THE OXFORD GUIDE TO THE

Romance Languages

EDITED BY

Adam Ledgeway & Martin Maiden

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

© editorial matter and organization Adam Ledgeway and Martin Maiden 2016
© the chapters their several authors 2016

The moral rights of the authors have been asserted

First Edition published in 2016

Impression: 1

All rights reserved. No part of this publication may be reproduced, stored in
a retrieval system, or transmitted, in any form or by any means, without the
prior permission in writing of Oxford University Press, or as expressly permitted
by law, by licence or under terms agreed with the appropriate reprographics
rights organization. Enquiries concerning reproduction outside the scope of the
above should be sent to the Rights Department, Oxford University Press, at the
address above

You must not circulate this work in any other form
and you must impose this same condition on any acquirer

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2015936770

ISBN 978-0-19-967710-8

Printed and bound in
Slovakia by Neografia

Links to third party websites are provided by Oxford in good faith and
for information only. Oxford disclaims any responsibility for the materials
contained in any third party website referenced in this work.

Short Contents

Detailed Contents xii
Series Preface xli
Abbreviations xliii
The Contributors lii

Introduction liii
ADAM LEDGEWAY AND MARTIN MAIDEN

PART I: The Making of the Romance Languages

1. Latin as a source for the Romance languages 3
JAMES CLACKSON
2. Latin and Romance in the medieval period: a sociophilological approach 14
ROGER WRIGHT
3. Early evidence and sources 24
BARBARA FRANK-JOB AND MARIA SELIG

PART II: Typology and Classification

4. A structural comparison of Latin and Romance 37
NIGEL VINCENT
5. Romance: a typological approach 50
PAOLO RAMAT AND DAVIDE RICCA
6. Classifications 63
GEORG BOSSONG
7. Romance linguistic geography and dialectometry 73
HANS GOEBL

PART III: Individual Structural Overviews

8. Romanian, Istro-Romanian, Megleno-Romanian, and Aromanian 91
MARTIN MAIDEN
9. Dalmatian 126
MARTIN MAIDEN
10. Friulian 139
PAOLA BENINCÀ AND LAURA VANELLI
11. Ladin 154
GIAMPAOLO SALVI

12. Romansh (Rumantsch) 169
STEPHEN R. ANDERSON

13. The dialects of northern Italy 185
PAOLA BENINCÀ, MAIR PARRY, AND DIEGO PESCARINI

14. Italian, Tuscan, and Corsican 206
ADAM LEDGEWAY

15. The dialects of central Italy 228
MICHELE LOPORCARO AND TANIA PACIARONI

16. The dialects of southern Italy 246
ADAM LEDGEWAY

17. Sardinian 270
GUIDO MENSCHING AND EVA-MARIA REMBERGER

18. French and northern Gallo-Romance 292
JOHN CHARLES SMITH

19. Southern Gallo-Romance (Occitan) 319
MICHÈLE OLIVIÉRI AND PATRICK SAUZET

20. Francoprovençal 350
ANDRES KRISTOL

21. Catalan 363
ALEX ALSINA

22. Spanish, Astur-Leonese, Navarro-Aragonese, Judaeo-Spanish 382
DONALD N. TUTEN, ENRIQUE PATO, AND ORA R. SCHWARZWALD

23. Galician and Portuguese 411
FRANCISCO DUBERT AND CHARLOTTE GALVES

24. Creoles 447
ANNEGRET BOLLÉE AND PHILIPPE MAURER

PART IV: Comparative Overviews

A. Phonology 469

25. Segmental phonology 471
STEPHAN SCHMID

26. Prosodic structure 484
GIOVANNA MAROTTA

B. Morphology 495

27. Inflectional morphology 497
MARTIN MAIDEN

28. Derivational morphology 513
FRANZ RAINER

29. Compounding 524
FRANCESCA FORZA AND SERGIO SCALISE

C. Syntax 539

30. The structure of the nominal group 541
GIULIANA GIUSTI

31. The structure of the clause 556
SILVIO CRUSCHINA AND ADAM LEDGEWAY

D. Semantics and pragmatics 575

32. Lexical stability and shared lexicon 577
STEVEN N. DWORKIN

33. Onomasiological differentiation 588
INGMAR SÖHRMAN

34. Information and discourse structure 596
SILVIO CRUSCHINA

E. Sociolinguistics 609

35. Sociolinguistic variation 611
MARI C. JONES, MAIR PARRY, AND LYNN WILLIAMS

36. Diglossia 624
JOHANNES KABATEK

37. Standardization 634
CHRISTOPHER J. POUNTAIN

PART V: Issues in Romance Phonology

38. Diphthongization 647
MARTIN MAIDEN

39. Palatalization 658
LORI REPETTI

40. Sandhi phenomena 669
RODNEY SAMPSON

41. Writing systems 681
THOMAS FINBOW

PART VI: Issues in Romance Morphology

42. Number 697
MARTIN MAIDEN

43. Morphomes 708
MARTIN MAIDEN

44. Tonic pronominal system: morphophonology 722
CHIARA CAPPELLARO

45. Clitic pronominal systems: morphophonology 742
DIEGO PESCARINI

PART VII: Issues in Romance Syntax

46. Functional categories — 761
ADAM LEDGEWAY

47. Subject clitics: syntax — 772
CECILIA POLETTO AND CHRISTINA TORTORA

48. Object clitics — 786
IAN ROBERTS

49. Auxiliary selection and participial agreement — 802
MICHELE LOPORCARO

PART VIII: Issues in Romance Syntax and Semantics

50. Split intransitivity — 821
DELIA BENTLEY

51. Negation — 833
CECILIA POLETTO

52. Copular and existential constructions — 847
DELIA BENTLEY AND FRANCESCO MARIA CICONTE

PART IX: Issues in Romance Pragmatics and Discourse

53. Illocutionary force — 863
ION GIURGEA AND EVA-MARIA REMBERGER

54. Deixis — 879
ADAM LEDGEWAY AND JOHN CHARLES SMITH

55. Address systems — 897
RICHARD ASHDOWNE

PART X: Case Studies

A. The nominal group — 909

56. Case — 911
ADINA DRAGOMIRESCU AND ALEXANDRU NICOLAE

57. Gender — 924
MICHELE LOPORCARO

B. The verbal group — 937

58. Tense and aspect — 939
PIER MARCO BERTINETTO AND MARIO SQUARTINI

59. Mood — 954
JOSEP QUER

0. Voice — 967
Michela Cennamo

1. Complex predicates — 981
Michelle Sheehan

. The clause — 995

2. Word order — 997
Giampaolo Salvi

3. Clausal complementation — 1013
Adam Ledgeway

54. Relative clauses — 1029
Elisabeth Stark

References — 1041

Index — 1169

CHAPTER 21

Catalan

Alex Alsina

21.1 Introduction: external elements

21.1.1 Territory, geographical dialects, and demography

Catalan is the native language of portions of four different present-day states in Europe. The following territories constitute the area in which Catalan is the ancestral language: (1) Andorra; (2) the département of Pyrénées-Orientales in France; (3) Catalonia, excluding the Occitan-speaking Val d'Aran; the eastern strip of Aragon; most of the Valencian Country; and the Balearic Islands, in Spain; and (4) the town of Alghero in Sardinia, in Italy. See Map 21.1 of the Catalan-speaking area, which shows the classification of the geographical dialects of Catalan into the western and eastern groups.

The population of the Catalan-speaking area is eleven million, more than seven million of whom can speak the language. The vitality of Catalan varies considerably from region to region. Catalan has limited presence at both the northern and southern tips of its linguistic area, but is alive and visible elsewhere in this area.

21.1.2 History

Catalan originated and developed in the northern strip of the current Catalan-speaking area on both sides of the Pyrenees, to which the pre-Catalan Romance-speaking area was reduced as a result of the Muslim occupation of most of the Iberian Peninsula in 711. Over the following six centuries, Catalan gradually expanded southwards. The last major expansion of the language took place in the thirteenth century with the conquest of the kingdoms of Majorca (the Balearic Islands) and Valencia.

Catalan is first attested in the written record in the ninth century as words or short phrases in texts otherwise written in Latin, and we find the first texts fully written in Catalan around 1150 (cf. §3.1). Catalan is the main written language in the Catalan-speaking area from the twelfth century, as shown by the large number of texts written in Catalan (see the *Corpus Informatitzat del Català Antic* (*CICA*)).

During the sixteenth and seventeenth centuries Catalan remained practically the only language used in the Catalan linguistic area. However, two events in the seventeenth and eighteenth centuries led to its subsequent subordination to foreign languages: the dismemberment of the portion of Catalonia north of the Pyrenees from the rest of Catalonia and its annexation to France in 1659, and the annexation of the Catalan-speaking territories south of the Pyrenees to Castile after the War of the Spanish Succession in 1714. The disappearance of a Catalan state was followed by a series of laws prohibiting the use of Catalan in all official and public situations, making French north of the new border and Spanish south of it the languages for written and formal uses.

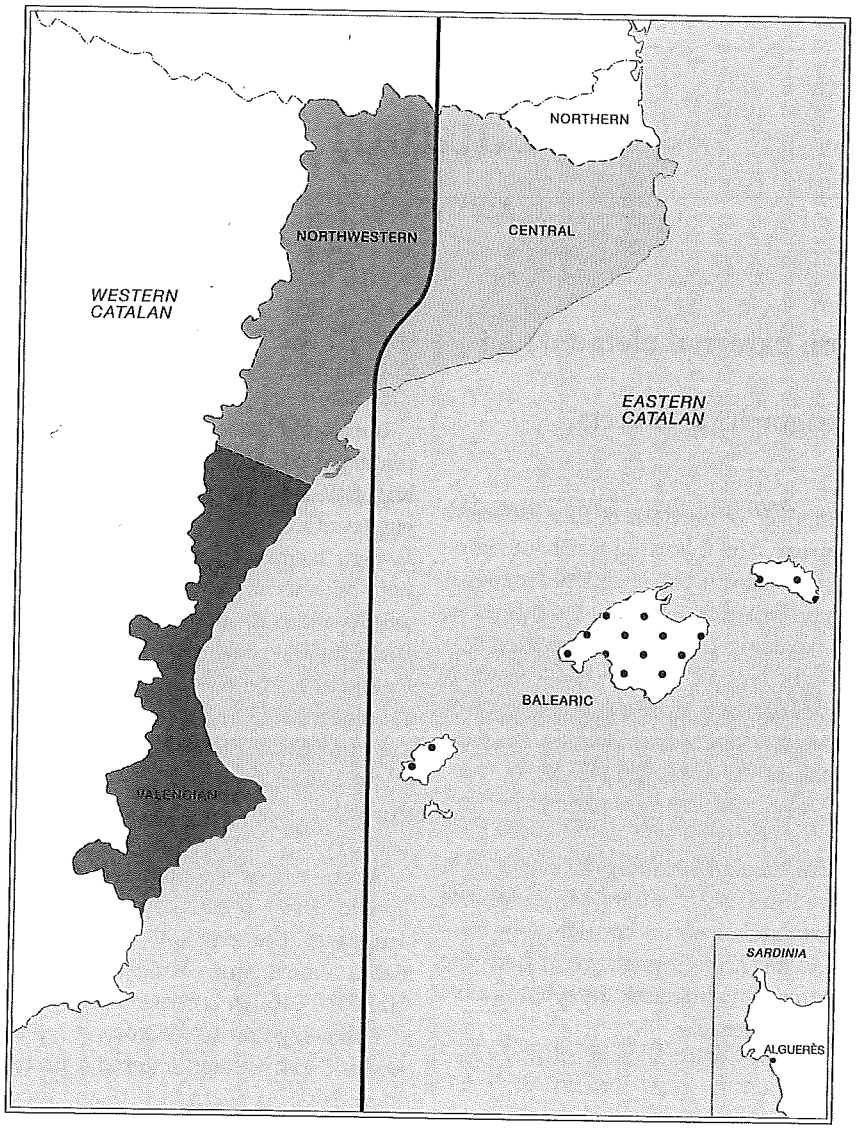
At the end of the nineteenth century, there was a widespread social demand for a unified standard form of the language. The engineer and self-made linguist Pompeu Fabra almost single-handedly developed and proposed the standard Catalan norms currently in use.

Following Franco's victory at the end of the Spanish civil war in 1939, Catalan entered a period of complete prohibition. With his death in 1975 and the collapse of the regime, the Catalan territories under Spanish administration have implemented language policies designed to increase the knowledge and use of Catalan. As a result, Catalan is again used in all communicative contexts, even though in some (e.g. the administration of justice) its presence is still very limited (see Melchor and Branchadell 2007 for more information about the history and sociology of the language).

21.2 Salient diachronic features

This section reviews some of the salient diachronic features of Catalan (Badia i Margarit 1951; Moll 1952), many of which are shared with Occitan, to which it is closely related.

