



HAL
open science

Syntactic and semantic aspects of Romance complementizers

Benjamin Fagard, Paola Pietrandrea, Julie Glikman

► **To cite this version:**

Benjamin Fagard, Paola Pietrandrea, Julie Glikman. Syntactic and semantic aspects of Romance complementizers. *Complementizers in European Languages*, De Gruyter, pp.75-130, 2016, 978-3-11-041651-0. hal-01481926

HAL Id: hal-01481926

<https://hal.science/hal-01481926>

Submitted on 3 Mar 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

Syntactic and semantic aspects of Romance Complementizers

Benjamin Fagard, Paola Pietrandrea & Julie Glikman.

1. Introduction

This article aims at describing formal and functional properties of Romance complementizers. The paper is organized as follows: we provide an overview of Romance languages (sections 1.1. and 1.2.) as well as of the methodology employed in our study (Section 1.3.). We propose functional and formal definitions of complementation, complementizers and ‘canonical’ complementizers (Section 2.1.); we briefly sketch an overview of Romance complementizers and we justify our focusing on the Romance complementizers corresponding to the English *if* and *that* (Section 2.2.). We describe the syntactic behavior of these complementizers (sections 2.3. and 2.4.). We then focus on their semantic aspects and distribution; we show that a general function of illocutionary subordination is common to all complementizers, as is proven by the semantic effects of complementizer omission (Section 3). We study the semantic function of Romance ‘that’ complementizers (basically unmarked for factuality) (Section 4) and of Romance ‘if’ complementizers (basically non-factual) (Section 6). A semantic analysis of these complementizers then enables us to specify which kind of non-factuality is encoded in complements introduced by ‘if’ complementizers (Section 7). We go on to examine the diachronic sources of Romance complementizers (Section 8) and we finally focus on the marginal deviations found in Romanian on the one hand and French on the other hand as compared to other Romance languages (Section 9).

1.1. Romance languages – sociolinguistic and geographical features

Romance languages are a subgroup of Indo-European languages originating from Latin. They are mostly spoken in Europe (where all Romance languages are attested), in the Americas (mainly French, Spanish and Portuguese), and in Africa (mainly French and Portuguese), but there are also smaller, isolated communities elsewhere. The total number of mother-tongue speakers is around 700 million, among which around 200 million in Europe (by Ethnologue¹ counts, which we compiled). There are also Creoles based on Romance languages, with a total of around 11 million speakers. For instance, Creole languages based on French are found in Martinique, Guadeloupe, Haiti, French Guyana, Réunion and Louisiana, with close to 10 million speakers in total (according to the Ethnologue website, these are the following varieties: Saint Lucian, Guianese, Haitian, Karipúna Creole French, Louisiana Creole French, Guadeloupean, Morisyen, Réunion Creole French, San Miguel Creole French, Seselwa and Tayo). Chavacano and Palenquero are Spanish-based Creoles and have close to 300,000 speakers in total. The Portuguese-based Creoles are Angolar, Cafundo Creole, Crioulo, Fa D’ambu, Indo-Portuguese, Kabuverdianu, Korlai Creole Portuguese, Macanese, Malaccan Creole Portuguese, an extinct Pidgin in East Timor, Principense, Sãotomense, and Ternateño (with close to 700,000 speakers in total). However, from a grammatical point of view at least –which is our concern here– Creoles in general, and even Romance-based ones, are very different from European Romance, and it does not seem justified to address them jointly: we believe Romance-based Creoles deserve a separate,

¹ In its web version (www.ethnologue.com). Note that some numbers seem low; for French, for instance, by the 2010 counts of the ‘Organisation internationale de la francophonie’, the actual number of native speakers in the world is 115 million. The numbers in Table 1 should therefore be taken with caution; however, they do provide a rough idea of the number of speakers and the relative (numerical) importance of each Romance language.

dedicated study. For questions of simplicity, we shall henceforth use the term “Romance” to refer exclusively to European Romance languages, to the exclusion of Romance-based Creoles. Besides, although Romance languages proper are also spoken in South America, Africa and marginally elsewhere, our study focuses on European varieties.

Table 1: Number of speakers for the main Romance languages (in millions)

Variety	Number of speakers in the country of origin	Total number of speakers
Romanian	19,7	23,4
Romansch	0,035	0,035
Italian	55	61,7
French	53,2	67,8
Catalan	11,2	11,5
Occitan	1,9	2
Spain	28,2	328,5
Portuguese	10	178
Galician	3,2	3,2

1.2. Romance languages – major typological characteristics

The major typological characteristics of Romance languages include the fact that most of them are “pro-drop” languages, i.e. that subject pronouns tend to be omitted, or can be omitted. Note that this is not the case of French, nor of some Romance varieties in Northern Italy and Switzerland. A partly correlated feature of Romance is its relatively free word order, in pro-drop Romance languages. Romance languages are by and large inflectional as far as verbs are concerned, with a heavy Tense-Aspect-Modality system, but not so much for nouns and adjectives, which display variable genre and number marking and no case marking, with the exception of Romanian. Romance languages generally have preposed definite and indefinite articles, a large set of prepositions and conjunctions, and few modal particles comparable to the German modal particles such as *ja*, *auch*, *denn*, *doch* (cf. Diewald and Kresić 2010; Squartini 2013; Schoonjans 2013), relying instead on discourse markers.² Finally, they have infinitive clauses and a synthetic future tense. French differs from “Central Romance” on a number of points besides word order and pronouns; it displays for instance an analytic future tense and two distinct paradigms for demonstratives (demonstrative pronouns vs. demonstrative determiners). Romanian, on the other hand, presents a set of distinctive Balkan features. These include, for the verbal sphere, an analytic future tense (actually, two: [*o să* + subjunctive], based on the complementizer *să*, and [*voi* + infinitive], based on the modal verb *a vrea* ‘to want’), a wide use of the subjunctive mood and (correlated) avoidance of the infinitive. Romanian also has doubled clitics (e.g., *am văzut-o pe Maria* ‘I saw (her) Maria’). For the nominal sphere, the most notable features are the postposed definite article (e.g. *lupul* ‘the wolf.NOM/ACC’) and a residual set of nominal case markers with genitive-dative fusion (e.g. *lupule* ‘the wolfe.VOC’, *lupului* ‘the wolfe.GEN/DAT’). Finally, the lexicon has an important Slavic component absent from other Romance languages. This specificity of French on the one hand (see e.g., Koch 2002) and Romanian on the other can be observed in the case of complementizers, too, as we will see in this paper.

1.3. Methodology

² For the distinction between modal particles and discourse markers, we refer to Degand, Cornillie & Pietrandrea (2013: 23-4). The authors define modal particles as “a class of distributionally constrained words dedicated to semantically qualify the speech act with regard to the pragmatic presupposed context” and they define discourse markers as “linguistic elements endowed with the indexical function of relating the utterance in their scope to the linguistic context” (Degand, Cornillie, and Pietrandrea 2013: 23-24, but see also Diewald 2013 for a fine discussion).

In the following sections we investigate formal and functional aspects of Romance complementizers. In order to investigate their uses, we combined different methodologies. In a first phase, we based our hypotheses on the existing literature (for instance Jeanjaquet 1894; Herman 1963; Bat-Zeev Shyldkrot & Kemmer 1988), and on traditional philological instruments, i.e. dictionaries and grammars (see the Reference section). In a second phase, we called on native speakers' judgments to confirm these claims. In a third phase, we relied on corpora to check the validity of our results. The use of corpora for determining how speakers make use of language is now quite well established, as one can see from the flourishing subfield of corpus-based grammars (e.g., for English: Biber et al. 1999; for French: Abeillé, Godard & Delaveau in press, and in diachrony Marchello-Nizia et al. in prep. etc.). It is based on the idea that linguistic systems are best viewed as language(s) in use, and that grammar is shaped by frequency effects (see e.g., Bybee & Hopper 2001) – in other words, it seems to us that it is more interesting to look at language use and find out from there what rules the speakers are following than to rely exclusively on introspection. The corpora we used are the following (links to the corresponding databases are given in the Reference section):

- French: Valibel, PFC, Clapi, Frantext, frTenTen;
- Italian: Badip, itTenTen;
- Portuguese: Corpus do Português (Davies & Ferreira 2006), ptTenTen11;
- Romanian: CORV (Dascălu-Jinga 2002), Ruxândoiu, Romanian web corpus;
- Spanish: CREA, CORDE, Corpus del Español (Davies 2002), esTenTen.

We believe that these 'cross-examinations' guarantee the validity of our results³.

2. General description of Romance complementizers

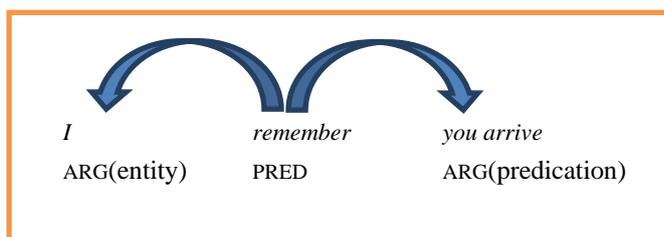
2.1. The semantic relation of complementation and its syntactic realizations in Romance

Before giving an overview of Romance complementizers, we briefly introduce the semantic notion of complementation. We define *complementation* as

the semantic situation that arises when a predicate takes a predication as one of its semantic arguments.

Fig. 1 illustrates the semantic relation linking the predicate 'remember' and the predication [*you arrive*] in a complementation situation:

(1) The semantic relation of complementation



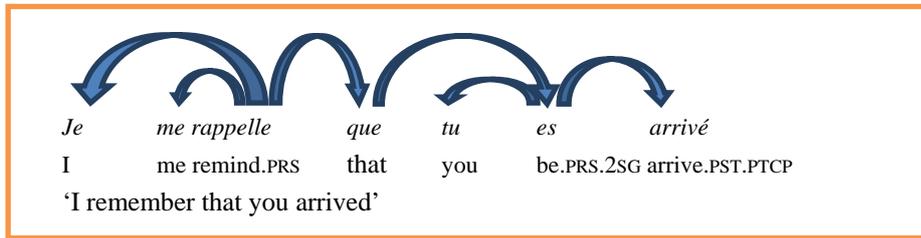
This semantic situation has different syntactic realizations in Romance languages: sentential complementation (2); integration via an infinitive complement (3); integration via a WH-element which functions as a shared complement (4); integration via nominalization (5); integration with complementizer omission (6); and integration by juxtaposition (7). Let us examine them in detail.

³ We would like to thank here the friends and colleagues who accepted to collaborate, especially Alexandru Mardale, José Carlos de Hoyos, Lennys Ospino, Daniel Rojas-Plata, Ana Fonseca, Sonia Pereira, Ana Margarida Abrantes and Giovanni Depau.

Sentential complementation is the syntactic situation which arises when the argument is realized as a sentence, i.e. a finite clause, as in (2):

(2) Sentential complementation

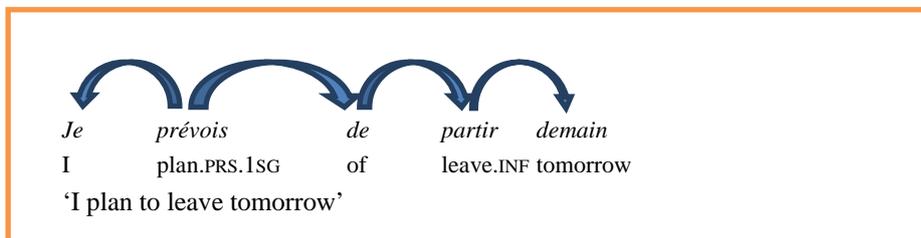
French



When the argument predication is realized as an infinitive clause, we speak of *integration via infinitive complements*:

(3) Integration via infinitive complements

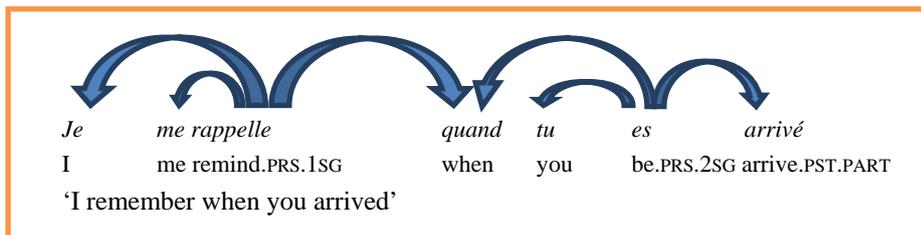
French



A complementation phenomenon which is quite frequent in Romance languages is the integration via a *wh*-element which functions as a complement shared by both the predicate and the argument predication. Let us take, for example, the French sentence in (4), *Je me rappelle quand tu es arrivé* 'I remember when you arrived': as shown in our graphic representation of the dependency relations, the *wh*-element *quand* 'when' can be analysed as realizing both the head of the object of the verb *se rappeler* 'to remind oneself, to remember' and an adjunct of the verb *arriver* 'to arrive'. In such cases, we speak of *integration via a shared complement*:

(4) Integration via a shared complement

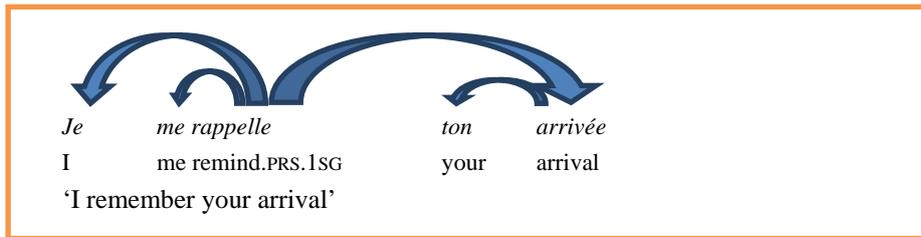
French



The argument predication can also be realized as a noun rather than as a verb. It is the case in instances of integration via nominalization (5):

(5) Integration via nominalization

French



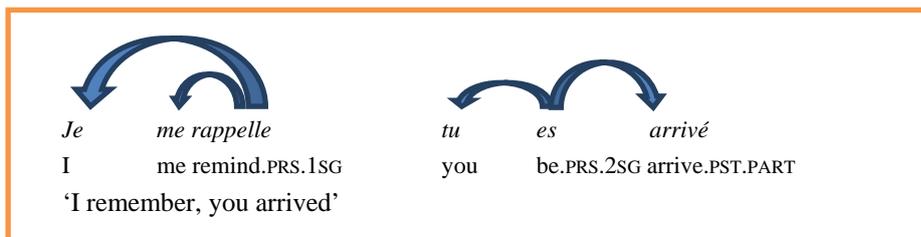
Finally, complementation can be realized through two quite similar albeit non-coincident mechanisms of syntactic integration, which make no use of complementation markers: omission of the complementizer (6) and juxtaposition (7). An example of complementizer omission is given in (6), where the main predicate *credo* 'I believe' syntactically governs the verb *arrivare* 'to arrive' by imposing a subjunctive inflection: *sia arrivato* [be.SBJ.3SG arrived]. The two sentences are linked by a dependency relation, in spite of the fact that there is no complementizer linking them (see below for a more detailed analysis).

(6) Integration with complementizer omission
 Italian



The mechanism of juxtaposition is illustrated in (7): the argument predication *tu es arrivé* 'you arrived' is simply juxtaposed to the main predicate *je me rappelle* 'I remember', without any dependency marker linking the two parts of the sentence.

(7) Integration via juxtaposition
 French



The syntactic difference between integration with complementizer omission and integration via juxtaposition can be shown by applying a test of mobility. Such a test shows that while the postposition of the main predicate is not possible for the former (8), it is entirely possible for the latter (9):

Italian

- (8) **Sia* *arrivato*, *credo*
 be.SBJ.PRS.3SG arrive.PST.PTCP believe.PRS.1SG
 'He has arrived, I believe'

French

- (9) *Tu es arrivé, je me rappelle*
 you be.PRS.2SG arrive.PST.PART I me remind.PRS.1SG
 ‘you arrived, I remember’

It should be noted, incidentally, that the constructions illustrated so far show different degrees of *syntactic* dependency between the main verb and the complement. The structures (2) through (6) are clear examples of microsyntactic dependency. By microsyntactic dependency, we mean the syntactic situation which arises when the existence, the category or the position of a word is determined by another word (see for example Tesnière 1959; Mel’cuk 1988; Kahane 2001).

