1

Introduction

I start by looking at the external history of French (§1.1), the distribution of the language around the world today (§1.2), the internal syntactic history of the language and the major typological features of the modern language (§1.3). In §1.4 I give a taste of what's syntactically interesting about French, both theoretically and cross-linguistically. Finally, I provide a brief overview of the syntactic framework in which the rest of the book is couched (§1.5).

1.1 Development and spread of French

Like all Romance languages, French has its roots in Latin, more particularly the vernacular spoken by the Romans who, in the first and second centuries BCE, colonised Gaul, at the time a predominantly Celtic-speaking area. Over the next five hundred years Celtic gradually gave way to Latin, the language of power, which therefore survived the demise of the Roman Empire towards the end of the fifth century CE. However, as was the case with varieties spoken in other regions of the Empire, the variety of Latin spoken in Gaul had begun to diverge from the Latin of Rome, and this process of divergence accelerated following the loss of the centralising influence of the Empire. Thus, while written Latin remained stable, the vernacular did not.

One major factor determining how the Latin of Gaul developed after the fall of Rome was the invasion of Germanic speakers, who by the end of the sixth century CE controlled most of Gaul. In contrast to the Romans, though, these Visigoths, Burgundians and Franks didn't impose their language on the indigenous peoples. On the contrary, they were willing to adopt much of what they found in their conquered lands, language and religion alike. The period of Germanic–Latin bilingualism which preceded the adoption of the local Romance variety played a significant role in the way the Latin of Gaul developed. A number of features of the invaders' Germanic tongues rubbed off on the local varieties of Latin.

This was most noticeable in the north, which was peripheral to the Empire and occupied by the Germanic hordes first. The social disruption caused by occupation was therefore greatest, while the influence of Rome had been weakest. Conversely, the influence of the Germanic tongues on the local varieties of Latin was least noticeable in the south, where the impact of Roman civilisation had been greatest and lasted longest. It's for this reason that the development of Latin in Gaul formed

two distinct dialect areas, a linguistically innovative (that is, more Germanic-influenced) one in the north (the Langue d'Oïl) and a conservative (that is, less Germanic-influenced) one in the south (the Langue d'Oc). And the contrast in terms of innovation and conservation continued into the second millennium CE: the changes which have taken place as the Langue d'Oc has developed from Old Occitan into Modern Occitan aren't as great as those which have taken place as the Langue d'Oïl has developed from Old French (OF) into Modern French (ModF).

The history of Gallo-Romance and French is usually divided in four/five stages. The OF period stretches from 842 (the oldest extant 'French' text, the Serments de Strasbourg, being dated then) to around 1300, Middle French (MidF), from around 1300 to around 1500. Early Modern French (EModF) covers the sixteenth century, only, while ModF stretches from around 1600 to the present day. To capture some of the more recent developments in the language, some linguists recognise a further stage of Contemporary French (ConF). I suggest in §1.4 that ModF and ConF are in fact two contemporary varieties spoken in a diglossic situation.

There was much variation, both dialectally and diachronically within the 450-year-long OF period. The unifying influence of Francian – the variety of the Langue d'Oïl spoken in the Ile-de-France that ultimately developed into what we know today as French – didn't come until the late OF period, around the turn of the thirteenth century. The significant turning point in the history of the language, and the one which arguably led to ModF being as lacking in characteristically Romance features as it is, came at the beginning of the MidF period, as Francian spread throughout Gaul. By the Renaissance, this spread was complete. The EModF period saw the beginning of political unity and a centralised monarchy, and is the time when French was first felt to be a national language and a reflection of national unity. It was also when French was first exported to North America and parts of Africa. Interest in, and concern for, the state of the national language continued and became more systematic in the ModF period. During the twentieth century, the development of a common French, independent of sociolects and dialects can be attributed to the social and geographical mobility which followed World Wars 1 and 2, as well as the development of (tele)communications, especially television.[1]

1.2 French in the world today

Counting the number of French speakers in the world today isn't easy. There are two reasons for this, and neither is specific to French. First, in most places around the planet, people aren't actually asked which language they speak; few countries – not even France! (see Rowlett 2006a) – include questions on language use and proficiency in their censuses. Second, even when people *are* asked which language(s) they speak, their answers aren't always straightforward. Certainly, it's

