Non-finite complementation in French L2:
A learner corpus approach

Hugues Peters

The University of New South Wales, School of International Studies, Sydney
h.peters@unsw.edu.au

Abstract. The Complementizer Phrase (CP) is mainly unexplored territory in L2 French acquisition studies (Herschensohn 2007:128). This study aims at partially filling this gap by exploring the structure of non-finite complementation using data from a longitudinal oral learner corpus of 10 Jamaican learners of French (Peters 2005, 2006). It specifically explores the realization of the Complementizer (COMP) functional category, and analyses the structure of the non-finite embedded clauses with control and raising structures and embedded interrogative. The influence of the native languages and of the French input on non-target uses will be evaluated. The present study, therefore, answers White's (2003:36) call "to probe quite intricate properties of the interlanguage representation, in order to understand the nature of the grammar that the learner creates to account for the L2." Furthermore, this presentation explores a methodological interface (Rankin 2009) between the framework of the Principle and Parameter framework, and the Minimalist Program (Chomsky 1995, 1999) as applied to SLA research (Lardière 1998, 2000, 2009a, b) and the methods of learner corpus linguistics (Granger *et al.* 2000). The corpus approach, although unlikely to answer all relevant question of structure when crucial data is missing in naturally occurring data, will prove useful in clarifying the issues and delineate further areas of investigation.

Keywords. acquisition of L2 French, control, embedded interrogatives, non-finite complementation, raising

1. Introduction

This study analyses instances of non-finite complementation in a longitudinal corpus of oral productions by ten Jamaican adult learners of French as a foreign language.[1] The goal is to describe the grammar of this structure created by these learners. Drawing on the syntax of non-finite complementizers proposed in Kayne (1981) (see also Jones (1996), Rowlett (2007), for more recent approaches), I focus specifically on the distinction between control and raising predicates and on the expression of embedded infinitival interrogatives. This study therefore uses a corpus approach to investigate a specific issue of morpho-syntax (Granger *et al.* 2000, MacWhinney 2000) and is informed by a comparative analysis of the respective grammars of the native (L1: Jamaican English and Jamaican Creole) and target (L2: Standard French) languages of these learners, within the Principle and Parameters framework, and the Minimalist Program (Chomsky 1995, 1999). It exploits an interface between the method of corpus linguistics and the theory of Universal Grammar (UG) in second language acquisition (SLA) (Rankin 2009).

Assuming, as is customary, that the native L1 grammar of the learners and their approximation of the L2 grammar play a crucial role in determining the structure of the interlanguage (IL), this study adopts a contrastive analysis of the structure of non-finite complementation in English and French, and makes hypothesis on how the structure is manifested in the IL on the basis of this comparison (Haegeman 1992). However, this approach also takes seriously the hypothesis that IL is a UG constrained system in its own right, and might therefore have systematic characteristics that are not reducible to the native or to the target languages.

By providing a study of the realization of a precise morpho-syntactic category, the complementizer phrase (CP), this study concurs with the assessment expressed in White: "it seems clear that we (…) are now probing quite intricate properties of the interlanguage representation, in order to understand the nature of the grammar that the learner creates to account for the L2" (2003, 36) and, by

---

focussing on the syntax of CP in L2 French, this study addresses a gap in the research identified in Herschensohn: "As for the CP realm, it is mainly unexplored territory in L2 French" (2007, 128). While there have been studies on the acquisition of properties associated with the extended CP, such as interrogation or relativisation (Myles 1996, Hawkins 1989, Prévost 2009), no specific study is available on the constraints operating on the realization of the head of the non-finite CP itself in L2 French.

The choice of the phonetic realization of the embedded non-finite COMP (*'à'*, *'de'*, or a phonetically empty variant), and the choice between finite and non-finite object complementation, are well-known difficulties for Anglophone learners of French L2, one of those so-called "unteachables" (Cox 1983). It is characterized by many idiosyncrasies: in some cases several COMPs are allowed with a single matrix verb: *'commencer à/de faire quelque chose'* (to begin to do something), etc.; in other cases the choice of COMP is modified by a change in the thematic grid: active vs. passive: *'forcer quelqu'un à faire quelque chose'* vs. *'être forcé de faire quelque chose'* (to force someone *'à'* to do something vs. to be forced *'de'* to do something), or intransitive vs. transitive: *'décider de faire quelque chose'* vs. *'décider quelqu'un à faire quelque chose'* (to decide *'de'* to do something vs. to persuade someone *'à'* to do something), etc.; additionally, diachronic variation can occur: *'espérer _ faire quelque chose'* vs. *'espérer de faire quelque chose'* (to hope (*'de'*) to do something), the latter marked as archaic / literary, etc. A clarification of the issue has therefore potential pedagogical import for L2 teaching practices.

This issue has been treated in various theoretical frameworks: Delattre (1964) provides a useful typology of infinitival constructions on the basis of their superficial distribution from a structuralist point of view; Cox (1983) develops an aspectual account of the choice between *'de'* and *'à'* to help learners grasp the difference between the two COMPs; in a cognitive linguistics approach, Achard (2000) associates different types of finite and non-finite complementation strategies to subtle differences of conceptualization (using the notion of 'perspective'); and finally, Gross (1968) provides the most detailed description of non-finite structures and of the transformations that can be applied to these structures. In contrast to previous linguistically motivated studies focussing on the target L2, however, the present study starts from an analysis of the actual

productions of second language learners. Furthermore, set within a UG approach to language acquisition, it does not rely on apparent word order similarities, but strives at discovering underlying structural and hierarchical convergences in order to describe the grammar the learners attribute to the target language. I will place more emphasis on the realization of the COMP category itself, than on the phonetic form attributed to that category. Finally, focussing on the task of determining the underlying structural specification, I will leave aside issues of meaning in this paper, although I recognize that the meaning issue is essential for communicative approaches to second language teaching.

In section 2, I briefly present the learner corpus used in this study; in section 3, outline the framework of SLA adopted; in section 4, present a brief overview of the syntax of non-finite complementation in French as compared to English, and add some remarks on Jamaican Creole; in sections 5-6, analyse and discuss learner corpus data to determine the structure of CP in their IL grammar, identify in the French input the possible source of the data observed in learners' production, and evaluate the validity of the method, and finally I draw conclusions, and outline possible avenues for further research.

2. Description of the French learner corpus

My learner corpus is composed of semi-guided conversations with ten Jamaican learners of French over a four semester period at the University of the West Indies (UWI), Mona (Peters 2005, 2006). The learners were studying French within the framework of their B.A. in a purely classroom instructed setting, none having spent any significant period of time in a French speaking country. These learners were receiving French language instruction courses (as well as French studies courses taught in English on French civilisation, linguistics and literature) at UWI. The native languages of these learners are Jamaican Creole and Jamaican English. Each learner was recorded individually during six or seven interviews from semester 2 of year 1 (Intermediate level) in April 2003 to semester 1 of year 3 (Advanced level) in November 2004. It is therefore an oral and longitudinal corpus of learners of French in an instructed environment.