In the structures (2) through (4) the very existence of the complementation markers *que* (in [2]), *de* (in [3]), *quand* (in [4]) is determined by the verbs of the main clause. In (5) the position of the object noun *ton arrivée* ‘your arrival’ is also determined by the main verb. In (6) the sentential complement is governed by the verb *credo* ‘I believe’, which imposes the subjunctive inflection on the verb *essere* in *sia arrivato* ‘that he has arrived (lit. *he be arrived*)’. The syntactic dependency of a complement on a verb can be tested with the help of simple diagnostic tests, such as the equivalence of the dependent complement with a clitic pronoun (10):

- | | | | | |
|---------|--|--|-----|---|
| (10) a. | <i>Je me rappelle</i>
‘I remember | <i>que tu es parti</i>
that you left’ | } ≡ | <i>je me le rappelle</i>
‘I remember that ’ |
| | | <i>quand tu es parti</i>
when you left’ | | |
| | | <i>ton départ</i>
your departure’ | | |
| b. | <i>Je prévois</i>
‘I plan to leave’ | <i>de partir</i> | } ≡ | <i>je le prévois</i>
‘I plan that ’ |
| c. | <i>Credo sia arrivato</i>
‘I believe he has left’ | | } ≡ | <i>lo credo</i>
‘I believe that ’ |

or the fact that the dependent complement falls under the scope of the negation of the main verb (11):

- (11) a. *Je ne me rappelle pas quand tu es arrivé/que tu es arrivé/ton arrivée*
 ‘I don’t remember when you arrived/ that you arrived/ your arrival’
- b. *Je ne prévois pas de partir demain*
 ‘I don’t plan to leave tomorrow’

- c. *Non credo sia arrivato*
'I don't think he has arrived'

The syntactic integration of the structure in (7) is slightly more problematic and it has been a hotly debated issue in the literature (Ross 1973; Schelfhout, Coppen & Oostdijk 2004; Dehé & Kavalova 2006; Ifantidou-Trouki 1993; Espinal 1991; Haegeman 1988). Following Kahane & Pietrandrea (2012), we claim that the integration between the main verb *je me rappelle* 'I remember' and the complement *tu es arrivé* 'you arrived' is not merely a matter of semantics or discourse, but a truly syntactic cohesion phenomenon. Kahane & Pietrandrea (2012) have proposed that, in spite of the fact that it is not possible to speak of a proper microsyntactic dependency (since the main verb does not control the existence, the position or the form of the complement clause), some syntactic constraints exist on the main verb – such as condition of adjacency or non-modifiability – which leads them to claim that a syntactic integration operates at a macrosyntactic level.

All in all, a number of syntactic realizations of the complementation situation are possible in Romance and complement markers and complementizers are not the only elements marking complementation in Romance. But let us further refine the notion of complementizer adopted in this article and thus specify the scope of our analysis.

2.2. Canonical and non-canonical complementizers

Following Noonan ([1985] 2007: 55), we define a complementizer as “a word, particle, clitic or affix, one of whose functions it is to identify [a clause] as a complement”.

Two refinements to this notion are necessary, however.

Firstly, following the convention adopted in this volume, we define *canonical complementizers* as expressions that both conform to Noonan's definition and occur only in finite complements. Markers of integration such as the preposition *de* in (3), whose function is to mark the integration of infinitive complements, are therefore not taken into account in our paper.

Secondly, let us further examine Noonan's definition: Noonan states that complementizers are expressions *one of whose functions* is to identify a clause as a complement. Expressions functioning as complementizers can indeed have other functions (as relativizers, adverbializers, etc. – see below, Section 2.3.). This taken into account, we restrict further our definition of complementizers as follows:

Canonical complementizers are expressions displaying the two following properties:

- (i) they have among other possible functions, the function of identifying a clause as a complement.
- (ii) their *only syntactic function in the sentence* is to identify a finite clause as a complement.

The second part of this definition leads us to exclude shared wh-complements such as 'when', 'where', 'how', and 'what' from the repertoire of canonical Romance complementizers. As we have seen in (4), these elements have two syntactic functions in complementation constructions. They are thus non-canonical complementizers: *complementizers*, because they are indeed used as complementation markers; *non-canonical*, because they have a syntactic function in the subordinate clause. Like canonical

complementizers, they also have non-complementizer uses, as we will see in Section 2.3. These non-canonical complementizers make up a much larger category, with slightly more formal and functional variation between languages. Table 2 below lists a few non-canonical complementizers in Romance.

Table 2 – ‘How’, ‘what’, ‘when’ and ‘where’ (non-canonical) complementizers in a few Romance languages

Complementizer form:	how	what	when	where	why
French	<i>comment</i>	<i>(ce) que / quoi</i>	<i>quand</i>	<i>où</i>	<i>pourquoi</i>
Italian	<i>come</i>	<i>che, che cosa, cosa</i>	<i>quando</i>	<i>dove</i>	<i>perché</i>
Portuguese	<i>como</i>	<i>o que</i>	<i>quando</i>	<i>onde</i>	<i>por quê</i>
Romanian	<i>cum</i>	<i>ce</i>	<i>cînd</i>	<i>unde</i>	<i>de ce</i>
Sardinian ⁴	<i>comenti</i>	<i>ita</i>	<i>candu</i>	<i>du, aundi, inui...</i>	<i>poita</i>
Spanish	<i>cómo</i>	<i>qué</i>	<i>cuando</i>	<i>donde</i>	<i>por qué</i>

These non-canonical complementizers globally have similar distributions in each language, and occur in similar contexts in different languages, as illustrated by (12) below.

- (12) *je (ne) sais (pas) quand/comment/pourquoi il est arrivé* (French)
(non) so quando/come/perché è arrivato (Italian)
(no) sé cuando/cómo/por qué llegó (Spanish)
(naõ) sei quando/como/por quê chegou (Portuguese)
(nu) ştiu cînd/cum/de ce a sosit (Romanian)
(no) iscu candu/comenti/poita est arribau (Sardinian)
 ‘I (don’t) know when/how/why he arrived’

The question is open as to whether relative pronouns, like the equivalents of *what* in (13) below, should be included in this category. Indeed, relative pronouns share some properties with non-canonical complementizers: they both identify the clause as a complement and have a syntactic function in the subordinate clause. Unlike non-canonical complementizers, though, relativizers often have an anaphoric relation to their so-called “antecedent” in the main clause, except in case of “headless relativizers” as in free relatives like (13) below. This is the most common argument for a distinction between relativizers and non-canonical complementizers, which is, besides, not unanimously subscribed to by all scholars (see for example Kayne 2005: 13 for the Generative approach, but also Le Goffic 1993: 525 in another syntactic framework).

- (13) *je sais ce que je veux* (French)
je sais quoi je veux (Substandard French)
so (che) cosa voglio (Italian)
so che voglio (Colloquial Italian)
sé qué quiero (Spanish)
sei o que quero (Portuguese)

⁴ For all languages, we refer as much as possible to the most widely accepted ‘standard’ variant. This is of course more problematic for some languages than others; very much so for Sardinian (see e.g. Mensching 2004:14). The Sardinian data presented here do not render justice to the enormous variation between Sardinian dialects, but that is not our goal.

știu ce vreau (Romanian)
iscu ita bolu (Sardinian)
 ‘I know what I want’

In almost all Romance languages, canonical complementizers are restricted to two series, one functionally close to the ‘that’ complementizer in English, the other to the ‘if’ complementizer (Table 3 below).

Table 3 – ‘that’ and ‘if’ complementizers in Romance languages (except Romanian)

	‘that’ complementizer	‘if’ complementizer
Spanish	<i>que</i>	<i>si</i>
Portuguese	<i>que</i>	<i>se</i>
Sardinian	<i>chi</i>	<i>si</i>
Italian	<i>che</i>	<i>se</i>
French	<i>que</i>	<i>si</i>

Typical examples of the ‘that’ complementizer are given in (14) below, and in (15) for the ‘if’ complementizer.

- (14) *je sais qu’il est arrivé* (French)
so che è arrivato (Italian)
sé que llegó (Spanish)
sei que chegou (Portuguese)
știu că a sosit (Romanian)
iscu chi est arribau (Sardinian)
 ‘I know that he has arrived’

- (15) *je me demande s’il est arrivé* (French)
mi chiedo se è arrivato (Italian)
me pregunto si llegó (Spanish)
pergunto-me se chegou (Portuguese)
mă întreb dacă a sosit (Romanian)
mi preguntu si est arribau (Sardinian)

‘I wonder whether he has arrive Note that Romanian presents a different configuration, with three canonical complementizers, two ‘that’ complementizers and one ‘if’ complementizer (Table 4 below): only Romanian has a different complementizer in (14) and (16).

Table 4 – ‘that’ and ‘if’ complementizers in Romanian

	‘that’ ₁ -complementizer	‘that’ ₂ -complementizer	‘if’ complementizer
Romanian	<i>că</i>	<i>să</i>	<i>dacă</i>

- (16)
- | | | | | | |
|-----------|---------------|------------|-----------|----------------|--------------|
| <i>je</i> | <i>veux</i> | <i>qu’</i> | <i>il</i> | <i>vienne</i> | (French) |
| | <i>voglio</i> | <i>che</i> | | <i>venga</i> | (Italian) |
| | <i>quiero</i> | <i>que</i> | | <i>venga</i> | (Spanish) |
| | <i>quero</i> | <i>que</i> | | <i>venha</i> | (Portuguese) |
| | <i>vreu</i> | <i>să</i> | | <i>vină</i> | (Romanian) |
| | <i>bolu</i> | <i>chi</i> | | <i>bèngiat</i> | (Sardinian) |
| (I) | want | COMP | (he) | come.SBJ.3SG | |

‘I want him to come’

2.3. Non-complementizer use of canonical complementizers

It should be noted that canonical complementizers can have other uses in Romance besides complementation.

‘That’ forms have a relativizer function in all Romance languages (except Romanian), as illustrated by the Portuguese example below:

Portuguese
(17)

<i>a</i>	<i>cura</i>	<i>de</i>	<i>austeridade,</i>	<i>que</i> (...)	<i>irá</i>	<i>cortar</i>
DEF.F	cure	of	austerity,	that	go.FUT.3SG	cut.INF
<i>nos</i>	<i>subsídios</i>	<i>de</i>	<i>férias</i>	<i>e</i>	<i>de</i>	<i>Natal</i>
in.DEF.M.PL	subsidy.PL	of	vacation.PL	and	of	Christmas

‘The austerity measures, **which** will cut into the vacation and Christmas bonuses’

They can also be used to introduce adverbial clauses, for example purposive clauses (18), (19):

Italian
(18)

<i>vieni</i>	<i>qua</i>	<i>che</i>	<i>ti</i>	<i>spiego</i>
come.IMP.2SG	here	that	you.DAT	explain.PRS.1SG

‘Come here and I’ll explain’

Portuguese
(19)

<i>anda</i>	<i>cá,</i>	<i>que</i>	<i>quero</i>	<i>falar</i>	<i>contigo</i>
come.IMP.2SG	here	that	want.PRS.1SG	speak.INF	with.you

‘Come here, I want to talk to you’

Another function of complementizer forms is to introduce non-assertive de-subordinated main sentences (Frajzyngier 1995), as in the French, Romanian and Italian examples below. These sentences can be of various types, jussive, interrogative (apparently only in Italian) and optative:

jussive, French
(20)

<i>Qu’</i>	<i>il</i>	<i>viene!</i>
That	he	come.SBJV.3SG

‘Let him come!’

jussive, Romanian
(21)

<i>Să</i>	<i>vină!</i>
That	come.SBJV.3SG

‘Let him come!’

interrogative, Italian

(22)

Che *hai* *sonno?*
That have.2SG sleep
‘Are you asleep?’

optative, French

(23)

Si *seulement* *je* *pouvais* *lui* *manquer*
If only I can.PST.1SG 3sg.DAT miss.INF
‘If only he could miss me’

optative, Romanian

(24)

Dacă (*măcar*) *ar* *veni*
If only have.COND.3SG come.INF
‘If only he would come’

A *quotative* function of ‘that’ forms is found apparently only in French (see (25) below):

French

(25)

*Cette femme est incroyable! “Allez, Allez!” **qu**’elle dit à la journaliste*
‘This woman is incredible. “Go ahead”, **COMP** she says to the journalist’.
(Tweet)

‘If’ forms are used in hypothetical constructions, more precisely, they introduce conditional protasis clauses. This adverbializer use, illustrated here for Spanish (26), is found in all Romance languages:

Spanish

(26)

Si *tú* *no* *vuelves,* *se* *secarán*
If you NEG come.back.PRS.2SG REFL dry.FUT.3PL

todos *los* *mares*
all.M.PL the.M.PL sea.PL
‘If you don’t come back, all the seas will dry up’

2.4. Combinability issues

The combination of complementizers with other subordination markers is a marginal phenomenon in Romance languages. It appears mostly in regional and/or substandard, informal variants, in which the ‘that’ complementizers can be combined with various complementation markers, such as ‘when’ markers. Not all combinations of non-canonical complementizer + ‘that’ complementizer are attested, and most of them are clearly substandard and very rare. However, we found in our corpora constructions such as ‘when

that’ (French *quand que*, Italian *quando che*, Portuguese *quando que*, Spanish *cuando que*, Sardinian *candu chi*) or ‘how that’ (French *comme que*, Portuguese and Spanish *como que*, Romanian *cum că*, Sardinian *comenti chi*), albeit often with very low frequencies. Examples below illustrate the constructions ‘when’ + ‘that’ for French (27) and Italian (28), and ‘how’ + ‘that’ for Romanian (29) and Sardinian (30):

Louisiana French

(27)

Je le ferai quand que j’irai à la ville.
 I CL do.FUT.1SG **when that** I.go.FUT.1SG to the town
 ‘I’ll do it when I’ll go into town’
 (Ditchy 1997: 173)

Regional Italian (Rome)

(28)

Quanno che ’ncomincia la partita ogni tifosetta
when that begin.PRS.3SG the.F match every fan.DIM.F.SG

se fà ardità
 REFL make.PRS.3SG bold.F.SG
 ‘When the match begins, every little fan becomes bolder’

Romanian

(29)

Mi-a spus cum că ar
 me.DAT-have.AUX.3SG tell.PST.PTCP **like that** have.AUX.COND.3SG

fi plecat
 be.AUX.INF leave.PAST.PTCP
 ‘(s)he told me that (s)he would be gone’

Sardinian

(30)

Bivi comenti chi no depis mòrriri cras
 live.IMP.2SG **like that** NEG must.IND.2SG die.INF tomorrow
 ‘Live as if you were to die tomorrow’

It should be mentioned that the semantic change induced by the adjunction of a ‘that’ complementizer in these constructions is not necessarily the same in different instances, including clear semantic contrasts such as the use of *when* + *that* to introduce conditionals (in the Emilian dialect of Italian, cf. Mazzoleni 1997: 202): these changes are thus not easily identified. Such constructions seem to suggest that, in some cases, the ‘that’ complementizer has grammaticalized to the point of being semantically neutral, and a obligatory marker of subordination (see Section 8), while in others it has retained a semantic value.

3. General semantic functions of Romance complementizers

Having isolated the class of canonical complementizers and studied their syntactic properties, we focus in the next sections on their functional aspects: we first observe the

general illocutionary (Section 4.1.) and semantic properties (Section 4.2.) of complementizers, and then analyse the semantic difference between ‘if’ and ‘that’ complementizers (Section 5).

3.1. Complementizers as markers of illocutionary dependency

The first question to be answered is: what is the semantic contribution of a complementizer to the sentence? In other words, we want to determine to what extent the presence vs. absence of a complementizer (irrespective of its nature) may change the meaning of a sentence.

We claim that *complementizers have the general function of marking the argument predication as illocutionarily dependent on the main predicate*. In order to further specify this notion, we refer to Cristofaro’s (2005) functional definition of subordination. According to her, all cases of subordination presuppose an “asymmetry assumption”: “a cognitive asymmetry is established between linked SoAs⁵, such that the profile of one of the two (henceforth, the main SoA) overrides that of the other (henceforth, the dependent SoA). This is equivalent to saying that the dependent SoA is (pragmatically) non-asserted, while the main one is (pragmatically) asserted” (Cristofaro 2005: 35). Since complementation is a particular case of subordination, following Cristofaro we may say that, in a complementation situation, the argument predication is non-asserted, whereas the situation depicted by the main predicate is asserted.

This asymmetry has been diversely addressed in the literature (see for example Cristofaro 2005; Boye & Harder 2009; Kahane & Pietrandrea 2012) and a battery of tests has been elaborated to distinguish between illocutionarily independent and illocutionarily dependent situations representations. The test in (31), for example, shows a canonical semantic situation of complementation, in which the main predicate bears the illocutionary force of assertiveness, whereas the argument predication lacks it: in the discourse negotiation (B, in example 31) the interlocutor challenges the truth of the assertion made by (A), and it is clear that his challenge focuses on the main predicate, *je t’ai dit* ‘I told you’, rather than on the argument predication *il est venu* ‘he came’.

French

- (31) A: *Je t’ai dit qu’il est venu*
‘I told you that he has come’
B: *Ce n’est pas vrai ! (que tu me l’as dit)*
‘It’s not true (that you told me)’
? Ce n’est pas vrai ! (qu’il est venu)
‘It’s not true (that he came)’

Similarly, a tag question such as *n’est-ce-pas ?* ‘isn’t it?’, whose function is to modify the illocutionary force of the anchor sentence, has scope on the main predicate – i.e. the part of the sentence that has an illocutionary force which can be modified – rather than on the argument predication, which lacks such illocutionary force – see (32):

French

- (32) *Je t’ai dit qu’il est venu, n’est-ce-pas ?*
‘I told you that he has come, didn’t I?’
‘*I told you that he has come, didn’t he?’