[1] For a sketch of the external history of French see Battye *et al*. (2000: 9–50) and Marchello-Nizia (2003).

difficult to define the notion 'a French speaker' in any meaningful way. In France it's true that 82% of the population are monolingual French speakers with native-speaker competence. Elsewhere in the Francophone world, however, this is the exception rather than the rule. Often speakers don't have native-speaker competence, and might more usefully be called French users rather than French speakers. Significantly, the status and function of French vary widely from one place to another: it may be an official language, vehicular language or vernacular language. So while French may well be used in numerous countries, it's often one of many languages within a multilingual setting, and often not even the dominant language. Thus, the notion 'Francophone country' is doubly problematic. On the one hand, what on the surface might look like a French speaker might in truth have a rudimentary competence in the language, only. On the other hand, it's not even the case that everyone living in a 'Francophone country' has any competence in French at all: there are some 500 million people living in the fifty or so member states (and six observers) of the *Organisation internationale de la Francophonie* (OIF), some three times the size of the world's French-speaking population. The use of the notion 'Francophone country' in estimating the size of the world's French-speaking population isn't therefore as straightforward as it might seem.

Despite the difficulty inherent in the enterprise, estimates of the number of French speakers in the world have been produced. The most recent edition of the report *La francophonie dans le monde* (Haut conseil de la francophonie 2005) speaks of 175 million *francophones* worldwide of whom sixty million are *francophones partiels*. Gadet (2003: 146–8, using sources dating from 1997) catalogues 142 million *francophones réels* and a further sixty-three million *francophones partiels*. Between 100 and 110 million people are learning French as a foreign language. French is thus the tenth or eleventh most widely spoken language in the world. Within international organisations like the United Nations (14% of speeches delivered to the General Assembly in 2001) and the European Union (30% of original documents produced by the European Commission), French is second only to English.

Nevertheless, official circles are clearly concerned at the potential international decline of the language, and the OIF has launched a *Plan d'urgence pour la relance du français dans les organisations internationales* 'Emergency plan to re-establish French within international organisations' and a *Plan pluriannuel d'action pour le français en préparation de l'élargissement de l'Union européenne* 'Multiyear plan of action in support of French in preparation for the enlargement of the European Union'. The OIF is also endeavouring to work closely with organisations promoting the use of Spanish and Portuguese, in order to defend multilingualism within international organisations, and cultural diversity more generally (Rowlett 2006a).[2]

[2] For details of the geographical distribution of French around the world see, for example, Battye *et al*. (2000: 2–9), Rossillon (1995) and Walter (1988). For recent trends see Haut conseil de la francophonie (2005). On the linguistic situation of French within France see

1.3 Evolution of French syntax

Classical Latin (CL) was, like Proto-Indo-European, overwhelmingly (S)OV. While CL had long been thought to have free word order, non-(S)OV orders were in fact marked and used for pragmatic effect. As expected from a typological perspective, CL used (synthetic, morphological) postdetermination: it had three noun classes (M, F, N), five nominal declensions, six nominal cases, but no prenominal articles; in verb syntax CL had four verbal inflection classes (plus one mixed class), fifteen simple verb paradigms (giving non-defective verbs up to eighty-five distinct finite forms and at least another nine non-finite forms), but was pro drop. Synthetic future, perfect and passive verb paradigms were available; comparative and superlative adjectives bore synthetic suffixes. However, even within CL, there were indications of typological changes to come: the adverbial sentential negative marker NON was preverbal rather than occupying its typologically expected postverbal position.

In Vulgar Latin (VL) innovation was widespread, with shifting patterns of basic word order and a move from postdetermination to (analytic, syntactic) predetermination. The three-way noun-class system was simplified into a two-way distinction, with the loss of the neuter; two of the five nominal declensions were also lost; the case distinctions were weakened by phonetic erosion, leading to an increased reliance on prepositions, especially DE and AD, to mark case distinctions. VL also developed prenominal articles, derived from the demonstratives and the numerals. CL's synthetic future/perfect/passive verbal paradigms and comparative/superlative adjectives were replaced by analytic ones. The verbs had a postposed auxiliary; the adjectives, a preceding adverbial.