The Oxford Guide to the Romance Languages. Adam Ledgeway and Martin Maiden (eds)
This chapter © Alex Alsina 2016. Published 2016 by Oxford University Press.



Map. 21.1 The Catalan-speaking area

21.2.1 Vowels

One of the diachronic features that distinguishes Catalan (see also §14.2.1.1 for Corsican) in the Romance context is the 'switch' that has occurred with the mid front vowels: spoken (late) Latin /e/ has evolved into central Catalan /ɛ/ and proto-Romance /ɛ/ has evolved into central Catalan /e/, leaving aside a few instances of /ɛ/ that have remained as such. It is clear that this switch cannot have occurred directly: if /ɛ/ had changed into /e/, while /e/ remained /e/, the two vowels would have merged. Whereas /ɛ/ became /e/ in all Catalan dialects (1a), apart from certain contexts (1b), the treatment of Latin /e/ shows a dialectal split: /ə/ in most of the Balearic Islands, /ɛ/ in central Catalan, and /e/ in the western dialects, as shown in (1c).

(1)

a. SĔLLAM > s/e/lla 'saddle' LĔPOREM > ll/e/bre 'hare'
BRĔUEM > br/e/u 'brief' SĔMPER > s/e/mpre 'always'
SĔCAT > s/e/ga 'reaps; cuts' BĔNE > b/e/ 'well'

b. FĔRRUM > f/ɛ/rro 'iron' APĔRTUM > ob/ɛ/rt 'open'
CAELUM > c/ɛ/l 'sky' PRĔTIUM > pr/ɛ/u 'price'

c. PLĒNUM >	pl/ə/ (Bal.)	pl/ɛ/ (central)	pl/e/ (western)	'full'
PĬRA >	p/ə/ra (Bal.)	p/ɛ/ra (C)	p/e/ra (W)	'pear'
BĬBIT >	b/ə/u (Bal.)	b/ɛ/u (C)	b/e/u (W)	'drinks'
CATĒNAM >	cad/ə/na (Bal.)	cad/ɛ/na (C)	cad/e/na (W)	'chain'

Table 21.1 Stages in the evolution of the front mid vowels

Vulgar Latin	ProtoCatalan Stage 1	ProtoCatalan Stage 2	Central Catalan	Western Catalan
e	e → ə	e ə	e ə	e ← ə
ɛ	ɛ	↑ ɛ/(ɛ)	↙ ɛ	ɛ

A possible explanation (although by no means the only one) is schematized in Table 21.1. First, /e/ became the central or back vowel /ə/ (stage 1). Once /e/ had disappeared from the vowel inventory, /ɛ/ closed becoming /e/, except in a few words and contexts in which it remained /ɛ/ (stage 2). At this point, the front mid vowels had given way to a three-way distinction: /ɛ–e–ə/. This is the stage that is preserved in present day Balearic Catalan. In the rest of the eastern Catalan dialects stressed /ə/ changed to /ɛ/ (Badia i Margarit 1951:142; Mascaró 2002:120f), restoring the seven-vowel system, except that original /ɛ/ and /e/ had switched positions. In western Catalan dialects stressed /ə/ changed back to /e/: here the original seven vowel system is restored as well, except that most of the instances of original /ɛ/ are now /e/ and only a few of the original cases of /ɛ/ remain as such.

The original mid open vowels /ɛ/ and /ɔ/ (from CLat. ĕ and ŏ, respectively) underwent a process of diphthongization, but only when immediately followed by a front glide (cf., however, §38.3). Examples of non-diphthongized /ɛ/ are given in (1a,b). As for non-diphthongized /ɔ/, we can cite HŎSPITEM > h/ɔ/ste 'guest', BŎUEM > b/ɔ/u 'ox', BŎNAM > b/ɔ/na 'good.FSG'. When followed by the front glide, the mid open vowels are hypothesized to have undergone diphthongization and subsequent raising of the mid vowel /ɛj/ > /jej/ > /jij/ > /i/ and /ɔj/ > /woj/ > /wuj/ > /uj/, e.g. SPĔCULUM > *espill* 'mirror', MĔDIUM > *mig* 'half', LĔCTUM > *llit* 'bed', SĔX > *sis* 'six'; and ŏCULUM > *ull* 'eye', CŎCTUM > *cuit* 'cooked', PŎDIUM > *puig* 'hill', FŎLIA > *fulla* 'leaf'. Evidence that the process went through the stages indicated is provided, according to Coromines (1971:249), by the words HŎDIE > *(a)vui* 'today' and ŏCTO > *vuit* 'eight', where the word-initial back glide in the stage /wuj/ became the voiced fricative [v], and by the word *ˈvɔkitu > *buit* 'empty', where the initial /v/ fused with the following /w/ into /b/. This diphthongization of the mid open vowels before the front glide is a feature that Catalan shares with Occitan and French, in contrast with Spanish, where these vowels fail to diphthongize only before the front glide.

The Latin diphthong AU coalesced into the monophthong /ɔ/: PAUCUM > *poc* 'little, few', CAUSAM > *cosa* 'thing', AURUM > *or* 'gold'. The early western Romance diphthong /aj/, which originated from the transformation of /k/ in coda position into /j/, closed to /ej/ and subsequently coalesced to /e/: FACTUM > *['fajto] > *[fejt] > *fet* 'done', FRAXINUM > *['frajsinu] > ['frejʃen] > *freixe* 'ash (tree)'.

Another salient feature of Catalan diachronic phonology, which also groups Catalan with Occitan and French, is the generalized loss of vowels other than /a/ after the stressed syllable. In words with penultimate stress, the loss of word-final /e/ and /o/ is the norm (CABALLUM > *cavall* 'horse', DENTEM > *dent* 'tooth', PANES > *pans* 'loaves', DICO > *dic* 'I.say'), although a vowel (-*e*) is retained following consonant groups that do not qualify as codas (LATRO > *lladre* 'thief', NOSTRUM > *nostre* 'our', UENTREM > *ventre* 'belly'). In words with antepenultimate stress, one of the two post-tonic vowels is generally lost: the penultimate vowel is lost in preference (as in DUODECIM > *ˈdodeke > *dotze* 'twelve', HEDERAM > *heura* 'ivy', LEPOREM > *llebre* 'hare', HOSPITEM > *hoste* 'guest', EPISCOPUM > *bisbe* 'bishop'). In certain contexts, however, the non-final vowel is retained and the last vowel is dropped: when the unstressed penultimate vowel precedes an /n/ (HOMINEM > *['ɔmen] > *home* 'man', IUUENEM > *['ʒoven] > *jove* 'young', FRAXINUM > *['frejʃen] > *freixe* 'ash (tree)'), as well as in other contexts (*ˈessere > *ésser* 'be.INF', PLANGERE > *plànyer* 'pity.INF', FLEBĬLEM > *frèvol* 'weak').

21.2.2 Consonants

As in western Romance in general, a voiceless obstruent in onset position after a vowel becomes voiced (cf. §25.2.5): OPACAM > *obaga* 'shady side of the mountain', CAPRAM > *cabra* 'goat', UITAM > *vida* 'life', SECURUM > *segur* 'safe', DEFENSAM > *devesa* 'meadow'. Geminate obstruents were shortened: CUPPAM > *copa* 'cup', GUTTAM > *gota* 'drop', SICCAM > *seca* 'dry.FSG', ABBATEM > *abat* 'abbot'. Intervocalic voiced obstruents undergo lenition, leading to a fricative (CABALLUM > *cavall* 'horse', PLORABAT > *plorava* 'cried.3SG', ALAUDAM > *alosa* 'lark'), to a glide in coda position (BIBIT > *beu* 'drinks', PEDEM > *peu* 'foot', IUGUM > *jou* 'yoke'), or to the loss of the consonant (SABUCUM > *saüc* 'elder (tree)', *aˈbutu > OCat.*haüt* 'have.PTCP', LAUDARE > *lloar* 'praise.INF', IUDAEUM > *jueu* 'Jew', RUGAM > *rua* 'wrinkle').

When an /l/ was adjacent to a front glide, the two segments coalesced into the palatal /ʎ/, as in other Romance languages, represented by the digraph *ll* in modern Catalan: FILIUM > *fill* 'son', PALEAM > *palla* 'hay', OCULUM > *ull* 'eye', MELIUS > OCat. *mills* 'better'. At a later stage, two different sound changes increased the number of words with /ʎ/. On the one hand, initial *l* was palatalized as /ʎ/: LANAM > *llana* 'wool', LUNAM > *lluna* 'moon', LINGUAM > *llengua* 'tongue', LOCUM > *lloc* 'place', *ˈlakte > *llèt* 'milk'; on the other hand, geminate *ll* was palatalized as /ʎ/: CABALLUM > *cavall* 'horse', GALLINAM > *gallina* 'hen', COLLEM > *coll* 'mountain pass'. The palatal /ʎ/ from these two origins converged in most Catalan dialects with earlier /ʎ/ that had originated from the /lj/ or /jl/ sequences. In some dialects—Balearic and the northern half of central Catalan—this merger or neutralization did not happen because earlier /ʎ/ lost its lateral component and became a glide, /j/, before the new palatal /ʎ/ emerged.

The palatal nasal /ɲ/, with the spelling *ny*, also has various origins. The /nj/ group gave /ɲ/, as in UINEAM > *vinya* 'vineyard', EXTRANEUM > *estrany* 'strange', AGNELLUM > *anyell* 'lamb', LIGNA > *llenya* 'fire wood'. Geminate *nn* is sometimes palatalized as /ɲ/ (ANNUM > *any* 'year', CANNAM > *canya* 'reed', PINNAM > *penya* 'rock') and sometimes simplified as /n/ (CANNABEM > *cànem* 'hemp', CAPANNAM > *cabana/cabanya* 'hut'). The group *mn* also yielded these two alternative results: /ɲ/ , as in DAMNARE > *danyar* 'hurt.INF', and /n/ as in DOMINAM > *ˈdomna > *ˈdonna > *dona* 'woman'.

Word-final *n* is lost: MANUM > *man > *mà* 'hand', UINUM > *vin > *vi* 'wine', HOMINEM > *ˈomen > *home* 'man'. This loss fails to occur whenever an affix follows the *n*, including the plural *s*, and in words that tend to be used proclitically (MANUS > *mans* 'hands', UNUM > *un* 'one', NEC + UNUM > OCat. *negún* 'no/any'). After the loss of word-final *n*, a new sound change created more intervocalic and final /n/: the sequence *nd* between vowels is simplified as /n/ (UENDIT > *ven* 'sells', ROTUNDAM > *rodona* 'round', RESPONDEBAT > *responia* 'answered.3SG').

The back glide in coda position has several different origins. Latin intervocalic B (> *v*) and U which came to be in coda position, after the loss of post-tonic vowels, became the back glide /w/: BIBERE> *beure* 'drink.INF', CLAUEM > *clau* 'key', NOUUM > *nou* 'new'. Intervocalic /d/ and /g/ also became /w/ when in a coda, as a result of vowel deletion: CREDIT > *creu* 'believes', CADERE > *caure* 'fall.INF', IUGUM > *jou* 'yoke'. Latin /k/ became a palatal affricate before a front vowel and was subsequently fronted to /ts/; in intervocalic position, like all obstruents, it became voiced and subsequently /ð/; when it occurred in coda position through the loss of posttonic vowels it became the back glide: DICIT > *diu* 'says', NUCEM > *nou* 'walnut', PACEM > *pau* 'peace'. The Latin group /tj/ converged in its evolution with Latin /k/ before a front vowel at the /ts/ stage: PRETIUM > *preu* 'price', PUTEUM > *pou* 'well', PALATIUM > *palau* 'palace'.

21.2.3 Morphology

Plural forms of nouns and adjectives have the suffix *-s* added to the default singular stem. When the stem ends in a sibilant, the addition of the plural suffix requires the presence of an epenthetic vowel between the two sibilants. In old Catalan this epenthetic vowel was the default vowel *e*; so, the plural forms of *ós* (< URSUM) 'bear', *pes* (< PENSUM) 'weight', or *cors* (< CORPUS) 'body' were *ósses*, *peses*, and *corses*, respectively. However, by the fifteenth century a new plural form was established in which the epenthetic vowel was replaced by *o*, giving the forms *óssos*, *pesos*, and *cossos*, respectively. This special epenthetic vowel was restricted to masculine forms, thereby distinguishing the masculine from the feminine form of lexemes that had previously not distinguished gender morphologically: instead of *franceses* 'French.PL' for both genders, with the introduction of *francesos* as the masculine form, *franceses* became exclusively feminine. This specialized epenthetic vowel originated in forms in which final *-e* changed to *-o* through assimilation to the back rounded stressed vowel in forms like SUBERUM > *ˈsure > *suro* 'cork', ROTULUM > *rotlle* > *rotllo* 'role', TAURUM > *ˈtaure > *toro* 'bull', *ˈmoniku > *monge* > *monjo* 'monk'.