Concerning Romance complementation via juxtaposition, we claim that the absence of the complementizer changes the direction of the illocutionary dependency between main

⁵ SoA(s): state-of-affairs.

predicate and argument predication. Let us examine a sentence such as (33). In (33), there is no syntactic integration between the two parts of the sentence. However, a semantic integration between the two parts can be detected: the predicate *je t'ai dit* 'I told you' takes in fact as semantic object the predication *il est venu* 'he came'. From a semantic point of view, *il est venu* is thus dependent on *je t'ai dit*. But from an illocutionary point of view, we cannot say that the semantic object *il est venu* is dependent on the main predicate *je t'ai dit*. Rather, as the test shows, it is the semantic object that naturally bears the illocutionary force of the utterance:

French

- (33) A: *Il est venu, je t'ai dit*
 'He has come, I told you'
 B: *Ce n'est pas vrai ! (qu'il est venu)*
 'It's not true (that he has come)'
 B ?**Ce n'est pas vrai ! (que tu me l'as dit)*
 'It's not true (that you told me)'

A similar test can be applied to complementizer omission in Italian. Let us compare for example (34) and (35):

Italian

- (34) *Penso che è simpatico*
 'I think that he is funny'
 (35) *Penso sia simpatico*
 'I think he is funny'

In (34), the use of a so-called *recteur faible* ('weak head', cf. Blanche-Benveniste 1989; Blanche-Benveniste & Willems 2007), *penso* 'I think', followed by the complementizer *che* 'that', leaves open the ambiguity concerning the direction of the illocutionary dependency, as is shown by the test in (34'):

Italian

- (34') A: *penso che è simpatico*
 'I think that he is funny'
 B *non mi sembra (che lo pensi)*
 'It does not seem to me (that you think this)'
 B *non mi sembra (che è simpatico)*
 'It does not seem to me (that he is funny)'

In (35), instead, the use of the same *recteur faible*, *penso* 'I think', without any complementizer imposes the reading according to which it is the semantically dependent part of the sentence (the semantic object *è simpatico* 'he is nice') that bears the illocutionary force of the utterance, as is shown in (36):

Italian

- (36) A *penso sia simpatico*
 'I think he is nice'

B: **non mi sembra (che lo pensi)*
'It does not seem to me (that you think that)'

B *non mi sembra (che è simpatico)*
'It does not seem to me (that he is nice)'

In sum, we claim that one of the main general functions of complementizers in Romance is to mark the argument predication as not only syntactically and semantically, but also illocutionarily dependent on the main predicate.

3.2. Complementizers as markers of modality

Along with the general function of markers of illocutionary dependency, it is well known that at least some complementizers may have modal functions. According to Frajzyngier (1995), the focal function of complementizers is actually a modal one (which has been diachronically eroded in many languages). Japanese, for example, has three complementizers: *no*, *koto* and *to*. In many cases the choice among the three forms is determined by the verb of the matrix clause. In contexts in which the choice is open, *to* indicates that the speaker does not believe in the truth of the proposition represented in the complement. By contrast, neither *no* nor *koto* has this modal value:

Japanese

(37)

John wa Mary ga sin-da to sinzi-na-katta
John TOP Mary SBJ die-PST COMP believe-NEG-PST

'John did not believe that Mary was dead' (she might or might not have been)

(Kuno 1973: 216 – quoted in Palmer 1986: 148)

(38)

John wa Mary ga sin-da koto o sinzi-na-katta
John TOP Mary SBJ die-PST COMP OBJ believe-NEG-PST

'John did not believe that Mary was dead' (she was)

(Kuno 1973: 216– quoted in Palmer 1986: 148)

In Jacaltec (Mayan), different complementizers can mark different degrees of certainty. As noted by Craig (1977: 267), "the use of [the complementizer] *chubil* denotes a high degree of credibility or certainty, and the use of [the complementizer] *tato* introduces a notion of disbelief or reservation about a hearsay". She provides the (much quoted, see e.g. Noonan 2007: 48) examples below to illustrate this opposition, and remarks that in (39) "the source of information (...) is taken to be an unreliable source of information", while in (40) "the source of information is felt to be a reliable one" (Craig 1977: 268):

Jacaltec

(39)

Xal naj tato chuluj naj presidente
said DET COMP will.come DET president

'He said that the president is going to come' [unreliable source]

(Craig 1977: 268)

(40)

Xal naj alcal chubil chuluj naj presidente
said DET alcalde COMP will.come DET president
'The alcalde said that the president would come' [reliable source]
(Craig 1977: 268)

In Kinya Rwanda, a Bantu language, there are three complementizers: *ko*, *ngo* and *kongo*. *Ko* indicates that the speaker has a neutral attitude with respect to the truth of the complement, while *ngo* and *kongo* indicate that the speaker doubts the truth of the complement (on the basis of some direct and indirect evidential justification, respectively) (Givón 1982: 30):

Kinya Rwanda

(41)

yavuze ko azaaza
he.PST.say that he.FUT.come
'He said that he'd come'
(Givón 1982: 30)

(42)

yavuzu ngo azaaza
he.PST.say that he.FUT.come
'He said that he'd come' (but I doubt it)
(Givón 1982: 30)

(43)

yavuze kongo azaaza
he.PST.say that he.FUT.come
'He said that he'd come' (but I have indirect evidence which makes me doubt it)
(Givón 1982: 30)

In addition, Kinya Rwanda has a complementizer, *-niiba-*, marking the speaker's genuine (i.e. not evidentially supported) uncertainty about the truth of the complement clause (Givón 1982: 30).

As for European languages, it is well known that in some Slavic languages as well, some complementizers can have an epistemic meaning. Jakobson (1957), for example, reports their presence in Russian, where for instance *budto* 'as if' can be used as a complementizer and mark the speaker's doubts concerning the truthfulness of a statement. Similarly, Frajzyngier shows that the Polish construction *niby* – formed by the epistemic marker *ni* and the irrealis enclitic *by*, always occurring in clause-initial position, i.e. in the typical position of Polish complementizers – expresses the speaker's doubt about the truth of the proposition:

Polish

(44)

Ni-by był w Warszawie
COMP be.3M.SG.PFV in Warsaw.DAT
'Apparently he was in Warsaw'
(Frajzyngier 1995: 495)

This complementizer can co-occur with the *de dicto* complementizer *że* and introduce complements concerning situations presented as uncertain by the speaker:

(45)

Powiedział że ni-by był w Warszawie
say.PFV.PST.3SG.M COMP COMP be.3SG.M.PFV in Warsaw.DAT
'He said that he was in Warsaw' (I have my doubts about it)
(Frajzyngier 1995: 496)

Nordström (2010) has shown that, in Germanic languages, 'that' complements can depict either logically realis or logically irrealis situations, i.e., either situations in an "actualized or certain fact of reality" (Elliott 2000: 66–67), or situations portrayed "as purely within the realm of thought, knowable only through imagination" (Mithun 1995: 173), whereas 'if' complements are always irrealis.

Nordström & Boye (this volume) refine this proposal, by characterizing the contrast between 'that' and 'if' complementizers in Germanic languages as an epistemic modal contrast: 'that' complements are epistemically neutral, whereas 'if' complements encode the "uncertainty about the proposition expressed by the complement". In the next section, we closely examine the difference between 'that' and 'if' complements in Romance languages. We argue that the semantic contrast between the two complementizers is better described as a contrast in factuality than as a modal contrast.

We first focus our analysis on French, Italian, Spanish, Portuguese and Sardinian, and go on to examine closely the case of Romanian, which has a system of three complementizers quite different from the other Romance languages.

4. Semantics of 'that' and 'if' complements in Romance languages

4.1. Distribution: Predicates taking either 'that' or 'if' complements

The semantic difference between 'if' and 'that' is visible in their distribution, especially in the semantic contrast between predicates taking 'that' complements and predicates taking 'if' complements.

In order to examine this contrast, we searched, through the Sketch Engine query system (<http://www.sketchengine.co.uk>; Kilgariff et al. 2004), five large (written-language) corpora: the TenTen Corpora for French (frTenTen 10,666,617,369 words), Italian (ItTenTen 2,588,873,046 words), Spanish (esTenTen 2,103,770,763 words) and Portuguese (ptTenTen 2,757,635,105 words) and the Romanian Web as a Corpus for Romanian (RoWaC, 44,729,032 words). These corpora are created by Web crawling, processed with cleaning and de-duplication tools and lemmatized; the French, Spanish and Italian corpora are also tagged for part of speech. Although these corpora cannot be regarded as representative of the

languages in focus, their important size allowed us to identify and measure the salience of some very general patterns. This was particularly interesting for our purpose, i.e., identifying and comparing general tendencies in Romance languages.

In these corpora we searched for the most frequent verbs immediately preceding ‘that’ and ‘if’ complementizers, in all five languages. We could not search for all verbs governing ‘that’ and ‘if’ complementizers, as dependency relations are not encoded in our corpora and we could only rely on a positional criterion. Nevertheless, we verified and eliminated all the false positives (i.e., the verbs immediately preceding ‘that’ and ‘if’ complementizers that do not syntactically govern them) and the results we obtained are quite clear and consistent for the different languages under examination.

Let us consider in detail the following tables, which show the 12 most frequent verbs immediately preceding and governing ‘that’ and ‘if’ complementizers in Italian, French, Spanish and Portuguese. We study the distribution of Romanian, which is quite different from the others, in Section 6.1.

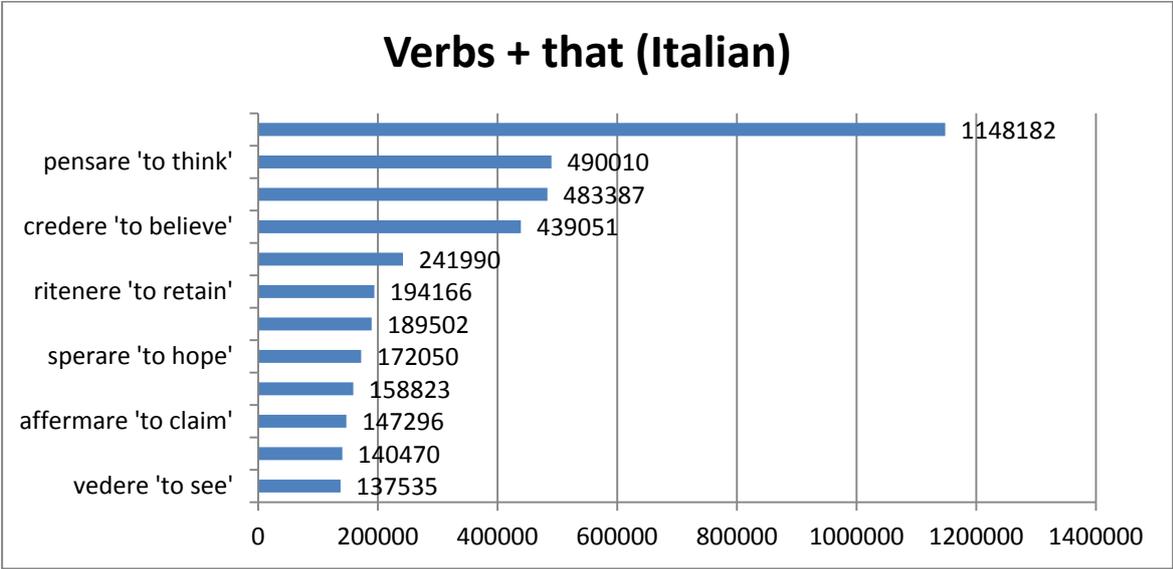


Figure 1: Most frequent verbs introducing complement clauses with ‘that’ in Italian

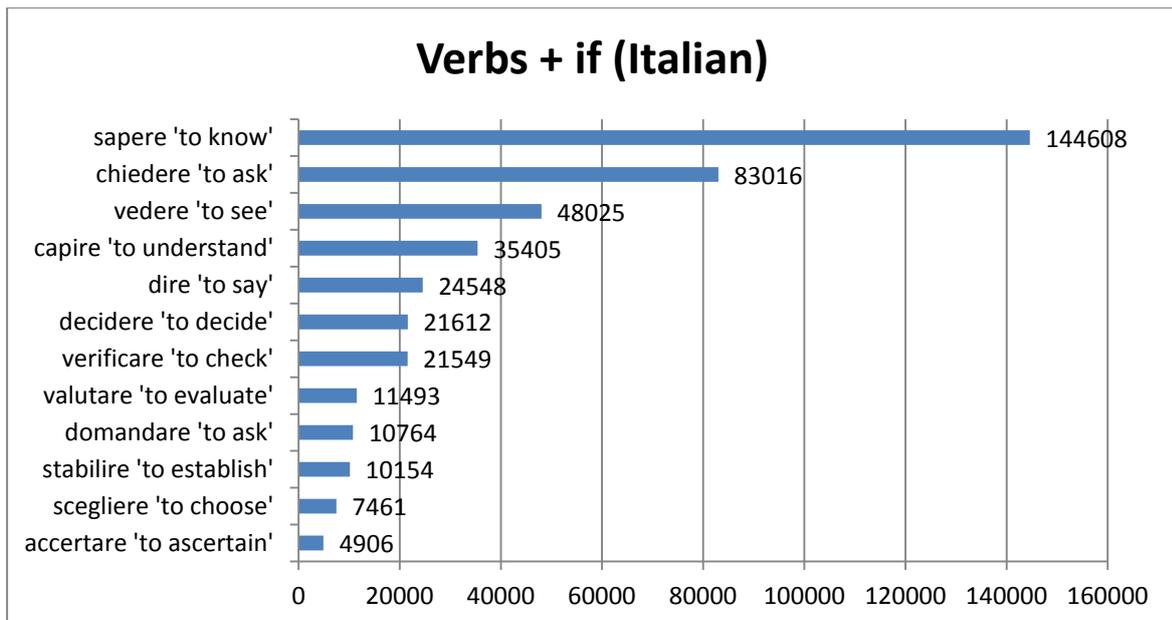


Figure 2: Most frequent verbs introducing complement clauses with 'if' in Italian

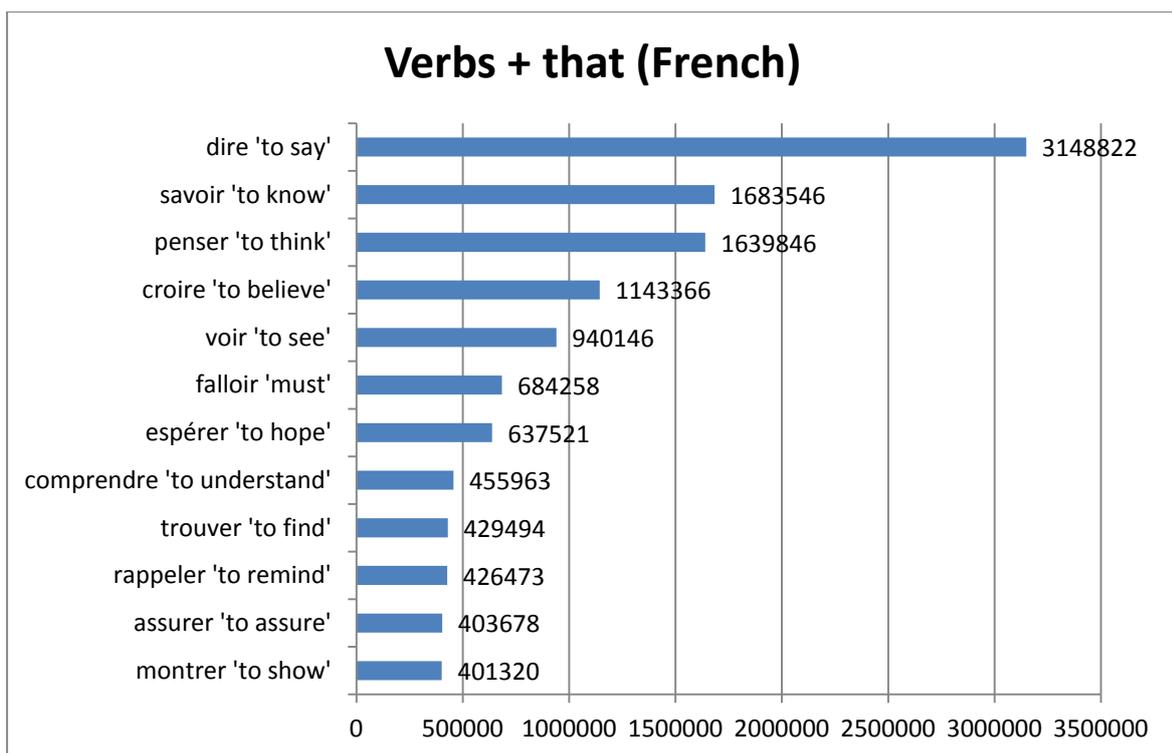


Figure 3: Most frequent verbs introducing complement clauses with 'that' in French