VL and early Romance shifted away from CL in terms of basic word order, too. First, (S)OV moved to a TVX pattern of sentence-initial topics (rather than subjects), and verb-second (in the case of Gallo-Romance possibly a Frankish influence). Clause-initial phrasal constituents of various classes could provide the pragmatic link to the preceding discourse. Unless it was a topic, the object formed part of a pragmatically ordered postverbal sequence of constituents. However, TVX was vulnerable because the kind of evidence required by children to ensure its acquisition wasn't readily available. For a child unambiguously to arrive at a TVX model of clause structure, clauses need to be available with (a) overt subjects and (b) non-subject topics (Roberts 1993). Yet OF was still optionally pro drop, and topics and subjects often coincided. Thus, the TVX status of early Romance was far from robust and ultimately doomed. Consequently, the preverbal topic position regrammaticalised as the subject position: TVX → SVX (see Li and Thompson's 1976 notion of the subject as the grammaticalised topic). By the fifth century CE, SVO was widespread (particularly in subordinate contexts). Thus, verb-final had become verb-medial and SVO provided the unmarked word order for early

Rowlett (2006a).

Romance. Much of the broad picture of the evolution of late Latin into early Romance and modern Romance – the functional load shifting from morphology to syntax, from synthesis to analysis – can be attributed more or less directly to the typological OV → VO shift in basic word order (Marchello-Nizia 2003).

By the MidF period, the typological shift from postdetermination to predetermination had taken firm hold. The preverbal position had become increasingly regrammaticalised as the position of the subject. The spoken language had lost several person/number markers on finite verb forms, meaning that preverbal subject proforms were now an essential marker of subject φ features (a later resurgence in pro drop was due to Italian or Latin influence). Topicalised constituents could still precede the preverbal subject, but this was via the innovative device of left dislocation, which meant that they had a separate intonational contour and often co-occurred with a core-clause-internal resumptive proform. Crucially, there was no longer any systematic inversion, so the finite verb no longer occupied its characteristic second position (except residually in clauses introduced by such adverbials as *peut-être* 'maybe' and *sans doute* 'doubtless') (Kroch 2001). Unmarked TVX word order was thus lost, replaced by SVO core-clause word order, with a pragmatically activated left periphery. Small levels of residual verb-final structures are attributed to Latin influence.

As for nominal structure, the ongoing development within Latin and into Gallo-Romance saw the decline of the nominal case system (Vincent 1997). OF had just a two-way NOM–OBL distinction (*li chevaliers* NOM ~ *le chevalier* OBL 'the knight'; *li chevalier* NOM ~ *les chevaliers* OBL 'the knights'). The loss of case distinctions was accompanied by the rise of articles: definite *le* and *la* developed from demonstrative ILLUM and ILLAM and indefinite *un* and *une* from the numerals UNUM and UNAM during the fourth and fifth centuries.

Apart from in the pronominal system, the NOM–OBL distinction showed up on M nouns and adjectives, only, and later (around 1200 onwards) even this was lost, with the NOM being discarded in favour of the now multi-purpose OBL. Concomitantly, basic word order became increasingly fixed, and the use of determiners spread further. Determinerless nouns had been possible in generic or vague contexts, but definite and, later, indefinite articles were used here, too. From the fourteenth century onwards the articles were grammaticalised as default nominal markers rather than semantic markers of (in)definiteness. The spoken language lost the PL *-s* and F *-e* suffixes; prenominal articles were thus important markers of number and gender.

The loss of case distinctions also led to changes in the expression of dependency relations. Within nominals, the morphological genitive gave way to preposed dependants (*l'autrui joie* 'the joy of others'), as well as various strategies involving postposed dependants (*la fille le duc* 'the duke's daughter', *la fille a un roi* 'the daughter of a king', *la mort de Rollant* 'R.'s death').

