

JANUA LINGUARUM

STUDIA MEMORIAE
NICOLAI VAN WIJK DEDICATA

edenda curat

C. H. VAN SCHOONEVELD
Indiana University

Series Practica, 198

THE LEXICAL STRUCTURE OF SPANISH

by

WILLIAM PATTERSON
Texas Tech University

and

HECTOR URRUTIBÉHEITY
Rice University

1975
MOUTON
THE HAGUE · PARIS

© Copyright 1975 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: 73-77744

Printed in Hungary by Franklin Printing House, Budapest

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	7
1. INTRODUCTION	9
2. GENEALOGY	11
2.1 Genealogical Classes	11
2.2 Data	21
2.3 Functional Sub-Classes	22
2.4 Physical Sub-Classes	25
2.5 Statistical Sub-Classes	29
2.6 Chronological Sub-Classes	32
3. CHRONOLOGY	36
3.1 Chronological Classes	36
3.2 Data	38
3.3 Functional Sub-Classes	39
3.4 Physical Sub-Classes	46
3.5 Statistical Sub-Classes	54
3.6 Genealogical Sub-Classes	63
4. FUNCTION	71
4.1 Functional Classes	71
4.2 Data	72
4.3 Physical Sub-Classes	73
4.4 Statistical Sub-Classes	80
4.5 Chronological Sub-Classes	86
4.6 Genealogical Sub-Classes	93
5. LENGTH	100
5.1 Physical Classes	100
5.2 Data	101
5.3 Functional Sub-Classes	102
5.4 Statistical Sub-Classes	107

5.5 Chronological Sub-Classes	113
5.6 Genealogical Sub-Classes	119
6. USAGE	125
6.1 Statistical Classes.....	125
6.2 Data.....	125
6.3 Functional Sub-Classes	126
6.4 Physical Sab-Classes	132
6.5 Chronological Sub-Classes	139
6.6 Genealogical Sub-Classes	146
7. CONCLUSIONS	153
7.1 Genealogical Classes	153
7.2 Chronological Classes	154
7.3 Functional Classes	156
7.4 Physical Classes	158
7.5 Statistical Classes	159
BIBLIOGRAPHY	161

ACKNOWLEDGEMENTS

We take this opportunity to express our appreciation to Professor Alphonse Juilland of Stanford University who helped us to formulate and develop this study. We are also indebted to Professors Aurelio M. Espinosa Jr. and Robert Politzer for their valuable suggestions during the course of this project.

INTRODUCTION

The purpose of this study is to establish the lexical structure of Modern Spanish in terms of the relationships which obtain between certain fundamental properties of Spanish words.

The properties of Spanish words considered relevant in this study are the following:

(1) Functional, i.e., the structural properties underlying parts of speech classifications, such as nouns, verbs, adjectives, etc.

(2) Physical, i.e., word length measured in number of syllables, such as monosyllabic, bisyllabic, etc.

(3) Statistical, i.e., frequency classes, such as first five hundred, second five hundred, etc.

(4) Genealogical, i.e., etymological properties which underlie such classifications as inherited, borrowed, or internally created words, by a change in function, composition, derivation, etc.

(5) Chronological, i.e., historical characteristics which underlie classifications based on the period or century of first attestation, such as 12th century, 13th, etc.

With each basic Spanish word characterized in terms of these five properties, the lexical structure of Spanish can be conceived as the network of relations which obtain between word classes determined in terms of etymology, chronology, function, length and frequency.

The Spanish vocabulary is represented by its manifestations in modern Spanish texts rather than by the inventory listed in a modern dictionary. In other words, we aim to account for, and make predictions about certain fundamental properties of Modern Spanish texts rather than dictionaries. While a standard dictionary lists some 100,000 entries, the first 100 most frequently used words account for more than 30% of the lexical materials that constitute any Spanish text; the first 1000 account for more than 50%; the first 5000 for more than 90%. In other words, the occurrences of most entries registered in a Spanish dictionary account for only a minute part of representative Spanish texts. Therefore, instead of basing our conclusions on a study of all lexical items, which would give a distorted picture of lexical structure by placing on the same footing the many words likely to occur once every 1000 pages with the very few likely to occur ten times per page, it is preferable to consider only a few thousand words, taken from the top of a frequency hierarchy.