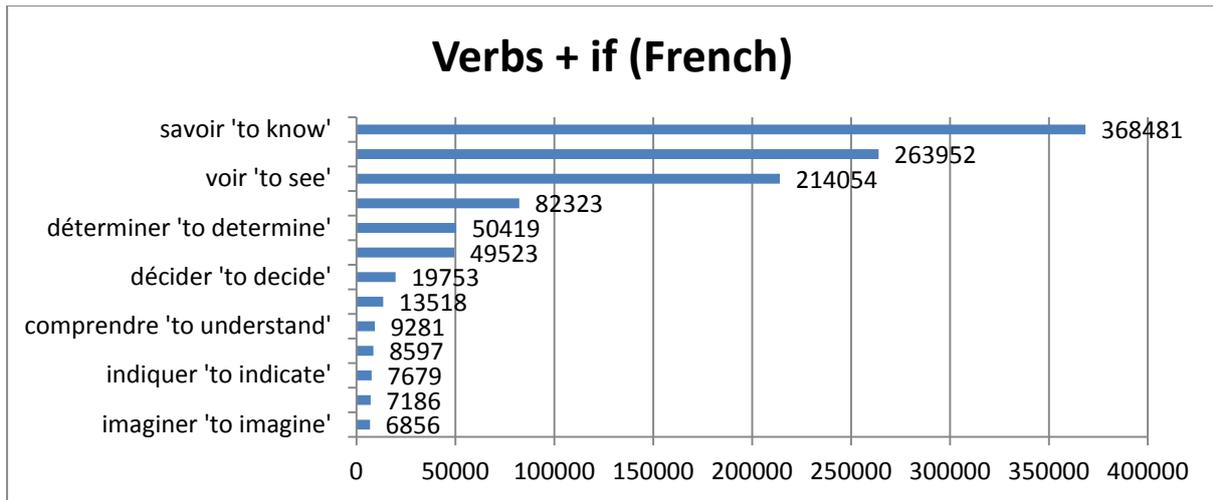


Figure 4: Most frequent verbs introducing complement clauses with ‘if’ in French

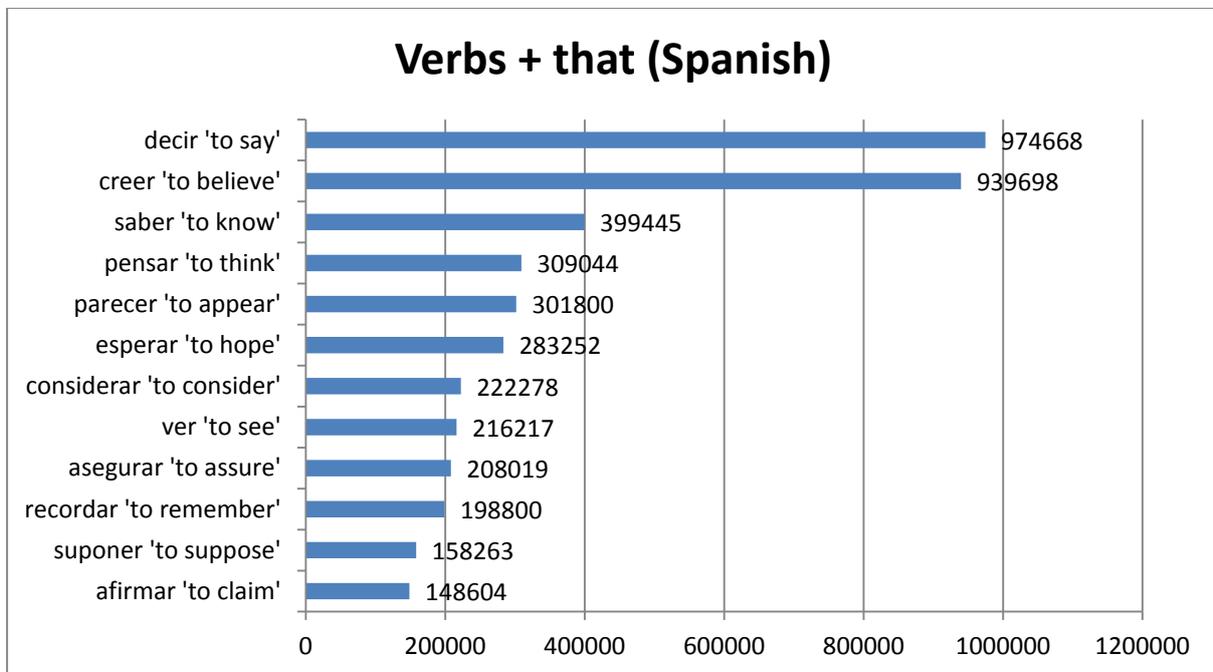


Figure 5: Most frequent verbs introducing complement clauses with ‘that’ in Spanish

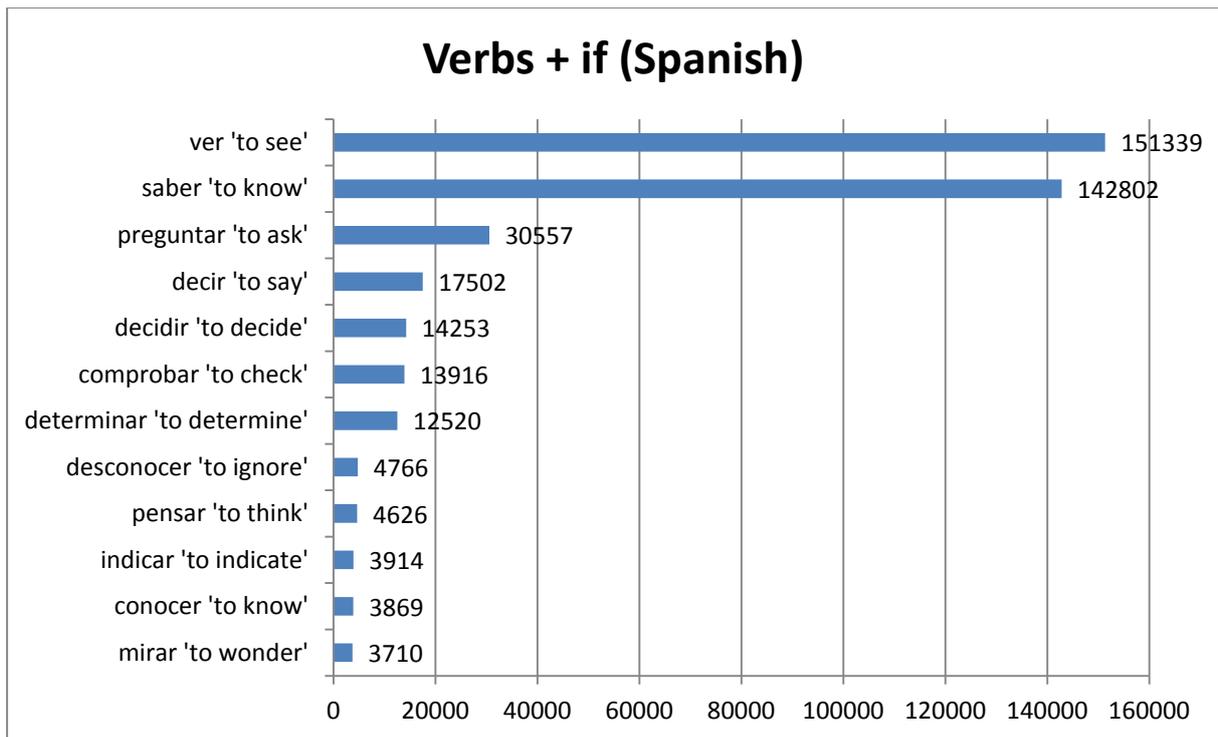


Figure 6: Most frequent verbs introducing complement clauses with ‘if’ in Spanish

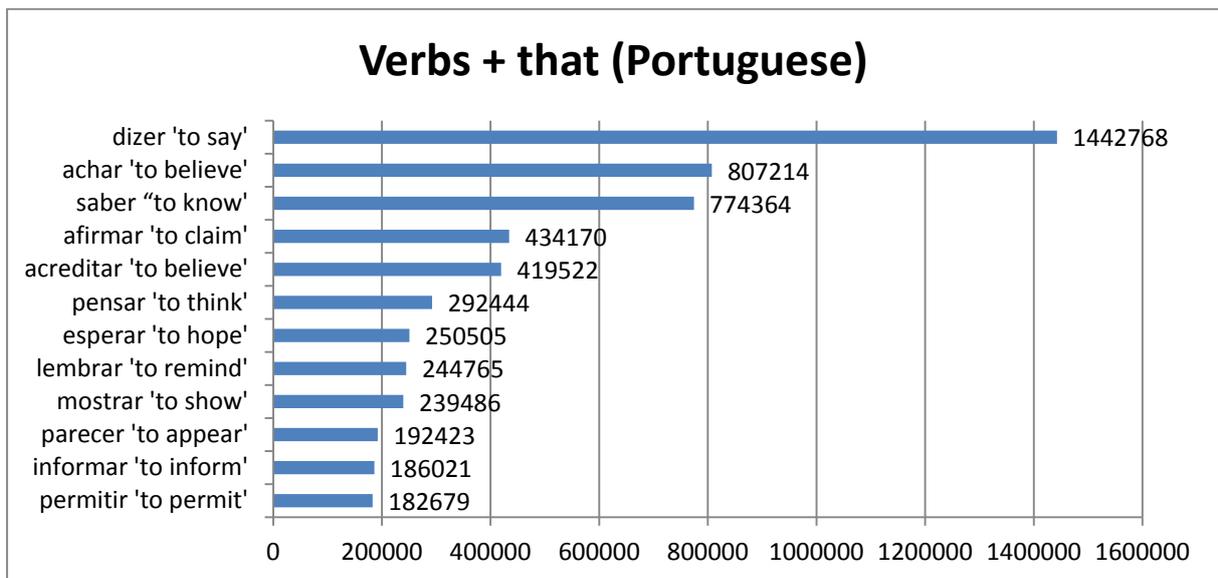


Figure 7: Most frequent verbs introducing complement clauses with ‘that’ in Portuguese

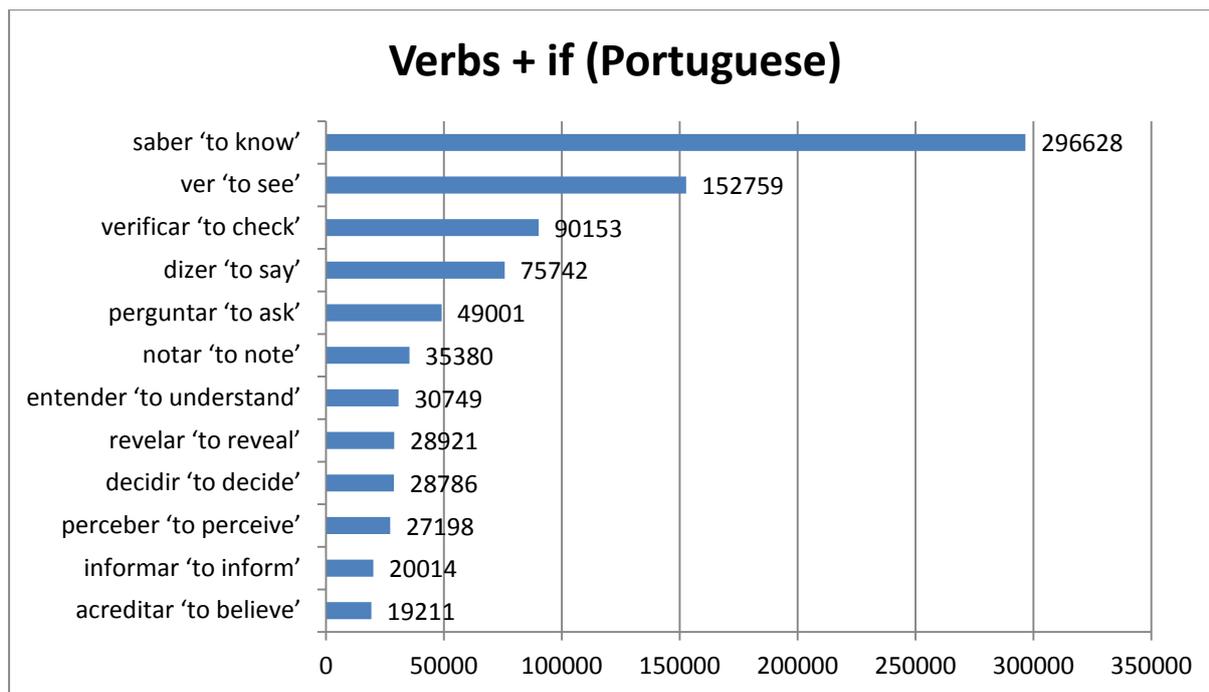


Figure 8: Most frequent verbs introducing complement clauses with ‘if’ in Portuguese

The first fact immediately observable is that, in Romance languages, there are different series of predicates. One series takes only ‘that’ complements, and another one takes only ‘if’ complements.

‘Say’, ‘think’, ‘know’, ‘believe’ are the four most frequent verbs governing ‘that’ complements in Italian, French, Spanish and Portuguese (the situation of Romanian is slightly different, see Section 6.1.). Whereas the verbs ‘ask’, ‘see’, ‘know’ are the most frequent verbs governing ‘if’ complements.

At first sight, it may seem that some predicates such as the verbs ‘know’, ‘say’, ‘understand’, ‘see’ govern, albeit unevenly, both ‘if’ and ‘that’ predicates. A closer look at the data shows, however, that these highly polysemic predicates are employed in different contexts or with different meanings depending on whether they govern ‘that’ or ‘if’ complementizers. For example, the verb ‘know’ only appears in negative (46) or non-factual contexts (47) when governing ‘if’ complements, whereas it can occur in factual contexts (48) when governing ‘that’ complements:

Spanish

(46)

<i>no</i>	<i>sé</i>	<i>si</i>	<i>puedo</i>	<i>viajar</i>	<i>con</i>	<i>un</i>	<i>visado</i>	<i>de</i>	<i>turista</i>
NEG	know.PRS.1SG	if	can.PRS.1SG	travel.INF	with	a	visa	of	tourist

‘I don’t know if I can travel with a tourist visa’

Spanish

(47)

<i>queria</i>	<i>saber</i>	<i>si</i>	<i>puedo</i>	<i>viajar</i>	<i>con</i>	<i>un</i>	<i>visado</i>	<i>de</i>	<i>turista</i>
want.PST.1SG	know.INF	if	can.PRS.1SG	travel.INF	with	a	visa	of	tourist

‘I wanted to know if I can travel with a tourist visa’

Spanish

(48)

yo sé que puedo viajar con un visado de turista
I know.PRS.1SG that can.PRS.1SG travel.INF with a visa of tourist
‘I know that I can travel with a tourist visa’

Similarly, the predicate ‘say’ occurs in negative or non-factual contexts (49–50) when governing ‘if’ complements and in factual contexts when governing ‘that’ complements (51):

Italian

(49)

Non dice se esiste una cura in tal senso
NEG say.PRS.3SG if exist.PRS.3SG a cure in such meaning
‘It (/he) doesn’t say if there exists a cure in that sense’

Italian

(50)

Potrebbe gentilmente dirmi se esiste
can.COND.3SG kindly say.PRS.3SG.me.DAT if exist.PRS.3SG
una cura in tal senso
a cure in such meaning
‘Could you kindly tell me if there exists a cure in that sense’

Italian

(51)

Dice che esiste una cura in tal senso
say.PRS.3SG that exist.PRS.3SG a cure in such meaning
‘He says that there exists a cure in that sense’

The verbs ‘understand’ and ‘see’ are cognitive verbs which can be classified, as for Aktionsart (Vendler 1967), as accomplishment verbs, i.e., as verbs whose internal temporal nature can be analysed as being durative and telic. Interestingly, when these verbs govern ‘if’ complements, it is the durative phase that is semantically selected (52), (54), whereas when they govern ‘that’ complements it is the telicity that it is selected (53), (55):

French

(52)

il est en train de voir si cela est possible
he be.PRS.3SG PROG see.INF if that be.PRS.3SG possible

et pas dangereux
 and not dangerous
 ‘he is trying to see if this is possible and not dangerous’

French

(53)

j’ ai vu que c’était possible
 I see.PST.1SG that it be.PST.3SG possible

et pas dangereux
 and not dangerous

‘I have seen that this is possible and not dangerous’

French

(54)

j’ essaie de comprendre si vous êtes un souverainiste
 I try.PST.1SG to understand.INF if you be.PRS.2PL a *souverainiste*

comme on dit
 like INDF.PRO say.PRS.3SG

‘I am trying to see if you are a partisan of sovereignty, as people say’

French

(55)

j’ ai compris que vous êtes un souverainiste
 I have.PRS.1SG understand.PST.PTCP that you be.PRS.2PL a *souverainiste*

comme on dit
 like INDF.PRO say.PRS.3SG

‘I have understood that you are a partisan of sovereignty, as people say’

In light of these results, we may propose a first rough generalization concerning the semantics of the predicates governing ‘if’ and ‘that’ complementizers.

Predicates governing ‘if’ complementizers can be regarded as a specific type of cognitive predicates, i.e. uncertainty predicates. An uncertainty predicate presents its complement as *not known* by the speaker (or the subject) of the utterance: either the speaker questions its truth value, or he claims not to know it, or he is depicted as being in the middle of a search for truth (like, for example, when a cognitive verb in its durative phase is used). The complements of these predicates correspond to what traditional grammar calls “embedded questions”: these complements are always presented as non-factual by the speaker.

Predicates governing ‘that’ complementizers belong, instead, to the larger class, already identified by Givón ([1984] 2001:154), of PCU predicates: perception (‘see’),

cognition ('think', 'know') and utterance predicates ('say', 'tell'). Verbs belonging to the class of perception, cognition and utterance predicates may or may not cast non-factuality on the embedded clause (e.g., the verb 'know' entails the factuality of its complements, whereas the verb 'think' does not)⁶.

Concerning the modal value of 'that' and 'if' complementizers, we may therefore conclude that in Romance languages 'if' complements are always non-factual, whereas 'that' complements are unmarked for factuality.

