as *marrone* or *viola* are also identical, *libro marrone* and *libri marrone* but *libro bianco* and *libri bianchi*, *cappello viola* and *cappelli viola* with *cappello bianco* and *cappelli bianchi*. For instances of syntactic homography between adjectives, compare *grand'uomo* and *grand'isola*.

## INTRODUCTION

3.22233 *Nouns*. — The singular and plural of several Italian nouns, especially those ending in a stressed vowel, are homographic:

=à: abilità, attività, bontà, brutalità, carità, cavità, celebrità, città, civiltà, collettività, comodità, difficoltà, diversità, entità, eredità, età, facilità, felicità, inferiorità, località, maestà, metà, modalità, necessità, neutralità, novità, opportunità, ospitalità, ostilità, papà, personalità, possibilità, probabilità, profondità, proprietà, qualità, realtà, semplicità, sensibilità, serenità, serietà, sincerità, società, solidarietà, sovranità, stabilità, universalità, utilità, varietà, velocità, verità, volontà

=e: re, tè, specie

=i: crisi, estasi, ipotesi, mezzodi, oasi, tesi

=ù: gioventù, servitù, tribù, virtù

Add *film*, with three plurals, *film*, *filmi* and also *films*.

3.22234 *Pronouns*. — On the basis of our interpretation, morphologic homography between variants of the same pronouns (expressing different categories through the same shape) are between the direct object and prepositional forms of the personal pronouns, e.g., *vede me* and *lavora per me*, of **me**, **te**, **lui**; between the subject and prepositional form of **essi** (*essi vedono*, *lavora per essi*) **esso**, **essa**, **esse**, **lei**, **noi**, **voi**, **loro**; between the direct, indirect, and reflexive forms of **mi** (*mi vede*, *mi parla*, *mi lavo*), **ti**, **si**, **ci**, **vi**; between the direct and reflexive forms of **mi** (*me lo dice*, *me lo lavo*).

3.22235 *Verbs*. — Instances of verbal homography involve forms which express different numbers and/or person within the paradigm of the same tense, or forms belonging to different modes and/or tenses.

A. *Person/Number*: The first, second, and third person singular forms of the present subjunctive forms are homographic in the three conjugations, e.g., *ami*, *venda*, *parta*; the first and second persons of the imperfect subjunctive of all verbs, e.g., *amassi*, *vendessi*, *partissi*. The first and sixth persons of the present indicative of **essere**: (*io*, *essi*) *sono*.

B. *Mode/Tense*

1. *Indicative/Imperative*. Instances of homography occur between the third and the fifth persons of the present indicative and the second person of the imperative in the first conjugation verbs: (*egli*) *ama* and *ama!*, (*voi*) *amate* and *amate!* Other instances involve the second person singular and the second person plural of the present indicative and the parallel forms of the imperative in the second and third conjugation verbs: (*tu*) *vendi*, *vendi*, (*voi*) *vendete*, *vendete* and (*tu*) *parti*, *parti*, (*voi*) *partite*, *partite*.

2. *Indicative/Subjunctive*. Homography occurs between the fifth persons of the past absolute indicative and the imperfect subjunctive: (*che*) *voi amaste*, (*che*) *voi vendeste*, (*che*) *voi partiste*.

3. *Subjunctive/Imperative*. The first, second and third person and the third person plural forms of the present subjunctive and the forms of the imperative of all conjugation verbs: (*che io*, *tu*, *egli*) *ami*, (*che io*, *tu*, *egli*) *venda*, (*che io*, *tu*, *egli*) *parta*, (*che essi*) *amino*, (*che essi*) *vendano*, (*che essi*) *partano*.

4. *Indicative/Subjunctive/Imperative*. The third singular imperative form is homographous with the second singular form of the present indicative and the three singular persons of the present subjunctive in first conjugation verbs: *ami*, (*tu*) *ami*, (*che io*, *tu*, *egli*) *ami*. The first plural imperative form is homographous with the parallel forms of the present indicative and the present subjunctive: *amiamo*, (*noi*) *amiamo*, (*che noi*) *amiamo*; *vendiamo*, (*noi*) *vendiamo*, (*che noi*) *vendiamo*; *partiamo*, (*noi*) *partiamo*, (*che noi*) *partiamo*.

To distinguish cases of homography between forms of the indicative, the subjunctive and the imperative, the superscripts <sup>2,3</sup> have been used for the second and third person forms of the present subjunctive and the abbreviation *i*. has been post-scribed to the imperative forms.

## 4. COEFFICIENTS

The results of word counts are usually presented as graded inventories in which words are ranked in decreasing order of frequency or dispersion, with the resulting hierarchies mechanically divided into numerical classes: first hundred or thousand, etc.

■ The value of such hierarchies and classifications depends on the validity of the sample and on the appropriateness of the technique used to extract indices of word usage from the raw frequencies. In earlier studies, usage was equated with frequency on the assumption that the number of occurrences of a word in the sample reflects its use in the language as a whole. Words were ranked in terms of coefficients equal or proportional to the number of their occurrences in the sample. Quantifying techniques were refined when it appeared that results based on this assumption often failed to correspond to the intuitive estimates of specialists, or to satisfy the intended purpose. To remedy the shortcomings of straight frequency, statistical lexicographers have relied on a number of techniques based essentially on dispersion, as a corrective or as an alternative for raw frequency. As an alternative, one may equate the usage of a word with the evenness of its occurrences, e.g., by assuming a constant relationship between the number of genres or units or subsamples in which a word occurs, and its occurrences in the language as a whole. As a corrective, dispersion may be used to adjust frequency coefficients, just as frequency might be used to adjust coefficients of usage based on dispersion; or the two may be combined in various formulas designed to yield more accurate coefficients.

Indeed, regardless of the technique, the ultimate objective remains the same: to give the best possible insight into the properties of representative texts, spoken or written. The formulas aim to approximate the statistical properties of the various orders of constituents in the "language as a whole", an intuitive concept best understood relative to an ideal sample: as large, as representative, and as homogeneous as possible. In other words, rankings based on word counts of necessarily limited samples are considered adequate when it can be assumed that similar counts of other or larger samples are likely to yield similar rankings and similar classes with similar memberships. Since weighted statements about representative texts are based on the occurrence of the various types of constituents, frequency is the fundamental property, and dispersion is secondary: dispersion may be preferred to frequency only insofar as one assumes that dispersion in a smaller sample provides a better approximation of frequency in a larger sample, one which comes closer to the ideal. The same principle underlies formulas which determine word usage by combining frequency and dispersion.

Techniques for quantifying word usage are of three basic types: those which rely on frequency, those which prefer dispersion, and those which resort to both. The statistical lexicographer chooses the one he assumes to come closest to the ideal; claims that this choice makes little difference have not been verified. For our part, we have selected the basic vocabulary from the top of a list in which words were ranked in decreasing order of a coefficient combining frequency and dispersion.

4.1 *Frequency*

In ranking and classifying words according to frequency, the investigator relies on the validity of the sample: he tries to make sure that the sample is sufficiently representative to yield, within a small margin of error, coefficients equal or proportional to those that could be extracted from other samples equal or larger in size.

## INTRODUCTION

### 4.11 Frequency Weighting

Our textual counts were exhaustive, so that the frequency index of every word represents the number of all its occurrences in the entire sample; no word, however frequent or infrequent, was discounted before or during the count, either on the basis of a preliminary study or after the count had reached a pre-established total. As a result, no adjustments were needed to equalize the results obtained in potentially unequal samples.

The hierarchy established in terms of frequency is given in the second column of Part Two, and the results of a brief statistical analysis of the most frequent 5,000 words are provided in 5.1, where it appears that the increase in rank is accompanied by the type of decrease in frequency known as the Zipf law.

### 4.12 High Frequency Words

For an insight into the type of words which populate the high frequency ranges, consider the 100 most frequent words, four articles, eleven prepositions, eight conjunctions, 23 pronouns, eleven (mostly functional) adverbs, fourteen (mostly functional) adjectives, nineteen verbs, nine nouns, and one numeral:

articles:	il, la, lo, uno
prepositions:	a, come, con, da, di, fra, in, per, senza, su, tra
conjunctions:	che, come, e, ma, o, perchè, quando, se
pronouns:	altro, che, chi, ci <sup>1</sup> , ci <sup>2</sup> , cui, egli, io, la, le, li, lo, lui, mi, ne, noi, quale, quello, si, ti, tu, tutto, vi
adverbs:	anche, ancora, come, così, già, no, non, ora, più, poi, sempre
adjectives:	altro, grande, loro, mio, nostro, nuovo, ogni, più, primo, qualche, quello, questo, suo, tutto
verbs:	andare, avere, avere <i>aux.</i> , dare, dire, dovere, essere, essere <i>aux.</i> , fare, parere, parlare, potere, sapere, stare, stare <i>aux.</i> , trovare, vedere, venire, volere
nouns:	anno, casa, cosa, giorno, parte, sud, uomo, vita, volta
numeral:	due