Reliance upon the weighted information provided by a frequency dictionary provides a deeper insight and a more adequate picture of the lexical structure of a language. Since there are no less than four frequency dictionaries of Spanish,¹ our first task was to choose the most adequate one for our purpose. We selected the *Frequency Dictionary of Spanish Words* by Alphonse Juilland and E. Chang-Rodríguez because the corpus underlying it was determined by more reliable sampling techniques and the results were weighted through exhaustive scanning combined with improved methods for determining word usage.

Unlike previous studies in which words were ranked and classified in terms of their frequency, the Juilland-Chang-Rodríguez dictionary ranks the basic vocabulary according to three coefficients: usage, frequency and dispersion. In earlier investigations, usage was simply equated with frequency on the assumption that, provided the sampling is adequate, a one-to-one correspondence holds roughly between the number of occurrences of a word in the sample and its occurrences in the language as a whole. Subsequently, as weighting techniques were refined, it appeared that results obtained on the basis of the assumed frequency-usage correlation often failed to correspond to intuitive expectations based on the specialist's knowledge of the language. True, intuitive expectations have been wrong in more than one instance, but equally often, objective evidence could be marshalled to prove that this was not the case. Take a count yielding, for example, 24 occurrences for a word such as *químico*, and only 12 occurrences for a word such as *consultar*, which suggests that the word *químico* is twice as important as the word *consultar*. The linguist's doubts about this assumption may be reinforced by differences in the 'dispersion' of the word's occurrences in the five equal sub-samples of the lexical universe. Whereas the fewer occurrences of *consultar* are fairly evenly distributed, 4 2 2 1 3, the more numerous occurrences of *químico* cluster almost exclusively in two genres, 0 0 1 10 13. 'Usage' attempts to account for both frequency and dispersion by combining them in a formula which yields a somewhat different coefficient: due to its uneven dispersion, *químico*, with a frequency of 24, has a coefficient of usage of 10.09, while *consultar* with a frequency of 12, has a coefficient of 9.45.

¹ L. Rodríguez Bou, *Recuento de vocabulario español* (Baltimore, 1952); Milton A. Buchanan, *A Graded Spanish Word Book* (Toronto, 1927); V. García Hoz, *Vocabulario usual, vocabulario común y vocabulario fundamental* (Madrid, 1953); and Alphonse Juilland and E. Chang-Rodríguez, *Frequency Dictionary of Spanish Words* (The Hague, 1964).

GENEALOGY

2.1 GENEALOGICAL CLASSES

We recognize three major genealogical classes of words: inherited from Latin; borrowed from other languages; and created in Spanish by various formative processes.¹

In assigning words to genealogical classes, we have relied primarily on Joan Corominas' *Diccionario crítico etimológico de la lengua castellana*,² and on his *Breve diccionario etimológico de la lengua castellana*.³

2.11 *Inherited*

By 'inherited' we mean the words which used to constitute the vocabulary of the Vulgar Latin of Spain and which became the vocabulary of the earliest form of Spanish. Consisting essentially of Latin words, it also contains a number of elements borrowed by Vulgar Latin from the various substrata and adstrata of the peninsula.

2.111 *Latin*

Spanish words inherited from Latin are of two types:

(1) POPULAR, whose shape conforms to the phonological rules which characterize the development of Vulgar Latin into Spanish: *marido* < MARITUM, *diez* < DECEM, *bueno* < BONUM, *otro* < ALTERUM;

(2) SEMI-POPULAR, whose phonological shape, adjusted under learned influence, does not completely conform to the rules of Spanish sound change: *fruto* < FRUCTUM, *fiesta* < FESTA, *siglo* < SAECULUM.

¹ We have not added a separate class for words of unknown or uncertain origin, for, while their particular etyma may be in question, the genealogical class to which they belong is unquestioned: they are indisputably borrowings; therefore, such words have been assigned to a special sub-class of borrowed words.

² (Berne, 1954–57); hereinafter referred to as *Diccionario crítico*.

³ (Madrid, 1961); hereinafter referred to as *Breve diccionario*.