In the following sections, we refine this general statement by showing

- (i) which type of non-factuality is encoded by 'if' complements;
- (ii) the role played by mood in determining the factuality status of 'that' complements;
- (iii) the areal exceptions to the general Romance rule constituted by French on one side and Romanian on the other side.

4.2. The non-exclusion of factuality value of 'if' complements

The occurrence of 'if' complements within the scope of uncertainty predicates entails that 'if' complements have a specific type of non-factual meaning, i.e. what has been called a non-exclusion-of-factuality meaning (Masini & Pietrandrea 2010; Pietrandrea 2011). Let us examine closely this conceptual notion, thoroughly described by Pietrandrea (2011).

As argued by Pietrandrea (2011: 186), non-factuality can be seen as a supercategory encompassing several domains concerned with qualifying the actualization of a situation. A situation may be presented as not grounded in the perceivable reality for at least three different semantic purposes:

- (a) The situation is qualified as not real, i.e. as not occurring or not having occurred in the actual world: a negated situation, for example. This sense of non-factuality may be labeled *counterfactuality*.
- (b) The situation is introduced without specifying its spatio-temporal references in the actual world. The situation may be an a-temporal non-referential or a habitual situation: this sense may be labeled *non-referentiality* or, more generally, *non-factuality* (for the relation between non-referentiality and non-factuality, see Givón 2001: 303).
- (c) The situation is presented as being just an eventuality, in other terms the factuality of the situation is not to be excluded among other options. This third sense of non-factuality has been labeled by Masini & Pietrandrea (2010) and Pietrandrea (2011), *non-exclusion of factuality*, or *NEF*.

Non-exclusion of factuality is a defining component of some typically non-factual situations: hypothetical conditionals, concessive conditionals, some optatives, some recommendations and alternative relations (Pietrandrea 2011). In all these situations, the non-factual situation is portrayed as being an eventuality, i.e., as just one of a set of mutually exclusive alternatives.

Hypothetical conditional

- (56) *If it rains I will stay at home (and if it does not rain I will go out)*
(Prandi 2004; Dancygier & Sweetser 2005)

⁶ For the notion of factuality we refer to Elliott (2000), Mithun (1995).

Concessive conditional

- (57) *Even if we do not get any financial support, we will go ahead with our project*
(Haspelmath & König 1998)

Alternative relations

- (58) *Either I walk or I go by bus*
(Mauri 2008)

From a communicative point of view, non-exclusion of factuality expressions are not modal in nature: they suspend the commitment of the speaker rather than qualifying it (cf. Pietrandrea 2011; see also Bellert 1977: 344-345, who introduced this important distinction which had been quite neglected in the literature). In this sense, non-exclusion of factuality expressions convey a type of non-factuality that is quite easily distinguishable from the non-factuality encoded by epistemic, deontic and volitional predicates. Epistemic, deontic, volitional predicates *qualify* the commitment of the speaker toward a non-factual situation (i.e., toward a non-referential situation), whereas non-exclusion of factuality meanings do not qualify the speaker commitment.

From a semantic and syntactic point of view, it is important to stress that non-exclusion of factuality meanings in general and ‘if’ complements in particular depict semantic situations implying alternatives. We have mentioned indeed that an eventuality can be defined as one of a set of mutually exclusive options. It is always possible for ‘if’ complements (as well as for any other non-exclusion of factuality expression, Pietrandrea 2011) to make explicit these other options, as shown by the following examples:

- (59) *je me demande s’il est arrivé **ou non*** (French)
*mi chiedo se è arrivato **o no*** (Italian)
*me pregunto si llegó **o no*** (Spanish)
*pergunto-me se chegou **ou não*** (Portuguese)
*mă întreb dacă a sosit **sau nu*** (Romanian)
*mi preguntu si est arribau **o no*** (Sardinian)
‘I wonder whether he has arrived **or not**’
- (60) *je ne sais pas s’il est arrivé **ou non*** (French)
*non so se è arrivato **o no*** (Italian)
*no sé si llegó **o no*** (Spanish)
*não sei se chegou **ou não*** (Portuguese)
*nu știu dacă a sosit **sau nu*** (Romanian)
*no ischo si est arribau **o no*** (Sardinian)
‘I don’t know whether he has arrived **or not**’
- (61) *je n’ai pas compris s’il est arrivé **ou non*** (French)
*non ho capito se è arrivato **o no*** (Italian)
*no entendí si llegó **o no*** (Spanish)
*não percebi se chegou **ou não*** (Portuguese)
*n-am înțeles dacă a sosit **sau nu*** (Romanian)

no appo cumpresu si est arribau o no
'I didn't understand whether he has arrived **or not**'

(Sardinian)

The data from corpora show that the alternatives are indeed often made explicit within 'if' complements.

Italian

(62) *non ho capito se rispondono tramite email oppure contattano telefonicamente*
'I didn't understand whether they answer by e-mail or (if) they contact you by telephone'

French

(63) *je ne sais pas si tu parles à moi ou bien a skander09*
'I don't know if you are talking to me or to skander09'

Spanish

(64) *no he visto si es optico o laser*
'I didn't see if it is optical or laser'

Portuguese

(65) *uma outra que não sei se é cobra ou se é peixe*
'another one which I don't know if it is a cobra or a fish'

It should be said that the literature often regards 'if' complements as expressing the modal notion of doubt (Nordström 2010). The Romance data, however, lead us to question such an interpretation. As the following examples show, predicates of doubt in Romance generally cannot be followed by 'if' complements.

(66)	<i>*je doute s'il est arrivé</i>	(French)
	<i>*dubito se è arrivato</i>	(Italian)
	<i>dudo si llegó</i> but also ^{ok} <i>dudo que llegase</i>	(Spanish)
	<i>*dúvido se chegou / chegasse</i>	(Portuguese)
	<i>*mă îndoiesc dacă a sosit</i>	(Romanian)
	<i>*dubbido si est arribau</i>	(Sardinian)
	'I doubt if he has arrived'	

Dubitative situation, indeed, do not imply alternatives:

(67)	<i>*je doute s'il est arrivé ou pas</i>	(French)
	<i>*dubito se è arrivato o no</i>	(Italian)
	<i>*dudo si llegó o no</i>	(Spanish)
	<i>*dúvido se chegou ou não</i>	(Portuguese)
	<i>*mă îndoiesc dacă a sosit sau nu</i>	(Romanian)

**dubbido si est arribau o no*

(Sardinian)

*‘I doubt that he has come or not’

Dubitative situations are modal in nature and qualify the commitment of the speaker rather than suspending it.

In order to corroborate our hypothesis that ‘if’ complements serve the purpose of suspending the commitment of the speaker by putting forward the depicted situation as but one of a set of mutually exclusive situations, let us examine the possibility of modification of ‘if’ complements by non-exclusion of factuality and epistemic adverbs.

4.3. Constraints on the adverbial modification of ‘if’ complements

Let us start by making a semantic distinction between modal epistemic and non-exclusion of factuality adverbs. Semantically, we may define a modal epistemic adverb as an adverb which qualifies the commitment of the speaker toward what he says. For example, let us take the English epistemic adverb *probably*: by using this adverb (65), the speaker qualifies his/her commitment by saying that he/she puts forward the content of the utterance as probable.

(68) *This means that he is **probably** interested in her, which is why he hangs around so much.*

A non-exclusion of factuality adverb, instead, serves the purpose of suspending, rather than qualifying, the commitment of the speaker. For example, by using the English non-exclusion of factuality adverb *by chance* in (66), the speaker does not qualify his commitment, but simply presents the depicted situation as one option which is not to be excluded:

(69) *If **by chance** you have already visited Craigendarroch under the terms of this special invitation, I very much regret that we are unable to make the offer a second time (BNC corpus)*

Using a non-exclusion of factuality adverb, the speaker can present a situation as non-factual without having to qualify his commitment as to the likelihood of the situation.

In order to better appreciate the distinction between non-exclusion of factuality and epistemic adverbs, let us examine the compatibility of both kinds of adverbs with protases of hypothetical constructions.

As observed by Bellert (1977), the protasis of a conditional construction serves the purpose of suspending the commitment of the speaker, therefore, these constructions cannot host epistemic adverbs, since it is not possible for the speaker to suspend and qualify his/her commitment at the same time. On the contrary, NEF adverbs can occur in the protases of hypothetical constructions.

So for example, we have for Romance languages the following situation: the Italian adverb *magari* ‘perhaps’ (Pietrandrea 2010), the French adverb *des fois* ‘perhaps’, the Spanish adverb *a lo mejor* ‘perhaps’ (Cornillie 2010), the Portuguese adverb *por ventura* ‘perhaps’, the Romanian adverb *cumva* ‘perhaps’ can all occur in protases, whereas the modal epistemic

adverbs corresponding to *probably*, i.e., *probabilmente* (Italian), *probablement* (French), *probablemente* (Spanish), *provavelmente* (Portuguese), *probabil* (Romanian) do not occur in similar contexts:

Italian

(70) *un cellulare gli potrebbe tornare utile, così, { nel caso in cui / se } magari un giorno non riuscisse ad alzarsi dal letto per chissà quale motivo.*

‘a cell phone might turn out useful to him, like, { in case / if } perhaps one day he didn’t manage to get up for whatever reason’

(71) **un cellulare gli potrebbe tornare utile, così, { nel caso in cui / se } probabilmente un giorno non riuscisse ad alzarsi dal letto per chissà quale motivo.*

‘a cell phone might turn out useful to him, like, { in case / if } probably one day he didn’t manage to get up for whatever reason’

French

(72) *Super, vous pouvez ramener ça { au cas où / si } des fois mon disque implose ou une autre bêtise du genre*

‘Great, you can take this back { in case / if } perhaps my disk implodes or some other crazy thing happens’

(73) **Super, vous pouvez ramener ça { au cas où / si } probablement mon disque implose ou une autre bêtise du genre*

‘Great, you can take this back { in case / if } probably my disk implodes or some other crazy thing happens’

Spanish

(74) *Si a lo mejor/ *probablemente no tiene puesto ningún carrete, me doy por satisfecha.*

‘If { perhaps / *probably } there’s no film inside, I’ll consider myself satisfied’ (online dictionary example, cf. <http://en.bab.la/dictionary/english-spanish/film-festival>)

Portuguese

(75) *Recordo-me de uma aula. Nessa altura nós tínhamos falta e havia um contínuo que no fim da aula ia ler as faltas, para alguém poder reclamar se { porventura / ?provavelmente } lá estivesse.* (Corpus do Português, Spoken section)

‘I remember a class... At the time we were absent and there was an attendant who at the end of the class came to read out the absents, so anyone could protest if perhaps/??probably they were present’

Romanian

(76) *o să fac așa dacă ?cumva/*probabil va veni Ion*

‘let’s do it this way if ?perhaps/*probably John comes’ (created example)

Having identified a series of non-exclusion of factuality adverbs (adverbs presenting a situation as non-factual without qualifying the commitment of the speaker) for Romance languages, let us examine their distribution within the scope of ‘if’ complementizers.

While epistemic adverbs cannot modify ‘if’ complements, non-exclusion of factuality adverbs can; this is illustrated in the examples below with pairs of sentences in each language: ‘I didn’t understand if perhaps/*probably he arrived’. In these sentences, there is a clear contrast between the first variant with a non-exclusion of factuality adverb (‘perhaps’: *magari, a lo mejor, porventura, cumva* and *des fois*), which is generally quite acceptable, and the second variant with an epistemic adverb (‘probably’: *probabilmente, probablemente, provavelmente, probabil, probablement*), which is clearly non-grammatical. We tested the difference systematically in our corpora, with pairs of identical sentences, and the results are quite clear-cut, the first variant being always at least ten times more frequent than the second; we then confirmed these results with native speakers.

Italian

(77)

Non ho capito se
 NEG have.PRS.1SG understand.PST.PTCP COMP

*magari /
 probabilmente è arrivato
 perhaps be.PRS.3SG arrive.PST.PTCP
 ‘I didn’t understand if perhaps/*probably he arrived’

Spanish

(78)

*No entendí si a lo mejor / *probablemente llegó*
 NEG understand.PST.1SG COMP perhaps /probably arrive.PST.3
 SG

‘I didn’t understand if perhaps/*probably he arrived’

Portuguese

(79)

*Não percebi se porventura/*provavelmente chegou*
 NEG understand.PST.1SG COMP perhaps/*probably arrive.PST.3SG

‘I didn’t understand if perhaps/*probably he arrived’

Romanian

(80)

*N-am înțeles dacă cumva/*probabil*
 NEG-be.PRS.1SG understand.PST.PTCP COMP perhaps/*probably

a sosit
 be.PRS.3SG arrive.PST.PTCP

‘I didn’t understand if perhaps/*probably he arrived’

French

(81)

je n'ai pas compris s'il est
I NEG.have.PRS.3SG NEG understand.PRS.1SG COMP-he.cl be.PRS.3SG

*arrivé des fois / *probablement*
arrive.PST.PTCP perhaps/*probably
'I didn't understand if perhaps/*probably he arrived'

If such configurations crop up in speech, they seem invariably linked to deviations from the standard, and can be found for instance in children's productions. In the example below, Jules clearly does not master the epistemic content of *à mon avis* 'in my opinion'.

French

(82)

Je ne sais pas si maman sera rentrée à mon avis
I NEG know.PRS.1SG NEG COMP mommy be.FUT.3SG return.PST.F.SG **in my opinion**
'I don't know if mommy will already be home, in my opinion (sic)'
(Jules, 2 years and 10 months)

These results plead for considering Romance 'if' complementizers as NEF markers, i.e. markers of non-exclusion of factuality.

4.4. The value of 'that' complements

We have seen above that the class of predicates taking 'that' complements is quite heterogeneous (it includes many perception, cognition and utterance predicates), and seems not to be specialized for a particular modal meaning.

'that' complements can indeed either represent factual situations endowed with a specific spatio-temporal and referential anchoring:

Italian

(83) *Ho visto che c'era un tipo che rideva*
'I saw that there was (be.IND.PST.3SG) a guy who laughed'
(> a particular person)

or non-factual situations, i.e., situations that can be interpreted as either referring or non-referring (see above)

Italian

(84) *Vorrei che ci fosse un tipo che ride*
'I wish there were a guy who laughs'
(> either a particular person or anybody)

In this sense we might agree with Nordström & Boye's (this volume) claim that 'that' complementizers are epistemically neutral (this volume).

Nevertheless, we think that, at least for Romance languages, this first generalization needs to be refined. It is well known in fact that subjunctive and indicative moods may alternate in Romance ‘that’ complements. For example, the verbs ‘want’, ‘be strange’, ‘regret’ tend to select the subjunctive mood:

French

- (85) *Je voudrais qu’il vienne*
‘I wish he would come’ (come.SBJV.PRS.3SG)
- (86) *C’est bizarre qu’il ne vienne pas*
‘It is strange that he does not come’ (come.SBJV.PRS.3SG)
- (87) *Je regrette qu’il ne soit pas venu*
‘I regret he has not come’ (be. SBJV.PRS.3SG come.PTCP)

Whereas the verbs ‘say’, ‘see’, ‘discover’ tend to select the indicative mood:

French

- (88) *Je vois qu’il est venu*
‘I see that he has come’ (be. IND.PRS.3SG come.PTCP)
- (89) *Je découvre qu’il est venu*
‘I discover that he has come’ (be. IND.PRS.3SG come.PTCP)
- (90) *J’ai dit qu’il est venu*
‘I said that he has come’ (be. IND.PRS.3SG come.PTCP)

This opposition has often been interpreted as an opposition in modality. Hooper (1975), Klein (1975), and Bybee & Terrell (1990) have proposed that the selection of mood is related to a speech-act opposition: according to these authors, asserted situations are in the indicative, while non-asserted situations are in the subjunctive. Palmer (1986) claims that the opposition between indicative and subjunctive has to do with the degree of belief being conveyed: strongly believed situations are in the indicative, weakly believed situations are in the subjunctive. Farkas (1992) and Giannakidou (1994) have argued that the indicative is selected by veridical verbs, i.e., verbs that allow the inference that their complements are taken to be true at least by one epistemic agent (the speaker, or the subject of the main verb); whereas the subjunctive is selected by non-veridical verbs. All these interpretations have advantages as well as shortcomings, as thoroughly illustrated by Marques (2004; 2009). Marques has proposed one of the most recent and convincing interpretations for this opposition. He claims that, at least for Western Romance, the subjunctive is the unmarked choice for the great majority of ‘that’ complements. We do find, in fact, the subjunctive both within the scope of non-factual verbal predicates,⁷

Italian

⁷ Note that here we are employing the term “factive” in Kiparsky and Kiparsy (1971) and Karttunen’s (1971) sense: a predicate such as regret does not only cast reality over its complement (it is not merely factual) but it also presupposes it as true (it is truly factive).