### 4.13 Low Frequency Words

For an insight into the type of words which populate the low frequency ranges, consider the 261 words with five occurrences (ranks 4701 to 4961), four numerals, ten adverbs, 48 verbs, 77 adjectives, 120 nouns, one pronoun, and one interjection:

numerals:	diciassette, quattordici, tredici, ventiquattro
adverbs:	colà, dappertutto, egualmente, felicemente, giornalmente, miracolosamente, quaggiù, sinceramente, solidamente, temporaneamente
verbs:	abbagliare, addomesticare, affiorare, alleggerire, allestire, alloggiare, annodare, appassionare, ardere, assimilare, astenere, bastonare, calcare, chiarificare, collaborare, contraddire, declinare, dispensare, dissetare, forzare, gravare, imboccare, impallidire, incendiare, inclinare, issare, languire, licenziare, maledire, modellare, modulare, ondeggiare, raffigurare, ravvolgere, ributtare, ricomporre, rinchiudere, rinfrancare, rispecchiare, scarmigliare, scatenare, sgambettare, sottolineare, squadrare, suddividere, trionfare, vigilare, violare
adjectives:	abbassato, accennato, accorato, adiacente, adorabile, africano, agevole, aggressivo, alato, allineato, altrui, archeologico, arioso, armonioso, arruffato, atterrito, attiguo, barbarico, casalingo, contiguo, corrosivo, cronologico, decaduto, dimenticato, disgraziato, disseminato, distribuito, frenetico, fuggibile, gagliardo, gremito, immutabile, inaudito, inconsistente, ingrato, inquietante, insensato, insufficiente, intrinseco, involontario, irraggiungibile, lecito, mancato, meschino, miserabile, mostruoso, nasale, operoso, penetrante, piemontese, poggiate, postumo, prolungato, rialzato, riconoscibile, rilevato, riunito, rude, rurale, scheletrico, sciolto, scrupoloso, smisurato, soddisfatto, soffocato, soggetto, spaventevole, strambo, stupefacente, stupendo, sublime, sudato, sveglio, venerabile, virtuoso, visto, zelante
nouns:	adolescente, aggruppamento, agricoltore, altipiano, ammaestramento, ammontare, appetito, apprensione, ara, architettura, arsenale, ascolto, aurora, automobilismo, becco, benda, bottone, boxe, braccetto, cabina, cagione, caminetto, carriera, cautela, certo, cesta, chiamata, chiaro, cittadina, coincidenza, collana, colpevole, compensazione, concittadino, condannato, continuità,

corno, deposizione, difensore, discrezione, diva, esagerazione, fiaba, fiammella, fibra, finezza, forestiero, gentilezza, gentiluomo, ghiacciaio, giovinetto, granduca, grosso, indomani, insidia, in-tonaco, lanterna, levante, licenza, limo, litigio, mandibola, manoscritto, marchio, messo, monaco, monastero, moro, murata, necessario, opuscolo, orazione, palpito, parentesi, pasqua, pensatore, pensione, picco, precedenza, pregiudizio, presa, principale, profezia, prua, pulviscolo, razza, reden-zione, rena, resa, rifiuto, ritegno, ritirata, rogo, sarto, segretaria, seguente, sentinella, sifilide, sintesi, smeraldo, società<sup>2</sup>, sospensione, stabilità, stima, superiore, sussidio, svago, tempra, timone, tiro, toga, trasparenza, turco, vacanza, valuta, venerazione, vigilia, vivo, zitella, zolla

pronoun: sola

interjection: olà

## 4.2 Dispersion

The preceding list suggests that pure frequency is not sufficiently discriminating, inasmuch as it assigns the same coefficient to 261 words with five occurrences (from rank 4701 to rank 4961, the order is alphabetical). Dispersion may therefore be used to discriminate between words characterized by the same frequency.

### 4.21 Dispersion Weighting

In order to emphasize that simple frequency is not sufficiently discriminating, notice that the ten occurrences of eight words are dispersed in two genres, those of 73 words in three, those of 94 in four, and those of 42 in five. To be sure, thirty different distributions are possible for ten occurrences in a maximum of five genres, as follows:

one	two	three	four	five
10	9:1	8:1:1	7:1:1:1	6:1:1:1:1
	8:2	7:2:1	6:2:1:1	5:2:1:1:1
	7:3	6:3:1	5:3:1:1	4:3:1:1:1
	6:4	6:2:2	5:2:2:1	4:2:2:1:1
	5:5	5:4:1	4:4:1:1	3:3:2:1:1
		5:3:2	4:3:2:1	3:2:2:2:1
		4:4:2	4:2:2:2	2:2:2:2:2
		4:3:3	3:3:3:1	
			3:3:2:2	

The various techniques for measuring dispersion, both simple and complex, have been examined in section 1.22 of the *Frequency Dictionary of Spanish Words*.<sup>19</sup>

Our own technique, discussed in section 1.223 of the same study,<sup>20</sup> relies on the ratio of the standard deviation to the mean. The dispersion index used is:

$$D = 1 - 1/2(\sigma/\bar{X}) = 1 - \sqrt{\frac{n\sum X_i^2 - T^2}{2T}},$$

where  $\sigma/\bar{X}$  is also known as the coefficient of variation. This, of course, associates higher values of  $D$  with even allocations over genres, and lesser values of  $D$  with clustering in one or two genres. In this formula,  $n$  represents the number of genres, here 5; and  $T$  is the sum of the frequencies in the genre:

$$X_1 + X_2 + \dots + X_n, \text{ or } T = \sum X_i, \text{ while } \sum X_i^2 = X_1^2 + X_2^2 + \dots + X_n^2.$$

<sup>19</sup> Cf. *FDSW*, pp. xlv-lv.

<sup>20</sup> Cf. *FDFW*, pp. li-lvii.

## INTRODUCTION

Let  $\bar{X}$  be the mean or average value of the  $X$  frequencies,

$$\bar{X} = \frac{1}{n} \sum X_i = \frac{T}{n}.$$

Then the variance  $\sigma^2$ , which is the average of the squared differences from the mean, can be written:

$$\sigma^2 = \frac{1}{n} \sum (X_i - \bar{X})^2.$$

The square root  $\sigma$  of this expression is the standard deviation. By converting the term defining  $\sigma^2$ , we also have:

$$\sigma^2 = \frac{\sum X_i^2 - n\bar{X}^2}{n}.$$

If we multiply by  $n^2$ , we arrive at:

$$n^2 \sigma^2 = n \sum X_i^2 - (n\bar{X})^2$$

or, using the symbols introduced above:

$$n^2 \sigma^2 = n \sum X_i^2 - T^2.$$

The term on the right side of this equation is the expression under the radical sign in the above definition of  $D$ . Thus, we can rewrite:

$$D = 1 - \frac{n\sigma}{2T} = 1 - \frac{\sigma}{2\bar{X}}.$$

This means that in calculating  $D$ , one first determines the standard deviation from the mean. The mean is doubled, so that the quotient to be subtracted from 1 will typically vary between 0 and 1. While the quotient increases with the size of  $\sigma$  relative to  $\bar{X}$ , the total coefficient  $D$  conversely increases from 0 to 1 with the decreasing size of  $\sigma$  relative to  $\bar{X}$ . A value of  $D=0$  may be regarded as dispersion pessima, a value of  $D=1$  as dispersion optima. A word whose occurrences cluster in one category will have a dispersion of 0 regardless of frequency; a word whose occurrences are equally distributed among the genres has a dispersion of 1 regardless of frequency.

Calculated according to this formula, the dispersion coefficients of the thirty possible distributions of ten occurrences in a maximum of five genres, together with the words characterized by these coefficients, are as follows:

<b>0.00</b>	10:0:0:0:0	none
<b>7.50</b>	9:1:0:0:0	none
<b>22.50</b>	8:2:0:0:0	none
<b>24.17</b>	8:1:1:0:0	adibire, anomalia, convertire, emigrato, estratto, incendio, incompleto, mangiatoia, milite, pellegrinaggio, presidenza, proiezione, simile <i>n.</i> , traffico
<b>31.08</b>	7:3:0:0:0	catacomba, esito, inghiottire, nonostante <i>prep.</i> , polemica, prefetto, risorgimento, scheletro, se <i>inter.</i>
<b>34.81</b>	7:2:1:0:0	alunno, bimba, chioma, divano, intorno <i>n.</i> , numerico, odierno, passaporto, schiena, trasporto, tuttavia <i>conj.</i> , urlo
<b>36.75</b>	7:1:1:1:0	concreto, rimpiangere, sgomento
	6:4:0:0:0	impartire, modificazione, ottocento, parziale, qualitativo, radiotelegrafico, telefonico, utilizzazione
<b>38.76</b>	5:5:0:0:0	sor, spasimo
<b>42.99</b>	6:3:1:0:0	appello, direttivo, esaltazione, esaurire, esercitazione, germe, ironico, parlamento, rivendicazione, sfogare, tenda, varcare