To preserve anonymity, each learner has been attributed a code: "L" followed by a two-digit number: L08 – L12 – L14 – L16 – L17 – L18 – L20 – L31 – L33 – L38.[2] The learners had varied proficiency in French as some had previously studied the language at high school up to four years (L17, L18, L38) or up to six years (L14, L20) while others had studied the language through intensive one-year tracks at the University level (L08, L12, L16), or in a private establishment (L33), or a mix of both high school and private school (L31).

Each utterance produced by the learners has been transcribed on a main tier following the CHILDES transcription and encoding protocol (MacWhinney 2000) and each learner's utterance has been morphologically decomposed on a separate morphological tier (%mor) to allow for deeper analysis. In total, 28,798 words (tokens) of learners' speech have been encoded to this date (excluding repetitions, reformulations in the count).[3]

3. Overview of the UG model of second language acquisition

There is some debate as to the extent of access to Universal Grammar (UG) enjoyed by adult learners (White 2003), but UG models generally assume that, at the initial stage of acquiring a second language, all parameters are set to the value of the learners' L1. Within this research, I adopt the *'Full Transfer Full Access'* model (Schwartz & Sprouse 1996, 2007). This model assumes that learners of a foreign language have full access to UG, and that this faculty helps learners restrict the range of possible grammars they create on the basis of the input they are exposed to. However, the task of adult second language learners, often characterized by incomplete acquisition and optionality, is rendered more difficult than for acquisition of the first language by the fact that a particular association between packages of formal (grammatical) features and lexical items has already been parameterized in their native tongue. The selection of new features that

---

[2] The gap in numbering is due to the fact that some learners not included here were part of a pilot group, that other learners didn't continue the full program of study, were not recorded over a four semester period, and therefore are not part of this longitudinal study.

[3] Transcription and encoding have been realized by the author: the transcription has been checked several times by the author over a five year period, and finally checked thoroughly by research assistants in 2010 thanks to a grant from the Faculty of Arts and Social Sciences, UNSW.

might not be present in their L1, or the particular reassembly of features in functional categories and their phonetic realization in specific vocabulary items, characteristic of a second language (Lardière 1998, 2000, 2009a, b) is a difficult task for adult learners, and therefore acquisition will be influenced by features of their native language. The influence of the first language on second language acquisition is known as the phenomenon of transfer (Gass & Selinker 1992). Positive and negative transfers occur when features of the L1 respectively facilitate or impede acquisition of a construction in the L2. Within this framework, transfer is not primarily caused by superficial word order similarities / dissimilarities, but, at a more abstract level, by consideration of the value and lexical realization of grammatical categories across languages, and the necessity of a reorganization of these parameterized features.

In summary, one general research question, and two sub-questions come into light:

- How can we account for L2 acquisition by cognitively and linguistically mature learners of (some forms of) non-finite complementation?

- What is the respective influence of the L1 (via parametric transfer) and UG (via universal principles) on L2 acquisition of the French CP system, as manifested in the learners' IL?

- What specific L2 French input might be the source of the correct/incorrect analysis produced by learners?

4. The syntax of non-finite verbal complementation

From a descriptive point of view, complementation covers cases in which an embedded clause functions as the selected argument of a verbal, a nominal or an adjectival predicate. In this paper, I specifically focus on verbal non-finite object complementation, that is, on the production of non-finite embedded clauses as internal argument of a matrix verb.

In the case when the infinitival Verb Phrase (VP) is the sole complement of a verbal predicate, as in the following examples, non-finite completives are

introduced with the morphemes '*de*' (of, from), '*à*' (to, at), or their phonetically empty counterpart, depending on the matrix predicate:

(1)a.   Jean décide **de/*à/***_ lire   un livre

John decides   *de*   to-read a book

'John decides to read a book'

  b.   Jean cherche ***de/à/***_ lire   un livre

John looks   *à*   to-read a book

'John is looking forward to reading a book'

  c.   Jean veut ***de/*à/** _ lire   un livre

John wants   _   to-read a book

'John wants to read a book'

In cases of verbal predication with two complements: one overt Determiner Phrase (DP) and one infinitival completive, the DP is either a direct object complement (endowed with Accusative case) or an indirect object complement introduced by the preposition/case marker '*à*' (endowed with Dative case) of the matrix predicate. [4] In this di-transitive context, the choice of the phonetically empty COMP is restricted: Accusative DPs can be followed by a completive introduced with '*à*' or '*de*', but Dative DPs can only be followed by a non-finite completive introduced by '*de*':

(2)a.   Jean encourage Marie ***de/à/***_ chanter

John encourages Mary   *à*   to-sing

'John persuades Mary to sing'

  b.   Jean empêche Marie **de/*à/***_ chanter

Jean prevents Mary   *de*   to-sing

'John prevents Mary from singing'

---

[4] In the present paper, I leave aside several constructions: causatives: '*je laisse Marie partir*' (I let Marie go), and perception verbs: '*je vois Marie partir*' (I see Marie go) as well as the periphrastic future '*je vais partir*' (I'm going to leave) and deictic verbs of movement '*je vais au supermarché acheter des pommes*' (I'm going to the supermarket and/to buy apples).

    c.   Jean rappelle à Marie **de/*à/*_** chanter

          John reminds to Mary   *de*   to-sing

          'John reminds Mary to sing'

The only exceptions to these rules are the impersonal verbs '*falloir*' (to be necessary) and '*sembler*' (to seem) used with a Dative clitic and an empty COMP (as in 3.a-b) and "*apprendre / enseigner à DP à VP*" (to teach '*à*' DP '*à*' to do something) which take an infinitival complement introduced by '*à*': "*Jean apprend / enseigne à Marie à chanter*" (John teaches Mary to sing).[5]

(3)a.   Il   lui   faut   partir

        Expletive him/her-DAT is-necessary to-leave

        'He/She has to leave'

    b.   Il   lui   semble avoir   raison [6]

        Expletive him/her-DAT seems to-have right

        'It seems to him/her that he/she is right'

Delattre (1964) provides a classification of contexts of non-finite verbal complementation that offers us a useful point of departure. He creates a typology of 24 sub-structures based on their superficial distribution, but when clustering in the same categories causatives, movement, and perception verbs, and control or raising structures, he fails to distinguish deeper structural properties, and, he does not deal with embedded interrogatives which are, as we will see, at the core of the present study.