2.112 *Celtic*

The Spanish vocabulary contains a few words of Celtic origin, known for the most part only in a Latinized form. The language brought into Spain by Celtic invaders was presumably that of Gaul, but “it is difficult to say whether a Celtic word was there before the Romans, was brought to Spain by Roman legionaries, or came in later with the Germanic invasions”.⁴ There is no clear evidence of a direct contact between the Celtic of Spain and the later Peninsular languages, except for two words: *tona* (Galician, Portuguese), *ton* (Welsh), *tonn* (Irish); and *arapende* < ARAPENNIS.⁵ But scholars agree that the inhabitants of the Iberian Peninsula retained a few Celtic words, either because they were deeply rooted or because there were no equivalents in Latin.⁶ This is especially true of the names of plants, agricultural tools, clothing, food, and of words related to differences in climate, rural customs, and regional traditions. From the rather small contribution of Celtic to the Spanish vocabulary, we might mention *carro* < CARRUM, *camino* < CAMMINUS, *cambiar* < CAMBIARE, *camisa* < CAMISIA, *legua* < LEUCA, *pieza* < PETTIA, *pico* < BECCUS, and *mina* < *MINA (which may have come into Spanish by way of French).

2.113 *Basque, Iberian*

The Spanish vocabulary shows elements in common with Basque and, as far as we know, Iberian.⁷ Like Celtic words in French, the pre-Romance remnants are well represented in place names and in certain personal names. Most of these words have been given a pre-Romance Iberian ancestry because no adequate derivation could be found in Latin: e.g., Spanish *vega* (Portuguese *veiga*), attested in medieval texts as *vaica* or *vaiga*, seems to be a cognate of Basque *ibaiko* ‘bank’, derived from *ibai* ‘river’.⁸ Pliny used ARRUGIA ‘subterranean channel’, which seems to be the origin of *arroyo*. Other words of presumed pre-Romance origin include *gordo* < GURDUS; *caparazón* (perhaps related to Catalan *carabassa*); *cama* < CAMA; *manteca* < MANTECCA (which may be of Indo-European origin and related to Slavic *smetana*); *pestaña* (common to Spanish, Portuguese, Catalan, and Gascon, from PISTANNA, probably related to the Basque *pizta*).

Basque is the only pre-Romance language of the Iberian Peninsula which has survived. The most common pre-Romance word is *izquierdo*, Basque *ezkerr*, Catalan *esquerre*, and Portuguese *esquerdo*. According to Elcock, “its meaning of ‘left’ was

⁴ J. B. Trend, *The Language and History of Spain* (London, 1953), 17, hereinafter cited as *Language*.

⁵ W. J. Entwistle, *The Spanish Language* (London, 1965), 40; hereinafter cited as *Spanish Language*.

⁶ Erich Auerbach, *Introduction to Romance Languages and Literature* (New York, 1961), 23; hereinafter cited as *Romance Languages*.

⁷ Entwistle, *Spanish Language*, 33.

⁸ Trend, *Language*, 15.

still conveyed by *siniestro* < SINISTER, but *izquierdo* is early attested, and suggestions that it might be a relatively recent borrowing from Basque are weakened by the widespread use of cognate forms in southern France and Portugal".⁹

Barro (Aragonese *bardo*) shows the same *-rr-~ -rd-* alternance. Similar in structure, and unexplained, is *perro*. Corominas discounts the Celtic or Iberian etymologies and suggests that *perro* may derive from the onomatopoeic *prrr*, *brrr*, which shepherds use to arouse the dog.¹⁰ The words *pata* and *tirar* are probably pre-Roman. Bloch and Wartburg consider that *pata* (French *patte*), unknown in Gaul, belongs to a pre-Celtic stratum.¹¹ It was adopted by the Franks, who carried it into Gaul. *Tirar*, common to all the Romance languages but Rumanian, may come from an Indo-European root **(s)tig*.¹²