<b>45.23</b>	6:2:2:0:0	berretto, delirio, firma, giornalista, mondano, revisione, smarrito
<b>47.56</b>	6:2:1:1:0	adempire, anzichè, confidare, delitto, devozione, ignoranza, navigazione, nettamente, perchè <i>n.</i> , proprio <i>adject.</i> , scontare, serra
	5:4:1:0:0	dinanzi <i>adv.</i> , pazienza, prolungare, rado, spaventare, topo, travaglio
<b>50.00</b>	6:1:1:1:1	apparente, miniera
<b>52.57</b>	5:3:2:0:0	borgo, comunque, coronare, dodici, furore, grossolano, imminente, incarico, inganno, irritare, lombardo, mostro, seccare, selvatico, su <i>n.</i> , tremante, vecchiaia, velluto
<b>55.28</b>	5:3:1:1:0	asserire, custode, fortemente, generico, guida, umiltà
	4:4:2:0:0	consistenza, insegnante, italico, lutto, opposizione, smarrire, straccio, suora
<b>58.17</b>	5:2:2:1:0	accesso <i>n.</i> , aggettivo, anzitutto, buono <i>n.</i> , canna, circoscrivere, freschezza, giovinetta, incesante, lega, rimproverare, sempre <i>adverb.</i> , senso <sup>2</sup> <i>n.</i> ( <i>common sense</i> ), soffio
	4:4:1:1:0	delicatezza, esteriore, gaio, greco <i>n.</i> , investire
	4:3:3:0:0	commentare, credito, filare, fremere, illustrazione, misero, strettamente, veduto, voluttà
<b>61.27</b>	5:2:1:1:1	cambiamento, cosicchè, fiancheggiare, regolamento, rio, stabile, viceversa
<b>64.64</b>	4:3:2:1:0	allegria, caduta, capanna, custodire, decina, definitivamente, donare, elenco, elogio, frattanto, gonna, gratitudine, implorare, invenzione, maestro, <i>adj.</i> maturare, oscillare, pelo proiettare, protestare, raccolto, rotto, sboccare, scampo, scossa, seconda <i>pron.</i> , sotto <i>n.</i> , urgente, veloce, vocazione
<b>68.37</b>	4:3:1:1:1	convinzione, fiera, ispirazione, serenità, sicchè
	4:2:2:2:0	cena, due <i>n.</i> , meglio <i>adj.</i> , salone
	3:3:3:1:0	liberazione, presentire, splendente
<b>72.61</b>	4:2:2:1:1	accingere, contenuto <i>adj.</i> , funebre, governare, rivoltare, soggiorno
	3:3:2:2:0	bensì, depositare, metallo, perfezionare, preda, tremendo
<b>77.64</b>	3:3:2:1:1	compenso, evoluzione, gabbia, lucente
<b>84.19</b>	3:2:2:2:1	accumulare, disco, discreto, dovunque, impressionare, nube, rimedio

Note that 26 different distributions of 10 occurrences correspond to 18 dispersion coefficients, which illustrates the capacity of our formula to assign the same dispersion coefficient to different distributions of the same number of occurrences. To be sure, a more even dispersion in a smaller number of genres may compensate for a less even dispersion in a larger number of genres: the coefficient **36.75** corresponds to 7:1:1:1:0 as well as 6:4:0:0:0, **47.56** to 5:3:1:1:0 and to 5:4:1:0:0, **72.61** to 4:2:2:1:1 and 3:3:2:2:0; even more, **58.17** is the coefficient of three different dispersions, 5:2:2:1:0, 4:4:1:1:0, and 4:3:3:0:0, just as **68.37** characterizes 4:3:1:1:1, 4:2:2:2:0, and 3:3:3:1:0.

Essentially determined by the distribution evenness/unevenness of occurrences, the coefficient of dispersion is slightly biased by certain properties of the formula and by the five-fold division of the sample, which favor certain coefficients at the expense of certain others. Regardless of language, words with the same frequency tend to cluster around certain dispersion coefficients, as illustrated by the following tabulation of French, Rumanian, Spanish, and Italian words with ten occurrences:

	<i>Fr</i>	<i>Rum</i>	<i>Sp</i>	<i>It</i>
<b>100.00</b>	1	0	0	0
<b>84.18</b>	7	4	7	7
<b>77.63</b>	8	7	5	4
<b>72.61</b>	23	9	15	6
<b>68.37</b>	14	11	16	12
<b>64.64</b>	26	20	28	31
<b>61.27</b>	8	4	5	7
<b>58.16</b>	22	28	29	29
<b>55.27</b>	23	15	24	14
<b>52.56</b>	14	15	18	18
<b>50.00</b>	1	1	0	2
<b>47.55</b>	15	28	23	19
<b>45.22</b>	4	0	7	7
<b>42.99</b>	21	22	20	12
<b>38.76</b>	1	0	2	2

## INTRODUCTION

36.75	8	6	11	11
34.80	17	18	17	12
31.08	6	3	6	9
24.17	0	10	0	14

Certain coefficients appear to be favored, e.g., 72.61, 68.37, 64.64, 58.16, 55.27, 52.56, 47.55, 42.99, and 34.80, while others, such as 100.00, 50.00, or 38.76, are not. Favored coefficients are not necessarily those which characterize more than one possible distribution: if both 68.37 and 58.17 apply to three different distributions, 72.61, 64.64, 52.56, 42.99, and 34.81 correspond to one only.

It goes without saying that our formula, which can assign different dispersion coefficients to words with the same number of occurrences, can also assign the same dispersion coefficient to words with different frequencies, as illustrated by the following 121 words with dispersion coefficient 59.18 ranked in decreasing order of frequency:

<i>Freq.</i>	<i>Dispersion</i>	
33	16:7:7:3:0	costume
21	10:6:3:1:1	pittura
	9:7:4:1:0	marchese
18	9:4:3:1:1	nascita
	9:3:3:3:0	aereo, nostro <i>pron.</i>
15	7:4:3:1:0	bontà, migliorare, sottrarre
12	6:3:1:1:1	accertare, impegnare, spogliare
	6:2:2:2:0	ambizione, apprezzare, autorevole, moneta, risveglio
	4:4:4:0:0	preferenza
9	4:3:1:1:0	agitato, alquanto <i>adj.</i> , americano <i>n.</i> , angosciato, arnese, assassinio, benefico, catalogo, ceppo, crescente, discendente, distruzione, gradito, lama, marciare, nastro, prevenire, proiettile
		ribellione, saldo, serrato, tondo, vivamente
	3:3:3:0:0	gemma, interminabile, norvegese, sia <i>conj.</i>
6	3:1:1:1:0	abate, annotare, assiduo, attestare, avvertimento, cinquecento <i>num.</i> , colonico, delizia, eccitazione, effimero, esemplare, fieno, fissazione, impastare, incubo, indubbiamente, ingiustizia, liberamente, lustro <i>n.</i> , novello, orientare, pantalone, pari <i>n.</i> , pausa, pienamente, possibilmente, premura, presto <i>adj.</i> , quinto, rinascita, rosa <i>adj.</i> , sanità, somiglianza, supremazia, svenire, tempo <sup>2</sup>
	2:2:2:0:0	affissare, allineare, ampiezza, ancora <i>n.</i> , annegare, arbitrario, auspice, campanile, centesimo, circolare <i>v.</i> , culto, desiderato, disponibile, dotto, ducato, entusiasta, espandere, estivo, evo, freccia, guizzare, inerzia, insignificante, laborioso, legislazione, mescolanza, miracoloso, oscurare, previsione, quercia, ricetta, riordinare, riversare, severamente, sfogliare, sospeso, spaccare, talento, vallata, vegliare, vezzo.

Note that only seven different frequencies correspond to the eleven different distributions of words with the same coefficient of dispersion, which illustrates in reverse the fact that the formula can assign the same dispersion coefficient to different distributions of the same number of occurrences.

It may be of interest to compare, according to the number of occurrences, the 122 French, the, 123 Spanish, and the 121 Italian words having the same dispersion coefficient of 59.18:

<i>Frequency</i>	<i>Fr</i>	<i>Sp</i>	<i>It</i>
81	0	1	0
45	0	0	0
33	0	0	1
30	2	2	0
27	1	0	0
21	0	2	2
18	1	3	3
15	8	2	3



12	1	10	9
9	26	23	27
6	83	80	77

#### 4.22 High Dispersion Words

For an insight into the type of words which populate the high dispersion ranges, consider the 100 most evenly dispersed words, three articles, two numerals, nine pronouns, three conjunctions, seven prepositions, 19 adjectives, nine adverbs, 24 nouns, and 24 verbs:

articles:	il, la, uno
numerals:	quattro, tre
pronouns:	altra, altro, che, loro, ognuno, quello, questa, una, uno
conjunctions:	come, e, mentre
prepositions:	a, da, di, con, fra, per, tra
adjectives:	abbastanza, altro, certo, chiaro, dimenticato, forte, grande, intimo, molto, ogni, più, questo, raro, riconoscibile, stesso, troppo, tutto, ultimo, zelante
adverbs:	altrettanto, anche, anzi, continuamente, già, lontano, pur, sempre, solamente
nouns:	affetto, anno, campagna, certezza, collega, corpo, cura, eleganza, giro, lettura, mare, mese, miglio, nome, ora, peso, rogo, sangue, segno, sensibilità, splendore, vacanza, volta, vuoto
verbs:	accompagnare, avere, avvertire, cessare, comandare, comprendere, condurre, congiungere, dare, discendere, incontrare, mancare, offrire, preparare, provare <sup>1</sup> , ricordare, riflettere, ripetere, riuscire, scrivere, sembrare, servire, solere, trovare

If we compare the 100 most evenly dispersed words with the 100 most frequent words, we notice that only 23 are common: *a, altro, anche, anno, avere, da, dare, di, fra, già, grande, il, la, ogni, per, più, questo, sempre, tra, trovare, tutto, uno, volta*.