The definite article in Catalan has two sets of forms: one set derived from Latin ILLE 'that'—*lo/el*, *la*, etc.—and another derived from Latin IPSE 'self'—*so/es*, *sa*, etc. The two forms coexisted in all the Catalan speaking territory up to the twelfth century. At that point the written language adopted the ILLE forms; the IPSE forms disappeared from the written documents but remained the everyday forms of part of the territory and nowadays are the normal form in the spoken language of the Balearic Islands and have a residual use in the coastal area from Blanes to Cadaqués (for further discussion, see §46.3.1.1 and Ledgeway 2012a:§4.2.2.1). The IPSE forms are attested in many place names, particularly, in the eastern dialect area: *Collserola* (< *coll* 'mountain pass' + *s'* 'the' + *erola* 'clearing'), *Sant Quirze Safaja* (< *Sant Quirze'* Saint Quiricus' + *sa* 'the' + *faja* 'beech tree'). As for the ILLE forms, old Catalan had an allomorphy in the masculine singular between *lo*, the etymological form, and the reduced form /l/, used whenever the article immediately followed or was preceded by a vowel. This system is still in use in the northwestern dialect. From the reduced form, a reinforced form *el*, pronounced [el] or [əl] depending on the dialect, emerged for use before consonant-initial words, which replaced the etymological *lo*. Before vowel-initial words, the reduced form, written *l'*, continued to be used. This system, with an alternation between *el* and *l'* for the masculine singular, is now standard.

Some of the most salient diachronic facts of verb morphology concern the form of past participles, infinitives, first person singular verb forms, and present subjunctives. Catalan, like Occitan, French, and Italian, but in contrast to Spanish and Portuguese, has several irregular participial forms in /z/ and /t/ and forms that regularly continue Latin -UTUM in the second conjugation, in addition to the regular first and third conjugation past participles descended from -ATUM and -ITUM respectively. Examples of the irregular forms of the second conjugation are FUSUM > *fos* 'melted', CLAUSUM > *clos* 'closed', INCENSUM > *encès* 'lit' – with -s, phonemic /z/, as shown by the feminine forms *fosa*, *closa*, etc. – and DICTUM > *dit* 'said', TRACTUM > *tret* 'removed', SCRIPTUM > *escrit* 'written' – with -t, phonemic /t/, as shown by the feminine forms *dita*, *treta*, etc. Among the regular forms of the second conjugation we can cite *begut* 'drunk', *mogut* 'moved', *crescut* 'grown', *temut* 'feared', of analogical origin.

Catalan has some infinitives with stress on the stem, again like Occitan, French, and Italian, and unlike Spanish and Portuguese. The Latin first, second, and fourth conjugation forms of the infinitive were stressed on the ending and have remained so in Catalan: PLORARE > *plorar* 'to weep', HABERE > *haver* 'to have', AUDIRE > *oir* 'to hear'.[1] But the Latin third conjugation forms were stressed on the stem, as are their Catalan descendants, ending either in *-re* (e.g. BIBERE > *beure* 'to drink', PERDERE > *perdre* 'to lose', MOLERE > *moldre* 'to grind'), or in *-er* (e.g. PLANGERE > *plànyer* 'to pity', CRESCERE > *créixer* 'to grow', or CURRERE > *córrer* 'to run').

The loss of final unstressed vowels other than /a/ yielded first person singular forms in the present indicative either without a vowel or with the epenthetic vowel *-e* when needed to syllabify the preceding consonant cluster: AMO > *am* 'I.love', CANTO > *cant* 'I.sing', PORTO > *port* 'I.carry', COMPARO > *compre* 'I.buy', SEMINO > *sembre* 'I.sow', INTRO > *entre* 'I.enter'. The insular dialects (Balearic and Alguerès) generalized the absence of an ending to all the first person singular forms of the present indicative, such that today, in addition to etymological forms like *cant*, these dialects also have *compr*, *sembr*, and *entr*. According to Coromines (1971:206), by the fifteenth century the peninsular dialects had generalized the *-e* ending to all first person singular forms of the present indicative of first conjugation verbs; this means that, alongside forms like *compre* and *sembre*, these dialects also had *ame*, *cante*, and *porte* for the first person singular. This is the situation currently found in Valencian.

In the dialects of Catalonia, this situation created widespread homophony between first and third person singular forms: in the eastern dialect, because unstressed /e/ and /a/ were neutralized to /ə/, and in most of the northwestern dialects because etymological *-a* in the third person singular forms of the present indicative of the first conjugation was replaced by *-e*. The sound change mentioned at the beginning of §21.2.3 by which final *-e* changed to *-o* through assimilation to the back rounded stressed vowel created some forms that avoided this ambiguity (Coromines 1971:204-7): *compre* > *compro* 'I.buy', *done* > *dono* 'I.give'. In central and northwestern dialects, this ending was generalized to all first person singular forms of the present indicative, except for those forms (of the second conjugation) that were different from the corresponding third person singular: this produced new forms such as *canto*, *porto*, *sembro*, *entro*, whereas forms such as *dic* 'I.say', *vinc* 'I.come', *crec* 'I. believe' and *veig* 'I.see' (distinct from *diu* 'says', *ve* 'comes', *creu* 'believes', or *veu* 'sees') remained unchanged. In northern dialects, *-i* became the mark of first person singular forms, e.g. *canti*, *porti*, *sembri*, *entri*.

The present subjunctive forms with stress on the verb stem, due to regular phonetic changes in old Catalan, had no vowel following the stem in the first conjugation (DURET > *dur* 'last.3SG.PRS.SBJV', CANTEM > *cant* 'sing.1SG.PRS.SBJV') or had an epenthetic vowel (INTRET > *entre* 'enter.3SG.PRS.SBJV'), while /a/ is found in the other conjugations (DICAM > *diga* 'say.1SG.PRS.SBJV', DEBEAT > *deja* 'owe.3SG.PRS.SBJV'). The epenthetic vowel *-e* was generalized to all first conjugation verbs, giving forms like *dure* and *cante*, which is the situation preserved in Valencian. In most other dialects, the *-i* ending is found for the present subjunctive: *duri*, *canti*, *entri*. The origin of this ending is presumed to be from forms like *sàpia* 'know.1/3SG.PRS.SBJV', *dòrmia* 'sleep.1/3SG.PRS.SBJV', *témia* 'fear.1/3SG.PRS.SBJV', *òbria* 'open.1/3SG.PRS.SBJV', *còbria* 'cover.1/3SG.PRS.SBJV', where the *-ia* ending was the result of regular phonetic laws; from these forms it was generalized giving forms like *tròbia* (or *tròpia*) 'find.1/3SG.PRS.SBJV' and *càntia* 'sing.1/3SG.PRS.SBJV'; the *-ia* ending was then reduced to *-i* (Coromines 1971:272f.).

21.3 Phonology

21.3.1 Stressed vowels

The inventory of stressed vowels in most continental varieties of Catalan consists of seven vowels, illustrated in (2) for central Catalan:

(2) [diw] *diu* 'says' [us] *ús* 'use'
[dew] *déu* 'god' [os] *ós* 'bear'
[dɛw] *deu* 'owes' [daw] *dau* 'die' [ɔs] *os* 'bone'

[1] There is an alternative form *heure* from HABERE in the non-auxiliary use. The conflation of the 2nd and 3rd conjugations has resulted in some stress shifts and the emergence of doublets like *caber/cabre* 'fit', *valer/valdre* 'be worth'.

There are two dialects in which the inventory of stressed vowels is different from this: north Catalan and most varieties of Balearic Catalan. North Catalan lacks the contrasts in the mid vowels: historically, [e] and [ɛ] have merged as [e], whereas [o] and [u] have merged as [u] and [ɔ] has closed to [o]. Consequently, north Catalan has the simpler five-vowel inventory: /i, e, a, o, u/. Most of Balearic has [ə] in addition to the seven vowels illustrated in (2) (cf. §21.1). In some reduced areas of the Balearic Islands, [ə] has evolved into [ɛ], as in central Catalan, but independently of it, as in Mahón and a region in the centre of Majorca.

Mid front vowels show a great degree of variability within the same dialect and even in the same area and are not completely free in their distribution. Within the same dialect zone, a given word may be pronounced with an [ɛ] in some areas and with an [e] in other areas. But even in the same area, certain words are used with both mid front vowels. This is the case with words like *consell* 'council', *primavera* 'spring', or *rem* 'oar' in Barcelona, whose stressed vowel varies between [e] and [ɛ] (see Badia i Margarit 1970). And, finally, there is a phonological process, discovered by Mascaró (1984a), by which mid vowels are half-open ([ɛ] or [ɔ]) when followed by an unstressed derivational suffix, which explains contrasts such as (3):

(3)

[plə'to]	/ [plə'tɔnik]	*Plató* 'Plato'	/ *platònic* 'platonic'
[ku'lon]	/ [ku'lɔniə]	*colon* 'settler'	/ *colònia* 'settlement'
[sis'temə]	/ [sis'tɛmik]	*sistema* 'system'	/ *sistèmic* 'systemic'
[du'sen]	/ [du'sɛnsjə]	*docent* 'teacher'	/ *docència* 'teaching'

21.3.2 Unstressed vowels: vowel reduction and its exceptions

The fundamental difference between eastern and western dialects concerns the extent of vowel reduction in unstressed syllables. In western dialects the mid vowels neutralize to the half-close vowel, so that /ɛ/ becomes [e] and /ɔ/ becomes [o] in unstressed syllables merging with /e/ and /o/, respectively. In eastern dialects /ɛ/, /e/, and /a/ become [ə] in unstressed syllables, and, with the exception of Majorca, /ɔ/ and /o/ merge with /u/ as [u] in unstressed syllables. Majorca follows the western pattern of neutralization of the back vowels in unstressed syllables. Thus, in western Catalan there are five contrasting vowels in unstressed syllables, whereas in central Catalan and in Menorca and Ibiza there is a three-way contrast in unstressed syllables with just [i], [ə], and [u], and Majorca has [o] in addition to these three vowels.

As a result, eastern Catalan has many more homophones caused by vowel reduction than western Catalan. For example, western Catalan distinguishes the feminine and the masculine forms of certain adjectives and nouns such as ['maɣrə] *magra* (F) and ['maɣre] *magre* (M) 'lean', ['tɛndrə] *tendra* (F) and ['tɛndre] *tendre* (M) 'tender', ['sɔɣrə] *sogra* 'mother-in-law' and ['sɔɣre] *sogre* 'father-in-law'. In eastern Catalan these pairs are homophonous: ['maɣrə] *magra/magre*, ['tɛndrə] *tendra/tendre*, ['sɔɣrə] *sogra/sogre*, ['kaβrə] *cabra/cabre*, ['parə] *para/pare*, and ['ʎaðrə] *lladra/lladre*.

There are some very restricted environments in which unstressed [e] and [o] are retained as such in central Catalan. On the one hand, there is a small group of words, which are obvious loans from Greco-Latin, Spanish, or English, in which these vowels are retained in posttonic position, as in (4) from Wheeler (2005:70f):

(4)

àlg[e]*bra* 'algebra'	*class*[e] 'class'
bàsqu[e]*t* 'basketball'	*rèc*[o]*rd* 'record'
Bóst[o]*n*	*tún*[e]*l* 'tunnel'

The other situation in which we find exceptions to the vowel reduction rule is when underlying /e/ is adjacent to /a/ and, less categorically, when it is adjacent to /o/ or /ɔ/, as in (5) and (6):

(5)

[iðe'al] *ideal* 'ideal'	[sərɛ'al] *cereal* 'cereal'
[irəel'itə] *israelita* 'Israelite'	[reəβil'itə] *rehabilita* 'rehabilitates'
[əerɛu'pɔrt] *aeroport* 'airport'	['areə] *àrea* 'area'

(6)

[te'ɔlək] *teòleg* 'theologian'	[teu'krasjə] *teocràcia* 'theocracy'
[urfe'o] *orfeó* 'choral society'	[kureuɣrə'fiə] *coreografia* 'choreography'

21.3.3 Consonant inventory, dialectal variation, and allophones

The inventory of consonant phonemes is shown in Table 21.2.

In most dialects, /v/ has merged with /b/ as [b]~[β], and so there is only one phoneme, /b/, corresponding to /v/ of older stages of the language and to orthographic *v*. /v/ is retained as a separate phoneme, phonetically distinct from /b/, in Alguerès, Balearic, and southern and northern Valencian. Central Valencian—the area including and

Table 21.2 Inventory of consonant phonemes

	LABIAL		DENTO-ALVEOLAR		ALVEOLO-PALATAL		VELAR	
	−VOICE	+VOICE	−VOICE	+VOICE	−VOICE	+VOICE	−VOICE	+VOICE
Plosive, affricate	p	b	t	d	tʃ	dʒ	k	g
Fricative	f	(v)	s	z	ʃ	(ʒ)		
Nasal		m		n		ɲ		
Tap				ɾ				
Lateral				l		ʎ		

surrounding the city of Valencia—has lost the /v/–/b/ distinction. The area around the city of Tarragona, including towns such as Reus and Valls, retained the distinction until the first half of the twentieth century, and today it is only found in very old speakers in this area, if at all.