(91)

Voglio che sia bello
want.PRS.1SG that be.SBJ.PRS.3SG beautiful.M
'I want him to be beautiful'

and within the scope of factual predicates such as regret:

Italian

(92)

Mi dispiace che sia bello
me.DAT displease.PRS.1SG that be.SBJ.PRS.3SG beautiful.M
'I am sorry he is beautiful'

Besides, we find the subjunctive within the scope of both assertive and non-assertive (presuppositional) verbs⁸:

Italian

(93)

Spero che sia bello
hope.PRS.1SG that be.SBJ.PRS.3SG beautiful.M
'I hope he is beautiful'

(94)

Sono felice che sia bello
be.PRS.1SG happy that be.SBJ.PRS.3SG beautiful.M
'I'm happy he is beautiful'

On the contrary, the indicative appears in 'that' complements only in one specific semantic situation: when main predicates "express someone's [certainty]⁹ concerning the truth of their complement proposition" (Marques 2004). Refining Marques's proposal and combining it with Bybee and Terrel's (1990), we claim that such a certainty should be asserted, and not merely presupposed, by the speaker (or by the subject of the sentence). In order to clarify the difference between factual complements, on the one hand, and complements asserted as true, on the other, let us contrast (95) and (96):

Italian

(95)

È bello che sia venuto
'It is great that he has come'

(96)

È chiaro che è venuto
'It is clear that he has come'

⁸ By "assertive" verbs we intend, following Simons (2007), verbs that have an asserted complement within their scope.

⁹ Marques (2004) writes "express someone's belief" rather than certainty, but we prefer to rephrase his formulation in order to avoid ambiguity.

The complements in (95) and (96) are both presented as true, they are both factual. But while in (95) the truth is presupposed, in (96) the truth of the complement is asserted. We can test this by showing that while in (95) the challengeable part of the sentence is the main predicate (97), in (96) the complement is also negotiable (98):

Italian

(97)

A: *È bello che sia venuto*
 ‘It is great that he has come’ (be.SBJV.PRS.3SG come.PTCP)

B: *Non sono d’accordo (non è bello)*
 ‘I don’t agree (this is not great)’

(98)

A: *È chiaro che è venuto*
 ‘It is clear that he has come’ (be.IND.PRS.3SG come.PTCP)

B: *ma non mi pare proprio* (che è chiaro but also che è venuto)
 ‘but I really don’t think so’

Interestingly, while predicates taking subjunctive ‘that’ complements cannot have a parenthetical use, predicates taking indicative ‘that’ complements can have parenthetical uses:

Italian

(99) *Ho visto/ ho scoperto/ ho detto che è rientrato*
 ‘I have seen/ discovered/ said that he has come back’

(100) *È rientrato, ho visto*
 ‘He has come back, I have seen’

(101) *È rientrato, ho scoperto*
 ‘He has come back, I discovered’

(102) *È rientrato, ho detto*
 ‘He has come back, I said’

(103) *Vorrei/ era strano/ mi dispiaceva che fosse rientrato*
 ‘I wish/it was strange/ I regretted that he had come back’

(104) **È rientrato, vorrei*
 ‘He has come back, I wish’

(105) **È rientrato, era strano*
 ‘He has come back, it was strange’

(106) **È rientrato, mi dispiaceva*
 ‘He has come back, I regretted’

The parenthetical use of a predicate entails in fact that it loses illocutionary autonomy and becomes illocutionarily dependent on its complements (Kahane & Pietrandrea 2012):

only asserted (and not presupposed) complements can escape the illocutionary dependency on the main predicate and fulfill the function of nucleus of the illocution (see also Simons 2007, for a similar analysis on asserted factual complements).

We therefore claim that, in spite of the fact that ‘that’ complements seem not be marked for epistemicity, mood selection in ‘that’ complements nevertheless leads to a distinction between unmarked ‘that’ complements and ‘that’ complements asserted as believed by the speaker: while the former select the subjunctive mood, the latter only select the indicative mood. Table 5 below illustrates this generalization:

Table 5: General distribution of functions between that + subjunctive and that + indicative in Romance

	That + subjunctive	That + indicative
Illocutionary status of the complement	+/- <i>asserted</i>	+ <i>asserted</i>
Factual status of the complement	+/- <i>factual</i>	+ <i>factual</i>

The analysis of corpus data, though, shows that the picture is still a little bit more complex. Table 6 illustrates the uses of subjunctive and indicative in Western Romance languages. It is clear that the choice of mood is quite oscillating in language use and a diachronic generalization of the indicative seems at play. This situation is quite advanced in French, where the indicative was extended to non-factual asserted complements such as the complements of verbs such as ‘think’ and ‘hope’. We attempt an explanation of this divergence in Section 6.2.

Table 6: Verb types and mood of the verb in the subordinate clause (based on corpus data and speaker judgments)

Verb type	French	Italian	Sardinian	Portuguese	Spanish
<i>I say / tell / see / discovered</i>	(almost always) Indicative				
<i>I hope / think</i>	<i>both (more or less equally) available</i>				
<i>I would like / wish / fear; It's possible / strange</i>	(almost always) Subjunctive				

All in all, while it is clear that the ‘that’ complementizer does not have any modal value by itself, it is possible to identify for some uses, in most Romance languages, a modal value encoded by the construction [‘that’+indicative], which conveys asserted belief. From a diachronic point of view, though, there seems to be a trend toward the loss of this modal value, in almost all Western Romance languages. This tendency is at its strongest in French.

4.5. Distribution of ‘if’ and ‘that’ complementizers in Romance: summing up

To sum up what we have discussed in the previous sections, we may say that Romance ‘if’ and ‘that’ complements do contrast as to their modal value.

- ‘if’ complements are always non-factual, namely they represent a specific type of non-factual situations, i.e. non-exclusion of factuality situations. These complements are presented as but one of a set of mutually exclusive options, and speakers do not qualify their epistemic commitment toward the

situation, suspending instead their own commitment. In this sense, we claim that non-exclusion of factuality situations depicted by ‘if’ complements are not modally qualified. For the same reason, we also reject, for Romance, the characterization of ‘if’ complements as depicting dubitative situations.

- ‘that’ complements never represent non-exclusion of factuality situations. They may depict either factual or non-factual situations.

It may be roughly proposed that mood selection plays a role in the modality of ‘that’ complements, in that indicative ‘that’ complements tend to express a particular type of modal situation, i.e. a situation asserted as believed by the speaker (or another epistemic agent). Nevertheless, this generalization has to be qualified in light of language use, which shows that the indicative might be drifting toward a generalization of its uses, in the whole Romance area and especially in French.

These general considerations, however, do not mean that ‘if’ and ‘that’ complements behave exactly in the same way across Romance languages. There are a number of small differences between languages that we cannot comment on here; we focus in the next section on two specific cases of divergence, which are mostly linked to areal phenomena.

5. The boundaries of the Romance world: The case of Romanian and French

As we noted above, for many verbs introducing complement clauses all Romance languages apply the same pattern, but for other classes of verbs there are slight differences, with either Romanian only or both French and Romanian standing apart from the others – albeit always with distinct phenomena in the two languages. In the two following sections, we focus on differences that can be at least partly explained by taking into account language-external considerations such as the very strong influence of Germanic languages on French, and of Balkan features on Romanian.

5.1. Romanian

The following figures illustrate the distribution of the complementizers *că*, *să* and *dacă* in Romanian. The most striking difference with respect to Western Romance languages is the existence of three complementizers: the first ‘that’ complementizer, *că*, is selected by perception, cognition and utterance predicates in a similar way to other Romance languages. The ‘if’ complementizer, *dacă*, is selected by uncertainty predicates and, just like ‘if’ complementizers in Western Romance, it projects a non-exclusion of factuality meaning on its complements. The second ‘that’ complementizer, *să*, is selected by predicational predicates, i.e. predicates which encode “either inception, termination, persistence, success, failure, attempt, intent, obligation or ability – by the subject of the main clause – to perform the action or be in the state that is depicted in the complement clause” (Givón 2001: 150).

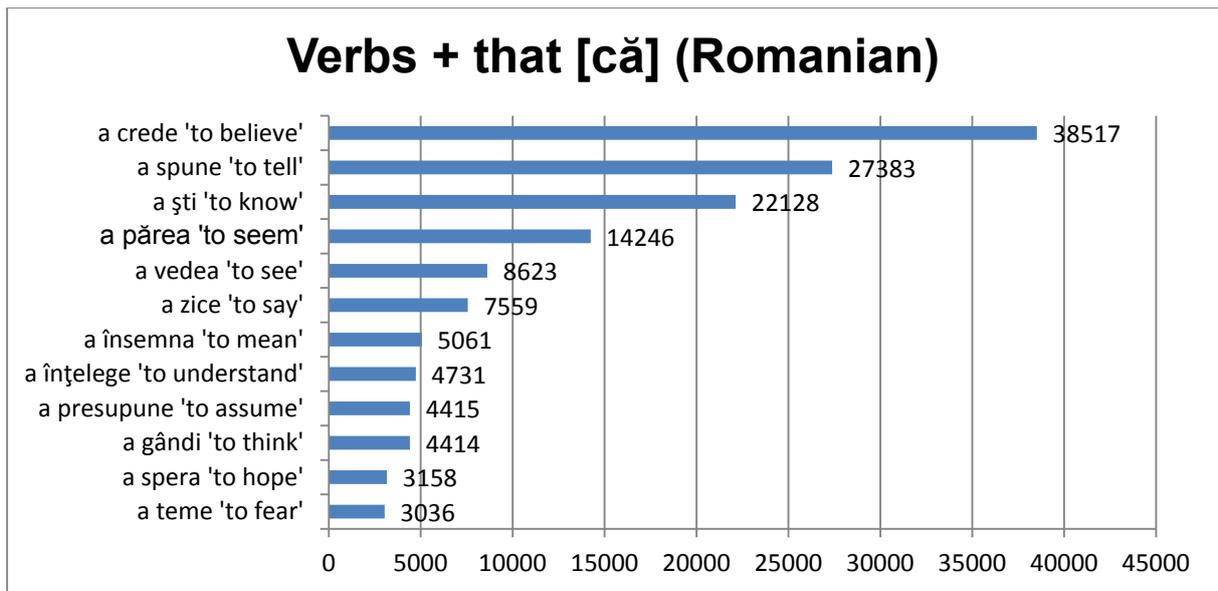


Figure 9: Most frequent verbs introducing complement clauses with ‘that₁’ (că) in Romanian

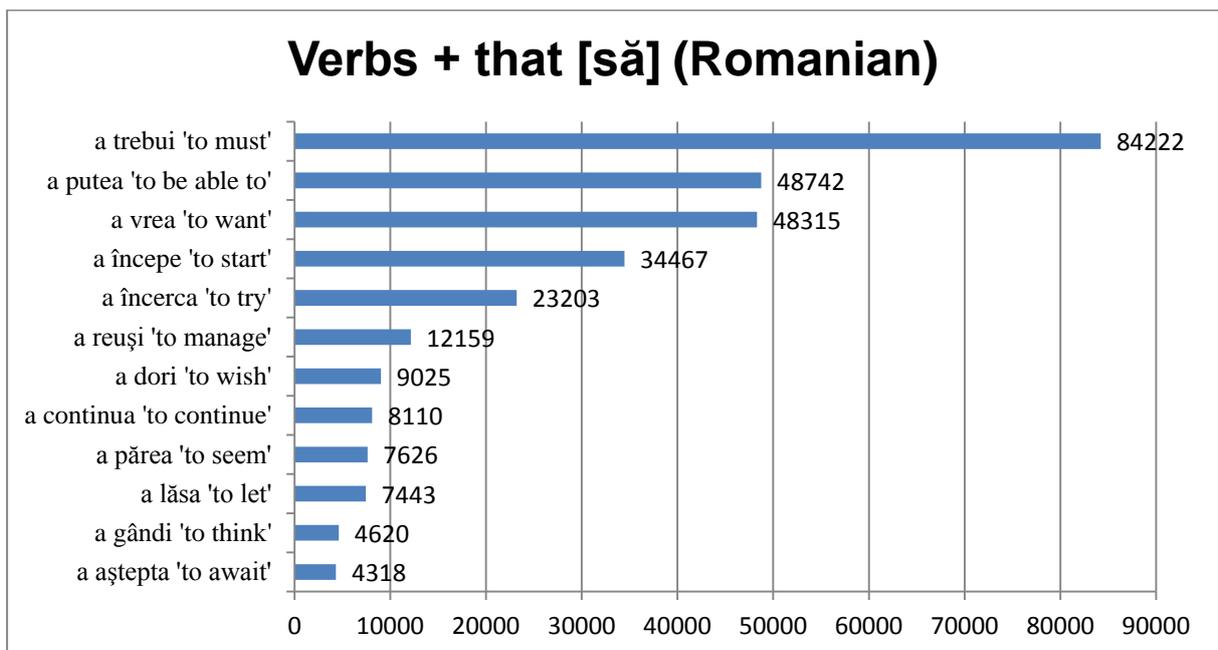


Figure 10: Most frequent verbs introducing complement clauses with ‘that₂’ (să) in Romanian

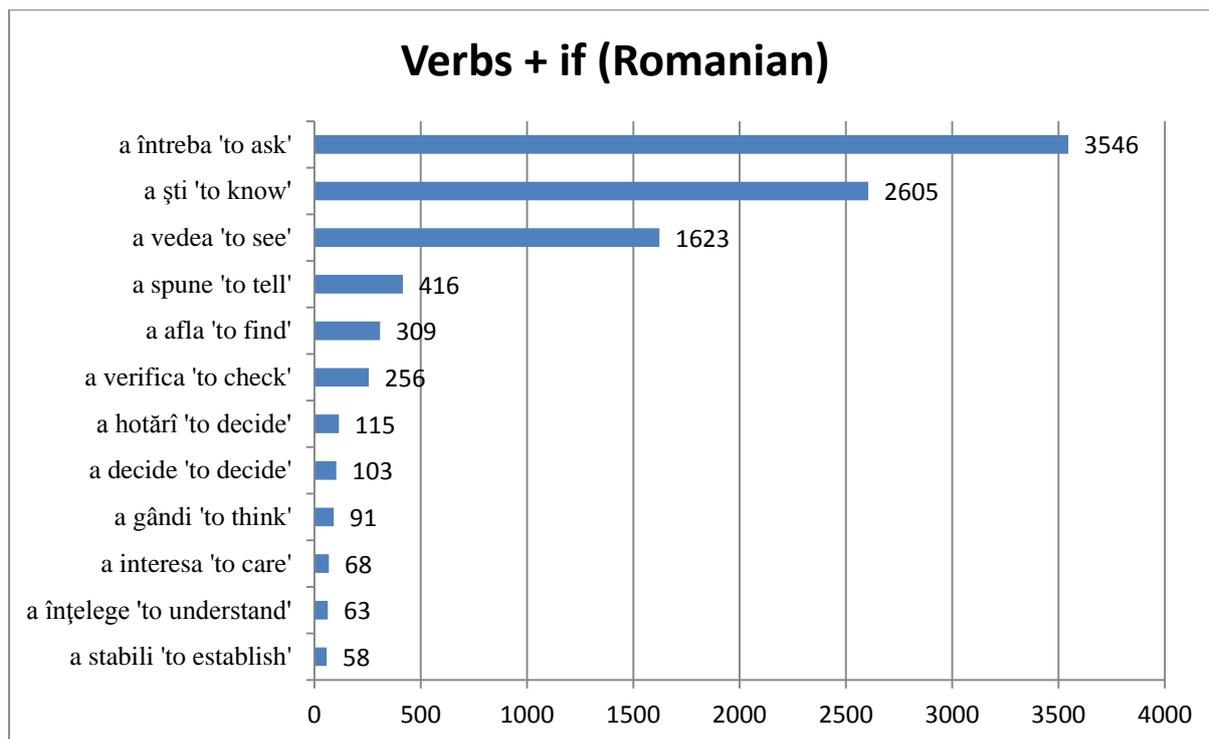


Figure 11: Most frequent verbs introducing complement clauses with ‘if’ (*dacă*) in Romanian

The differences between Romanian and other Romance languages are not haphazard, but linked to other language-internal features – themselves linked to language-external considerations. As we have seen, the *să* complementizer is used to introduce predicational complements; in these contexts other Romance languages use infinitive clauses (107), although the subjunctive is also possible (108):

French
(107)

Marie a demandé à Jean de lui écrire
 Marie have.PRS.3SG ask.PST.PTCP to Jean COMP DAT.PRO write.INF
 ‘Marie asked John to write to her’

(108)

Marie a demandé à Jean qu’ il lui écrive
 Marie have.PRS.3SG ask.PST.PTCP to Jean COMP he DAT.PRO write.SBJV.PRS.3SG
 ‘Marie asked John to write to her’

This specificity of Romanian is linked to a Balkan feature, the weakness of infinitives, i.e. their highly ‘nominal’ behavior. In Romanian, such utterances are expressed with the subjunctive (i.e. with *să*), to the exclusion of the infinitive (109):

Romanian

(109)

Maria i-a cerut lui Ion să-i scrie

Maria CL-have.PRS.3SG ask.PST.PTCP DAT.CL Ion COMP-CL write.SBJV.PRS.3SG
 ‘Maria asked Ion to write to her’ (Farkas 2003: 2).

Concerning the distribution of indicative and subjunctive in Romanian complements, it should be said that Romanian has a discontinuous subjunctive; the complementizer *să* is thus part of the subjunctive marking (Giannakidou 2009). This is clearly a Balkan feature, present in other languages of the region, including ‘outliers’ like Serbian (Bulatović 2008). As Table 7 shows, Romanian tends, like French, to generalize the use of the indicative to some optative and epistemic predicates (‘hope’, ‘think’, ‘be certain’); the *să* complementizer is also used with the subjunctive, for the complements of volitive predicates.