#### 4.23 Low Dispersion Words

For an insight into the types of words which populate the lower end of the dispersion scale, consider the last 100 words, one preposition, one conjunction, two pronouns, two interjections, four adverbs, six verbs, 17 adjectives, and 67 nouns:

preposition:	appunto
conjunction:	dove
pronouns:	lei, tu
interjections:	ah, eh
adverbs:	ebbene, no, pertanto, stanotte
verbs:	applicare, approvare, constare, dedurre, differenziare, situare
adjectives:	apostolico, cospicuo, eminente, ereditario, esteso, giapponese, letterario, negativo, neutrale, normale, polacco, primario, sperimentale, successivo, tuo, vegetale, zitto
nouns:	acido, aeroplano, agenzia, alleato, asse, apparato, approssimazione, assistenza, associazione, aviazione, benedizione, bordo, britannico, caduto, calcolo, cardinale, cassetto, cellula, cento, ciclo, cilindro, commemorazione, comportamento, conflitto, conservatore, console, contessa, cripta, curva, dea, decreto, degenerazione, determinazione, diffusione, fedele, giapponese, giunta, inaugurazione, incidente, laboratorio, lesione, monsignor, museo, neutralità, nucleo, particella, pendolo, presidente, pressione, prosa, radice, ricerca, rottura, salotto, santità, scirocco, seme, strato, superficie, tessuto, trasmissione, urto, variazione, vecchia, vescovo, voce <sup>2</sup> .

### 4.3 Usage

Exclusive reliance on either frequency or dispersion is not likely to provide the ranking, which best reflects word usage in the language as a whole. Indeed, frequency based predictions about the occurrence of words in representative texts or in conversation would lead us to assume that *elevato* for instance, is as important a word as *panorama*, an anomaly which could be mechanically cor-

## INTRODUCTION

rected with reference to differences in dispersion; and vice versa, scholars who think that pure dispersion is a more reliable index of word usage would have to assume that a word such as *oratorio* for instance, is among the most useful of Italian words, an anomaly which could be corrected with reference to differences in frequency. Obviously, a formula relying on both frequency and dispersion is needed, which would exploit both in a meaningful way.

### 4.31 *Usage Weighting*

We have indicated elsewhere that techniques which rank words with reference to both frequency and dispersion depend on a number of alternatives, the most important of which seem to be: use the two factors separately or in combination; if in combination, coordinate or subordinate the two, by assigning them equal weight in determining ranking, or by correcting rankings established in terms of one with reference to the other; the correction may be made by expansion, by reduction, or by both; correction by expansion may be an addition or a multiplication; correction by reduction may be a subtraction or a division; the alternative factors may be used exhaustively or partially, to select the entire basic vocabulary or part of it, e.g., by relying on frequency to select the top of the list, on dispersion for the bottom, and a combination of both for the middle ranges; etc.<sup>21</sup>

Our own formula<sup>22</sup> coordinates frequency and dispersion inasmuch as the coefficients it determines are proportional to both: for the same frequency, the more or less even the dispersion, the higher or lower the coefficient of usage; and vice versa, for the same dispersion, the higher or lower the frequency, the higher or lower the coefficient of usage:

$$U = \frac{F \times D}{100}$$

This means that for  $D=1$ , which is the case when occurrences are equally dispersed among the categories of the sample,  $U=F$ ; for  $D=0$ , which is the case when occurrences cluster in one single category,  $U=0$ ; for  $D=0.5$ ,  $U=F/2$ , etc. In other words,  $U$  varies between a high equal to  $F$  and a low equal to 0, depending on the magnitude of  $D$ . Consequently,  $U$  is a dispersion percentage of  $F$ , which means that a less frequent word can be ranked ahead of a more frequent one, provided that its occurrences are more evenly dispersed in the genres of the sample.

The idea behind this formula is that evenly dispersed frequencies are “dense” and unlikely to “shrink” in an enlarged or improved sample, just as unevenly dispersed frequencies are “puffy” and likely to shrink if checked against a more representative sample. We assume, then, that occurrences which have maintained themselves at a steady level throughout the genres of a sample (which is what a “high dispersion coefficient” actually means) will continue to hold steadily at the same level, without climbing or declining abruptly if the sample is improved by adding new genres or by enlarging existing ones. Conversely, we assume that occurrences which fluctuate widely from genre to genre have not yet found their natural level and will tend to stabilize on a lower level in an enlarged and homogenized sample.

In turn, these assumptions are justified with reference to other, more concrete ones, since steadily frequent occurrences throughout the genres of a sample must have either a linguistic or an extra-linguistic justification: they must be associated with some constantly recurrent characteristic of the substance (meaning) communicated, or of the form (medium) of communication.

<sup>21</sup> Cf. *FDSW*, pp. lxi-lxix.

<sup>22</sup> Cf. *FDSW*, pp. lxxvii-lxix.

Thus, for evenly dispersed *functional* words, high frequency is conditioned by the recurrence of a relatively small number of grammatical modes which express relations between the terms of grammatical propositions regardless of the semantic content conveyed by the roots of the lexical words; for evenly dispersed *lexical* words, high frequency must be determined by the recurrence of certain constant conditions in the natural, physical, social, or cultural environment, which are likely to be present regardless of the topics which characterize a particular genre. However, the high frequency of unevenly dispersed lexical words can be explained only as a result of the hazards of random selection, which may favor terms associated with a topic characteristic of a particular genre, e.g., medical terms in scientific literature.

Such assumptions explain why even dispersion is an indication that frequency has reached its natural level and in view of the linguistic or extralinguistic constants it reflects, is likely to maintain itself regardless of sample alterations; conversely, uneven dispersion is an indication that a high frequency word has not yet reached its natural level and is likely to stabilize at a lower level in a more homogeneous sample in which the hazard of random selection has been reduced. This partial correlation between frequency and dispersion (or between "rank" and "range") underlies techniques of vocabulary selection which stop counting words once their occurrences have reached a sufficient total to qualify for the basic total or after their presence has been attested in a certain number of genres.

To illustrate how the coefficient of usage varies relative to both frequency and dispersion consider the 15 words with the same usage coefficient of 10.00:

Frequency		Dispersion
26	<i>ammazzare</i>	38.46
23	<i>trasformazione</i>	43.48
18	<i>muto</i>	55.56
16	<i>allargare, verde n., vivente</i>	62.50
15	<i>apprendere, compreso, isolato, olio, penna, serbare</i>	66.67
14	<i>affondare, cinque n., trono</i>	71.43

It may be interesting to tabulate according to frequency and dispersion the 32 French, 31 Spanish, 34 Rumanian, and 15 Italian words having a coefficient of usage of 10.00:

Frequency	Fr				Dispersion
	Fr	Sp	Rum	It	
26	0	0	0	1	38.46
23	0	0	0	1	43.48
22	0	2	0	0	45.45
21	0	1	0	0	49.05
20	1	0	1	0	50.00
18	3	5	3	1	55.56
17	1	7	3	0	58.82
16	2	1	4	3	62.50
15	4	2	7	6	66.67
14	8	7	7	3	71.43
13	7	3	4	0	76.92
12	3	4	2	0	83.33
11	2	2	0	0	90.90
10	1	0	0	0	100.00

#### 4.32 High Usage Words

For an insight into the nature of rankings based on usage, consider that, except for the nouns *ora*, *tempo*<sup>1</sup>, and *modo*, the verbs *passare*, *prendere*, and *portare*, the pronoun *quella*, the adverb

## INTRODUCTION

*dove*, and the preposition *dopo*, the first 100 words in decreasing order of usage are the same as the first 100 words in decreasing order of frequency. The ten words which do not appear in the usage list are the pronouns *ti*, *tu*, *vi*, *li*, and *chi*, the adverbs *no* and *come*, the noun *casa*, and the auxiliary (modal) *stare*, whose coefficients of usage are reduced by uneven dispersions.