Within a UG framework, Kayne (1981, see also Kayne & Haik 1980) proposes to treat the morpheme '*de*', sometimes phonetically null, as a non-finite Prepositional Complementizer (COMP) heading its own Complementizer Phrase (CP).[7]

---

[5] These verbs meaning 'to teach, to learn' are quite important in the context of classroom instructed lingo, and are likely to be used / known by students.

[6] See below for the significance of this use of '*sembler*' (to seem) as a control verb rather than a raising verb.

[7] This hypothesis is compatible with a proposal generating '*de*' in Spec. of CP as in Kayne (1991), but not compatible with the subsequent revision to the theory proposed in Kayne (1999). This paper does

Similarly, I consider '*à*' to be a variant of '*de*', as suggested by Kayne (1981), but maybe selected by a distinct aspectual value of the matrix predicate (Cox 1983).[8]

First, Kayne (1981) reminds us that the complementizer '*de*' is distinct from the preposition '*de*' (*of, from* + DP): DP arguments fulfilling the same thematic role as the infinitival object completive, for instance, with the verb '*dire*' (to say), are not introduced by a preposition '*de*':

(4)   J'ai dit à Paul de partir   / (*de) quelque chose

       I told   to Paul *de* to-leave / (*of) something

       'I told Paul to leave / something'

This conception of '*de*' as a COMP could mean at first sight that '*de*' would be the non-finite equivalent to the finite declarative COMP '*que*' (that):

(5)   J'ai dit   à Paul de partir   / qu'il parte

       I have told to Paul *de* to-leave / that he leave-SUBJUNCT

       'I told Paul to leave / that he (should) leave'

However, in a Split CP framework (Rizzi 1997), '*que*' and '*de*' (or English '*for*' and Jamaican '*fi*' (Durrleman-Tame 2008)) are shown to occupy different layers within the CP: respectively Force P and Finite P.

Treating '*de*' as a COMP also means that it is of the same syntactic category as '*for*'. However, a crucial difference between the two is that '*de*' (or its other realizations) is unable to licence overt subjects:

---

not intend to take position on this theoretical issue, but see arguments in Borsley (2001) against what he and one reviewer believe to be unnecessary complications to account for empirical data.

[8] As pointed out by a reviewer, while there seems to be a clear agreement as to the complementizer nature of '*de*' in some contexts, the complementizer status of '*à*' is not uncontroversial (starting with Huot 1981, see Carnac-Marquis (1996) in the context of *tough*-constructions, Rowlett (2007:157ff.) for an interesting discussion of '*de*' and '*à*' as case markers, and Martineau & Motapanyane (2000) for a historical overview of the French infinitival complementizer system. I leave this issue of the possible differences between '*de*' and '*à*' in the learner's grammar for further research.

(6)a.   I prefer (for John) to read a book

      Je préfère (*pour/*de/*_ John) lire un livre

  b.   Je préfère que John lise   un livre

      I prefer   that John read-SUBJUNCT a   book

      'I prefer for John to read a book'

When there is a need for an overt embedded subject, French must resort to a finite embedded sentence, as in (6.b).

The main argument presented in Kayne (1981) supporting the hypothesis that '*de*' belongs to the COMP category is that it is syntactically distinct from the English pre-verb '*to*' generated at the inflectional level (IP). This is clearly shown by the fact that the overt COMP '*de*', as opposed to '*to*', is ungrammatical in co-occurrence with an overt *wh*-constituent (such as *where, when,* etc.) in the Specifier of CP:

(7)   Je lui   ai   dit   où   (*d') aller

      I   him-DAT have told where (*de*) to-go

      'I told him where to go'

This has to do with an observational generalisation: the "doubly filled COMP filter" (Chomsky & Lasnik 1977) that prevents both the head and the Spec. of CP to be overtly filled. The filter operates in both French and English, and maybe universally at a more abstract level.

(8)a.   I asked him where (*that) you go

  b.   Je lui   ai   demandé où   (*que) tu vas

      I   him-DAT have asked   where that you go

      I asked him where you go'

This filter has been shown to have apparent exceptions, for instance, in Belfast English (Henry 1995) and in Quebec French, as far as the finite clauses are concerned. However, the generalisation holds without exception for infinitival

embedded interrogatives. As the generalization preventing the co-occurrence of an overt COMP with a *wh*-phrase in Spec. of CP seems robust in both Standard English and Standard French, we would logically expect the learner to take this constraint to hold as well in the grammar of the L2 they are creating to make sense of the French input.

Another argument presented in Kayne (1981, see also Rizzi (1978) for Italian) is that COMP '*de*' (or its phonetically empty counterpart) can be found in *control* contexts, but never in *raising* ones:

(10)a. Jean a   essayé/oublié/décidé   de partir
       John has tried/forgotten/decided *de* to-leave
       'John tried/forgot/decided to leave'

   b. Jean semble/parait/se trouve/s'avère   (*d') être parti
       John seems/appears/happens/turns out (*de*) to-be left
       'John seems/appears/happens/turns out to have left'

Although Kayne himself suggests a range of syntactic options to account for this difference, and rather focusses on the respective case properties of Prepositional Complementizers in French and English, I will adopt the by-now traditional proposal (Haegeman 1991, among others) which assumes that raising verbs like '*seem*' select an Inflectional Phrase (IP) complement, while control predicates like '*try, decide*' select a CP complement.

In the case of control of an understood embedded subject by a DP in a matrix clause: "*John decided to sing*", both the matrix verb '*decide*' and the embedded verb '*sing*' assign their own agentive thematic role: there is a decider and a singer that happen to be co-referential. The structural subject position in the embedded infinitival clause is filled with a phonetically empty PRO 'controlled' by a DP in the matrix clause, either the subject or the object.

(1).a. John tried / forgot / decided [CP PRO to leave]
      (Subject Controller 'John')

   b. Mary asked Paul [CP PRO to leave]
      (Object Controller 'Paul')

In some cases, when no controller is available (as in 12), the embedded PRO subject is assigned an arbitrary generic interpretation.