Table 7: Verb types, complementizer and mood of the verb in the subordinate clause, in Romanian (based on corpus data and speaker judgments)

Verb type	Romanian
1. <i>I know / say / see / tell; I discovered</i>	<i>că</i> + Indicative (also <i>dacă</i>)
2. <i>I hope; I think; It is certain</i>	<i>că</i> + Indicative / <i>să</i> + Subjunctive
3. <i>I fear; It's possible / strange</i>	<i>că</i> + Indicative / <i>să</i> + Subjunctive
4. <i>I would like / wish</i>	<i>să</i> + Subjunctive

5.2. French

As we noted for Romanian, the differences between French and other Romance languages are linked to other language-internal features, which are themselves linked to language-external considerations. The main difference observed is the use of the indicative for non-asserted ‘that’ complements such as those of ‘think’ and ‘hope’. It should be pointed out that this is a recent development: for instance, the verbs *penser* ‘to think’ and *espérer* ‘to hope’ could be followed by a ‘that’ clause in the subjunctive in Old and Middle French, a possibility which gradually disappeared during the 17th and 18th centuries.

This evolution can be explained by the diachronic weakening of the morphological distinction between indicative and subjunctive in French regular verbs. In other Romance languages the choice between subjunctive and indicative remains meaningful in a variety of contexts: the subjunctive mood still has a semantic value. On the contrary, in French, the choice of subjunctive vs. indicative often results from grammatical constraints: for French speakers, more than for instance for speakers of Italian or Spanish, there is an important loss of paradigmatic variability in the use of the subjunctive, as noted by De Mulder and Lamiroy (2012). In this perspective, the subjunctive can be said to have lost (much of) its semantic value in French; in other words, it is more grammaticalized than in ‘Central Romance’ – in Lehmann’s terms, the loss of (here paradigmatic) “variability” is indeed one of the “three main aspects of grammaticalization” (Lehmann 1995: 123).

The fact that the subjunctive, or more precisely the use of the subjunctive in complementation, can be seen as more grammaticalized in French can be brought back to a tendency for French to have, for many constructions, more grammaticalized variants than other Romance languages. This is the case for instance of determiners, with three paradigms in French and only two in almost all other Romance varieties (Carlier 2007): French has a partitive article which is absent or not completely grammaticalized elsewhere (in Lehmann’s terms, the degree of grammaticalization is higher in French on account of paradigmaticity, Lehmann 1995: 123). Similar accounts can be given for demonstratives (Marchello-Nizia

2006: 178), auxiliaries and tense, aspect and modality markers (Lamiroy 1999; 2001; 2003; 2011; De Mulder 2001; De Mulder & Lamiroy 2012); this analysis could even be extended to other areas of language, such as conjunctions, prepositions and discourse markers (Fagard 2010; 2011; Fagard & Mardale 2012).

Of course, this is a global trend and there are counterexamples. However, it is quite pervasive, and could be accounted for by language-external considerations, namely the influence of Germanic languages on French (though not, of course, in a direct way). There is some evidence of a more important impact on the linguistic system at large, and some authors claim that the Germanic superstratum triggered a global phonetic shift in French (Marchello-Nizia 1995: 177). This has been reported quite early on, and Pope (1934: 15) for instance writes that “[t]he Frankish system of accentuation was a strong expiratory one and it was in the intensifying of the weak Latin tonic stress that the Germanic speech-habits, and in particular the Frankish, exercised their strongest influence in pronunciation”, with sufficient phonetic consequences (such as diphthongization, reduction of unstressed vowels, closing of some secondary stressed vowels) to explain that these could have resulted in more extended morpho-syntactic changes (see also Von Wartburg [1934] 1965; Zink 1986; for a contradictory analysis, see Noske 2009). This applies to the issue at hand, since the partial loss of contrast between indicative and subjunctive in French is not only semantic, but also formal: in many instances, French verbs have identical forms for both moods while other Romance languages present a formal contrast. An example among others is the form of the verb ‘eat’: French *mange* (eat.PRS.IND/SBJV.3SG) vs Italian *mangia* (eat.PRS.IND.3SG)/*mangi* (eat.PRS.SBJV.3SG), Catalan *menja* (eat.PRS.IND.3SG)/*mengi* (eat.PRS.SUBJ.3SG), Portuguese and Spanish *come* (eat.PRS.IND.3SG)/*coma* (eat.PRS.SBJV.3SG), Romanian *mănâncă* (eat.PRS.IND.3SG)/*mănânce* (eat.PRS.SBJV.3SG), Sardinian *mandigat* (eat.PRS.IND.3SG)/*mandighet* (eat.PRS.SBJV.3SG), etc.

5.3. A synopsis

To sum up, comparing verb types across Romance languages and their behavior (complementizer + mood) when introducing complement clauses, we identified the following cases:

- Similar pattern across all Romance languages:
 1. Perception, cognition and utterance predicates trigger the use of ‘that’ + indicative, whatever the language. Uncertainty predicates trigger the use of ‘if’ + indicative; the subjunctive is also possible.
- Specific pattern for French and Romanian:
 2. Optative verbs and some epistemic verbs (‘hope’, ‘think’, ‘be certain’) trigger the use of ‘that’ + subjunctive or indicative in most Romance languages, but in French and Romanian the subjunctive is excluded.
- Specific pattern for Romanian only:
 3. Some epistemic verbs as well as verbs of fearing (‘be possible’, ‘be strange’, ‘fear’) trigger the use of ‘that’ + subjunctive in all Romance languages, except Romanian where *să* + subjunctive alternates with *că* + indicative (see Tables 6 and 7).
 4. Verbs of volition trigger the use of ‘that’ + subjunctive in all Romance languages, but in Romanian the complementizer is *să* (also + subjunctive) (see Tables 6 and 7).

Table 8 below sums up these findings. Note that some of these are a tendency rather than a rule, and that there is some variation, both regional and diachronic. For instance, as we saw in section Section 5.2., in French verbs of type 2 could introduce complements in the subjunctive until the 18th century.

Table 8: The use of ‘that’ with indicative and subjunctive and verb types in Romanian and French vs. other Romance languages (Italian, Sardinian, Portuguese, Spanish) (based on corpus data and speaker judgments)

Verb type	‘Central’ Romance	French	Romanian
1. <i>I know / say / see / tell; I discovered</i>	<i>That + Indicative (also If for the <u>first group</u>)</i>	<i>That + Indicative (also If for the <u>first group</u>)</i>	<i>Că + Indicative (also <u>dacă</u> for the <u>first group</u>)</i>
2. <i>I hope; I think; It is certain</i>			
3. <i>I fear; It’s possible / strange</i>	<i>That + Subjunctive</i>	<i>That + Subjunctive</i>	<i>Că + Indicative / Să + Subjunctive</i>
4. <i>I would like / wish</i>			<i>Să + Subjunctive</i>

6. Origins of Romance complementizers

All standard Romance ‘if’ forms (except Romanian *dacă*, see below) have the same origin – the ‘if’ complementizer in Latin, *si* (see i.a. De Dardel 1978). And, as we have seen, they retain its conditional meaning. Note that other constructions can be used to express condition, such as ‘when (that)’, or even paratactic constructions. Mazzoleni (1997) reports this possibility for various Italian dialects, but it is also found elsewhere (see, for French, Corminboeuf 2008).

Most ‘that’ forms have the same origin, Latin *quem* ‘whom, which, that’ (relativizer, in the accusative), but not Sardinian *ki*, which comes from Latin *qui* ‘who, which, that’ (also a relativizer, in the nominative). Concerning Romanian, the first ‘that’ complementizer, *că*, comes from Latin *quod* ‘that’ (already a complementizer). The second ‘that’ complementizer, *să*, with more limited uses, comes from the Latin ‘if’ complementizer (*si*) mentioned above, which remained an ‘if’ complementizer in all other Romance languages. Actually, it was still used this way in the first literary texts of Old Romanian, but was progressively replaced by *dacă*, which is now the ‘if’ complementizer in Romanian, and is a compound of the first ‘that’ complementizer and the adposition *de* ‘of, from’ (from Latin *dē* ‘from (above)’).

If Romanian, as we saw above, can be said to diverge from other Romance languages in many respects, there are two main historical reasons. The first is that Romanian has been cut off from the rest of the Romance world for centuries, and thus many innovations which are found in almost all Romance languages did not reach Romanian. This might account for the fact that Romanian kept *quod*, a Late Latin complementizer which was replaced everywhere by the relativizer *qui*, *quem* (Table 3). The second is that Romanian has been heavily influenced by other Balkan languages. This might account for the presence of two paradigms of ‘that’ complementizers, possibly linked to the fragility of the infinitive and the frequency of use of the subjunctive (of which *să* is considered a part, see Giannakidou 2009). In most Romance languages, ‘how’ and ‘what’ complementizers come respectively from Late Latin *quo modo* ‘in what way’ and *quem* (already seen above); Sardinian *ita* comes from the homonymous Latin adverb meaning ‘thus’.

7. Conclusion

In almost all Romance languages, canonical complementizers are restricted to two series, one close to the ‘that’ complementizer in English, the other to the ‘if’ complementizer. Romanian presents a different configuration, with three canonical complementizers, two ‘that’ complementizers and one ‘if’ complementizer.

Diachronically speaking, most ‘that’ forms have the same origin, Latin *quem* ‘whom, which, that’ (relativizer, in the accusative). Almost all ‘if’ forms have the same origin – the ‘if’ complementizer in Latin, *si*.

Canonical complementizers can have other uses in Romance besides complementation. ‘If’ forms are used in hypothetical constructions: this adverbializer use is found in all Romance languages; ‘that’ forms have a relativizer function in all Romance languages except Romanian and they have a quotative function in French. Both ‘if’ and ‘that’ forms can introduce non-assertive main sentences such as jussives, interrogatives and optatives. The combination of complementizers with other subordination markers is a marginal phenomenon; however, it is found in all Romance languages, and could be indicative of a global trend toward a use of the ‘that’ complementizer as a general subordinator.

From a functional standpoint, Romance complementizers have one main function and an accessory function. The main function is marking the argument predication as illocutionarily dependent on the main predicate (as is proven by the semantic effect of complementizer omission); the accessory function is qualifying the factuality status of the dependent predication. More specifically, ‘that’ complements, selected by perception, cognition and utterance predicates, may be either factual or non-factual, whereas ‘if’ complements, selected by uncertainty predicates, are always non-factual.

We proposed that ‘if’ complements convey a particular type of non-factuality, i.e., non-exclusion of factuality. Speakers present the predication introduced by ‘if’ as but an eventuality, suspending rather than qualifying their commitment toward its truth. In this sense, we maintained that ‘if’ complements do not convey epistemic modality.

Concerning ‘that’ complements, we argued that, at least for Central Romance, mood selection seems to play a role in the modality of ‘that’ complements, in that indicative ‘that’ complements tend to express a specific type of modal situation, i.e. a situation asserted as believed by the speaker (or another epistemic agent), whereas subjunctive ‘that’ complements are unmarked.

Nevertheless, the observation of language use shows that the indicative may be following a trend toward a generalization of its uses in the whole Romance area; this phenomenon is already well under way in French.

Indeed, some areal considerations have led us to refine this general picture. The most important deviation from this generalization is the existence of three complementizers in Romanian: the first ‘that’ complementizer, *că*, is selected by perception, cognition and utterance predicates; the ‘if’ complementizer, *dacă*, is selected by uncertainty predicates and, just like ‘if’ complementizers in Western Romance, it projects a non-exclusion of factuality meaning on its complements; the second ‘that’ complementizer, *să*, is selected by modality predicates. Modality predicates select infinitive clauses in other Romance languages; the specificity of Romanian is linked to a Balkan feature, the weakness of infinitives.

Concerning French, we have seen that this language seems to be more advanced in the generalization of the usage of indicative in complement clauses. We explained this change with the diachronic weakening of the morphological distinction between indicative and subjective in French regular verbs, which might in turn be partly due to the influence of

Germanic languages on French and the tendency of French to have grammatical paradigms which underwent more renewal and differentiation from Latin than in other Romance languages, as has been claimed by several authors.

References

- Abeillé, Anne, Danièle Godard & Annie Delaveau. In press. *Grande grammaire du français*. Paris: Actes sud.
- Bat-Zeev Shyldkrot, Hava & Suzanne Kemmer. 1988. Le développement sémantique des conjonctions en français : quelques concepts généraux. *Revue Romane* 23(1). 9–20.
- Bellert, Irena. 1977. On Semantic and Distributional Properties of Sentential Adverbs. *Linguistic Inquiry* 8(2). 337–351.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman grammar of spoken and written English*. Harlow, United Kingdom: Pearson Education.
- Blanche-Benveniste, Claire & Dominique Willems. 2007. Un nouveau regard sur les verbes faibles. *Bulletin de la société linguistique de Paris* 102(1). 217–254.
- Blanche-Benveniste, Claire. 1989. Constructions verbales en incise et rection faible des verbes. *Recherches sur le français parlé* 9. 53–73.
- Blanche-Benveniste, Claire. 1997. *Approches de la langue parlée en français*. Paris: Ophrys.
- Boye, Kasper & Peter Harder. 2009. Evidentiality: Linguistic categories and grammaticalization. *Functions of Language* 16(1). 9–43.
- Briscoe, Ted. 1994. Prospects for practical parsing of unrestricted text: Robust statistical parsing techniques. In Nelleke Oostdijk & Pieter de Haan (eds.), *Corpus-based research into language*, 67–95. Amsterdam: Rodopi.
- Bulatović, Andjelka Gina. 2008. *Modality, Futurity, and Temporal Dependency: the Semantics of the Serbian Perfective Nonpast and Future 2*. Chicago: University of Chicago dissertation.
- Bybee, Joan L. & Paul J. Hopper (eds.). 2001. *Frequency and the emergence of linguistic structure*. Amsterdam & Philadelphia: John Benjamins.
- Bybee, Joan L. & Tracy D. Terrell. 1990. Análisis semántico del modo en español. In Ignacio Bosque (ed.), *Indicativo y subjuntivo*, 145–163. Madrid: Taurus Universitaria.
- Carlier, Anne. 2007. From Preposition to Article. The grammaticalization of the French partitive. *Studies in Language* 31(1). 1–49.
- Corminboeuf, Gilles. 2008. ‘Tu m’embrasses encore, et c’est mon pied dans les pompons!’ Comment construit-on le sens ? *Discours* 3. <http://discours.revues.org/4173> (accessed 1 September 2013).
- Cornillie, Bert. 2010. On conceptual semantics and discourse functions. The case of Spanish modal adverbs in informal conversation. *Review of Cognitive Linguistics* 8(2). 300–320.
- Craig, Colette Grinevald. 1977. *The Structure of Jacaltec*. Austin: University of Texas Press.
- Cristofaro, Sonia. 2005. *Subordination*. Cambridge: Cambridge University Press.
- Dancygier, Barbara & Eve Sweetser. 2005. *Mental Spaces in Grammar: Conditional Constructions*. Cambridge: Cambridge University Press.
- Dascălu-Jinga, Laurenția. 2002. *Corpus de română vorbită (CORV) : eşantioane* [Corpus of spoken Romanian (CORV): samples]. București: Editura Oscar Print.
- Davies, Mark & Michael J. Ferreira. 2006. *Corpus do Português: 45 million words, 1300s-1900s* [Portuguese Corpus]. Available online at <http://www.corpusdoportugues.org>.
- Davies, Mark. 2002. *Corpus del Español: 100 million words, 1200s-1900s*. Available online at <http://www.corpusdelespanol.org>.