2.114 Germanic

Even before the occupation of portions of the Roman Empire by Germanic tribes, certain Germanic words had been introduced into Vulgar Latin. According to Spaulding, more than 300 Romance words are a result of the intimate contact on several frontiers between German tribes and Roman soldiers and colonizers. Among the legionaries, too, many were Germans. Words of Germanic origin were in use among Romans long before the fall of the Empire, and some, e.g., *burgus*, *harpa*, *sapon* > Spanish *jabón*, appear in late Latin writers.¹³ Although there is a dearth of documentary testimony concerning the Germanic element in Vulgar Latin, Romance evidence suggests that the following terms were adopted by Vulgar Latin prior to the conquest of Spain by the Goths in the fifth century:¹⁴ *banda* < BANDWO, *blanco* < BLANK, *banco* < BANK, *bruno* < BRUN, *espuma* < SKUMS (English *scum*), *gris* (borrowed by Spanish from Provençal GRIS), *guardar* < WARDON, *guarnecer* < *guarnir* < WARNJAN, *guerra* < WERRA, *guisa* < WISA, *marca* < MARKA, *robar* < RAUBON (German *rauben*).

Among the Germanic words which came to the Peninsula during the main Germanic invasions from the fifth century onwards are *brote* < *BRUTE, *espía* < *SPAIHA, *esquina* < *SKINA, *guiar* < WIDAN;¹⁵ *fresco* < FRISK, *franco* < FRANK, *ganar* < GAINON, whose meaning 'to earn' evolved under the influence of another *guadanar*; *ganso* < GANS, *ropa* < RAUPA, *rico* < REIKS, *sala* < SAL, *sacar* < SAKAN. Some Germanic words came into Hispano-Romance through Catalan, French, or Provençal, e.g., *orgullo*, *bosque*.

⁹ W. D. Elcock, *The Romance Languages* (London, 1960), 178.

¹⁰ Joan Corominas, *Breve diccionario*, 449.

¹¹ Bloch and W. von Wartburg, *Dictionnaire étymologique de la langue française* (Paris, 1964), 469.

¹² Corominas, *Breve diccionario*, 554.

¹³ Robert K. Spaulding, *How Spanish Grew* (Berkeley, 1943), 49.

¹⁴ Elcock, *Romance Languages*, 207–10.

¹⁵ Corominas, *Breve diccionario*, 302.

2.12 *Borrowed*

The second major class consists of words borrowed from other languages. Note that non-Latin words were considered 'inherited' if borrowed by and present in Vulgar Latin before the formation of the Spanish language (see 2.11); on the other hand, Latin words were considered 'borrowed' if taken from Latin writings after the formation of the Spanish language. Corominas uses small italics and the expression *tomado del latín* to distinguish Latin borrowings from the Latin inherited words, which he prints in capital letters.

2.121 *Latin*

Latin contributes the greatest number of borrowings to the Spanish lexicon. Usually adjusted to the rules which characterize Spanish historical phonology, Latin borrowings (*cultismos* or *voces cultas*) differ from the strictly popular or inherited words inasmuch as they do not undergo the phonological changes reflected by inherited words. In general, learned borrowings changed little; e.g. *público* < PUBLICUS (10th century); *virtud* < VIRTUTIS (11th century); *inclinarse* < INCLINARE (12th century); *violencia* < VIOLENTIA (13th century); *original* < ORIGINALIS (14th century); *presencia* < PRAESENTIA (15th century); *provenir* < PROVENIRE (16th century); *museo* < MUSEUM (17th century); *uniforme* < UNIFORMIS (18th century); *frase* < PHRASIS (19th century); *medieval* < MEDIUM AEVUM (20th century).

During the Renaissance, a return to classical sources reintroduced many words which had fallen into desuetude. From then on, the Latin lexicon is constantly explored for words which may express new meanings. The Spanish vocabulary exhibits such a vast number of apparently 'learned' words which are, and always seem to have been, in popular use, that linguists have wondered whether a considerable portion of the learned layer is not in reality contemporaneous with the 'popular' vocabulary, representing the conservative influence of the more cultured classes from classical times, an influence which frequently overcame popular tendencies and imposed itself upon the popular language.¹⁶

Corominas, who distinguishes between truly LEARNED and SEMI-LEARNED words, between *cultismos* and *semi-cultismos*, seems to use the latter with two different meanings: INHERITED WORDS, whose phonological shape has been altered under learned influence (see 2.11); and BORROWINGS, whose shape has been adjusted to popular forms, e.g., *regla* < REGULA, *milagro* < MIRACULU, *peligro* < PERICULU, *virgen* < VIRGINE. Menéndez-Pidal distinguishes the *voces cultas* and *semi-cultas* from the popular forms, "pues tienen un desarrollo distinto de las voces