(12)   It is necessary [CP PRO to study more]

In the case of raising from the embedded subject position to the matrix clause subject: "*John seems (to me) to have left*", the matrix clause subject '*John*' is not thematically selected by the matrix verb '*seem*', but semantically linked to the embedded predicate '*leave.*' This is shown by the possibility of an 'unraised' variant: "*it seems (to me) that John has left*" with an expletive '*it*' inserted because of the need to have an overt subject in the matrix clause in English. Such an expletive subject could not fulfil the thematic role requirements of a control predicate, as shown by the ungrammaticality of "**It tried to be a dog in the garden.*"

Finally, one also can find 'Raising to Object' (RtO) structures with *believe*-type verbs:

(13)   John believes Mary$_i$ [IP t_i to be late]

Interestingly, the corresponding French verb '*croire*' (to believe) behaves more like a control verb, taking a PRO subject, than a raising verb: we therefore observe a perfect contrast between French and English, and this contrast is accounted for by assuming that '*croire*' takes a CP complement (making it possible for a PRO to be generated) and that *believe* takes an IP complement (allowing raising from the embedded to the matrix clause):

(14)a. *Je crois Jean être le plus intelligent de tous

    b.   I believe John to be the most intelligent of all

    c.   Je crois PRO être le plus intelligent de tous

    d. *I believe PRO to be the most intelligent of all

There are however contexts of *wh-* or clitic movement in which French *croire*-type has been shown to manifest raising properties (see Kayne 1981, Rooryck 1997, Boskovic 1997), and the third argument advanced by Kayne (1981) in favour of

the complementizer status of *'de'* is that, when such raising to object occurs, no *'de'* is allowed, as opposed to English *'to'*:

(15)   Quel   garçon crois-tu   (*de) avoir   oublié   ses clés?
        Which boy   believe you (*de*) to-have forgotten his keys
        'Which boy do you believe to have forgotten his keys?'

The important point to notice here is that the displacement of the embedded subject to the matrix clause would be prevented by an overt complementizer while the presence of a complementizer in a control structure does not interfere with the control of the understood PRO subject by a matrix controller, and furthermore, might be necessary to prevent the empty PRO to incorrectly receive case from the matrix verb.

What we have said for control structures in French and English extends to Jamaican Creole (Bailey 1966, Durrleman-Tame 2008). However, Jamaican Creole does not seem to have typical raising to subject, as there does not seem to be a raised equivalent of unraised *'komiin laik'* (is seems like), *'fieba'*, *'tanka'* (it seems) for example:

(16)a. (I) komiin laik se di pikni a go ron we (Durrleman-Tame 2008:108)
        Expl seem like *se* the child prog prosp run away
        'It seems like the child is going to run away'
    b. tanka se dem gaan aredi (Bailey 1966:40)
        seem *se* they gone already
        'They seem to be gone already'

In conclusion, we have seen that the task of the second language learner is not a simple one as he/she must realize (i) that non-finite French COMPs in French do not allow overt subjects, (ii) that the French COMP category has several phonetic realizations, whose selection by the matrix verb is marked by idiosyncrasies, and (iii) that croire-type verbs usually behaves like control verbs, except in cases of *wh-* or clitic movement. It is therefore interesting to observe how learners actually

manage these constraints through the utterance they produce in oral conversation, and what kind of grammar of French they create in order to make linguistic sense of the input they are exposed to.

5. Observation of Learners data in the UWI corpus

Following the constructions emphasized in Kayne (1981) to argue in favour of the COMP status of '*de*', I focus on the opposition between control and raising predicates, as well as on the structure of non-finite embedded sentences.

5.1 Embedded non-finite interrogatives

The first issue to deal with is the ungrammaticality of '*de*' in co-occurrence with a *wh*-phrase in the Spec of CP in embedded non-finite interrogatives. Within the entire corpus, one finds seven instances of embedded infinitival *wh*-interrogatives.

Five are correctly produced without '*de*' by L12 (2), L16 (1), L20 (1), L31 (1):

(17)a. je sais pas comment m' expliquer,   comment vous expliquer (L12 III1)[9] [10]

     I know not how   myself to-explain how   you to-explain

     'I don't know how to explain myself, how to explain to you'

  b. pour   faire comment? (L16 III1)

     in-order to-do how

     'to do how ?' (NB: with the intended meaning: to do what?)

  c. je dois apprendre comment, comment dit on, diriger mon temps (L20 II2)

     I must learn how,   how   says one, to-manage my time

     'I must learn how – how do you say? – to manage my time'

---

[9] In all examples, "L" followed by 2 digit refers to the learners, Roman numeral refers to the year and Arabic numeral to the semester when the interview took place, III1 = Year 3 semester 1.

[10] For ease of reading, I will usually reproduce the learners' utterances without the repetitions, hesitations, reformulations, self-corrections, filled pauses, and other encodings, unless relevant.

d. j' avais un besoin de [/] <de apprend> [//] de appris [/] appris comment
    parler la langue (L31 III1)

    I had a need *de* [/] <*de* learn> [//] *de* learned- PAST-PART [/] learned-PAST-
    PART how to-speak the language

    'I needed to learn how to speak the language'

In (17.c), two intertwined interrogatives are produced by L20, one parenthetic
(*comment dit-on? / how do you say?*) on the other (… *learn how to manage my time*). The
parenthetic interrogative intervenes between the *wh*-interrogative word and the
infinitival verb. In (17.d), L31 easily generates the embedded interrogative, but the
control verb is produced with hesitation concerning its morphology, manifested
by repetitions, marked by [/], and self-correction, marked by [//], from the part
of the learner looking for the correct form of the control verb before finally
settling on an incorrect past participial form. These five embedded infinitival
interrogatives all use the interrogative adverb '*comment?*' (how?), twice with the
verb '*apprendre*' (to learn how to do something) and twice with the verb '*savoir*' (to
know how to do something).

The last two utterances of embedded interrogatives in the corpus, produced by
L14 and L17, are incorrectly generated with an overt '*de*':

(18)a. j' ai appris comment d' utiliser mon xx (L14 I2)

        I have learned how *de* to-use my xx

        'I learned how to use my xx'

   b. je ne sais pas quoi <de pense> [//] &euh@fp de penser (L17 II1)

       I neg now not what <*de* think-TNS> [//]   *de* to-think

       'I don't know what to think'

Again the actual utterances produced manifest a fair amount of hesitation: 'xx'
indicating inaudible material, and '&euh@fp' indicating a filled pause. L17
produces an instance of self-correction as the embedded verb is first realized with
a bare stem (characteristic of a present tense form), and immediately corrected
into the correct infinitival form. This might indicate that the internal grammar of
L17 does not yet produce non-finite interrogatives without conscious self-
monitoring.

As the doubly filled COMP filter is effective both in the L1 and in the L2 of these learners, I logically assume that it must be effective in their IL as well. As an overt '*de*' is generated by these two learners, I must conclude that at that stage in their development, the morpheme '*de*' is not generated as the head of CP. So, the question becomes: where is it realized?

As far as learner L14 is concerned, there exists a robust cluster of arguments pointing towards the conclusion that she inserts '*de*' as the phonetic realization of a bundle of features corresponding to the English morpheme '*to*'.