The general OV → VO shift, together with the loss of word stress and its replacement with phrase stress, had a significant consequence for pronominal structures, specifically object proforms (Boucher 2003). Previously preverbal, the

stressed object proforms became postverbal, as expected. However, the fact that object proforms don't need to be stressed (typically, they encode old rather than new information) meant that the preverbal position didn't lose its object proforms altogether. Rather, a separate, *un*stressed set of object proforms survived preverbally, forming a phonological unit with the verb. Thus, the Latin object proforms survive as two distinct sets in Romance. This was particularly strong in French (for example, ME > unstressed preverbal *me*, stressed postverbal *moi*), as opposed to the other Romance languages (cf. Spanish and Italian where unstressed object proforms are sometimes postverbal), because French also lost pro drop: the preverbal position of unstressed *object* proforms allowed them to cluster with the increasingly compulsory unstressed *subject* proforms.

The development of sentential negation within French correlates nicely with the OV → VO shift, too. As we have seen, CL marked sentential negation using the negative adverbial NON, but in the unexpected preverbal position, a fact which suggests that a typological shift was already underway; certainly, there was no new shift in the behaviour of NON in VL. Rather, NON suffered the same morphosyntactic fate as the object proforms. Like the proforms, NON split into stressed and unstressed forms (*non* and *ne*, respectively); and as with the object proforms, unstressed *ne* was restricted to preverbal position (where it, too, formed a phonological unit with the verb), while its stressed counterpart (*non*) enjoyed considerable syntactic freedom. In OF *ne* was sufficient to mark sentential negation on its own. Increasingly, though, there was a problem: unlike the object proforms which, as replacements for discourse-familiar constituents, typically encode old as opposed to new information, the negative marker is very high in information content. This was problematic in that the division of labour between *ne* and *non* meant that sentential negation was marked by *ne*, the very negative marker which was incompatible with stress, and was squeezed in between a preceding (pro)nominal subject and a following object-proform(s)-plus-verb cluster. In order to highlight sentential negation, therefore, *ne* came increasingly systematically to be reinforced by postverbal elements, which could be stressed. This trend has now gone so far that *ne* is no longer capable of marking sentential negation on its own, and one particular postverbal negative reinforcer, *pas*, preferred in Francian, has become the default negative marker. Indeed, the weakening of *ne* is now such that its very presence in preverbal position is under threat, arguably since it prevents subject and object proforms from forming a single preverbal pronominal cluster.

Turning to the syntax of interrogatives, OF had no specific morpheme marking yes–no questions, although a marked verb-initial word order was available, involving inversion of the verb around a (pro)nominal subject. ('Inversion' is still possible with pronominal subjects, but was lost in the context of nominal subjects in the sixteenth century.) As for wh questions, fronting of the wh phrase, with the verb again inverted (but now in second position), was also possible early on. The interrogative marker *est-ce que* was available from the twelfth century in wh questions as an alternative to verb-second. Initially, *est-ce que* was perceived as a pragmatically marked, syntactically complex sequence, involving inversion (*c'est*

que → *est-ce que*). Uninverted wh + *c'est que* + SVO was also possible, as was wh + *que* + SVO. From the fourteenth century, however, *est-ce que* was seen as an atomic unit, and it expanded into yes–no questions around the fifteenth and sixteenth centuries. Significantly, atomic *est-ce que* has the attraction of allowing interrogation to be marked without disturbing SVO word order in the core clause. The French pattern known as complex inversion arose, not surprisingly, in parallel with the loss of simple inversion around a nominal subject, from the phenomenon dating from the fifteenth to seventeenth centuries of left dislocating the inverted pronominal subject (*Jean, est-il parti?* 'Has J. left?'). Reanalysis as a core-clause-internal phenomenon, and the loss of the comma intonation, resulted in the modern construction (*Jean est-il parti?*).

The EModF and ModF periods, because of the growing symbolic role of French as a reflection of national unity, saw growing concern about the state of the language spill over into interference with it. The official guardian of the language (within France, at least), the *Académie française*, was set up in 1635 and coincided with something of an obsession with *le bon Usage*. Of relevance here is the fact that some of the syntactic features of the modern standard language can, at least in part, be attributed to rulings by this artificial linguistic authority, rather than being the result of a natural evolution. For example, while concessive conjunctions previously happily introduced either IND or SUBJ subordinate clauses, the *Académie* decided that they should select the subjunctive, only. Conversely, while bridge verbs (of saying or thinking) originally also selected either IND or SUBJ dependent clauses, they were later ruled by the *Académie* to take the indicative, only (unless they appeared in negative or interrogative clauses). Theoretical syntacticians therefore need to be wary of how much relevance they attach to mood distinctions.