Most dialects distinguish between the fricative /ʒ/ and the affricate /dʒ/, but in most of Valencia (the region of La Marina Alta being an exception, cf. Colomina i Castanyer 1985) and the southern part of Catalonia the distinction is absent and [dʒ] is used, corresponding to both [ʒ] and [ddʒ]~[dʒ] in other dialects.

The central area of the Valencian Country, which includes the city of Valencia, and much of the Catalan-speaking strip of Aragon, lack the voiced strident phonemes /z/ and /dʒ/ which have merged with their voiceless counterparts and are pronounced [s] and [tʃ].

There is a debate about whether the affricates /dʒ/ and /tʃ/should be considered separate phonemes or the sequence of a stop and the fricative /z/ or /ʒ/ (Wheeler 2005:11-13). The interpretation of these affricates as a phonemic unit seems to be preferred for those dialects that lack [ʒ] and only have [dʒ]. The fact that both alveolo-palatal affricates occur word-initially, a position where obstruent clusters are not allowed, is an argument for assuming the unit interpretation of these affricates. However, in the dialects that have [ʒ] contrasting with [dʒ] and normally have the alveolo-palatal fricatives in word-initial position, the interpretation of these affricates as a sequence of two phonemes is more strongly motivated. The fact that the contrast between [ʒ] and [dʒ] and between [ʃ] and [tʃ] is only found in intervocalic position is an argument for the biphonemic proposal, given that a cluster of a stop and an alveolo-palatal fricative can only occur in intervocalic position.

Catalan, like Spanish, has a tap [ɾ] and a trill [r], which contrast intervocalically giving rise to such minimal pairs as *mirra* ['mirə] 'myrrh' and *mira* ['mirə] 'look' or *serra* ['sɛrə] 'saw' and *cera* ['sɛrə] 'wax'. On the basis of such contrasts one might be tempted to assume that the two rhotic sounds are distinct phonemes. However, there are several arguments that favour the analysis in which there is only one rhotic phoneme (Wheeler 2005:24f.). The main argument is the fact that the tap and the trill only contrast intervocalically: this follows from the assumption that the trill is a geminate (/rr/) in this position, which is the only position in which geminate consonants are allowed in Catalan. In the other positions in which the trill occurs, it is an allophone of the phoneme /r/.

A well-known source of allophonic variation, which Catalan shares with Spanish, Basque, and some Occitan varieties, is the phenomenon of voiced-stop lenition (Wheeler 2005:317f.). The pairs [b]~[β], [d]~[δ], and [g]~[γ] are in complementary distribution and considered to be allophonic realizations: the fricative, or approximant, allophone ([β], [δ], [γ]) is found in onset position following a continuous segment (vowel, liquid, or fricative), as in (7a); in all other contexts, the stop realization ([b], [d], [g]) is found, as in (7b). There are two exceptions to this rule: [d], rather than [δ], occurs after a lateral, and [b], rather than [β], after the labio-dental fricative, as in (7c).

(7) a. *agrada* [ə'γraδə] 'likes' *obliga* [u'βliγa] 'forces'
ho diu [u 'δiw] 'it= says' *es diu* [əz 'δiw] 'REFL= says'
el grau [əl 'γraw] 'the degree' *carbó* [kər'βo] 'coal'

b. *grau* [graw] 'degree' *diu* [diw] 'says'
un grau [uŋ 'graw] 'a degree' *et diu* [əd 'diw] 'you= says'
futbol [fud'bɔl] 'football' *regle* ['regglə] 'ruler'

c. *faldilla* [fəl'diʎə] 'skirt' *ell diu* [eʎ 'diw] 'he says'
filòsof banal [fi'lɔzuv bə'nal] 'banal philosopher'

The Catalan lenition rule differs from the Spanish and Basque rules in that Catalan lenition only occurs in onset position, whereas Spanish and Basque also have lenition in coda position (Mascaró 1984b; Hualde 1991:99-107).

21.3.4 Consonant deletion and assimilation

There are several alternations consisting of the presence or absence of the stem-final consonant, depending on whether

or not the consonant is in word-final position or in the final coda of the word: final-*n* deletion, final-*r* deletion, and cluster simplification.

Many stems end in a stressed vowel when the stem is word-final, but have an [n] following this vowel when the stem is followed by a suffix, as shown in (8). The orthography reflects this alternation, which is common to all Catalan dialects.

(8)

Bare stem	+Plural suffix	+Feminine suffix	Derived form
sa 'healthy'	*sans*	*sana*	*sanitat* 'health'
ple 'full'	*plens*	*plena*	*emplenar* 'fill out'
carbó 'coal'	*carbons*	–	*carbonitzar* 'carbonize'
matí 'morning'	*matins*	–	*matiner* 'early riser'

This alternation reflects the diachronic process by which word-final -*n* was lost. The classical generative analysis has been to assume that these *n*/Ø alternation words have an underlying representation with final /n/ and that there is a rule of word-final *n*-deletion, which derives the surface forms without [n]. However, this rule has many limitations: (a) it does not affect verbs, except for two verb roots (*ve* 'comes' and *té* 'has'); (b) it only affects words with stress on the last syllable, since there are many words like *origen* 'origin' or *òrgan* 'organ' that retain the word-final *n*; and (c) there are many lexical exceptions to this rule, even among non-verb forms with word-final stress such as *algun* 'some', *clon* 'clone', and *gran* 'big'.

An alternative to the n-deletion rule suggested by Wheeler (2005:330-32) consists in assuming that the alternating stems like those in (8) have two allomorphs, one with and one without final -*n*, and that a set of constraints favour the vowel-final allomorph just in case the allomorph is word-final.

A similar phenomenon to that just discussed is the r/Ø alternation, although the latter is different in that: (a) the historical loss of final /r/ is not reflected in the orthography and is not common to all dialects, since Valencian, excepting the northernmost and southernmost tips of the Valencian Country, maintains final [ɾ]; and (b) the r-less forms occur not only in word-final position but also when followed by the plural morph, as illustrated in (9) using central Catalan pronunciation.

(9)

Bare stem	+Plural suffix	+Feminine suffix	Derived form
dur [du] 'hard'	*durs* [dus]	*dura* ['durə]	*duresa* [du'ɾezə] 'hardness'
obrer [u'βɾe] 'worker'	*obrers* [u'βɾes]	*obrera* [u'βɾeɾə]	*obrerisme* [uβɾə'rizmə] 'workers' movement'
sencer [sən'sɛ] 'whole'	*sencers* [sən'sɛs]	*sencera* [sən'sɛɾə]	*senceret* [sənsə'ɾɛ 'whole (diminutive)'
por [po] 'fear'	*pors* [pos]	–	*poruc* [pu'ɾuk] 'fearful'

Many words retain final -r including most words with penultimate stress, except for infinitives. While some generalizations can be made about where retention or deletion of final -r is more likely (Wheeler 2005:333-8), the classical analysis involving a rule of final r-deletion requires an important number of exceptions (Wheeler 2005:338).

Certain consonant clusters containing a stop in second position are reduced word-finally through the loss of the stop (Lleó 1970:25-9; Wheeler 1979:280-87; 2005:220-49; Mascaró 1983:100-11; 1985; 1989; Iverson 1993a,b; Morales 1995; Bonet i Alsina and Lloret 1998:108-14). This stop-deletion process is specific to Catalonia and Ibiza; it is categorical when the stop follows a nasal and in many words in the groups /lt/ or /ld/ and also occurs in the groups /st/, /rt/, and /rd/, but only in less formal styles. (10) illustrates this process using central Catalan pronunciation.

(10)

	Bare stem	Derived forms	
a.	*junt* [ʒun] 'together	*junta* (F) ['ʒuntə]	*ajuntar* [əʒun'ta] 'join.INF'
	llamp [ʎam] 'lightning'	*llampec* [ʎəm'pɛk] 'flash of lightning'	
	franc [fraŋ] 'frank'	*franca* (F) ['fraŋkə]	*franquesa* [frəŋ'kɛzə] 'frankness'
	malalt [mə'lal] sick	*malalta* (F) [mə'laltə]	*malaltia* [mələ'tiə] 'sickness'
b.	*just* [ʒust] / [ʒus] 'just'	*justa* (F) ['ʒustə]	*ajustar* [əʒus'ta] 'adjust.INF'
	fort [fort] / [fɔr] 'strong'	*forta* (F) ['fɔrtə]	*enfortir* [əmfur'ti] 'strengthen.INF'

The requirement that the two consonants involved be homorganic (i.e. have the same point of articulation) is shared by all the forms on the left column of (10), but would not be satisfied by forms such as *corb* 'crow', *amarg* 'bitter', *calb* 'bald', *talc* 'talcum', *bosc* 'forest', *Casp* (place name), which retain the final stop. For obligatory stop deletion, as in (10a), the requirement that the first consonant of the cluster be non-continuant is clearly satisfied when it is a nasal and is satisfied by a lateral only when

the following consonant is dental-alveolar (Wheeler 2005:221f.). In (10b), the two consonants are homorganic, but the first one is continuant, which accounts for the variable application of the process.

21.3.5 Voicing and devoicing

Catalan neutralizes voice contrasts among obstruents in word-final and coda position, through the interaction of three different processes. While obstruents show contrasts in voice in onset position, all obstruents are voiceless in absolute final position before a pause.

(11)

	Obstruent word-final	Obstruent before a vowel
a.	*llop* [ʎop] 'wolf' *tap* [tap] 'stopper'	*lloba* (F) ['ʎoβə] *tapa* ['tapə] 'cover.3SG.PRS.IND'
b.	*mut* [mut] 'mute' *brut* [brut] 'dirty'	*muda* (F) ['muðə] *bruta* (F) ['brutə]
c.	*llec* [ʎek] 'ignorant' *rebec* [rə'βɛk] 'unruly'	*llega* (F) ['ʎeɣə] *rebeca* (F) [rə'βɛkə]
d.	*cus* [kus] 'sew.3SG.PRS.IND' *tus* [tus] 'cough.3SG.PRS.IND'	*cusi* ['kuzi] 'sew.3SG.PRS.SBJV' *tussi* ['tusi] 'cough.3SG.PRS.SBJV'
e.	*raig* [ratʃ] 'ray, jet' *feix* [feʃ] 'bundle'	*rajos* (PL) ['raʒus] *feixos* (PL) ['feʃus]
f.	*lleig* [ʎetʃ] 'ugly' *escabetx* [əskə'βɛtʃ] 'marination'	*lletja* (F) ['ʎedʒə] *escabetxar* [əskəβə'tʃa] 'marinate.INF'

These facts are analysed by positing a rule of word-final obstruent devoicing and an underlying representation of the stem of words like *llop/lloba* or *mut/muda* with a final voiced obstruent, which remains voiced if followed by a vowel-initial suffix, but is devoiced if final in the word.

In addition to devoicing, when the underlying voiced fricative /ʒ/ occurs word-finally, it becomes the affricate [tʃ], thereby undergoing a neutralization with the affricates /dʒ/ and /tʃ/, rather than with the voiceless fricative /ʃ/, as we see in (11e).

Word-final fricatives and affricates, whether underlyingly voiced or voiceless, become voiced when adjacent to a vowel in the following word, as shown in (12):

(12)

	Word-final before pause	Word-final before a vowel
a.	*cos* [kɔs] 'body' *cas* [kas] 'case'	*cos elàstic* [kɔz ə'lastik] 'elastic body' *cas únic* [kaz 'unik] 'unique case'
b.	*peix* [peʃ] 'fish' *calaix* [kə'laʃ] 'drawer'	*peix alat* [peʒ ə'lat] 'winged fish' *calaix obert* [kə'laʒ u'βert] 'open drawer'
c.	*roig* [rɔtʃ] 'red' *matx* [matʃ] 'match'	*roig i negre* [rɔdʒ i 'nɛɣrə] 'red and black' *matx igualat* [madʒ iɣwə'lat] 'balanced match'
d.	*bolígraf* [bu'liɣrəf] 'pen' *xef* [ʃɛf] 'chef'	*bolígraf espatllat* [bu'liɣrəv əspəʎ'ʎat] 'broken pen' *xef estrella* [ʃɛv əs'treʎə] 'star chef'

A version of this process is also found in Portuguese, Occitan, and French (at least) among the Romance languages (Wheeler 2005:162).