### 4.33 Low Usage Words

As for words which populate the lower usage ranges, consider the 146 words with a usage coefficient of 3.08, 6 adverbs, 31 verbs, 31 adjectives, and 78 nouns:

adverbs:	attentamente, fortunatamente, presso, simultaneamente, tutto, violentemente
verbs:	abbondare, affannare, analizzare, annoiare, avviluppare, conciliare, crollare, danneggiare, disprezzare, educare, eleggere, enunciare, implicare, imputare, occhieggiare, ostinare, perseguire, poggiare, premettere, professare, restaurare, riaprire, ridestare, rischiare, risiedere, schizzare, seminare, sfilare, spiccare, subentrare, tollerare.
nouns:	addio, adesione, amica, annunzio, ausilio, avamposto, avanti, avvento, avvilimento, beffa, benevolenza, branco, brano, breve, brigante, calzone, cantina, cappellino, carnevale, cartolina, castigo, cemento, cicala, cipresso, classico, clientela, crocifisso, deformazione, dieci, efficienza, grasso, guadagno, importazione, insegna, intento, interessamento, laurea, mascella, miglioramento, navata, nobile, obiezione, onestà, opposto, ostinazione, paga, paglia, passante, piccolo, pistola, premessa, preparativo, provinciale, purgatorio, requie, ricorso, riflessione, romanziera, rondine, scoglio, sdegno, sì, specialista, spinta, subbuglio, tappeto, triangolo, turchino, veduta, vibrazione.
adjectives:	alzato, amante, amato, arguto, conservato, corretto, democratico, diligente, equivoco, espressivo, insensibile, inumano, invincibile, lugubre, memorabile, mensile, mesto, obliquo, plastico, raffinato, redditizio, ricamato, ruvido, sanguigno, scialbo, sodo, sorridente, spietato, sporgente, vertiginoso, violaceo.

The parallelism of frequency and usage rankings raises an important question concerning the value of the latter. For low range words the significance of the usage coefficients is clear, since very few occupy the same position in the usage ranking as in the frequency ranking. As a matter of fact, the difference between usage and frequency rankings is actually greater than suggested by a comparison of the two, inasmuch as our three parallel columns provide, in decreasing order of usage, of frequency, and of dispersion, the same 5,000 words selected in terms of usage. Had we taken the first 5,000 in terms of frequency regardless of dispersion, the first 5,000 in terms of dispersion regardless of frequency, and the first 5,000 in terms of usage regardless of frequency and dispersion, differences between the three lists would be even more considerable (especially in the mid and low ranges). True, differences between words in the mid and low ranges have much less bearing on statements about the structural properties of representative texts, as the decline in the magnitude of the coefficients follows Zipf's law, thus considerably reducing their significance; the last 1,000 words total 6,196 occurrences, while the first word alone has 28,659 occurrences. This suggests that coefficients of usage are significant in making predictions about the textual occurrence of mid and especially low frequency words, but much less so in making predictions about the representative texts whose structure is essentially determined by the properties of high frequency words.

This impression can be at least partially corrected by verifying the rank occupied in the hierarchy of usage by the 50 most frequent words:

di	1	1	uno	7	7
il	2	2	essere	8	8
la	3	3	lo	9	9
e	4	4	che	10	10
a	5	6	per	11	13
in	6	5	da	12	14

si	13	11	stare	38	44
non	14	12	la	39	35
essere	15	15	ne	40	41
con	16	18	due	41	47
che	17	16	quale	42	45
avere	18	17	venire	43	51
avere	19	19	altro	44	54
questo	20	24	ci <sup>1</sup>	45	48
su	21	23	vedere	46	40
ma	22	21	più	47	56
lo	23	20	quello	48	59
fare	24	22	perchè	49	57
potere	25	26	volere	50	38
o	26	25	più	32	34
anche	27	30	dire	33	29
quello	28	27	dovere	34	39
suo	29	28	sud	35	37
tutto	30	31	dare	36	49
come	31	32	grande	37	50

Only ten words out of fifty occupy the same rank in the two lists: ranks 1 to 4 (*di, il, la, e*), 5 to 10 (*uno, essere, lo, che*), 15 (*essere*) and 19 (*avere*). Twelve words differ by one rank (*a, in, che, avere, ma, potere, o, quello, suo, tutto, come, ne*), nine by two (*per, da, si, non, con, su, fare, più, sud*), four by three (*lo, anche, ci<sup>1</sup>, quale*), three by four (*questo, dire, la*), one by five (*dovere*), three by six (*stare, due, vedere*), two by eight (*venire, perchè*), one by nine (*più*), one by ten (*altro*), one by eleven (*quello*), one by twelve (*volere*) two by thirteen (*dare, grande*).

We may conclude that the formula which determines usage by adjusting frequency in terms of dispersion is more effective among the lower ranked words. Except for those whose occurrences cluster in dialogue (interjections, second person pronouns and adjectives, greeting formulae, and affective words), the more numerous the occurrences, the more even their dispersion, so that the adjustment required by the computation of usage will be proportionally smaller; and the lower the frequency, the less even the dispersion (the more subject to the hazards of random selection), so that proportionally, the adjustment will be more considerable.

## 5. THE DICTIONARY

Once the statistical hierarchies have been established, the linguist still has to make several decisions: How large should the dictionary be? How many words should it contain? How are these words to be selected from the data? How should they be presented in the dictionary?

### 5.1 Size

Since the decrease in coefficients is gradual, there is no "natural" cutting-off point to determine the number of words to be included, and the author must choose one more or less arbitrarily. There are two ways of establishing the cut-off point, based either on a minimum coefficient, which determines the number of words, or on a predetermined number of words, which determines the minimum coefficient. The decision to include all items above a certain coefficient, such as a frequency of 20, is purely arbitrary. A somewhat less arbitrary procedure is to decide approximately how many words are appropriate to the purpose, and select as cut-off point the coefficient

## INTRODUCTION

boundary nearest that number. We decided that about 5,000 words would best suit our purpose, and we selected as cut-off point the coefficient which fell closest to 5,000. In the FDIW, this boundary happens to fall at the coefficient of 239, which gives us 5,014 words.

A dictionary of 5,000 words may appear somewhat large in view of our purpose, since about 90 % of the words in a representative text occur among the first 1,000; however, we decided that a more extensive list could serve a greater variety of purposes, especially studies of vocabulary and, although we did not include them among our goals, certain pedagogical uses. Extending the dictionary beyond 5,000 words soon reaches the point of diminishing returns, since the reliability of a hierarchy, its statistical classes, and their properties, declines considerably in the lower ranges.

### 5.2 Selection

The entries were selected from the top of a hierarchy established according to coefficients of usage: every word with a coefficient of 2.39 or more appears in our lists. This coefficient represents the first break after the desired total of 5,000 words, giving 5,014 entries. This simpler method of selection was adopted as a result of Muller's criticism<sup>23</sup> of the more elaborate techniques used in the *FDSW*, where the selection was undertaken in three steps: we first eliminated all words with a frequency of less than 4; we next discarded words dispersed in fewer than three genres; from the remaining list, we eliminated all words having a coefficient of usage of less than 3.00, thus isolating a basic vocabulary of 5014 words.<sup>24</sup>

### 5.3 Structure

The *Frequency Dictionary of Italian Words* consists of two parts, each containing the same 5,014 words with a coefficient of usage of 2.39 or more.

Part I, pp. 1-402, presents the entries in alphabetical order; each entry is accompanied by a certain amount of information which we shall explain by examining a typical entry:

<b>bello</b> <i>adj.</i>	<b>379</b>	<b>240.07</b>	<b>379</b>	<b>63.34</b>	
	379	147	132	64	32 4
begli	1	0	1	0	0 0
bei	20	7	7	5	1 0
bel	73	23	27	15	8 0
bell'	12	5	5	1	0 1
bella	141	59	54	19	9 0
belle	38	14	12	6	5 1
belli	18	9	2	4	3 0
bellissima	14	3	4	4	2 1
bellissime	9	0	5	3	1 0
bellissimi	2	1	1	0	0 0
bellissimo	8	1	5	1	1 0
bello	43	25	9	6	2 1

The first line provides the entry word, printed in bold type: **bello**, a part-of-speech abbreviation printed in italics: *adj.*, the coefficient of usage, 240.07, the coefficient of frequency, 379, and the coefficient of dispersion, 63.34.

<sup>23</sup> Ch. Muller, "Un Dictionnaire de fréquence de l'espagnol moderne," *Zeitschrift für Romanische Philologie*, xxxi (1965), 476-83; see also "Fréquence, dispersion et usage à propos des dictionnaires de fréquence," *Cahiers de Lexicologie*, vii (1965), 33-42.

<sup>24</sup> Cf. *FDSW*, pp. lxxiv-lxxvii.

The second line, printed in italics, provides the total number of occurrences, 379, followed by the number of occurrences in each genre; plays, 147; novels and short stories, 132; essays, 64; newspapers and magazines, 32; and technical and scientific literature, 4.

The third line consists of the word variant *begli*, followed by the total occurrences of that variant, 1, and by the number of its occurrences in each of the subsamples. The remaining lines list the other variants encountered in the sample, along with the same information concerning the number of occurrences of each variant.

Part II, pp. 403-519, consists of three parallel columns, which rank the words in decreasing order of usage (left), frequency (center), and dispersion (right), respectively.

In each column, the entry words are followed by a part-of-speech abbreviation, and by the appropriate coefficient. Words having the same coefficient are ordered alphabetically. Words variants are not listed in the second part.

The three lists are divided into groups of 500; the last group contains only 14 words, from rank 5,001 to rank 5,014.