First, '*de*' is systematically overgeneralized in instances of completives where we would expect a phonetically empty COMP in Modern Standard French (see below for exceptions to this rule): three times with the verb '*espérer*' (to hope) and once with the verb '*souhaiter*' (to wish) within the same interview:

(19) a. j' espère d' aller   au   bureau (L14 I2)

I hope   *de* to-go to-the office

'I hope to go to the office'

b. je souhaite de travailler (L14 I2)

I wish   *de* to-work

'I wish to work'

Although not strictly ungrammatical, these examples manifest a rather archaic / literary use of the language, i.e., a register to which learners are not likely to have been exposed to, not typical of colloquial French.[11]

Secondly, '*de*' is incorrectly realized with a non-finite relative clause, instead of the expected COMP '*à*':

---

[11] Of course, as we are interested in colloquial French only, we can simply say that the insertion of '*de*' is ungrammatical. One reviewer at least strongly agrees with this stance. I must point out however that the participants all took an introduction to French literature course that included classic texts such as *Dom Juan* by Molière or *L'étranger* by Camus and therefore may have encountered some archaic forms (although, as far as I can judge, the latter play does not include any form such as '*espérer*' or '*souhaiter*' with '*de*' and there is only one example of '*souhaiter de*' in Camus' novel (1942/1999 L'étranger Paris: Gallimard (Collection Folio Plus 10) p.119)).

(20)   je sais tous les choses que les hôtels (…) ont d' offrir aux touristes (L14 I2)

         I know all the things that the hotels   have **de* to-offer to-the tourists

         'I know everything that hotels have to offer to tourists'

Thirdly, '*de*' is realized in a probable '*for/to*' construction directly transferred from English (see Peters 2009):

(21)   c' est important pour les diplomates d' avoir   la capacité de parler   dans une autre langue (L14 I2)

         that is important for the diplomats   *de* to-have the capacity of to-speak in a-FEM other language

         'it's important for diplomats to have the ability to speak in a foreign language'

These characteristics of verbal complementation for L14 at level I2 show that '*de*' seems to be transferred from English with the features of '*to*'.[12] In fact, it is remarkable to note that, at that level, L14 systematically overgenerates '*de*' without exception (but see below).

The situation is a bit more confused for learner L17 who, at level II1, systematically uses a phonetically empty COMP with verbal complementation, as in (22) with '*commencer à / de*' (to begin to) not followed by the COMP '*à / de*'.[13] She also confuses verbal moods on several occasions, using infinitival verbs instead of expected past participle or finite tense verbs, or inversely using a finite tense verb instead of an infinitival verb. This confusion might mean a blurring of the line between finite and non-finite completives at that stage:

---

[12] A reviewer points out an interesting parallelism with English speaking learners of Spanish. Despite Standard Spanish not having prepositional complementizers, students tend to grab one preposition and overgeneralise it, usually '*a*' the literal translation of '*to*'.

[13] In fact, L14 is the only learner to show such a systematic overgeneralization of '*de*.'

(22)   *Je voudrais   continuer   étudier (L17 II1)[14]

       I would-like to-continue *_ to-study

       'I would like to continue to study'

Interestingly, with both L14 and L17, '*de*' is correctly never used with '*vouloir*' (want) or with the completives of modal verbs '*pouvoir*' (can) and '*devoir*' (must):

(23)a. je veux parler un+peu à propos de dans le tourisme (L14 I2)

       I want to-speak a little about   of in   the tourism

       'I want to speak a little bit about tourism'

    b. elle doit payer beaucoup de monnaie (L17 II1)

       she must pay   a-lot   of change (= money)

       'She must pay a lot of money'

Considering the possible influence of '*wanna*' contraction, one notices that these latter contexts where the learner correctly does not generate an overt '*de*' are precisely the contexts that would not take '*to*' in English either:[15] [16]

      Modals: *can, must, etc.*: 'I must (*to) try'

      Contracted verbs: *wanna, gonna, hafta, sposta*, etc.: 'I wanna (*to) try'

There is some evidence, but none conclusive, that at later stages the link between '*to*' and '*de*' might weaken, as these two learners introduce a greater variety of

---

[14] As suggested by a reviewer, in this case, the omission of an overt COMP could be akin to an instance of hypercorrection. I suppose that the learner being aware that what she believes to be the translation of 'to' is often not realized, she would overgeneralize the phonetically empty COMP. It is indeed important to distinguish hypercorrection from transfer. More research is needed on this topic.

[15] I also strongly agree with a reviewer's comment that one might speculate that the correct use of these verbs is due to the fact that they are usually taught and acquired earlier than the other verbs (in fact, even used as chunks in the conditional mood long before this mood is systematically taught) and have perhaps become more systematic: indeed, there is just one mistake in the choice of the null COMP with these verbs in the entire corpus.

[16] Again I leave for further research the case of causative verbs: '*faire*' (to make)*,* verbs of perception '*voir*' (to see), deictic verbs of motion '*venir*' (to come), nor the periphrastic future '*aller*' (*gonna*). For these verbs as well: a null COMP in French corresponds to the absence of 'to' in English.

COMPs. L14 no longer systematically over-generates '*de*' (She produces one example of '*souhaiter*' with a null COMP at level II1). This might indicate that learners develop several options over time. (See below for examples of intra-learner variability). However, as they produced only one instance of embedded interrogative each, it is impossible to determine conclusively whether they have managed to restructure their grammar at later stages.

Here we touch some inherent limitations of corpus research: the absence of relevant data, and the absence of starred sentences, as well as the necessity to distinguish performance mistakes from genuine systematic competence (see Peters (in preparation) for a more detailed analysis of the patterns of systematic errors).

In brief, it seems probable that for some learners (namely L14 and maybe L17 and others), a misanalysis of '*de*' treated as equivalent to English '*to*' in terms of formal features occurs.

5.2 Subject/object/arbitrary control and raising predicates

As shown in Table 1, Cases of control are relatively abundant in the corpus: considering all ten learners, one can find 441 instances of control: subject (417 instances), object (11 instances), or arbitrary (13 instances) control (see examples in 24):

(24)a. Philippe a   décidé d' aller à la plage (L16 III1) (Subject control)

        Philippe has decided *de* to-go to the beach

        'Philippe decided to go to the beach'

    b. le cours   m' a   aidé   à améliorer la prononciation. (L16 II2) (Object control)

        the course me has helped *à* to-improve the pronunciation

        'the course helped me to improve the pronunciation'

    c. il faut   pratiquer beaucoup le français. (L33 I2) (Arbitrary control)

        it is-necessary _ to-practice a-lot   the French

        'it is necessary to practice French a lot'

	L08	L12	L14	L16	L17	L18	L20	L31	L33	L38	Total
Token	43	96	35	36	64	22	53	4	48	40	441
	(2)	(2)	(1)	(1)	(0)	(0)	(3)	(0)	(15)	(0)	(24)
Type	7	14	11	12	10	8	12	4	12	7	36
	(1)	(2)	(1)	(1)	(0)	(0)	(3)	(0)	(1)	(0)	(7)

Table 1. Number of control verbs in the learner corpus.