Finally, obstruents in coda position assimilate in voicing to the following consonant, if there is one, whether in the same word or across words, as shown in (13):

(13)

	Voiceless obstruent	Voiced obstruent
a.	*cap* [kap] 'head, leader' *barret* [bə'rɛt] 'hat'	*cap de colla* [kab də 'kɔʎə] 'group leader' *barret girat* [bə'rɛd ʒi'rat] 'turned-around hat'
b.	*cops* [kɔps] 'blows' *sacs* [saks] 'sacks'	*cops amagats* [kɔbz əmə'ɣats] 'hidden blows' *sacs apilats* [sagz əpi'lats] 'piled up sacks'
c.	*subalpí* [supəl'pi] 'pre-Alpine' *sud-americà* [sutəməri'ka] 'South American'	*subdelegat* [subdələ'ɣat] 'subdelegate' *sud-vietnamita* [sudbjədnə'mitə] 'south Vietnamese'
d.	*obtenir* [uptə'ni] 'obtain.INF' *taxi* ['taksi] 'taxi'	*objecte* [ub'ʒektə] 'object' *examen* [əg'zamən] 'examination'

21.3.6 Syllable structure and stress

Catalan does not allow onsets consisting of /s/ and another consonant. In non-initial contexts, the sequence /sC/ is resolved by syllabifying the two consonants in different

syllables: *inscriure* [ins.ˈkɾiw.ɾə] 'inscribe.INF', *destrucció* [dəs.truk.ˈsjo] 'destruction', *caspa* [ˈkas.pə] 'dandruff', *constel·lació* [kuns.tə.lə.ˈsjo] 'constellation'. In word-initial contexts, the potential conflict is resolved by inserting the least marked vowel ([ə] in eastern Catalan) before the consonant group: *escriure* /skɾiwɾə/ [əs.ˈkɾiw.ɾə] 'write.INF', *estel* /stɛl/ [əs.ˈtɛl].

As a general rule, a consonant that can be syllabified either in the onset or in the coda of a syllable is preferentially syllabified as an onset. However, an exception to this rule occurs with many stem-final /bl/ and /gl/ sequences, as the stop is syllabified as the coda creating a geminate, in eastern Catalan and most of the northwestern varieties (Wheeler 2005:265-9), as in (14).

(14) *amable* [əˈmab.blə] 'kind' *possible* [puˈsib.blə] 'possible'
regla [ˈreg.glə] 'rule' *feble* [ˈfeb.blə] 'weak'
arreglar [ərəgˈgla] 'arrange' *afeblir* [əfəbˈbli] 'weaken'

The position of word stress in Catalan is subject to some general constraints and lexical specificities. The general constraints are of two kinds, as proposed in Wheeler (2005:276f.), depending on whether or not the word is a verb. In verb forms, stress falls on the last stressable affix, if there is one; otherwise, on the last syllable of the stem. In non-verb forms, stress can occur on any one of the last three syllables of a word, with the following condition: antepenultimate stress is only possible if the penultimate syllable is light (i.e. does not have a coda). Thus, since all vowels in a stem are stressable, verb forms consisting of a stem and an unstressable vowel in a suffix always have penultimate stress, even though the corresponding noun or adjective may have antepenultimate stress. This gives rise to contrasts such as *fabrica* [fəˈβɾikə] 'manufacture.3SG.PRS.IND' vs *fàbrica* 'factory' [ˈfaβɾikə], *castiga* 'punish.3SG.PRS.IND' [kəsˈtiɣə] vs *càstig* 'punishment' [ˈkastik], *copia* 'copy.3SG.PRS.IND' [kuˈpiə] vs *còpia* 'copy' (N) [ˈkɔpjə], etc.

21.3.7 Orthography

In the vocalic domain the letter *e* represents both phonemes /e/ and /ɛ/ (and also /ə/ in most of Balearic) and the letter *o* represents both /o/ and /ɔ/. Nevertheless, an orthographic distinction is made in a partial way between the two pairs of mid vowels through the use of accents. The position of word stress is represented by means of the presence and absence of accents. When a vowel needs to be written with an accent, the acute accent is used for the closed vowels /i/, /e/, /o/, and /u/ and the grave accent is used for the open vowels /ɛ/, /a/, and /ɔ/. The rule for accenting words is the following: stress on the last syllable requires an accent only if the word ends in a vowel, a vowel plus *s*, or the groups *en* and *in*, as in (15a); otherwise no accent is used to signal stress on the last syllable, as in (15b); stress on the penultimate syllable is only signalled by means of an accent if the word does not end in a vowel, a vowel plus *s*, or the groups *en* and *in*, as in (15c); elsewhere penultimate stress is unmarked, as in (15d); antepenultimate stress is always marked with an accent, as in (15e). For the purpose of accenting words, a high vocoid after a consonant counts as a syllabic nucleus (except in the *qu-/qü-* and *gu-/qü-* sequences), even if it is pronounced as a glide.

(15) a. *perdó* 'pardon' *arròs* 'rice'
homenàs 'big guy' *entén* 'understands'
b. *segon* 'second' *entens* 'you.SG.understand'
cabell 'hair' *finalitat* 'purpose'
c. *crèdit* 'credit' *orígens* 'origins'
cérvol 'deer' *apèndix* 'appendix'
d. *meravella* *arrossos* 'rices'
llapis 'pencil' *origen* 'origin'
e. *càndida* 'candid.F' *llúdries* 'otters'
dèficit 'deficit' *espécimen* 'specimen'

In addition to the cases in which the accent is used following the rule just presented, some words have what is called a 'diacritic accent' in order to distinguish them from another word. The list of words with a diacritic accent does not follow any systematic rule. The following are some pairs of words involving a diacritic accent: *és* 'is' and *es* (third person reflexive clitic pronoun), *sé* 'I.know' and *se* 'self-', *ús* 'use' (N) and *us* 'you.PL', *són* 'they.are' and *son* 'sleep' (N), *dóna* 'gives' and *dona* 'woman', *vénen* 'they.come' and *venen* 'they.sell'.

Another mark that is used on vowels is the diaeresis. This has two distinct uses in Catalan. On the one hand, if it is placed on the *u* in the groups *qu-* and *gu-* before *e* or *i*, it indicates that this letter is to be interpreted as a glide (as, without the diaeresis, it is mute), as in *qüestió* [kwəsˈtjo] 'question', *ambigües* [əmˈbiɣwəs] 'ambiguous.FPL', *pingüí* [piŋˈgwi] 'penguin'. On the other hand, it is placed on *i* and *u* to signal that these letters correspond to syllabic nuclei when they immediately follow a vowel and cannot be accented by the general accent rule: *veïna* 'neighbour.F' (but *veí* 'neighbour.M'), *veïnat* 'neighbourhood', *maleït* 'damned', *destruïm* 'we.destroy', *peülla* 'hoof'.

Western dialects have a phonological rule by which posttonic stem-final /a/ becomes [e] when not word-final; the effects of this rule are represented in the orthography. In

eastern varieties this rule is merely orthographic, as both unstressed /a/ and unstressed /e/ are phonetically [ə]. Some examples of this *a/e* alternation rule follow: *carta* 'letter' vs *cartes* 'letters', *fulla* 'leaf' vs *fulles* 'leaves, *canta* 'sings' vs *cantes* 'you.SG.sing' and *canten* 'they.sing', *sentia* 'felt.3SG' vs *senties* 'you.SG.felt' and *sentien* 'they.felt'.

Among consonant letters, *c* and *g* correspond to /s/ and /ʒ/ respectively, when followed by *e* or i, but correspond to /k/ and /g/ respectively in all other contexts. So, while *c* is /k/ in *cosa* 'thing', *crema* 'cream', or *opac* 'opaque', it is /s/ in *cel* 'sky', *fàcil* 'easy', or *opacitat* 'opacity'. Likewise, *g* is /g/ in *gust* 'taste', *grega* 'Greek.FSG', or *llarg* 'long', but is /ʒ/ in *gel* 'ice', *àgil* 'agile', or *lògic* 'logical'. Before *e* and *i*, the phonemes /k/ and /g/ are spelled *qu* and *gu* respectively, which gives rise to alternations with the letters *c* and *g* such as *seca* 'dry.FSG' vs *seques* 'dry.FPL', *marco* 'I.mark' vs *marques* 'you.SG.mark', or *formiga* 'ant' vs *formigues* 'ants', *amagar* 'hide.INF' vs *amagui* 'hide.3SG.PRS.SBJV'.

Ç is pronounced /s/ and appears in final or prevocalic position except before *e* and *i*, where it alternates with *c*: *puça* 'flea' vs *puces* 'fleas', *capaç* 'capable.SG' vs *capaces* 'capable.FPL', *venç* 'win.3SG.PRS.IND' or *vençut* 'win.PTCP' vs. *venci* 'win.3SG.PRS.SBJV' or *vèncer* 'win.INF'. The occurrence of *ç/c*, as opposed to *s* or *ss*, given the homophony of these graphemes, is determined on etymological grounds.

J is pronounced [ʒ] and the digraph *tj* is pronounced [dʤ] and they alternate with *g* and *tg* respectively with the same phonetic values when preceding *e* or *i*. The digraph *-ig* is used only in word-final position, optionally followed by *-s*, representing the sound [ʧ] corresponding to devoiced /ʒ/ and /ʤ/, which gives rise to alternations between intervocalic *g* or *j* and final *ig* (*roig* 'red.M' and *roja* 'red.F', *passeig* 'walk.MSG' and *passejos* 'walks', *veig* 'see.1SG.PRS.IND' and *vegi* 'see.1SG.PRS.SBJV') and between *tg* or *tj* and final *ig* (*enuig* 'annoyance' and *enutjar* 'annoy.INF', *mig* 'half.M' and *mitja* 'half.F', *lleig* 'ugly.M' and *lletges* 'ugly.FPL').

The palatal nasal /ɲ/ is *ny*: *bony* 'lump', *canya* 'cane', *nyicris* 'weakling'. The palatal lateral /ʎ/ is *ll*: *lloc* 'place', *milla* 'mile', *full* 'page'. Catalan has the exclusive letter combination *l·l* (known as *ela geminada* 'geminate l') to represent /ll/, as in *til·la* 'linden flower infusion' or *xarel·lo* 'type of wine', or the /l/ of words with etymological /ll/, as in *col·laborar* 'collaborate.INF', *il·limitat* 'unlimited', or *mel·liflu* 'mellifluous'.

The letter *x* has several values: the voiceless palatal fricative /ʃ/ (*xai* 'lamb', *xocolata* 'chocolate', *cuixa* 'thigh', *peix* 'fish', *carxofa* 'artichoke'), the voiceless palatal affricate /ʧ/ as part of the *tx* digraph (*motxilla* 'backpack', *despatx* 'office') and either /gz/ (*examen* 'examination', *exempt* 'exempt'), generally before the stressed syllable, or /ks/ (*lax* 'lax', *expert* 'expert').

21.4 Morphology

21.4.1 Nominal inflection

The inflectional morphology found on nominal categories (nouns, adjectives, and determiners) reflects the syntactic features of number and gender. There is no case morphology except in a vestigial way on pronouns (§21.5.1).

The suffix /z/ (written *-s*) is the exponent of plural number. As a general rule, this suffix is added to the stem of nouns, adjectives, and determiners to signal that the word is syntactically plural. The absence of this suffix indicates that it is syntactically singular. Examples of pairs of singular and plural forms are: *tap* 'stopper' ~ *taps* 'stoppers', *fàcil* 'easy.SG' ~ *fàcils* 'easy.PL', *quin* 'which.SG' ~ *quins* 'which.PL', *aquella* 'that.FSG' ~ *aquelles* 'those.FPL'. The presence of this suffix may cause the insertion of an epenthetic vowel before it: if the stem is of the masculine gender, has stress on the last syllable and ends in /s/, /z/, /ʃ/, /ʧ/, and, in some dialects, also in /ʒ/, /ʤ/, /sk/, or /st/, the vowel [u] (or [o], depending on dialect), written *-o*, is inserted between the stem and the plural suffix, as shown in (16) (see §21.2.3 for the origin of this epenthetic vowel):

(16) a.	*dolç* [dols] 'sweet.MSG'	*dolços* [ˈdolsus] 'sweet.MPL'
	cabàs [kəˈβas] 'basket.MSG'	*cabassos* [kəˈβasus] 'baskets.MPL'
	generós [ʒənəˈɾos] 'generous.MSG'	*generosos* [ʒənəˈɾozus] 'generous.MPL'
	calaix [kəˈlaʃ] 'drawer.MSG'	*calaixos* [kəˈlaʃus] 'drawer.MPL'
	empatx [əmˈpaʧ] 'indigestion.MSG'	*empatxos* [əmˈpaʧus] 'indigestion.MPL'
b.	*bateig* [bəˈtɛʧ] 'baptism.MSG'	*batejos* [bəˈtɛʒus] 'baptism.MPL'
	rebuig [rəˈbuʧ] 'rejection.MSG'	*rebutjos* [rəˈbuʤus] 'rejections.MPL'
	disc [disk] 'disk'	*discos* [ˈdiskus] 'disks.M.PL'
	llest [ʎest] 'smart.MSG'	*llestos* [ˈʎestus] 'smart.MPL'

The plural forms in (16a), with the epenthetic *-o*, are common to all dialects, but some dialects maintain the older plural forms corresponding to (16b), without the epenthetic vowel: *bateigs*, *rebuigs*, *discs*, and *llests*.