PART ONE

# TYPES AND TOKENS

ALPHABETICAL ORDER



<b>a prep.</b>	<b>10316.54</b>	<b>10901</b>	<b>94.64</b>			
	10,901	1,851	2,312	2,230	2,513	1,995
a	5,079	1,154	1,252	1.082	947	644
ad	832	69	143	143	183	294
agl'	2	0	0	1	0	1
agli	189	23	31	45	57	33
al	1,749	256	363	341	447	342
all'	496	62	81	110	118	125
all' f.	284	25	48	70	76	65
all' f.p.	2	0	0	1	1	0
alla	1,526	195	271	307	430	323
alle	619	58	106	112	213	130
allo	123	9	17	18	41	38

<b>abate n.</b>	<b>3.55</b>	<b>6</b>	<b>59.18</b>						
	6	0	2	2	0	2			
abate	5	0	2	1	0	2			
abati	1	0	0	1	0	0			

<b>abbagliare v.</b>	<b>2.76</b>	<b>5</b>	<b>55.28</b>						
	5	1	0	2	2	0			
abbagli	1	1	0	0	0	0			
abbaglia	1	0	0	1	0	0			
abbagliante	1	0	0	0	1	0			
abbagliati	1	0	0	0	1	0			
abbagliato	1	0	0	1	0	0			

<b>abbandonare v.</b>	<b>39.59</b>	<b>55</b>	<b>71.98</b>						
	55	12	21	13	5	4			
abbandona	1	0	0	0	1	0			
*abbandona	3	1	1	1	0	0			
abbandonando	2	0	1	1	0	0			
abbandonano	1	0	1	0	0	0			
abbandonar	4	2	0	1	1	0			
abbandonare	11	1	4	5	0	1			
abbandonarono	3	1	1	0	0	1			
abbandonata	5	1	1	0	1	2			
abbandonate	1	1	0	0	0	0			

abbandonati	2	1	0	1	0	0			
abbandonato	6	1	2	2	1	0			
abbandonava	6	0	4	2	0	0			
abbandonerà	2	1	1	0	0	0			
abbandoni	1	1	0	0	0	0			
abbandono	2	0	2	0	0	0			
*abbandono	1	1	0	0	0	0			
abbandonò	4	0	3	0	1	0			

<b>abbandono n.</b>	<b>13.88</b>	<b>20</b>	<b>69.38</b>						
	20	0	3	7	4	6			
abbandoni	2	0	1	0	1	0			
abbandono	18	0	2	7	3	6			

<b>abbassare v.</b>	<b>13.95</b>	<b>24</b>	<b>58.13</b>						
	24	1	12	5	1	5			
abbassa	2	1	0	0	0	1			
*abbassa	4	0	0	3	0	1			
*abbassano	1	0	0	1	0	0			
abbassar	2	0	1	1	0	0			
abbassare	5	0	4	0	1	0			
abbassarono	1	0	1	0	0	0			
abbassato	5	0	2	0	0	3			
abbassò	2	0	2	0	0	0			
*abbassò	2	0	2	0	0	0			

# ABBASSATO

<b>abbassato</b> <i>adj.</i>	<b>3.42</b>	<b>5</b>	<b>68.38</b>
	5	1	2 1 0 1
abbassata	1	0	1 0 0 0
abbassate	1	0	0 1 0 0
abbassato	3	1	1 0 0 1
<b>abbastanza</b> <i>adj.</i>	<b>20.68</b>	<b>24</b>	<b>86.18</b>
abbastanza	24	4	5 5 3 7
<b>abbastanza</b> <i>adv.</i>	<b>12.04</b>	<b>19</b>	<b>63.35</b>
abbastanza	19	5	1 6 0 7
<b>abbattere</b> <i>v.</i>	<b>9.22</b>	<b>16</b>	<b>57.61</b>
	16	4	2 2 8 0
*abbatte	2	0	0 0 2 0
*abbattè	1	0	0 0 1 0
abbattendo	1	0	0 0 1 0
abbattere	4	0	1 1 2 0
*abbatteva	1	1	0 0 0 0
abbattuta	2	1	0 1 0 0
abbattuti	1	0	0 0 1 0
abbattuto	3	1	1 0 1 0
*abbattuto	1	1	0 0 0 0
<b>abbisognare</b> <i>v.</i>	<b>5.00</b>	<b>7</b>	<b>71.43</b>
	7	0	2 2 2 1
abbisogna	3	0	0 2 0 1
abbisognamo	1	0	0 0 1 0
abbisognano	1	0	0 0 1 0
abbisognasse	1	0	1 0 0 0
abbisognava	1	0	1 0 0 0
<b>abbondante</b> <i>adj.</i>	<b>5.92</b>	<b>15</b>	<b>39.45</b>
	15	0	1 1 3 10
abbondante	6	0	0 1 2 3
abbondanti	7	0	1 0 1 5
abbondantissimo	2	0	0 0 0 2
<b>abbondanza</b> <i>n.</i>	<b>6.08</b>	<b>9</b>	<b>67.61</b>
abbondanza	9	0	3 3 1 2
<b>abbondare</b> <i>v.</i>	<b>3.08</b>	<b>6</b>	<b>51.41</b>
	6	0	0 2 1 3
abbonda	2	0	0 1 0 1
abbondano	2	0	0 1 1 0
abbondante	1	0	0 0 0 1
abbondanti	1	0	0 0 0 1

<b>abbracciare</b> <i>v.</i>	<b>22.95</b>	<b>38</b>	<b>60.39</b>
	38	5	19 8 4 2
abbraccerà	1	0	0 0 1 0
abbraccerei	1	0	1 0 0 0
abbracci s.	1	1	0 0 0 0
abbraccia	3	0	0 0 1 2
abbracciamo	1	0	0 1 0 0
abbracciando	2	0	1 0 1 0
abbracciar	7	2	3 2 0 0
abbracciare	5	1	2 2 0 0
abbracciarono	2	0	2 0 0 0
abbracciata	3	0	2 1 0 0
abbracciato	1	0	0 1 0 0
abbracciava	5	1	4 0 0 0
abbraccio	1	0	0 0 1 0
abbracciò	5	0	4 1 0 0
<b>abile</b> <i>adj.</i>	<b>4.13</b>	<b>6</b>	<b>68.82</b>
	6	2	1 2 1 0
abile	3	1	0 2 0 0
abili	3	1	1 0 1 0
<b>abilità</b> <i>n.</i>	<b>2.64</b>	<b>7</b>	<b>37.73</b>
abilità	7	3	0 4 0 0
<b>abisso</b> <i>n.</i>	<b>7.80</b>	<b>14</b>	<b>55.68</b>
	14	1	2 7 4 0
abissi	5	0	0 3 2 0
abisso	9	1	2 4 2 0
<b>abitante</b> <i>n.</i>	<b>4.00</b>	<b>7</b>	<b>57.14</b>
	7	0	0 2 3 2
abitante	1	0	0 0 1 0
abitanti	6	0	0 2 2 2
<b>abitare</b> <i>v.</i>	<b>21.94</b>	<b>30</b>	<b>73.13</b>
	30	9	10 5 5 1
abita	3	2	0 0 1 0
abitante	1	0	0 0 1 0
abitare	5	3	2 0 0 0
abitarono	1	0	0 0 1 0
abitassero	1	0	1 0 0 0
abitata	3	0	1 1 0 1
abitate	1	1	0 0 0 0
abitato	3	1	1 0 1 0
abitava	5	0	1 3 1 0

## ACCENDERE

*abitava	1	1	0	0	0	0
abitavamo	1	0	1	0	0	0
abitavan	1	0	1	0	0	0
abitavano	2	0	2	0	0	0
abiti <sup>8</sup> s.	1	0	0	1	0	0
abito	1	1	0	0	0	0
<b>abitazione n.</b>	<b>3.94</b>		<b>8</b>	<b>49.22</b>		
abitazione	8	0	3	0	4	1
<b>abito n.</b>	<b>31.77</b>		<b>45</b>	<b>70.60</b>		
	45	12	16	9	8	0
abiti	18	6	5	2	5	0
abito	27	6	11	7	3	0
<b>abituare v.</b>	<b>7.61</b>		<b>13</b>	<b>58.58</b>		
	13	4	6	1	2	0
*abituare	1	1	0	0	0	0
abituare	1	1	0	0	0	0
abituata	2	0	2	0	0	0
*abituata	1	0	1	0	0	0
abituata	1	0	1	0	0	0
abituato	6	1	2	1	2	0
*abituato	1	1	0	0	0	0
<b>abituato adj.</b>	<b>4.13</b>		<b>6</b>	<b>68.82</b>		
	6	2	1	1	2	0
abituata	3	1	0	1	1	0
abituati	3	1	1	0	1	0
<b>abitudine n.</b>	<b>16.39</b>		<b>22</b>	<b>74.49</b>		
	22	4	8	5	4	1
abitudine	12	4	4	3	0	1
abitudini	10	0	4	2	4	0
<b>abusare v.</b>	<b>4.13</b>		<b>6</b>	<b>68.82</b>		
	6	2	1	1	2	0
abusa	1	1	0	0	0	0
abusando	1	0	0	0	1	0
abusare	1	1	0	0	0	0
abusarono	1	0	0	0	1	0
abusato	2	0	1	1	0	0
<b>accademia n.</b>	<b>8.66</b>		<b>14</b>	<b>61.87</b>		
	14	1	2	7	2	2
accademia	8	1	2	4	1	0
accademie	6	0	0	3	1	2