In table 1, '*Token*' refers to the number of instances of control structures (subject, object and arbitrary), learner by learner, and '*Type*' refers to the number of different lexical verbs used at least once for each learner in control structures. The numbers in parentheses refer to the number of cases of non-subject control. The column '*Total*' refers to the total number of tokens / types across learners. Instances of embedded interrogatives are not included in the count.

Out of the 24 cases of object / arbitrary control, 13 manifest arbitrary control and 11, object control. All examples of arbitrary control are produced by one learner, L33, who uses one impersonal verb: "*falloir*" (to be necessary) (as in 24.c) (used also as object control verb).

5.2.1. Subject control

There are 29 different subject control predicates in the corpus, but about two third of all instances of subject control (288/417) are produced with just three verbs: the desiderative verb '*vouloir*' (to want): 122 times, and the modal verbs: '*pouvoir*' (can) and '*devoir*' (must): 83 times each. The modal verbs are always used with a root interpretation. '*Pouvoir*' (can) and '*vouloir*' (to want) are the only verbs used at least once by all learners, and '*devoir*' (must) is used at least once by eight of the ten learners (only L16 and L31 do not use it). I consider these to be control

verbs with a phonetically unrealized COMP, and there is (almost) no error in the corpus with respect to the choice of null COMP for these verbs.[17]

Some other major subject control verbs in terms of productivity are:

- the aspectual verbs: '*commencer à/de VP*' (to start '*à/de*' to do something): used 25 times by eight learners, almost always correctly with '*à*', except once used with a null COMP by L20; another one is '*continuer à/de VP*' (to continue '*à/de*' to do something) used 4 times by three learners: 3 times with the correct '*à*', and once with an incorrect empty COMP by L17; etc.;
- the desiderative verbs: '*aimer _ VP*' (to love/like to do something): used 23 times by six learners, always with the expected null COMP; some other desiderative verbs include: '*préférer _ VP*' (to prefer) used 6 times by four learners with the correct empty COMP, except once with an ungrammatical '*de*' by L14 [18]; '*désirer _ VP*' (to desire/want to do something) used 4 times by one learner, L18, but inconsistently: once with an incorrect archaic/literary '*de*', once with an expected empty COMP, and twice with the ungrammatical '*pour*' (for); '*souhaiter _ VP*' (to wish to do something) used twice by L14, once with the incorrect archaic/literary '*de*' and in a subsequent interview, once with the expected empty COMP, etc.;
- some typical control verbs: '*essayer de VP*' (to try '*de*' to do something): used 18 times by eight learners, mostly with '*de*', except by L20 who once uses an empty COMP and once uses '*à*'); '*espérer _ VP*' (to hope to do something): used 14 times by four learners: 9 times with the incorrect archaic/literary COMP '*de*' (5 times by L14, twice by L16, and twice inconsistently by L18), and 5 times with the correct empty COMP; and '*décider de VP*' (to decide '*de*' to do something): used 12 times by five learners: 8 times correctly with '*de*' and 4 times incorrectly with an empty COMP by L12, L16 and L18 (see below); etc.
- and a durative verb: '*passer (duration) à VP*' (to spend time '*à*' doing something) incorrectly used with '*pour*' (for) by L14, etc.

---

[17] I have identified one unsystematic mistake by insertion of '*à*' by L20 (with '*vouloir*') who otherwise produces 25 cases without COMP with these verbs.

[18] Remember that L14 is the learner who systematically associates '*de*' with '*to*'.

As illustrated by this sample of control structures, learners are mostly correct in choosing the COMP. In fact, in total, there are only 29 clear cases of ungrammatical choice of COMP over 417. Overall, these errors can be classified as follows:

- incorrect use of '*de*' instead of an empty COMP: 12 times (mainly with '*désirer*' (to desire), '*souhaiter*' (to wish) and '*espérer*' (to hope)); [19]

- ungrammatical use of an empty COMP instead of an expected overt COMP: 11 times, 8 times instead of '*de*' and 3 times instead of '*à*'; [20]

- ungrammatical use of '*à*' instead of '*de*': twice;

- and ungrammatical use of '*pour*' (for): 4 times, once instead of '*de*' (by L31), once instead of '*à*' (by L14), and twice instead of an empty COMP (by L18).

The majority of ungrammatical cases are therefore characterized by the overuse of the empty COMP (11 times), mainly to replace an expected '*de*' (8 times) and, inversely, by the overgeneralization of '*de*' (sometimes in what would be an archaic / literary register) (12 times). We notice also that there are errors of incorrect use of '*à*' instead of '*de*', but not one case of an incorrect use of '*de*' instead of '*à*'. Finally, we notice that the preposition '*pour*' (for), that is not a Prepositional Complementizer in French, seems to be reanalysed as COMP by some Jamaican learners, probably as a translation of '*for/fi*'. Interestingly, this use of '*for*' is not accompanied by the licensing of an overt subject with verbal complementation (although it might be with adjectival complementation (Peters 2009), as seen in example (21) above). (See Peters (in preparation) for an analysis of '*pour*' as transfer from the native Jamaican).

The use of '*décider de*' (to decide) '*espérer*' (to hope), '*desirer*' (to desire, to want) provides us with interesting examples of intra-speaker optionality: two learners, L16 and L18, change the COMPs they use within the same interview. L16 first correctly uses the expected '*de*' with '*décider*' (as seen above in 24.a repeated below in 25.a), and then incorrectly uses an empty COMP in (25.b):

---

[19] When considering ungrammatical the archaic / literary use of '*de*'. See note 11.
[20] Maybe as a form of hypercorrection, according to one reviewer. See note 14.

(25)a. Philippe a   décidé d' aller à la   plage (L16 III1)

      Philippe has decided *de* to-go to the beach

      'Philippe decided to go to the beach'

   b. il décide   aller chez soi (L16 III1)

      he decides *_   to-go at-one's-home

      'he decides to go home'

Note that, in both examples, there was a slight pause between '*décide*' and '*aller*' as well as between '*a décidé*' and '*d' aller.*' It is possible that the optionality would be caused by purely phonetic factors such as the difficulty associated with the production of two geminated [d]'s in '*décide d'aller*' (decides to go) resulting from the obligatory deletion of the final schwa in '*décide*'. Tranel (1987, 89) mentions that English speakers sometimes have trouble pronouncing the consonant clusters resulting from e-deletion in final position. This is not surprising as English does not have true consonant geminates: they occur only at word or morpheme boundaries, and are often not pronounced in informal speech (Kaye 2005).[21] In that case, the syntax of the choice of COMP would be essentially correct in both cases for that learner. The apparent ungrammaticality of (25.b) would only result from a failure to spell out the underlying '*de*' because of a pronunciation difficulty.