Words that show a gender opposition between feminine and masculine in the singular have a feminine form ending in *-a* (phonetically [a] or [ə] depending on the dialect) and, generally, a masculine form characterized by the absence of a gender suffix; the masculine form may either end in a consonant or in an epenthetic *-e* (phonetically [e] or [ə] depending on the dialect) if the word requires a vocalic

support and, in a few words, in the affix *-o* (phonetically [o] or [u] depending on the dialect) or *-u* after a vowel. Thus, we have masculine/feminine pairs like those shown in (17), transcribed according to the central Catalan pronunciation: in (17a), with the *-a*/∅ contrast; in (17b), with the *-a*/*-e* contrast; and, in (17c), with the *-a*/*-o* or *-a*/*-u* contrast.

(17)	a.	*dolça* [ˈdolsə] 'sweet.FSG'	*dolç* [dols] 'sweet.MSG'
		contenta [kunˈtentə] 'happy.FSG'	*content* [kunˈten] 'happy.MSG'
		diputada [dipuˈtaðə] 'representative.FSG'	*diputat* [dipuˈtat] 'representative.MSG'
	b.	*pobra* [ˈpɔβrə] 'poor.FSG'	*pobre* [ˈpɔβrə] 'poor.MSG'
		mestra [ˈmɛstrə] 'teacher.FSG'	*mestre* [ˈmɛstrə] 'teacher.MSG'
		ampla [ˈamplə] 'wide. FSG'	*ample* [ˈamplə] 'wide. MSG'
	c.	*flonja* [ˈflɔnʒə] 'fluffy. FSG'	*flonjo* [ˈflɔnʒu] 'fluffy. MSG'
		sonsa [ˈsonsə] 'dull.FSG'	*sonso* [ˈsonsu] 'dull.MSG'
		atea [əˈtɛə] 'atheist.FSG'	*ateu* [əˈtɛw] 'atheist.MSG'

A good number of words belong to both grammatical genders, with no overt marking to distinguish them (e.g. *estudiant* 'student', *diferent* 'different', *igual* 'equal', *amable* 'kind'). In most of these words, the absence of a gender contrast in the singular form is maintained in the plural form, so, for example, the plural form *estudiants* 'students' can be used as a masculine word (as in *aquests estudiants* 'these.M students') or as a feminine word (as in *aquestes estudiants* 'these.F students'), just like the singular form can. But there is a class of words that, while having no gender opposition in the singular, have distinct gender forms in the plural. These are words like *capaç* 'capable', *feliç* 'happy', or *precoç* 'precocious', which coincidentally all end in *-ç*: their plural form requires the insertion of an epenthetic vowel between the stem and the plural affix, which is *o* for the masculine form, such as the forms in (16), and the unmarked epenthetic vowel *e* for the feminine form. In the singular form, they are used for both genders, as in *un nen capaç* 'a boy.MSG capable.SG' and *una nena capaç* 'a girl.FSG capable.SG', but, in the plural, they have different forms for each gender: *uns nens capaços* 'some.MPL boy.MPL capable.MPL', *unes nenes capaces* 'some.FPL girl.FPL capable.FPL'.

21.4.2 Verb inflection

Verbs are grouped into three morphological classes, or conjugations, on the basis of the inflectional affixes they take. These three classes are clearly distinguished by the form of the non-finite forms of the verb—infinitive, gerund and participle—as shown in (18) using the regular verbs *plorar* 'to cry', *perdre* 'to lose', and *dormir* 'to sleep:

(18)	Conjugation	Infinitive	Gerund	Participle
	I	*plorar*	*plorant*	*plorat*
	II	*perdre*	*perdent*	*perdut*
	III	*dormir*	*dormint*	*dormit*

Verbs of the second conjugation show a great deal of variation and irregularity. This is seen even in the non-finite forms. Conjugation II infinitives come in three forms: those ending in unstressed *-er*, like *témer* 'to fear', *créixer* 'to grow', or *plànyer* 'to pity', those ending in unstressed *-re*, like *caure* 'to fall', *vendre* 'to sell', or *prometre* 'to promise', and those ending in stressed *-er*, like *saber* 'to know', *poder* 'to be able', or *voler* 'to want'. In addition, there are the irregular *dur* 'to take', *dir* 'to say', *fer* 'to do', and *ser*/*ésser* 'to be'. Conjugation II participles have various endings, besides the regular *-ut* (§21.2.3): *-s* (/z/) ending, as in *ofès* 'offended', *romàs* 'remained', *empès* 'pushed', and *-t* (/t/) ending, as in *dut* 'taken', *cuit* 'cooked', *estret* 'narrowed'. There are a few participles in *-st*: *vist* 'seen' and *post* 'laid' and their derived forms. Conjugation III also has a few irregular participles, ending in *-ert*: e.g. *obert* 'opened', *sofert* 'suffered', *complert* 'fulfilled'.

In conjugation I, many forms include the thematic vowel (TV) *-a* between the verb stem and the affixes that mark tense, mood, and person and number, as seen in (18) for the non-finite forms. It is present in some forms of the present indicative and imperative (*plora*, as well as *plores* and *ploren*, where the rule replacing unstressed *a* in closed syllable by *e* applies), in the imperfect indicative (*plorava*), in the simple past (*plorà*), in the future (*ploraré*), and in the conditional (*ploraria*). It disappears when adjacent to a vowel, as in the present subjunctive (*plori*). The Balearic dialects preserve the TV in the first and second plural forms of the present indicative (*ploram* and *plorau*, respectively), whereas the continental dialects have replaced it by *-e-* in these forms (*plorem* and *ploreu*). A similar situation arises with the imperfect subjunctive, which has the TV *a* in Balearic (*ploràs*), but has undergone a replacement of the TV by *-e-* in the continental dialects (*plorés*).

Conjugation II verbs have the TV *-e* in many forms where this vowel is stressed: gerund, first and second plural forms of the present indicative (*perdem* and *perdeu*), in the simple past (*perdé*), and in the imperfect subjunctive (*perdés*), but lack this morpheme in forms where stress is on another vowel. There is a class of conjugation II verbs that take a /g/ morph in many of their forms. For example, a verb like *prendre* 'take.INF' has the stem /prɛn/, which we find without the /g/ morph in many

forms such as *pren* 'take.3SG.PRS.IND', *prenia* 'take.3SG.IPFV.IND', or *prenent* 'take.GER', and with the /g/ morph in all of the subjunctive forms, as in *prengui* 'take.3SG.PRS.SBJV', *prengués* 'take.3SG.PST.SBJV', and in the first person singular of the present indicative, where the /g/, being final in the word, is devoiced: *prenc* 'take.1SG.PRS.IND'. Other verbs that behave like *prendre* in taking the /g/ morph, are *beure* 'to drink', *escriure* 'to write', *ofendre* 'to offend', *valdre* 'to be worth', *treure* 'to remove'. The /g/ morph is also added to the stem to form the participle if the verb in question has a regular participle in *-ut* (*begut* 'drunk', *valgut* 'been worth'), but not if the verb has an irregular participle in /z/ or /t/ (*pres* 'taken', *escrit* 'written').

Conjugation III verbs have the TV *-i* in many forms: non-finite forms, as in (18), first and second plural forms of the present indicative (*dormim* and *dormiu*), the simple past (*dormí*), the future (*dormiré*), the conditional (*dormiria*), and the imperfect subjunctive (*dormís*). A class of conjugation III verbs takes the *-eix* morph (a semantically empty morph, like the TV and the /g/ morph) in exactly those forms that would otherwise be stressed on the stem and, instead, are stressed on this morph. This morph has the variant *-ix* (and *-isc* or *-ixc*) in western dialects and the allomorphs *-eix* and *-esc* in Balearic. (19) shows the present indicative and present subjunctive forms of *dormir* 'to sleep', which does not take this morph, and of *servir* 'to serve', which does take it, in the central Catalan form.

(19)

	PRS.IND	PRS.SBJV	PRS.IND	PRS.SBJV
1SG	*dormo*	*dormi*	*serveixo*	*serveixi*
2SG	*dorms*	*dormis*	*serveixes*	*serveixis*
3SG	*dorm*	*dormi*	*serveix*	*serveixi*
1PL	*dormim*	*dormim*	*servim*	*servim*
2PL	*dormiu*	*dormiu*	*serviu*	*serviu*
3PL	*dormen*	*dormin*	*serveixen*	*serveixin*

Catalan has a rich system of forms that reflect person and number agreement with the subject in finite verb forms. The suffix *-s* on the verb signals a second person singular subject, except in the imperative form, where it lacks a specific affix. The suffix *-m* on the verb indicates a first person plural subject; the suffix *-u* is for a second person plural subject; and the suffix *-n* is for a third person plural subject. There is no specific morphology for a third person singular subject. As for the marking of first person singular subjects, it varies depending on the tense and mood involved. In the subjunctive forms, both present and past, there is no specific morphology for first person singular subjects and, so, there is a homonymy for forms agreeing with first and third person singular subject. This homonymy also occurs in imperfect indicative and conditional forms. (20) shows the various forms for present subjunctive, past subjunctive, imperfect indicative, and conditional of the first conjugation verb *plorar* 'to weep', given in the central Catalan standard:

(20)

	PRS.SBJV	PST.SBJV	IPFV.IND	COND
1SG	*plori*	*plorés*	*plorava*	*ploraria*
2SG	*ploris*	*ploressis*	*ploraves*	*ploraries*
3SG	*plori*	*plorés*	*plorava*	*ploraria*
1PL	*plorem*	*ploréssim*	*ploràvem*	*ploraríem*
2PL	*ploreu*	*ploréssiu*	*ploràveu*	*ploraríeu*
3PL	*plorin*	*ploressin*	*ploraven*	*plorarien*

The first person singular forms have a distinct morphology in the future, the simple past, and the present indicative, which are illustrated in (21) together with the compound past for *plorar*. The simple past is used in formal style and has been replaced in colloquial usage by the compound past in all dialects except for central Valencian. The compound past consists of an auxiliary (*vaig*, *vas*, *va*, etc.) and an infinitive. This auxiliary form is historically derived from the present indicative of *anar* 'go', but differs from it morphologically: 1PL and 2PL are *vam* and *vau* for the auxiliary, but *anem* and *aneu* for the lexical verb, and in some dialects the auxiliary has developed forms analogical of the simple past (*vàreig*, *vares*, *vàrem*, etc.). In the present indicative, the ending for first person singular varies depending on dialect, as noted in §21.2.3. Leaving aside the verbs with the /g/ morph and a few irregular verbs (which have first singular present indicative forms like *veig* 'I.see', *faig* 'I.do', *vaig* 'I.go'), regular verbs in central and northwestern Catalan have an *-o* ending for the first person singular of the present indicative, as shown in (21)

(21)

	FUT	PRS.IND	PRT	COMPOUND PAST
1SG	*ploraré*	*ploro*	*plorí*	*vaig plorar*
2SG	*ploraràs*	*plores*	*plorares*	*vas plorar*
3SG	*plorarà*	*plora*	*plorà*	*va plorar*
1PL	*plorarem*	*plorem*	*ploràrem*	*vam plorar*
2PL	*plorareu*	*ploreu*	*ploràreu*	*vau plorar*
3PL	*ploraran*	*ploren*	*ploraren*	*van plorar*

The imperfect indicative has the ending *-ia* (with stress on the *i*) for conjugations II and III (*perdia*, *domia*), but, in a few conjugation II verbs in which this ending immediately follows a vowel, stress has been displaced to the preceding vowel and the *i* is pronounced as a glide: *feia* 'do.1/3SG.IMPF.IND', *deia* 'say.1/3SG.IMPF.IND', *queia* 'fall.1/3SG.IMPF.IND'. Valencian has preserved the imperfect subjunctives in *-ra* (descended from the Latin pluperfect indicative): *plorara*, *perdera*, *dormira*. In the other dialects, only two forms in *-ra* have survived, but are equivalent in content to a conditional form: *fóra* 'be.1/3SG.COND' and *haguera* 'have.1/3SG.COND'.

See Perea (2002) and Wheeler (2002) for more information about verbal inflection.