The orthographic representation of morphology hasn't gone untouched, either. The phenomenon of present-participle agreement, for example, was the subject of interference in 1679, when the *Académie* decided (by ten votes to six!) that present participles like *aimant* 'loving' should agree with their subject in structures like *une femme aimante* 'a loving woman', but not in structures like *une femme aimant ses enfants* 'a woman loving her children' (Klare 1998: 136). Section 2.3.2 shows how the artificial (irrelevant) orthographic rules relating to adverbial *tout* mask an underlying (relevant) phonological simplicity. And §2.2.1.4 suggests that orthographic norms imposed on imperative verb forms introduce a red herring of a complication. With past-participle agreement, too, care is needed. The (mostly but not entirely exclusively orthographic) phenomenon was introduced during the EModF period, in line with the pattern found in Italian. Now, given the structural similarities between the two languages, it's plausible that the phenomenon of past-participle agreement introduced in French reflects a valid underlying syntactic feature. Indeed, theoretical syntacticians have used the (im)possibility of past-participle agreement to support analyses of syntactic structure. However, if we aren't careful, we can be led astray by spelling conventions: recent orthographic reform of the 'rules' governing past-participle agreement might otherwise be taken to indicate a change in the syntax of structures involving the causative verb *laisser*

and subnominal en_2 'of it'. Such a conclusion would clearly be implausible. Thus, to the extent that, in a language like French, syntacticians look to the orthographic representation of otherwise phonologically non-overt morphological agreement for clues to syntactic structure, they need to be cautious.[3]

1.4 Syntactic interest of French

Should linguists be particularly interested in French? One reason for concluding that they should comes from the fact that, over the last half-century, the language has provided much of the empirical base which has triggered developments within theoretical syntax, particularly among researchers with comparative interests:

- the phenomenon of rightward quantifier float inspired Sportiche's (1988) work on the VP-internal subject hypothesis, an approach now extended to all thematic lexical items;
- auxiliary selection in Italian and French was at the heart of Perlmutter's seminal work on unaccusativity in the late 1970s;
- the contrast between verb–adverbial order in French and adverbial–verb order in English, as well as the syntactic differences between finite and non-finite verb forms in French, led to Emonds' (1978) analysis of V movement and Pollock's seminal (1989) work on the split-INFL hypothesis, approaches to clause structure which led directly to Cinque's (1999) massively exploded and hierarchical analysis of core-clause structure;
- contrasting noun–adjective orders in various Germanic and Romance varieties, including French and English, led to the parallel approach to nominal structure in terms of a strictly ordered hierarchy of functional categories and cross-linguistically varied degrees of N movement (Bernstein 2001);
- broader issues having to do with nominal-internal architecture and the distribution of formal features have usefully been investigated on the basis of the behaviour of French determiners;
- Pollock's (1989) idea that polarity is associated with a dedicated functional head/projection, Neg(P), was based largely on French bipartite negation;
- Kayne's (1975) seminal work on French clitics, including his classic tests for clitichood, led to much subsequent work within generative syntax on clitics;
- finally, our understanding of clause-initial phenomena such as wh fronting and subject–verb inversion has benefited greatly from consideration of some very recalcitrant facts from French, first because French doesn't fit neatly within the traditional distinction between wh-movement and non-wh-movement languages, and second because French patterns of inversion appear very different to those found in modern Germanic.

[3] For more detailed overviews of the development of French syntax see Harris (1978), Posner (1997: 198–214, 344–418) and Rickard (1989: 8–17).

Thus, the syntax of French has much to offer linguists, even those not crucially interested in the language per se.