L18 first uses '*espérer*' (to hope) with an archaic / literary '*de*' ungrammatical in colloquial French, and then correctly with the expected empty COMPs in two other occasions, at level II1:

(26)a. j' espère de travailler dans une   profession sociale (L18 II1)

      I hope   **de* to-work   in   a-FEM profession social

      'I hope to work as a social worker'

   b. j' espère voyager d' Afrique (L18 II1)

      I hope   to-travel of Africa

      'I hope to travel to Africa'

---

[21] Thanks to Barbara Bullock (p.c.) for pointing this out.

   c. j' espère passer   le   temps avec ma famille (L18 II1)

      I hope   to-spend the time   with my family

      'I hope to spend time with my family'

She again produces the same alternation between '*de*' and a null COMP in a subsequent interview at level II2. This tends to show that two systems are competing inside L18's grammar of Complementation. [22]

L18 also shows another case of optionality when she uses '*désirer*' (to want, to desire) correctly with the expected empty COMP in one occasion, and the ungrammatical '*pour*' (for) in two other occasions at level II1:

(27)a. il désire bâtir   une   pays   l' arabe (L18 II1)

        he wants to-build a-FEM country the-Arab

        'he wants to build an Arab nation'

   b. je ne désire pas pour engager   l'attention des   autres étudiantes (L18 II1)

       I neg want not *for to-engage the attention of-the other students-FEM

       'I do not want to attract the attention of other students'

   c. il y a beaucoup de personnes (…) qui désirer pour étudier à Uwi (L18 II1)

       it there has a-lot   of people   who to-want *for to-study at UWI

       'There are many people (…) who want to study at UWI'

In the last example, the control verb '*désirer*' is an incorrect root infinitive used instead of a finite tense.

As a whole, the data shows that even though learners may be unsure of which phonetic realization they need to give to the COMP, there is little doubt that they are aware that some linking category is necessary in the structure. Of course, it is still an open issue whether this linking category is definitely categorized as a CP: these linking words might also be generated as free morphemes at the inflectional level by some learners, as we have seen.

---

[22] The use of '*pour*' here might be facilitated by 'wish for' as pointed out by a reviewer.

One further observation supporting a higher positioning within CP is offered by examples in which an overt linking word is separated from the infinitival verb by an overt clitic pronoun. There are four such cases in the corpus, namely the following from L12, L20 and L33 (see also (31.b) below):

(28)a. <j' essaie de> [/] j' essaie de le faire maintenant (L12 III1)

       <I try *de*> [/]   I try   *de* it to-do now

       'I'm trying to do it now'

   b. mais j' essaie à m' organiser   meilleur (L20 III1)

       but   I try   **à* myself to-organise better-ADJ

       'but I try to organize myself better'

   c. c' est une femme   qui a   essayé de s'   élever (L33 II1)

       that is a-FEM woman who has tried   *de* herself raise

       'It is a woman who has tried to raise herself'

If we assume that the pronoun is located at the inflectional level of the clause (as in Kayne 1991 and others), this data constitutes an argument in favour of a generation of '*de*' or '*à*' higher than IP, as modern French does not allow clitic climbing outside the infinitival clause (Martineau 1991).

5.2.2. Object control

As shown in table 1, instances of object control verbs are few, but attested: six learners produce 11 instances of object control. The corpus contains seven different object control verbs:

- '*aider DP à VP*' (to help someone '*à*' to do something) used twice by L08 with the incorrect '*de*', and once by L12 and L16 each with the expected '*à*';
- '*demander à DP de VP*' (to ask someone '*de*' to do something) used once by L14 with the expected '*de*';
- '*enseigner à DP à VP*' (to teach someone '*à*' to do something) used once by L12 with the correct '*à*';
- 'clitic *falloir _ VP*' (to be necessary for someone (= obligatorily in the form of a Dative clitic pronoun) to do something), used twice by L33 with a correct empty COMP, once with an incorrect pronoun (see below);

- '*intéresser DP à VP*' (to interest someone in doing something) used once by L20 with an incorrect '*de*';

- '*permettre à DP de VP*' (to allow someone '*de*' to do something) used once by L20 (see below);

- '*mener DP à VP*' (to lead someone '*à*' to do something) used once by L20 with the correct '*à*'.

The most productive object control verb is '*aider*' (to help) used four times by three different learners with a consistent, if not correct, choice of COMP: L08 systematically selects the ungrammatical COMP '*de*' to introduce the completive of '*aider*' (as in 29.a, b), while the other two learners, L12 (in (29.c)), and L16 (in (24.b)), use the correct COMP '*à*':

(29) a. je pense que ma linguistique m'   aide   de parler   français pour
         mieux (L08 I2)

         I believe that my linguistics   me-ACC helps *de* to-speak French for
         better-ADV

         'I believe that my linguistic course helps me to speak French better'

   b. je pense que ça   va   m'   aider de parler   français mieux (L08
       II2)

       I believe that that goes me-ACC helps *de* to-speak French better-ADV

       'I believe that this is going to help me to speak French better'

   c. je voudrais   recevoir   quelque chose que peuvent aider   moi à étudier .
       (L12 II1)

       I would-like to-receive some-thing   that can   to-help me *à* to-study

       'I would like to receive something that can help me to study'

Half the cases of object control (5 out of 11) strictly adhere to the following pattern, as shown in the previous (24.b, 29.a, b) and in the following examples (30.a, b):

'Inanimate subject + first person clitic pronoun + control verb + CP (*de/à*)'

(30)a l' expérience doive m' enseigner (…) d' avoir plus contrôle (L12 II2)

the experience must-SUBJUNCT me to-teach (…) de to-have more control

'The experience must teach me to have more control'

b. cette année me demande de faire une haut [^ nivel] de travail (L14 II1)[23]

This year me asks *de* to-do a-FEM high level of work

'This year requires me to produce a high quality of work'

Slight modifications to the basic pattern occur in four other utterances: one with the incorrect use of a strong object pronoun instead of a weak clitic (see example (28.c): '*aider moi*' instead of '*m'aider*' (to help me)), one with the use of an impersonal construction and an empty COMP (in (31.a)), one with an animate subject (in (31.b)), and one with the use of an empty *pro* object pronoun with a meaning that would be equivalent to the indefinite pronoun '*one*' in English (Haegeman 1992) (in (31.c)):

(31)a. il me faut parler maintenant au sujet de monsieur Name (L33 I2)

Expl. me-DAT is-necessary to-speak now about of Mr. Name

'it is necessary for me now to speak about Mr. Name_of_Politician'

b. vous savez si je m' intéresse de s' inscrire par les moyens virtuels ? (L20 III1)

you know if I myself interest *de oneself enrol by the means virtual

'Do you know whether I am interested in enrolling myself with virtual means?'

c. l' enseignement de la langue, il mène à pratiquer plus (L20 II2)

the teaching of the language, it leads *pro à* to-practice more

'Teaching the language leads one to practice more'

Note that in (31.b), the first person *phi*-features of PRO, inherited from the controller, are not transferred on the anaphoric clitic pronoun with '*se*' in '*s'inscrire*'

---

[23] In (30.b), [^ nivel] is probably a vocabulary interference from Spanish. The intended French word is '*niveau*' (level).