- De Dardel, Robert. 1978. La Forme de la conjonction latine *si* en roman. *Zeitschrift für romanische Philologie* 94. 257–265.
- De Mulder, Walter & Béatrice Lamiroy. 2012. Gradualness of grammaticalization in Romance. The position of French, Spanish and Italian. In Kristin Davidse, Tine Breban, Lieselotte Brems & Tanja Mortelmans (eds.), *Grammaticalization and Language Change*, 199–226. Amsterdam & Philadelphia: John Benjamins.
- De Mulder, Walter. 2001. La Linguistique Diachronique : grammaticalisation et sémantique du prototype. *Langages* 130. 8–32.
- Degand, Liesbeth, Bart Cornillie & Paola Pietrandrea. 2013. Modal Particles and Discourse Markers. Two sides of the same coin? Introduction. In Liesbeth Degand, Bart Cornillie & Paola Pietrandrea (eds.), *Modal Particles and Discourse Markers. Categorization and Description*, 1–18. Amsterdam & Philadelphia: John Benjamins.
- Dehé, Nicole & Yordanka Kavalova. 2006. The syntax, pragmatics, and prosody of parenthetical *what*. *English Language and Linguistics* 10. 289–320.
- Diewald, Gabriele & Marijana Kresić. 2010. Ein übereinzelsprachliches kontrastives Beschreibungsmodell für Partikelbedeutungen. *Linguistik online* 44. http://www.linguistik-online.de/44_10/diewaldKresic_a.html (accessed 2 July 2013).
- Diewald, Gabriele. 2013. “Same same but different” – Modalparticles, discourse markers and the art (and purpose) of categorization. In Liesbeth Degand, Bart Cornillie & Paola Pietrandrea (eds.), *Modal Particles and Discourse Markers. Categorization and Description*, 19–25. Amsterdam & Philadelphia: John Benjamins.
- Ditchy, Jay K. 1997. *Les acadiens louisianais et leur parler*. Montreal: Comeau & Nadeau.
- Elliott, Jennifer R. 2000. Realis and Irrealis: Forms and Concepts of the Grammaticalisation of Reality. *Linguistic Typology* 4. 55–90.
- Espinal, Theresa M. 1991. The representation of disjunct constituents. *Language* 67(4). 726–762.
- Fagard, Benjamin & Mardale Alexandru. 2012. The pace of grammaticalization and the evolution of prepositional systems: Data from Romance. *Folia Linguistica* 46(2). 303–340.
- Fagard, Benjamin. 2010. *É vida, olha...*: Imperatives as discourse markers and grammaticalization paths in Romance: A diachronic corpus study. *Languages in Contrast* 10(2). 245–267.
- Fagard, Benjamin. 2011. Conjunctions et grammaticalisation : le cas des langues romanes. *Mémoires de la Société de Linguistique de Paris* 19. 79–102.
- Farkas, Donka F. 1992. On the semantics of subjunctive complements. In Paul Hirschbühler & Konrad Koerner (eds.), *Romance Languages and Modern Linguistic Theory*, 71–104. Amsterdam & Philadelphia: John Benjamins.
- Farkas, Donka F. 2003. Notes on Mood. Presented at *ESSLLI, Conditional and Unconditional Modality Workshop*, Vienna, 2003.
- Frajzyngier, Zygmunt. 1995. A functional theory of complementizers. In Joan Bybee & Suzanne Fleischman (eds.), *Modality in Grammar and Discourse*, 473–502. Amsterdam & Philadelphia: John Benjamins.
- Giannakidou, Anastasia. 1994. The semantic licensing of NPIs and the Modern Greek subjunctive. *Language and Cognition. Yearbook of the Research Group for Theoretical and Experimental Linguistics*. 4. Groningen: University of Groningen.
- Giannakidou, Anastasia. 2009. The dependency of the subjunctive revisited: temporal semantics and polarity. *Lingua* 120. 1883–1908.
- Givón, Talmy. 1982. Evidentiality and Epistemic Space. *Studies in Language* 6(1). 23–45.
- Givón, Talmy. 2001 [1984]. *Syntax. A Functional-Typological Introduction*, vol. I. Amsterdam & Philadelphia: John Benjamins.

- Haegeman, Liliane. 1988. Parenthetical adverbials: the radical orphanage approach. In Shuki Chiba, Akira Ogawa, Yasuaki Fuiwara, Norio Yamada, Osamu Koma & Takao Yagi (eds.), *Aspects of Modern English Linguistics*, 232–254. Tokyo: Kaitakushi.
- Haspelmath, Martin & Ekkehard König. 1998. Concessive conditionals in the languages of Europe. In Johan van der Auwera (ed.), *Adverbial Constructions in the Languages of Europe*, 563–640. Berlin: Mouton de Gruyter.
- Herman, József. 1963. *La formation du système roman des conjonctions de subordination*. Berlin: Akademie-Verlag.
- Hooper, Joan B. 1975. On assertive predicates. In John P. Kimball (ed.), *Syntax and Semantics* 4, 91–124. New York: Academic Press.
- Ifantidou-Trouki, Elly. 1993. Sentential adverbs and relevance. *Lingua* 90. 69–90.
- Jakobson, Roman. 1957. Shifters, Verbal Categories, and the Russian Verb. *Russian Language Project*, Department of Slavic Languages and Literatures. Cambridge, MA: Harvard University Press.
- Jeanjaquet, Jules. 1894. *Recherches sur l'origine de la conjonction que et des formes romanes équivalentes*. Paris: H. Welter, Leipzig: Gustav Fock, Neuchâtel: Attinger.
- Kahane, Sylvain & Paola Pietrandrea. 2012. Les parenthétiques comme ‘Unités Illocutoires Associées’. Une perspective macrosyntaxique. *Linx* 61(2009). 49–70.
- Kahane, Sylvain. 2001. Grammaires de dépendance formelles et théorie Sens-Texte, Tutoriel. *Proceedings of TALN'2001*, Tours, France.
- Kayne, Richard S. 2005. Some notes on comparative syntax, with special reference to English and French. In Guglielmo Cinque & Richard S. Kayne (eds.), *The Oxford Handbook of Comparative Syntax*, 3–69. New York: Oxford University Press.
- Kilgarriff, Adam, Pavel Rychly, Pavel Smrz & D. Tugwell. 2004. The Sketch Engine. In Geoffrey Williams & Sandra Vessier (eds.), *Proceedings of the Eleventh Euralex Congress*, 105–116. Lorient: UBS Lorient.
- Kiparsky, Carol & Paul Kiparsky. 1971. Fact. In Danny D. Steinberg & Leon A. Jakobovitz (eds.), *Semantics: an Interdisciplinary Reader in Philosophy, Linguistics and Psychology*, 345–369. Cambridge: Cambridge University Press.
- Klein, Flora. 1975. Pragmatic constraints in distribution: the Spanish subjunctive. In *Papers from the 11th regional meeting*, 353–365. Chicago: Chicago Linguistic Society.
- Koch, Peter. 2002. Das Französische in typologischer und historisch-vergleichender Sicht. In Ingo Kolboom, Thomas Kotschi & Edward Reichel (eds.), *Handbuch Französisch*, 36–58. Berlin: Schmidt.
- Kuno, Susumu. 1973. *The Structure of Japanese Language*. Cambridge, MA & London: MIT Press.
- Lamiroy, Béatrice. 1999. Auxiliaires, langues romanes et grammaticalisation. *Langages* 135. 33–45.
- Lamiroy, Béatrice. 2001. Le syntagme prépositionnel en français et en espagnol: une question de grammaticalisation? *Langages* 143. 91–106.
- Lamiroy, Béatrice. 2003. The External Possession Structure and Grammaticalisation in Romance and Germanic languages. In Yves D'Hulst & Martine Coene (eds.), *From NP to DP. The expression of Possession in Noun Phrases*, Vol. II, 257–281. Amsterdam & Philadelphia: John Benjamins.
- Lamiroy, Béatrice. 2011. Degrés de grammaticalisation à travers les langues de même famille. *Bulletin de la Société de Linguistique de Paris* 19. 167–192.
- Le Goffic, Pierre. 1993. *Grammaire de la phrase française*. Paris: Hachette.
- Lehmann, Christian. 1995. *Thoughts on Grammaticalization*. München & Newcastle: Lincom Europa.

- Marchello-Nizia, Christiane, Bernard Combettes, Sophie Prévost & Tobias Scheer (eds.). In preparation. *Grande grammaire historique du français*.
- Marchello-Nizia, Christiane. 1995. *L'évolution du français : ordre des mots, démonstratifs, accent tonique*. Paris: Armand Colin.
- Marchello-Nizia, Christiane. 2006. *Grammaticalisation et changement linguistique*. Bruxelles: De Boeck Université.
- Marques, Rui. 2004. On the system of mood in European and Brazilian Portuguese. *Journal of Portuguese Linguistics* 3(1). 89–109.
- Marques, Rui. 2009. On the selection of mood in complement clauses. In Lotte Hogeweg, Helen de Hoop & Andrej Malchukov (eds.), *Cross-Linguistic Semantics of Tense, Aspect, and Modality*, 179–204. Amsterdam & Philadelphia: John Benjamins.
- Masini, Francesca & Paola Pietrandrea. 2010. Magari. *Cognitive Linguistics* 21(1). 75–121.
- Mauri, Caterina. 2008. The irrealty of alternatives: towards a typology of disjunction. *Studies in Language* 32(1). 22–55.
- Mazzoleni, Marco. 1997. The syntax of conditional sentences. In Martin Maiden & Mair Parry (eds.), *The Dialects of Italy*, 202–207. Abingdon & New York: Routledge.
- Mel'cuk, Igor. 1988. *Dependency syntax: Theory and practice*. Albany: SUNY Press.
- Mensching, Guido. 2004. *Einführung in die sardische Sprache*, 3rd edn. Bonn: Romanistischer Verlag.
- Mithun, Marianne. 1995. On the Relativity of Irreality. In Joan Bybee & Suzanne Fleischman (eds.), *Modality in Grammar and Discourse*, 367–388. Amsterdam & Philadelphia: John Benjamins.
- Noonan, Michael. 2007 [1985]. Complementation. In Timothy Shopen (ed.), *Language Typology and Syntactic Description*, vol 2, *Complex Constructions*, 2nd edn, 52–150. Cambridge: Cambridge University Press.
- Nordström, Jackie & Kasper Boye. this volume. Semantic functions of Germanic complementizers.
- Nordström, Jackie. 2010. *Modality and subordinators*. Amsterdam & Philadelphia: John Benjamins.
- Noske, Roland. 2009. Autonomous typological prosodic evolution versus the Germanic superstrate in diachronic French phonology. In Enoch Aboh, Elisabeth van der Linden, Josep Quer & Petra Sleeman (eds.), *Romance languages and Linguistic Theory. Selected papers from 'Going Romance' Amsterdam 2007*, 223–242. Amsterdam & Philadelphia: John Benjamins.
- Palmer, Frank R. 1987. *Mood and Modality*. Cambridge: Cambridge University Press.
- Pietrandrea, Paola. 2011. The conceptual structure of irrealty: a focus on non-exclusion-of-factuality as a conceptual and a linguistic category. *Language Sciences* 34. 184–199.
- Pope, Mildred. 1934. *From Latin to Modern French*. Manchester: Manchester University Press.
- Prandi, Michele. 2004. *The Building Blocks of Meaning*. Amsterdam & Philadelphia: John Benjamins.
- Ross, John Robert. 1973. Slifting. In Maurice Gross, Morris Halle & Marcel-Paul Schützenberger (eds.), *The Formal Analysis of Natural Language*, 133–169. Berlin: Mouton de Gruyter.
- Schelfhout, Carla, Peter-Arno Coppen & Nelleke Oostdijk. 2004. Finite Comment Clauses in Dutch: A Corpus-based Approach. *Journal of Germanic Linguistics* 16. 331–349.
- Schoonjans, Steven. 2013. Modal particles: Problems in defining a category. In Liesbeth Degand, Bart Cornillie & Paola Pietrandrea (eds.), *Modal Particles and Discourse Markers. Categorization and Description*, 133–162. Amsterdam & Philadelphia: John Benjamins.

- Simons, Mandy. 2007. Observations on embedding verbs, evidentiality, and presupposition. *Lingua* 117(6). 1034–1056.
- Squartini, Mario. 2013. From TAM to discourse: The role of information status in North-Western Italian *gia* ‘already’. In Liesbeth Degand, Bart Cornillie & Paola Pietrandrea (eds.), *Modal Particles and Discourse Markers. Categorization and Description*, 163–190. Amsterdam & Philadelphia: John Benjamins.
- Tesnière, Lucien. 1959. *Éléments de syntaxe structurale*. Paris: Klincksieck.
- Vendler, Zeno. 1967. *Linguistics in Philosophy*. Ithaca: Cornell University Press.
- Von Wartburg, Walther. 1965 [1934]. *Evolution et structure de la langue française*, 7th edn. Bern: Francke.
- Zink, Gaston. 1986. *Phonétique historique du français*. Paris: Presses Universitaires de France.

Corpora

- Romanian: CORV (Dascălu-Jinga 2002), Ruxândoiu (spoken 0.1 million words), Romanian web corpus (written, 44.7 million words).
- Spanish: CREA (Real Academia Española, <http://www.rae.es>), CORDE (Real Academia Española, <http://www.rae.es>), Corpus del Español (Davies 2002), esTenTen (spoken 7.9 million words, written 2.1 billion words).
- Portuguese: Corpus do Português (Davies & Ferreira 2006), ptTenTen11 (spoken 2.2 million words, written 2.8 billion words).
- Italian: Badip (De Mauro et al. 1993, accessed via <http://badip.uni-graz.at/>), itTenTen (spoken 0.5 million words, written 2.6 billion words).
- French: Valibel Valibel (Groupe de Recherches Valibel – Discours & Variation, Université catholique de Louvain), PFC (<http://www.projet-pfc.net/>), Clapi (<http://clapi.univ-lyon2.fr>), Frantext (Atilf, CNRS, www.frantext.fr), frTenTen (spoken 1.6 million words, written 10.7 billion words).
- Note that the Romanian web corpus and the “TenTen” corpora were accessed via the Sketch Engine query system (<http://www.sketchengine.co.uk>; Kilgariff et al. 2004).

Grammars and dictionaries (selection)

- Anglade, Joseph. 1921. *Grammaire de l’ancien provençal*. Paris: Klincksieck.
- Avram, Mioara. 1960. *Evoluția subordonării circumstanțiale cu elemente conjuncționale în limba română* [Evolution of circumstantial clauses with conjunctive elements in Romanian]. Bucharest: Editura Academiei Republicii Populare Române.
- Badia Margarit, Antoni. 1962. *Gramática catalana*. Madrid: Gredos.
- Blumenthal, Peter & Achim Stein. 2002. *Tobler-Lommatzsch, Elektronische Ausgabe*. Stuttgart: Steiner.
- Buridant, Claude. 2000. *Grammaire nouvelle de l’ancien français*. Paris: Sedes.
- De Lima Coutinho, Ismael. 1977. *Gramática histórica* [Historical grammar]. Rio de Janeiro: Ao Livro Técnico.
- De Mauro, Tullio, Federico Mancini, Massimo Vedovelli & Miriam Voghera. 1993. *Lessico di frequenza dell’italiano parlato*. Milano: Etaslibri.
- De Rochegeude, Henri Pascal. 1819. *Essai d’un glossaire occitanien pour servir à l’intelligence des poésies des troubadours*. Toulouse: Benichet Cadet.
- Densusianu, Ovid. 1938. *Histoire de la langue roumaine*. Paris: Leroux.
- Diccionario Medieval Español*. 1986. Salamanca: Universidad Pontificia de Salamanca.
- Dicționarul explicativ al limbii române* [Explanatory dictionary of Romanian](DEX). 1975. Bucharest: Academia Republicii Socialiste România.

- Ditzionàriu de sa limba e de sa cultura sarda* [Dictionary of the Sardinian language and culture]. 2000. Cagliari: Condaghes.
- Echenique Elizondo, María Teresa & María José Martínez Alcalde. 2003 [2000]. *Diacronía y gramática histórica de la lengua española*, 2nd edn. Valencia: Tirant Lo Blanch.
- Ernout, Alfred & François Thomas. 1964. *Syntaxe latine*. Paris: Klincksieck.
- Etienne, Eugène. 1980 [1895]. *Essai de grammaire de l'ancien français (IX^e-XIV^e siècles)*. Genève: Slatkine Reprints [1st edition Nancy, Paris: Berger-Levrault].
- Fabra, Pompeu. 2008 [1933], *Gramàtica catalana* [Catalan Grammar]. Barcelona: Institut d'Estudis Catalans.
- Gougenheim, Georges. 1974. *Grammaire de la langue française du seizième siècle*. Paris: Picard.
- Gramatica Limbii Române* [Romanian Grammar]. 2005. Bucharest: Editura Academiei Române.
- Jensen, Frede. 1994. *Syntaxe de l'ancien occitan*. Tübingen: Max Niemeyer.
- Koschwitz, Eduard. 1973 [1894]. *Grammaire historique de la langue des Félibres*. Marseille: Laffitte.
- Lathrop, Thomas A. 2002 [1984]. *Curso de gramática histórica española*. Madrid: Ariel.
- Martorell de Laconi, S. 1990. *Guías de estudio de gramática histórica española*. Salta: Universidad Católica de Salta.
- Menéndez Pidal, Ramón. 1929 [1904]. *Manual de gramática histórica española*, 5th edn. Madrid: Victoriano Suarez.
- Migliorini, Bruno. 1966. *Storia della lingua italiana*. Firenze: Sansoni.
- Moignet, Gérard. 1984 [1973]. *Grammaire de l'ancien français*, 2nd edn. Paris: Klincksieck.
- Moll, Francesc de Borja. 1952. *Gramática histórica catalana*. Madrid: Gredos.
- Musso, Giancarlo. 2004. *Gramática astèsan-a*. Torino: Gioventura Piemontèisa.
- Penny, Ralph J. 1993. *Gramática histórica del español*. Barcelona: Ariel Lingüística.
- Rohlf, Gerhard. 1954 [1949]. *Grammatica storica della lingua italiana e dei suoi dialetti*. Torino: Einaudi [*Historische Grammatik der Italienischen Sprache und ihrer Mundarten*. Bern: Francke].
- Tekavčić, Pavao. 1972. *Grammatica storica dell'italiano. II: Morfosintassi*. Bologna: Il Mulino.
- Toso, Fiorenzo. 1997. *Grammatica del genovese*. Genova: Le Mani.
- Väänänen, Veikko. 1981 [1963]. *Introduction au latin vulgaire*. Paris: Klincksieck.
- Vocabolario degli accademici della Crusca*. 1866. Accademia della Crusca.
- Wagner, Max L. 1950. *La lingua sarda. Storia, spirito e forma*. Bern: Francke.
- Zamora Vicente, Alonso. 1985. *Dialectología española*. Madrid: Gredos.
- Zauner, Adolph. 1921. *Altspanisches Elementarbuch*. Heidelberg: Carl Winter.