21.4.3 Articles

The standard form of the definite article is *el/l'* (MSG), *la/l'* (FSG), *els* (MPL), and *les* (FPL). The reduced form *l'* is used before words beginning with a vowel: *el nét* 'the grandson', *l'avi* 'the grandfather', *la lluna* 'the moon', *l'aurora* 'the dawn'. The prescriptive rule requires the use of the full FSG form *la* when it precedes an unstressed *i* or *u*, as in *la inèrcia* 'the inertia' or *la unitat* 'the unity', although in spoken usage there is some variation. The prepositions *a* 'to', *de* 'of', and *per* 'by/for', and *ca* 'at the house of', a reduced form of *casa* 'house', combine with the full forms of the singular and plural masculine article (not with the reduced form) as amalgamated, or contracted, forms: *al, als, del, dels, pel, pels, cal, cals*. Examples: *al sentit* 'to the sense', *dels antics* 'of the old ones', *pel carrer* 'along the street', *cals avis* 'at the grandparents' house', but *per l'amic* 'by the friend', *ca l'avi* 'the grandfather's house'. As noted in §21.2.3, the northwestern dialects (and to some extent the southern part of the central dialect) preserve the classical forms of the masculine article *lo* and *los*, alternating with the contracted forms just mentioned and with the asyllabic forms *l* and *ls*, which are used when the *l* can be syllabified with an adjacent vowel: *trau los llibres* 'bring out the books', *dels llibres* 'of the books', *agafa·ls llibres* 'pick up the books'.

The non-standard colloquial varieties have a so-called 'neuter' form *lo*, not accepted by prescriptive grammar, which is used without a noun and has an abstract value: *lo bo* 'that which is good, the good', *lo autèntic* 'the genuine stuff', *lo que m'agrada* 'what I like'. The standard norm requires the usual masculine article in such cases: *el bo, l'autèntic, el que m'agrada*.

The Balearic dialect maintains an active use of the article descended from IPSE (cf. §21.2.3). The forms are *es/s'* (MSG), *sa/s'* (FSG), *es* (MPL), and *ses* (FPL), where *s'* has the same phonological conditioning as the standard *l'*: *es nin* 'the boy', *s'homo* 'the man', *s'al·lota* 'the girl', *es peus* 'the feet'. The same elements as in standard varieties yield the contracted forms *as, des, pes*, and *cas*. The masculine forms *so* and *sos* are used by older people in Majorca only after the preposition *amb* 'with': *amb so cotxo* 'in the car'. A phonological dissimilation rule in Majorca and Menorca turns a sibilant into a stop before another sibilant, giving forms such as *es sac* [əd'sak] 'the sack', *ses joves* [səd'ʒovəs] 'the young women', *ets homos* [əd'zɔmos] 'the men'. Despite the general use of the IPSE article in Balearic, the ILLE article is used in a variety of syntactic contexts (cf. §46.3.1.1): to name the hours of the day (*les tres* 'three o'clock'), institutions (*l'església* 'the church'), and unique entities (*la mar* 'the sea), in phraseology (*a la dreta* 'on the right'), with place names (*l'Havana*), etc. So, in the same phrase it is possible to have instances of both articles: *el sol de s'horabaixa* 'the afternoon sun'.

Some Catalan dialects have what is known as the 'personal article' used before personal names. The classical system, preserved in the Balearic Islands, has *n'* before a vowel and, before a consonant, *en* for the masculine and *na* for the feminine: *en Fuster, na Maria, n'Adrià, n'Amàlia*. In a large part of central Catalonia, in the system considered standard, the masculine form *en* is preserved, sometimes in competition with the definite article, and elsewhere the forms of the definite article are used: *en Fuster, la Maria, l'Adrià, l'Amàlia*. In the rest of Catalonia, except for the southern portion, the definite article is used as the personal article. And in this southern portion of Catalonia and in the Valencian Country, there is no personal article, e.g. *Fuster, Maria*.

21.4.4 Word formation

Many of the derivational affixes found in Catalan are not used in the formation of new words. This is the case with the suffixes *-ill(a)* found in *forquilla* 'fork' or *cordill* 'string', or *-ença* in *naixença* 'birth' or *temença* 'fear'. On the other hand, many affixes are productive, as they can be used in forming new words, e.g. the diminutive suffix *-et(a)*, the augmentative *-as(sa)*, and the pejorative *-ot(a)*. Using these suffixes we can form *llibret, llibràs*, or *llibrot*, from *llibre* 'book', or *lletreta, lletrassa*, or *lletrota*, from *lletra* 'letter'. Other productive suffixes are *-isme* '-ism' and *-ista* '-ist', as in *budisme* 'Buddhism' and *budista* 'Buddhist', *-ada* (manifestation typical of someone), as in *pujolada* 'action typical of a person named Pujol', *-ció* '-tion' and *-dor* '-ator', as in *experimentació* 'experimentation' and *experimentador* 'experimenter', *-itzar* '-ize', as in *verbalitzar* 'to verbalize'. Among the more productive prefixes, we can mention *des-* 'de-' as in *desactivar* 'to deactivate', *in-* 'in-' as in *inhabilitar* 'to disqualify', and *re-* 're-' as in *recomptar* 'to recount'.

Another word formation process is compounding. A productive compound type is the V-N noun, in which the first member is a verb in its unmarked ending and the second, a plural noun: *parabrises* (*para* 'stop' + *brises* 'breezes') 'windscreen', *netejavidres* (*neteja* 'clean' + *vidres* 'window panes') 'window cleaner', *obreampolles* (*obre* 'open' + *ampolles* 'bottles') 'bottle-opener'. This process is recursive: *netejaparabrises* (*neteja* 'clean' + *para* 'stop' + *brises* 'breezes') 'windscreen wiper'. Another frequent compound is the N-A adjective: *panxacontent* (*panxa* 'belly' + *content* 'happy')

'carefree', *camacurt* (*cama* 'leg' + *curt* 'short') 'shortlegged', *pèl-roig* (*pèl* 'hair' + *roig* 'red') 'red-haired'.

Adverbs are formed quite productively by concatenating an adjective in the feminine singular form and the stem *-ment*: *dolçament* 'sweetly', *estúpidament* 'stupidly', *breument* 'briefly', *constantment* 'constantly'. The resulting structure is on the border between compounding and syntax. The two component elements are inseparable and function as a unit, as a word, and each one has an unreduced vowel, which is characteristic of compounds. However, unlike compounds, it has secondary stress on the first component element (Mascaró 2002:115). And, under coordination, it is possible to have the *-ment* ending on only the first of the two (or more) coordinated adverbs, as in *insistentment i repetida* 'insistently and repeated(ly)', with the same meaning as the more common construction with *-ment* on all coordinated adverbs: *insistentment i repetidament* (López and Morant 2002:1808).

21.5 Syntax

21.5.1 Verbal clitics

The affixal status of verbal clitics is shown by the fact that a clitic cannot be separated from the verb it attaches to, not even by the negative element *no* 'not', which in turn must stand adjacent to the verb. If we consider the relative position of *no* and other usually preverbal elements such as *ja* 'already, no longer' or *hom* 'indefinite personal subject' with respect to the verb, we find that *no* is always adjacent to the verb: *ja no cantem* vs **no ja cantem* 'we no longer sing', or *hom no canta* vs. **no hom canta* 'one does not sing'. When a clitic is present before the verb, it cannot precede *no*: *no la cantem* vs **la no cantem* 'we do not sing it'. Another argument for the affixal status of clitics is that they behave like affixes with respect to certain phonological processes. For example, the process of cluster simplification discussed in §21.3.4 applies word-finally, reducing a consonant cluster like /nt/ to [n] even if followed by a vowel in the same phrase, as in *anant a París* [ə'nan ə pə'ris] 'going to Paris'; yet it fails to apply when the verb is followed by a vocalic clitic, *hi* or *ho*, as in *anant-hi* [ə'nanti] 'going there'. Likewise with r-deletion (§21.3.4), as shown by contrasts such as *fer honors* [fe w'nors] 'do honours' vs *fer-ho* ['feɾu] 'to do it'. For most dialects, verbal clitics are a layer of affixes that do not affect stress (i.e. the rule of stress placement ignores clitics), which explains the possibility of forms with the stressed syllable followed by more than two unstressed syllables only when clitics are involved: *dóna-me-les* 'give them to me', *permetre-se-la* 'to allow himself it'. However, in Balearic and north Catalan enclitics attract stress on the last syllable (cf. §§40.3.3.1, 41.3.1), in some cases causing the otherwise stressed syllable of the verb to undergo vowel reduction, as in the Balearic forms *visitar-los* [vizitəl'los] 'to visit them', *entra-hi* [ən'trəj] 'go in' (Colomina i Castanyer 2002:579), which is clear evidence of the affixal status of verbal clitics.

Verbal clitics can be proclitic or enclitic. While medieval Catalan allowed verbal clitics to either follow or precede a finite verb form, depending on the syntactic contexts (cf. §48.3), modern Catalan requires clitics to be proclitic on verb forms except for positive imperatives and non-finite forms (except for participles, to which clitics cannot attach). Northern Catalan only allows enclitics with positive imperatives (Bonet 2002:937f). Thus, we have contrasts such as (22) in most dialects:

(22) a. La compraré. (cf. **compraré-la)
it= I.buy.FUT
'I will buy it.'

b. per comprar -la (cf. **per la comprar)
for buy.INF=it
'in order to buy it'

A large class of verbs that take a non-finite form, such as an infinitive, as their complement allow the syntactic dependents of the complement verb to become their own syntactic dependents (Aissen and Perlmutter 1983; Rizzi 1978; cf. also §31.2.2.3). Among these are modals like *voler* 'to want', *poder* 'may, can', *deure* 'must', movement verbs, aspectual verbs, and the auxiliary of the compound past. The most visible effect of this is the phenomenon of clitic climbing. Given the modal *pots* 'you.SG.can' and the infinitive *comprar* 'buy.INF', we can have the clitic *la*, corresponding to the direct object of the infinitive, attached to the infinitive, as in *pots comprar-la*, or attached to the finite verb, as in *la pots comprar*, both with the meaning: 'you can buy it'. When there is a sequence of verbs that allow clitic climbing, the clitic may appear on any of the verbs in the sequence, as shown in the following sentence, where the possible positions are indicated with the clitic in parentheses: *(la) vas poder(-la) començar(-la) a comprar(-la)* lit. '(it=) AUX.PST be.able. INF(=it) begin.INF(=it) to buy.INF(=it)'.

The form of a given clitic, leaving aside dialectal variation, depends on whether it is attached to a verb or to another clitic and whether it immediately precedes or follows a consonant or a vowel in the same V-clitic unit. We can distinguish three classes of clitics depending on whether their underlying representation consists of only one consonant (1Consonant), consists of a consonant and the plural morph (2Consonant), or includes a vowel (Syllabic). The 1Consonant class includes /m/ (1SG), /t/ (2SG), /s/

(3REFL), /n/ (PART), /l/ (3MSG.ACC). In the standard dialect, these clitics add the unmarked vowel spelled *e* to their right when preceding a clitic beginning with a consonant, as in (23a), or following a verb ending in a consonant and not before a vowel, as in (23b); otherwise, they add the vowel *e* to their left if not adjacent to a vowel, as in (23c), and remain unchanged when they are adjacent to a vowel, as in (23d,e). This is illustrated in (23) for the 1SG clitic:

(23) a. Me la dóna.
me= her= gives
'He gives it to me.'

b. per donar-me permís
for give.INF=me permission
'to give me permission'

c. Em dóna permís.
me= gives permission
'He gives me permission.'

d. M'ha donat permís.
me=has given permission
'He has given me permission.'

e. Dóna'm permís!
Give.IMP2SG=me permission
'Give me permission!'

Within the 1Consonant class, the 3MSG.ACC clitic differs in one of the contexts from the other members of the class. As expected, it is *el* before a consonant-initial verb (*el veig* 'I see him') and the single consonant *l'* or *'l* when adjacent to a vowel (*me'l dóna* 'he gives it to me', *l'he vist* 'I have seen him', *per veure'l* 'to see him'). But, following a consonant-final verb, it is *lo* (not *le*):

(24) a. per donar-lo
for give.INF=it
'to give it'

b. Deixeu-lo!
leave.IMP.2PL=him
'Leave him!'

Like the *e* that is added in the contexts illustrated in (23), the *o* in *lo* can also be analysed as an epenthetic vowel, but with the difference that it is a case of morphologically conditioned epenthesis. This epenthetic *o* is also found in the clitics of the 2Consonant class, which consists of three forms with the plural morph /z/, spelled *s*: /nz/ (1PL), /wz/ (2PL), and /lz/ (3PL). (The clitic /lz/ in the prescriptive norm is either the third person masculine plural accusative form, distinct from the corresponding feminine form *les*, or the third person plural dative form). In these forms, o is inserted before the plural morph when they are attached to a consonant-final verb, giving *nos*, *vos*, and *los* (a special allomorph is needed for *vos*), as in (25a); otherwise, if the initial consonant of the clitic is not adjacent to a vowel, the vowel *e* is added before it or the /w/ in /wz/ becomes syllabic *u*, as in (25b); and, elsewhere, no change occurs, as in (25c).