¹⁶ W. D. Elcock, Review of J. Corominas, *Diccionario crítico etimológico de la lengua castellana*, in *Modern Language Review* LII (1957), 290-91.

estrictamente populares".¹⁷ But the latter deserve more attention "por su complicado desarrollo, por ser en ellas donde se manifiestan en modo mas completo las leyes fundamentales de la vida del lenguaje y por formar el fondo mas rico del español y su herencia patrimonial".¹⁸

2.122 Romance Languages

French is the Romance language which has contributed most to the Spanish lexicon. Apart from Latin, the principal influence exerted on Spanish during the medieval period is that of French, both the *langue d'oc* and the *langue d'oïl*.¹⁹ Many words entered the language by way of the pilgrims on their way to Compostela, or by the minstrels who entertained them. The influence of the French monks made itself felt in the substitution of the Roman ritual for the Mozarabic, the introduction of the Carolingian handwriting, the restoration of Latin (with the presumable checking of the vernacular), and the translation and imitation of French literary works.²⁰ Some early borrowings from French and from Provençal are:

	FRENCH	PROVENÇAL
11th century	<i>pleito</i>	—
12th century	<i>granja, renta</i>	<i>fraile, jamás</i>
13th century	<i>joya, talle</i>	<i>jornada, laurel</i>
14th century	<i>flecha, galán</i>	<i>marqués</i>
15th century	<i>jardín, norte</i>	<i>embajada, patio</i>

French continued its influence during the 16th and 17th centuries, but the periods of greatest French prestige were the 18th and 19th centuries. In spite of reactions to the *galicismo*,²¹ French words have continued to become a part of Spanish: *marchar, trinchera, país* (16th); *emoción, jefe, taller, tren* (17th); *destacar, instalar, paisaje, realizar, revolucionario, romántico, ruta* (18th); *detalle, egoismo, entrevista, evolución, hotel* (19th); *avión* (20th).

Other languages of the Iberian Peninsula had a small but nonetheless noticeable influence. Over the centuries, Portuguese and Spanish exchanged a number of words, two of which occur in our word list: *traje* and *vera* (which may be of pre-Romance or Celtic origin). Spanish owes to Catalan *presa* (10th); *mercader* (12th); *bala, caja, falda, orgullo, linaje, papel, pólvora, zozobra* (13th); *farol, salvaje, vanguardia, viaje* (14th); *avanzar, bosque, correo, cuartel, festejar, forastero, guante, pantalla, plantel, prensa, retablo, semblante, sor, trozo* (15th).

¹⁷ Menéndez-Pidal, *Manual*, 9.

¹⁸ Menéndez-Pidal, *Manual*, 14.

¹⁹ Entwistle, *Spanish Language*, 206.

²⁰ Spaulding, *How Spanish Grew*, 130–31.

²¹ Tomás de Iriarte, *Los literatos en cuaresma* (1773); and Rafael María Baralt, *Diccionario de galicismos* (1855).

Like French, but to a lesser degree, Spanish borrowed a number of words from Italian, especially during the Renaissance:²² *aguantar, apoyar, asalto, balcón, batallón, bronce, campeón, capricho, caricia, coronel, charlar, equilibrio, escopeta, fachada, fanal, fracasar, grupo, guardia, ingeniero, lápiz, manejar, marcar, medalla, modelo, muralla, novela, piloto, relieve, retrato*. Although the Italian influence declined after this period, a few Italianisms entered the language during the following centuries: *esbelto, recinto* (17th); *café, folleto, pintoresco* (18th); *fascismo* (20th).

2.123 Arabic

Direct borrowings from Arabic were quite numerous; in fact, there are 36 Arabic loan words in the basic vocabulary, which includes terms related to administration, war, agriculture, clothing, plants, games, music, mathematics, and astronomy: *barrio* (10th); *alcalde, alcazar, aldea* (11th); *arrabal, hazañar, ronda* (12th); *aceite, achacar, alcoba, alcohol, almacén, alquiler, arroz, asesino, auge, azar, azúcar, garra, hasta, jinete, marfil, rincón, taza* (13th); *alfiler, alforja, andaluz, gabán, guitarra* (14th); *almanaque, azotea, azulejo, cifra, limón, ola, tarea* (15th). Arabic borrowings are mostly nouns, with one verb (*achacar*), one adjective (*andaluz*), and one preposition (*hasta*).