<b>accademico</b> <i>n.</i>	<b>3.94</b>	<b>8</b>	<b>49.22</b>
	8 0 0	4 3 1	
accademici	3 0 0	3 0 0	
accademico	5 0 0	1 3 1	
<b>accadere</b> <i>v.</i>	<b>46.41</b>	<b>59</b>	<b>78.66</b>
	59 19 15	11 10 4	
accada <sup>3</sup> <i>s.</i>	2 0 0	2 0 0	
accadde	11 1 8	2 0 0	
accade	18 7 1	3 4 3	
accadere	6 3 0	1 1 1	
accadesse	1 0 1	0 0 0	
accadeva	2 0 2	0 0 0	
°accadeva	1 0 0	1 0 0	
accadono	2 0 1	0 1 0	
accadrà	2 2 0	0 0 0	
accaduta	1 0 1	0 0 0	
accaduto	13 6 1	2 4 0	
<b>accanto</b> <i>adv.</i>	<b>12.68</b>	<b>16</b>	<b>79.27</b>
accanto	16 5 3	4 1 3	
<b>accanto</b> <i>prep.</i>	<b>41.03</b>	<b>49</b>	<b>83.74</b>
accanto	49 12 9	13 4 11	
<b>accarezzare</b> <i>v.</i>	<b>3.26</b>	<b>7</b>	<b>46.55</b>
	7 2 4	1 0 0	
accarezza	1 1 0	0 0 0	
accarezzare	2 1 1	0 0 0	
accarezzato	1 0 0	1 0 0	
accarezzava	3 0 3	0 0 0	
<b>accendere</b> <i>v.</i>	<b>27.25</b>	<b>42</b>	<b>67.28</b>
	42 15 15	5 5 2	
accende	1 1 0	0 0 0	
*accende	2 0 1	0 0 1	
accendendo	1 0 1	0 0 0	
*accender	1 1 0	0 0 0	
accenderà	1 1 0	0 0 0	
accenderanno	1 0 0	0 0 1	
accendere	8 4 2	0 2 0	
accenderemo	1 0 0	1 0 0	
*accendeva	1 0 0	1 0 0	
accendevano	2 1 1	0 0 0	
accendi <i>i.</i>	3 3 0	0 0 0	
accendiamo	2 2 0	0 0 0	

# ACCENNARE

accendo	1	0	0	1	0	0				accertar	1	0	0	0	1	0
*accendono	2	0	0	1	1	0				*accertar	2	1	0	0	0	1
accesa	4	0	3	0	1	0				accertare	3	0	0	1	1	1
accese	5	0	5	0	0	0				accertata	2	0	0	0	2	0
*accese	2	1	0	1	0	0				accertate	1	0	0	0	1	0
accesi	3	0	2	0	1	0				accertato	2	0	1	0	0	1
*acceso	1	1	0	0	0	0										
<b>accennare v.</b>	<b>31.61</b>	<b>44</b>	<b>71.84</b>							<b>accesso adj.</b>	<b>12.26</b>	<b>20</b>	<b>61.28</b>			
	44	3	8	7	8	18					20	1	9	6	3	1
accenna	8	0	1	3	4	0				accesa	2	0	2	0	0	0
*accenna	1	0	0	0	0	1				accese	3	1	1	1	0	0
accennando	3	0	0	1	0	2				accesi	6	0	2	4	0	0
accennano	1	0	0	0	0	1				accesso	9	0	4	1	3	1
accennare	3	0	0	0	0	3				<b>accesso n.</b>	<b>5.82</b>	<b>10</b>	<b>58.17</b>			
accennata	1	0	0	0	0	1					10	0	1	2	2	5
accennati	1	0	0	0	0	1				accessi	3	0	0	0	1	2
accennato	13	2	1	2	2	6				accesso	7	0	1	2	1	3
*accennato	1	0	0	1	0	0				<b>accettare v.</b>	<b>47.61</b>	<b>64</b>	<b>74.40</b>			
accennava	5	0	3	0	0	2					64	13	11	16	22	2
*accennava	2	0	1	0	1	0				accetta	7	1	0	4	2	0
accennavano	1	0	1	0	0	0				*accetta	1	1	0	0	0	0
accennavo	1	0	0	0	1	0				accettai	1	0	1	0	0	0
accennerò	1	0	0	0	0	1				accettando	4	0	0	2	2	0
accenni <sup>3</sup> s.	1	1	0	0	0	0				accettano	2	0	1	1	0	0
accennò	1	0	1	0	0	0				accettare	15	1	2	5	7	0
<b>accennato adj.</b>	<b>2.76</b>	<b>5</b>	<b>55.28</b>							accettarono	1	0	0	1	0	0
	5	0	0	2	1	2				accettata	5	1	0	0	3	1
accennata	2	0	0	0	1	1				accettato	15	1	6	2	5	1
accennate	1	0	0	0	0	1				accettava	2	0	1	1	0	0
accennato	2	0	0	2	0	0				accetterò	2	1	0	0	1	0
<b>accenno n.</b>	<b>4.00</b>	<b>7</b>	<b>57.14</b>							accettiamo	1	0	0	0	1	0
	7	0	2	2	0	3				accetto	7	7	0	0	0	0
accenni	3	0	1	0	0	2				accettò	1	0	0	0	1	0
accenno	4	0	1	2	0	1				<b>acciaio n.</b>	<b>9.26</b>	<b>13</b>	<b>71.22</b>			
<b>accento n.</b>	<b>3.10</b>	<b>8</b>	<b>38.76</b>								13	1	3	1	5	3
	8	0	4	4	0	0				acciai	1	0	0	0	0	1
accenti	3	0	1	2	0	0				acciaio	12	1	3	1	5	2
accento	5	0	3	2	0	0				<b>accingere v.</b>	<b>7.26</b>	<b>10</b>	<b>72.61</b>			
<b>accertare v.</b>	<b>7.10</b>	<b>12</b>	<b>59.18</b>								10	1	1	2	2	4
	12	1	1	1	6	3				*accinge	1	0	0	0	0	1
accerta	1	0	0	0	1	0				*accinger	1	0	0	1	0	0
										*accingete	1	1	0	0	0	0

## ACCORDO

*accingeva	2	0	1	0	0	1	accompagnava	6	0	3	1	2	0
*accingo	1	0	0	0	0	1	accompagnerà	2	0	1	0	1	0
*accinse	3	0	0	0	2	1	accompagnò	5	5	0	0	0	0
*accinto	1	0	0	1	0	0	accompanò	4	0	3	0	1	0
<b>accogliere v.</b>	<b>35.28</b>	<b>53</b>	<b>66.57</b>				<b>acconsentire v.</b>	<b>3.00</b>	<b>4</b>	<b>75.00</b>			
	53	4	6	14	23	6		4	1	1	1	1	0
accoglie	6	0	0	0	2	4	acconsentan	1	0	0	0	1	0
accogliendo	1	0	0	0	1	0	acconsente	1	0	0	1	0	0
accogliere	1	1	0	0	0	0	acconsenti	1	1	0	0	0	0
accogliere	12	1	0	6	4	1	acconsenti	1	0	1	0	0	0
accogliesse	1	0	0	1	0	0							
accoglieva	3	0	2	0	1	0	<b>accontentare v.</b>	<b>10.68</b>	<b>14</b>	<b>76.31</b>			
accolga	1	0	0	0	0	1		14	3	5	1	2	3
accolgono	1	0	1	0	0	0	accontentano	1	0	1	0	0	0
accolse	4	0	2	1	1	0	accontentare	9	3	3	1	1	1
accolsero	1	0	0	1	0	0	accontentato	1	0	0	0	1	0
accolta	4	1	0	0	3	0	accontentava	1	0	1	0	0	0
accolte	4	0	0	1	3	0	accontenterei	1	0	0	0	0	1
accolti	5	1	0	2	2	0	accontentiamo	1	0	0	0	0	1
accolto	9	0	1	2	6	0							
<b>accomodamento n.</b>	<b>4.13</b>	<b>6</b>	<b>68.82</b>				<b>accorato adj.</b>	<b>2.76</b>	<b>5</b>	<b>55.28</b>			
accomodamento	6	1	2	1	2	0		5	1	2	2	0	0
							accorata	3	0	2	1	0	0
							accorati	1	0	0	1	0	0
<b>accomodare v.</b>	<b>3.86</b>	<b>11</b>	<b>35.08</b>				accorato	1	1	0	0	0	0
	11	7	4	0	0	0							
*accomoda	1	1	0	0	0	0	<b>accordare v.</b>	<b>10.82</b>	<b>18</b>	<b>60.13</b>			
accomodar	2	1	1	0	0	0		18	2	1	1	6	8
accomodare	1	0	1	0	0	0	*accorda	2	0	0	0	0	2
accomodate	1	0	1	0	0	0	accordano	1	0	0	0	1	0
*accomodati	1	0	1	0	0	0	*accordano	1	0	0	0	0	1
*accomodato	1	1	0	0	0	0	accordar	2	0	0	0	0	2
accomoderà	1	1	0	0	0	0	accordare	1	0	0	0	1	0
*accomodi i.	3	3	0	0	0	0	accordarono	1	0	1	0	0	0
							accordata	3	1	0	1	1	0
<b>accompagnare v.</b>	<b>82.15</b>	<b>95</b>	<b>86.48</b>				accordate	2	0	0	0	1	1
	95	18	24	10	24	19	accordati	1	0	0	0	0	1
accompagna	9	4	0	2	0	3	accordato	1	0	0	0	1	0
*accompagna	5	0	0	1	1	3	accordi	2	0	0	0	1	1
accompagnai	1	1	0	0	0	0	accordiate	1	1	0	0	0	0
accompagnan	12	1	3	2	2	4							
accompagnar	8	3	3	1	1	0	<b>accordo n.</b>	<b>49.38</b>	<b>90</b>	<b>54.87</b>			
accompagnare	1	0	0	0	0	1		90	12	6	14	50	8
accompagnasse	3	0	2	1	0	0	accordi	18	0	1	2	15	0
accompagnato	39	4	9	2	16	8	accordo	72	12	5	12	35	8