(25) a. per donar-nos permís
for give.INF=us permission
'to give us permission'

b. Ens dóna permís.
us= gives permission
'He gives us permission.'

c. Dóna'ns permís!
Give.IMP2SG=us permission
'Give us permission!'

Finally, clitics of the Syllabic class are *la* (3FSG.ACC), *les* (3FPL ACC), *li* (3SG.DAT), *ho* (3NEUT.ACC), and *hi* (LOC). The only alternation they show is the deletion of the vowel of *la*, which occurs in the same context as with the homophonous article *la* (§21.4.3): *la dóna* 'he gives it', *l'hem donada* 'we have given it', *la intueix* 'he intuits it'.

Clitics can combine with each other and the order in which they appear follows the template in (26) (Bonet 2002:973):

(26) *es* > 2nd person > 1st person > 3rd person > *en* > *ho/hi*

Although sequences of more than four clitics are rare this template governs the order in cases like *se me l'hi va emportar* 'she took it away from me' or *se te n'hi va quedar un* 'one of them stayed there on you'. However, certain clitic combinations do not exist, and the content in terms of semantic and syntactic features that would correspond to them is expressed by means of another morpheme combination. This is what happens with the clitic cluster that corresponds to two third person pronouns, one accusative and one dative, in the non-Valencian standard. Since the third singular dative clitic is /li/ and the third person accusative clitics are /l/, /la/, /lz/, /laz/ in accordance with gender and number, the transparent combinations of a clitic from each of these two sets are *li'l*, *li la*, *li'ls*, and *li les*, which are the forms used in Valencian. Outside Valencia, the same content corresponds to the accusative clitic followed by *hi*: *l'hi*, *la hi*, *els hi*, *les hi*, which is not a transparent combination. If the third person dative is plural, its combination with another third person pronoun is the concatenation of the third person plural dative clitic /lz/ with a

third person accusative clitic in the standard common to Valencian and non-Valencian: *els el, els la, els els, els les*. Many non-standard varieties outside Valencian have extended the analysis of /i/ (*hi*) as the expression of third person dative and use it regardless of whether it is a singular or a plural third person dative. As a consequence, the phonological sequence /lzi/ (*els hi*) corresponds to either one or two third person pronouns one of which is dative and the other plural, with either gender being possible. Thus it can correspond to a third person plural dative pronoun, regardless of gender, as in *els hi donem nous* 'we give them walnuts'; or to two third person pronouns, of which the accusative is singular and the dative is plural, or of which the accusative is plural and the dative is singular, or both of which are plural. Thus, a sentence like *els hi donem* can mean 'we give it to them', 'we give them to him/her', or 'we give them to them' (Bonet 2002:954-61).

Many verbal clitics correspond to pronouns and fulfil a grammatical function in the clause, which can be the direct object, indirect object, and various kinds of prepositional obliques. However, verbal clitics sometimes function as agreement markers, as lexically required markers, or as passivization markers. We find clitics functioning as agreement markers when there is clitic doubling: the clitic agrees in features such as person, gender, and number with the NP or PP that fulfils the complement function of the verb. An accusative clitic may function as an agreement marker only when the object phrase is a pronoun (obligatorily, if the latter is a first or second person pronoun):

(27) a. T' he vist a tu
you= I.have seen ACC.MRK you
'I saw you'

b. L' he vista a ella.
her= I.have seen.FSG ACC.MRK her
'I saw her'

As for dative clitics, whereas the prescriptive norm recommends limiting their use as agreement markers to the same cases as accusative clitics, many dialects and speakers use them as agreement markers much more generally, almost obligatorily, as in Spanish:

(28) a. Què (li) vas dir a la Marta?
what her.DAT= you.PST say.INF to the Marta
'What did you say to Marta?'

b. Això no (li) servirà a ningú.
this not 3SG.DAT= will.serve to no.one
'This will not be any use to anybody.'

Some verbs require specific clitics that do not fulfil any grammatical function: these are lexically required markers. An example of this use is *hi* in *haver-hi* '(existential) be': *haver* with an existential meaning requires the locative clitic *hi*: *No hi ha remei* 'There is no solution'. The lexically required clitic behaves like any other clitic and undergoes clitic climbing: *Hi pot començar a haver un remei* 'There may begin to be a solution'. Other examples of lexically required or inherent, clitics are the reflexive clitics in so-called pronominal verbs: *queixar-se* 'to complain', *desempallegar-se* 'to free oneself', *endur-se* 'to take away'.

Finally, the reflexive clitic *se* may be the marker of passivization (personal or impersonal). Verbs of many different classes (transitive, intransitive, unaccusative) may take this marker, leaving the argument that would otherwise be the subject (the logical subject) unexpressed, but understood to be human, as shown in (29a) for two unaccusative verbs. If there is no subject, as in (29a,b), it is said to be an impersonal *se* construction. If there is a subject, as in (29c), as shown by agreement, it is said to be a passive *se* construction.

(29) a. Quan es torna de vacances, s' arriba cansat.
when SE= returns from holidays, SE= arrives tired
'When you come back from holidays, you arrive tired.'

b. No se l'espera fins demà.
not SE= him=expects until tomorrow
'He is not expected until tomorrow.'

c. S'han venut els millors seients.
SE=have.3PL sold the best seats
'The best seats have been sold out.'

21.5.2 Subject and objects

The grammatical function subject in Catalan has the following properties: it agrees with the verb; it can be omitted with a definite interpretation; it has no case marking; it cannot be expressed as (or replaced by) a pronominal verbal clitic; in a copular passive clause, it corresponds to the direct object of the active counterpart. A copular passive includes a form of the auxiliary *ser* and the participle of the lexical verb, as in (30a); the logical subject (corresponding to the active subject) is optionally expressed with a phrase introduced by *per* 'by'.

(30) a. Han estat descrits els símptomes
have.3PL been described.PL the symptoms
(per un doctorand).
(by a doctoral student)
'The symptoms were described (by a doctoral student).'

b. Ha descrit els símptomes un doctorand.
have.3SG described the symptoms a doctoral student
'A doctoral student described the symptoms.'

In the passive (30a), the subject NP *els símptomes* agrees with the finite verb form of its clause *han* and with the participle of the passive verb *descrits*. Its thematic role is that of the direct object of the active (30b).

A direct object does not agree with the verb, as in (30b). In addition, (a) it cannot be omitted with a definite interpretation; (b) it may be expressed as a pronominal clitic; (c) it sometimes has a specific case morphology; and (d) it sometimes agrees with the participle in compound tenses with *haver* 'to have'. A sentence like (30b) would be ungrammatical without the direct object (**Ha descrit* 'He described') or would be interpreted as having a generic or non-specific object of description. The sentence is grammatical if the object is expressed by means of an accusative object:

(31) Els ha descrits.
them= has described.MPL
'He described them.'

A human direct object is sometimes marked with the same preposition that marks indirect objects, namely *a* 'to'. This preposition has the colloquial variants *amb*, *en/an*, and *ana* in all of its uses (Sancho Cremades 2002:1730f.), including as a direct object marker. Prescriptive grammar restricts this use of *a* to strong personal pronouns (*mi* 'me', *tu* 'you.SG', *ell/-a* 'him/her', *nosaltres* 'us', *vosaltres* 'you.PL', *si* '-self')—see (27)—and other pronouns and quantifiers such as *tothom* 'everyone', *tots* 'all', *el qual* 'whom' and to constructions in which the direct object could be confused with the subject (as in *El perseguia com el gat a la rata* 'She chased him like the cat chases the rat'; Sancho Cremades 2002:1737f.). Colloquial Catalan makes a much more general use of *a* with human direct objects. Thus, while both sentences in (32) are found colloquially, only (32a) is accepted by prescriptive grammar.

(32) a. Ha descrit a tothom.
has.3SG described ACC.MRK everyone
'He described everyone.'

b. Ha descrit a tots els sospitosos.
has.3SG described ACC.MRK all the suspects
'He described all the suspects.'

Lastly, according to the prescriptive norm, a direct object expressed as a third person clitic in a compound tense (*haver* 'to have' + participle) agrees with the participle in gender and number, as illustrated in (31), where the third person clitic *els* and the participle *descrits* are both masculine plural. The reality of the spoken language is that, leaving aside some conservative dialects (Balearic and northern Catalan) which display participle agreement in more contexts than are allowed in the standard language, most speakers either do not have any participle agreement or restrict participle agreement to the feminine accusative clitics *la* (SG) and *les* (PL) or just to the FSG *la* (Bel 2002:1134-7; Rosselló 2002:1932–4). Thus, for many speakers, when referring to a feminine singular antecedent such as *la construcció* 'the construction', it is normal to say *l'ha descrita* 'he described it', with agreement on the participle, while the unmarked form of the participle would be used with masculine clitics, such as in *l'ha descrit* 'he described it.MSG' or *els ha descrit* 'he described them.M'.

The canonical, or syntactically unmarked, position of the phrase (NP, PP) corresponding to the object or other complements of the verb is after the verb in the same intonational phrase, as seen in (25) and (32) for the direct object and in (28) for the indirect object. Leaving aside instances of clitic doubling, in which clitics function as mere agreement markers, pronominal clitics are the expression of a complement of the verb; consequently, a complement expressed as a clitic cannot also be expressed as a phrase. Thus, the direct object of *descriure* 'describe' can be expressed as an NP following the verb, as in *He descrit la Marta* 'I described Marta', or as an accusative clitic, as in *l'he descrita* 'I described her', but not as both, as in **l'he descrita la Marta* 'I described her Marta'.

If a phrase appears at all co-occurring with a coreferential clitic, it appears in a dislocated position, where it fills a discourse function such as topic, but not the complement function filled by the clitic. Thus, the NP *la Marta* is a dislocated phrase in (33), as it appears either before the verb, as in (33a), or separated from the preceding verb by an intonational break, as in (33b):

(33) a. La Marta, l'he descrita.
the Marta her=I.have described.FSG
'Marta, I described her.'

b. L'he descrita, la Marta.
her= I.have described.FSG the Marta
'I described her, Marta.'

There is considerable debate as to whether the phrasal expression of the subject should be assumed to have its canonical position following the verb in the same intonational phrase, like the object and other complements, or should be assumed to have a canonical position before the verb. What we find is that the NP subject can be postverbal without an intonational break separating it from the verb,

with all kinds of verbs: unaccusatives, as in (30a), since passive structures are unaccusatives, and transitives, as in (30b). However, the subject can also be preverbal, without any other syntactic difference in the sentence and with the same truth-conditional semantics, as in (34), corresponding to (30). And it can also be postverbal, with an intonational break before it, as in (35).

(34) a. Els símptomes han estat descrits
the symptoms have been described.MPL
(per un doctorand).
for a doctoral.student
'The symptoms have been described by a doctoral student.'

b. Un doctorand ha descrit els símptomes
a doctoral.student has described the symptoms
'A doctoral student has described the symptoms.'

(35) a. Han estat descrits (per un
have been described.MPL by a
doctorand), els símptomes.
doctoral.student the symptoms

b. Ha descrit els símptomes, un doctorand.
has described the symptoms a doctoral.student

A plausible interpretation of these facts is that there is no difference between subject and complements with respect to their canonical position as phrases: it is in the verb phrase (following the verb in the same intonational phrase). The difference between the subject and the complements is that there are no subject clitics and the verb optionally licenses a null pronominal subject. In the same way that the presence of an object clitic implies that the object is pronominal and that there is no phrase corresponding to the object (31), and that, if there is a phrase coreferential with the object, it is a topic phrase (33), a subject may be pronominal with no phrase corresponding to it (31-33). If there is a phrase that provides referential content to the subject, it is a topic coreferential with the pronominal subject and appears in a displaced position, as in (34) and (35) (see Bresnan and Mchombo 1987). When the verb does not license a null pronominal subject, the subject is realized as a phrase in the verb phrase (30). A similar view is defended in Rosselló (1986), Bonet i Alsina (1990), and Vallduví (1992b; 2002), whereas the view that the canonical subject position can also be preverbal has been defended most recently by Forcadell (2013).

There is a special preverbal position in which negative and quantified expressions such as *ningú* 'nobody', *res* 'nothing', *poca cosa* 'not much', as well as interrogative phrases such as *qui* 'who', can appear in either subject or complement function (Vallduví 2002:1264-7):

(36) Alguna cosa direm /sortirà.
some thing we.will.say /will.come out
'We will say something'/'Something will come out.'

Focus fronting allows any dependent of a verb to appear in initial position in the sentence and is interpreted as a contrastive focus (37) (Vallduví 2002:1267-72):

(37) UN GOS, tenim, no un gat.
a dog we.have, not a cat
'It's a dog that we have, not a cat.'