Quite apart from factors like those set out above, there's another reason to be interested in the syntax of French: the phenomenon of syntactic variation. Relevant here isn't so much the use of *on* instead of *nous* for 1PL subjects, the omission of negative *ne* in the expression of sentential negation or the omission of impersonal *il*.[4] Rather, of relevance is the idea that there's something much more significant and syntactically interesting going on. It's sometimes claimed that there's been no significant syntactic change in French since the end of the seventeenth century, and that the label ModF reflects a three-century-long period of grammatical stability. However, as the book progresses, we'll see evidence that to talk of stability is to massively oversimplify the situation with a convenient sociopolitical fiction hiding a degree of variation which suggests that two distinct grammatical systems co-exist, each with its own properties, in a situation of diglossia. Thus, ModF (Massot's 2003 *français classique tardif*, Bernstein's 1991 *literary French*), the conservative variety taught in schools, is distinguished from ConF (Frei's 1929 and Zribi-Hertz's 1994 *français avancé*, Raymond Queneau's *néo-français*, Massot's 2003 *français démotique contemporain*, Bernstein's *colloquial French*), the more innovative vernacular learnt in the home. Gadet (1997) characterises ConF in terms of a *séquence progressive*, fixed word order, analyticity, invariability, but *not* simplification. Some linguists have gone so far as to suggest that the degree of innovation which has occurred in the vernacular is such that ModF is no longer a coherent or psychologically real variety (Bauche 1926; Côté 1999). For others, the variation found within French is to be explained by concluding that speakers switch, on the basis of sociosituational factors, between two grammars, which differ from each other in a number of quite specific ways, for example:

- the status of number marking within nominals (§3);
- the pragmatic status of the canonical subject position (§5.3);
- the locus of the feature marking yes–no interrogatives (§5.7).

From such a perspective, what looks superficially like sociolinguistic variation along a continuum is code-switching between the two grammars, and possibly amounts to an extended period of change in progress.

1.5 Theoretical framework

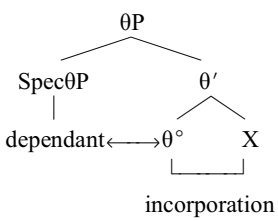
I round off this introductory chapter with an overview of the theoretical framework assumed in the book. The formal discussion throughout is couched within contemporary Chomskyan syntactic theory (Chomsky 1991; 1993; 1995a, b; 2000; 2001; 2005).

[4] For an accessible discussion of variation within French see Battye *et al*. (2000: 257–310) and Walter (1998). Particularly useful in this context is Gadet (1997: part three, syntax).

Lexical items are drawn from the (lexicogrammatical) lexicon as (more or less complex) bundles of phonological, semantic and formal (morphosyntactic) features. The grammar builds structures which allow the morphosyntactic requirements expressed by the formal features of lexical items to be satisfied. The grammar does this with two generalised, iterable, structure-building mechanisms, Merge and Move. Each combines two syntactic objects into one, allowing a feature of one (the dependant, or argument) to satisfy a requirement expressed by a feature of the other (the head, or functor). They differ with respect to the relationship between the functor and the argument: with Merge the functor and the argument are two independently existing syntactic objects; with Move the argument is (the copy of) a subpart of the functor. The formal mechanism relating a functor with an argument is Checking.

Thus, in (1) X is a predicate (a noun, a verb or an adjective[5]) associated with a lexical argument structure, that is, a number of θ roles each of which needs to project in syntax. This is possible by the intermediary of a θ head. X therefore needs a θ head, and this need is satisfied by merging with θ°. Since θ° is underspecified for any particular θ role, it needs access to a lexical argument structure. This need is satisfied by X moving to θ°. The [X θ°] complex can now assign a θ role, and does so by merging a dependant as a left-branching specifier. [X θ°] then checks its θ role against SpecθP. θP is an extended projection of X, in the sense of Grimshaw (1993). Phrases generated by Merge are thus binary branching, endocentric and antisymmetric (Kayne 1994).

(1)



Merge and Move are driven by (and therefore dependent on) the existence of a functor, that is, a syntactic object whose feature composition expresses a need (for an argument). If the inherent semantic structure of X includes no thematic grid, then no θ head is merged and no dependant either. More generally, a syntactic object whose formal-feature composition doesn't encode the need for an argument is, by definition, not a functor, and won't merge with an argument. The only way such a syntactic object is able to merge at all is as an argument (of some other functor). If the thematic grid of X contains more than one argument, then the structure in (1) is augmented by as many θP shells as are needed to provide a specifier position for

[5] There's an ongoing debate as to whether lexical items already bear categorial features when they are drawn from the lexicon, or whether categorisation is a by-product of the derivation (Borer 2005a, b). For ease of exposition, I assume that lexical items are marked for category from the outset.