# ACCORGERE

<b>accorgere v.</b>	<b>31.22</b>	<b>46</b>	<b>67.87</b>
	46	16	14 11 5 0
*accorga³	1	0	1 0 0 0
accorge	1	1	0 0 0 0
*accorge	3	1	0 2 0 0
*accorgendo	1	0	0 0 1 0
*accorger	8	1	3 3 1 0
*accorgerai	1	1	0 0 0 0
accorgere	2	0	0 1 1 0
*accorgesse	1	0	1 0 0 0
*accorgeva	1	0	1 0 0 0
*accorgevan	1	0	0 0 1 0
*accorgevi	1	1	0 0 0 0
*accorgi	1	1	0 0 0 0
*accorgiamo	1	0	0 1 0 0
*accorgo	4	1	2 1 0 0
*accorgono	4	2	2 0 0 0
accorta	1	0	0 0 1 0
*accorta	3	2	1 0 0 0
accorti	1	0	0 1 0 0
*accorto	10	4	4 2 0 0
<b>accorrere v.</b>	<b>25.45</b>	<b>42</b>	<b>60.59</b>
	42	6	20 6 10 0
accorre	1	1	0 0 0 0
accorreranno	1	0	0 0 1 0
accorrere	4	1	2 1 0 0
accorgete i.	2	2	0 0 0 0
accorreva	3	0	3 0 0 0
accorrono	1	0	0 0 1 0
accorsa	2	0	0 1 1 0
accorse	4	0	3 0 1 0
*accorse	13	0	12 0 1 0
accorsero	2	0	0 1 1 0
*accorsero	2	0	0 1 1 0
accorsi	3	1	0 0 2 0
*accorsi	1	1	0 0 0 0
accorso	3	0	0 2 1 0
<b>accostare v.</b>	<b>17.22</b>	<b>24</b>	<b>71.74</b>
	24	3	10 5 3 3
accosta	2	1	0 1 0 0
*accosta	1	0	0 0 1 0
*accosta i.	1	1	0 0 0 0
accostando	1	0	1 0 0 0
accostano	1	0	0 1 0 0

*accostano	1	0	0 1 0 0
*accostar	3	0	2 0 0 1
accostarono	1	0	0 1 0 0
accostata	1	0	1 0 0 0
accostati	1	0	0 0 1 0
accostato	1	0	0 0 0 1
*accostato	2	1	0 0 0 1
*accostava	1	0	1 0 0 0
*accostiamo	1	0	1 0 0 0
accostò	2	0	1 1 0 0
*accostò	4	0	3 0 1 0
<b>accrescere v.</b>	<b>10.66</b>	<b>16</b>	<b>66.63</b>
	16	2	1 4 2 7
accrebbero	1	0	1 0 0 0
accresce	2	1	0 0 0 1
*accresce	1	0	0 0 0 1
accrescendo	1	0	0 0 0 1
accrescer	2	0	0 2 0 0
accrescere	5	1	0 0 2 2
accresceva	1	0	0 1 0 0
accrescevan	1	0	0 1 0 0
accresciuta	1	0	0 0 0 1
accresciuto	1	0	0 0 0 1
<b>accumulare v.</b>	<b>8.42</b>	<b>10</b>	<b>84.19</b>
	10	1	2 3 2 2
accumulano	1	0	0 0 0 1
*accumulano	3	0	0 1 2 0
accumulare	1	1	0 0 0 0
accumularono	1	0	0 0 0 1
accumulate	1	0	1 0 0 0
*accumulati	1	0	1 0 0 0
accumulato	1	0	0 1 0 0
accumulò	1	0	0 1 0 0
<b>accusa n.</b>	<b>7.58</b>	<b>15</b>	<b>50.56</b>
	15	3	0 4 8 0
accusa	8	2	0 2 4 0
accuse	7	1	0 2 4 0
<b>accusare v.</b>	<b>15.00</b>	<b>19</b>	<b>78.95</b>
	19	4	4 6 4 1
accusano	1	0	0 1 0 0
accusar	2	1	0 1 0 0
accusare	2	1	0 1 0 0



# ADDORMENTARE

accusati	1	0	0	0	1	0
*accusati	1	0	0	1	0	0
accusato	5	0	1	0	3	1
accusava	1	0	1	0	0	0
accuserà	1	0	1	0	0	0
accuseranno	2	1	0	1	0	0
*accusino	1	0	0	1	0	0
accuso	1	1	0	0	0	0
accusò	1	0	1	0	0	0

<b>acerbo adj.</b>	<b>3.00</b>	<b>4</b>	<b>75.00</b>
	4	0	1 1 1 1
acerba	3	0	0 1 1 1
acerbo	1	0	1 0 0 0

<b>acido n.</b>	<b>2.48</b>	<b>30</b>	<b>8.26</b>
	30	1	1 0 0 28
acidi	1	0	0 0 0 1
acido	29	1	1 0 0 27

<b>acqua n.</b>	<b>137.45</b>	<b>173</b>	<b>79.45</b>
	173	15	43 37 23 55
acqua	134	14	39 27 13 41
acque	39	1	4 10 10 14

<b>acquistare v.</b>	<b>23.92</b>	<b>36</b>	<b>66.44</b>
	36	2	1 9 13 11
acquista	8	0	0 4 3 1
*acquista	1	0	0 1 0 0
acquistando	4	0	0 1 2 1
acquistano	2	0	0 0 1 1
acquistare	6	0	0 2 2 2
acquistate	1	0	0 0 0 1
acquistati	2	0	0 0 1 1
acquistato	7	2	0 0 3 2
acquistavan	1	0	1 0 0 0
acquisterà	1	0	0 1 0 0
acquisti	1	0	0 0 0 1
acquisto	1	0	0 0 0 1
acquistò	1	0	0 0 1 0

<b>acquisto n.</b>	<b>9.61</b>	<b>13</b>	<b>73.91</b>
	13	1	2 2 5 3
acquisti	2	0	1 1 0 0
acquisto	11	1	1 1 5 3

<b>acuto adj.</b>	<b>20.53</b>	<b>32</b>	<b>64.16</b>
	32	2	9 14 2 5
acuta	9	1	2 3 1 2
acute	3	0	0 2 0 1
acuti	5	0	1 3 0 1
acutissimo	2	0	1 0 1 0
acuto	13	1	5 6 0 1

<b>adattare v.</b>	<b>10.82</b>	<b>18</b>	<b>60.13</b>
	18	2	2 1 4 9
adatta	3	0	0 0 2 1
*adatta	2	0	0 0 1 1
*adattano	2	0	0 0 0 2
adattar	1	0	1 0 0 0
*adattar	3	1	0 1 1 0
adattare	1	0	0 0 0 1
adattata	1	0	0 0 0 1
adattate	1	0	0 0 0 1
*adattava	1	0	1 0 0 0
*adatterò	1	1	0 0 0 0
adatti	1	0	0 0 0 1
adatto	1	0	0 0 0 1

<b>adatto adj.</b>	<b>9.00</b>	<b>12</b>	<b>75.00</b>
	12	1	3 1 3 4
adatta	5	1	2 1 0 1
adatte	2	0	0 0 0 2
adatti	2	0	1 0 1 0
adatto	3	0	0 0 2 1

<b>addietro adv.</b>	<b>4.45</b>	<b>7</b>	<b>63.58</b>
	7	0	2 3 1 1

<b>addio n.</b>	<b>3.08</b>	<b>6</b>	<b>51.41</b>
	6	2	1 3 0 0

<b>addirittura adv.</b>	<b>21.20</b>	<b>30</b>	<b>70.66</b>
	30	6	3 12 2 7

<b>addomesticare v.</b>	<b>3.42</b>	<b>5</b>	<b>68.38</b>
	5	2	1 1 1 0
addomestica	5	2	1 1 1 0

<b>addormentare v.</b>	<b>8.90</b>	<b>22</b>	<b>40.47</b>
	22	6	14 1 1 0
addormenta	1	0	1 0 0 0