2.124 Greek

Greek has provided Spanish with a number of learned forms associated with literature, sports, scholarship, and education:²³ *monasterio* (11th); *cara* (12th); *astronomía, música, fantasma* (13th); *calma, monarca* (14th); *dramático, escolástico, prólogo, tema, trágico* (15th); *análisis, categoría, época, simpatía* (17th); *anécdota, crítica, farmacia, patriota, sistema* (18th); *autónomo, estético, panorama, teléfono* (19th).

2.125 Other Languages

Ignored on the Continent during the 16th and 17th centuries, English began to exert some influence, first through literature, then because of social prestige. Some English words reached Spanish directly, e.g., *lord, suicidio* (18th); *cheque, club, organismo, turista* (19th); *comité, interviú, racial* (20th), etc., but one word, *tranvía*, reached Spanish through the French *tramway*.

One Hungarian borrowing, *coche*, appears in the 16th century; from Chinese comes *te* (18th); *dique* (16th) and *escaparate* (17th) come from Dutch; *maíz* (15th), *chocolate* and *tabaco* (16th), *patata* (17th), and *butaca* (19th) from the American Indian languages; *amén* (12th) from Hebrew; *pistola* (17th), *regimiento* (18th) and *cultural*

²² Menéndez-Pidal, *Manual*, 25.

²³ Rafael Lapesa, *Historia de la language española* (Madrid, 1955), esp. p. 44.

(20th) from modern German. Many of these words, which represent but a few of the loan words from the various languages, indicate that interborrowing is not restricted to the languages of Europe.

2.13 *Created*

The third category consists of words created in Spanish by composition (or contamination), derivation, or change of function.

2.131 *Composition and Contamination*

Composition juxtaposes two existing words to form a third, e.g., *bienestar*, *mediodía*, *kilómetro*, *malograr*, *sobresalir*, *norteamericano*, *asimismo*, *cualquier*, *porque*.

Although Corominas considers as *compuestos* Spanish words derived from words created by this process in Classical or in Vulgar Latin, we have assigned them to the 'inherited' class. Corominas also treats as compounds words like *abajo* and *arriba*, which are created by a process of agglutination (see 2.325).

Included here are words created by CONTAMINATION, a process also called BLENDING. Both composition and contamination result from the combination of two different forms; however, the constituents of compounds are in sequence, e.g., *mediodía* < *medio* + *día*, whereas those of contaminations cannot be obtained by simple segmentation, e.g., *desparramar* < *esparcir* + *derramar*, *usted* < *vuestra* + *merced*, *sombrio sol* + *umbrio*.

2.132 *Derivation*

The process of derivation²⁴ is active throughout all periods of the language and affects only lexical words. We distinguish here prefixation, suffixation, prefixation-suffixation, back formation, agglutination, change of suffix, abbreviation, and expressive creation.

2.1321 *Prefixation*

The prefixes used to create new Spanish words are nearly all of Latin origin. The following are the more productive:

(1) *a-* forms mostly verbs, e.g., *acercar*, *aparecer* (12th); *asemejar* (13th); *acomodar* (16th).

(2) *in- im- ir-* account for a few nouns: *indiferencia*, *impotencia* (13th); *impaciencia* (15th); *impureza* (17th); one verb, *infiltrar* (19th); and, with the exception of *desigual*, all adjectives created by prefixation, e.g., *inesperado* (10th); *inagotable* (13th); *incomparable* (14th); *incierto*, *inconveniente* (15th); *inútil* (16th); *inverosímil* (17th); *insignificante*, *irresponsable* (18th); *inconsciente* (19th).

²⁴ For a discussion of Spanish derivation, cf. M. Ramsey and R. Spaulding, *A Textbook of Modern Spanish* (New York, 1956), 613–60; hereinafter referred to as *Textbook*.