(to enrol oneself) used instead of '*me*' (myself). However, the very presence of the pronoun is significant as it might confirm that '*de*' is generated above the inflectional level in the grammar of this learner. (See the discussion of (27) above).

The last two examples in (32) are ungrammatical: (32.a) has an inanimate object controller in a direct translation of "he allows it to happen." [24] As discussed in Rooryck (1988), the verb '*permettre*' (to allow), when used with an inanimate object, acquires a metaphorical meaning equivalent to '*let*' and behaves in a manner more akin to raising to object verbs than control verbs. The sentence produced by the learner is ungrammatical, because the correct subcategorization should select a Dative clitic (instead of an Accusative one) and a COMP '*de*' (instead of '*à*'): '*permettre à DP de VP*' (to allow '*à*' DP '*de*' to do something), and more fundamentally, because this type of metaphorical construction strictly rejects the pronominalisation of the inanimate object. The learner's attempt at producing a control structure as a consequence of direct interference from English therefore results in ungrammaticality.

(32) a. n'importe qui le dieu veut   pour moi, il le   permit@s à passer   pour
        moi (L20 I2)

        Whomever   the god wants for me, he it-ACC permits *à* to-happen for me

        'Whatever God wants for me, he lets it happen to me'

     b. il faut,   je pense, nous   profiter   de l'Internet (L33 III1)

        it is-necessary, I think,   we/us to-take-advantage of the Internet

        'it is necessary, I think, for us to take advantage of Internet'

Finally, in (32.b), an overt embedded subject '*nous*' (we, us), separated from the matrix verb by a parenthetical, is used instead of a clitic or instead of a PRO.[25]

In conclusion, we have seen that, in all cases of object control, the embedded CP is correctly realized with an overt COMP, not always the correct one of course,

---

[24] The direct translation from English is made more obvious by the fact that the pronunciation of the form 'permit' is a direct borrowing from English (marked by the @s symbol) used in place of the French verb '*permet*' (permits, allows).

[25] Therefore, this example should be treated as a case of Exceptional Case Marking or optional control rather than object control of a PRO, and we end up having only 10 true cases of object control instead of 11 in table 1.

when there needs to be such an overt COMP. It seems that the learners never make the mistake of introducing the non-finite clause with a phonetically empty variant of COMP, and therefore are aware of this important characteristic of French syntax which distinguishes object control verbs from perception or causative verbs. The impersonal verb '*falloir*' correctly takes a phonetically empty COMP and is used both as object control and arbitrary control predicate. Two ungrammatical sentences are however caused by direct transfer/translation from English.

5.2.3. Raising structure

In contrast to control structures, not one instance of raising to subject is found in the entire corpus. One could assume that there is simply little communicative need to use such a construction. Yet, we find three examples of unraised '*sembler*' (seem), or '*paraitre*' (appear) with finite CPs:

(33)a. il me semble que (…) les personnes dans le photo est (…) dans une hôtel pour les couples (L14 I2)

      it me-DAT seems that (…) the people in the picture *is   in a-FEM hotel for the couples

      'it seems to me that the people in the picture are in a hotel for couples'

   b. il paraît que c'est une scène (…) au restaurant de dans un hôtel (L31 III1)

      it appears that that is a-FEM scene at-the restaurant of in a hotel

      'it appears to be a scene in a hotel restaurant'

   c. il semble que ils   font beaucoup de confiance dans leurs médias (L12 III1)

      it seems   that they make a-lot   of confidence in   their medias

      'it seems that they trust their media a lot'

So, no transfer from English that would facilitate the acquisition of the raising to subject construction in French seems to take place (but there could be (negative) transfer from Jamaican Creole in which the structure is missing) (See Peters (in preparation), for an analysis in terms of transfer from Jamaican Creole).[26]

---

[26] There is however one utterance using '*censé*' (supposed to), usually analysed as raising to subject. But the learner, L20, uses it with an overt '*de*' instead of the expected IP complement: "si j' étais censée de [/] d' assister un autre université" (L20 III1) (If I was supposed/forced '*de*' to attend

Finally, as far as raising to object (with *believe*-type verbs) is concerned, we might have expected the structure to be incorrectly transferred from English even though it is ungrammatical in Standard French (without *wh*-movement). There is only one example in the entire corpus with such a *believe*-type verb followed by an infinitival clause:

(34) j' ai   prouvé être   prêt [//] ah ah   prête ouh: xx (L16 I2)

      I have proved to-be ready-MASC [//] ready-FEM xx

      'I proved (myself) to be ready'

Interestingly, '*prouver*' (to prove) is used correctly, in the sense that it is used as a control predicate taking a CP complement and a PRO instead of being used as a raising predicate with an overt subject. So, again no (negative) transfer from Standard English, that would have produced a raising structure, takes place.

6. Discussion

6.1. Data on the status of 'de' and control

I have hypothesized that, for certain Jamaican adult learners of French (namely L14 and L17), the pre-verb '*to*' might be transferred with its morpho-syntactic properties and phonetically realized as '*de*'. The source of this structural error seems obvious as both languages have the same basic structural representation and since superficial word order is consistent with both '*de*' in CP or IP, the learners therefore could reanalyse (35.a) as (35.c) on the model of (35.b):[27]

---

another university). In that case, it is possible that '*censé *de*' (supposed to) would have an interpretation of objective modal necessity, maybe in line with the grammaticalization affecting '*sposta*' in English (Collins 2009). Alternatively, '*censé*' could be treated as a case of adjectival complementation on the model of '*foutu de*' (capable of), etc., and for L20 '*de*' would be required by the adjective, rather than by the CP (See Kayne & Haik 1980:50, note 9).

[27] The generation of '*de*' in Infl must make special provision with respect to the checking of non-finite features of the infinitival verbs at the inflectional level in covert syntax.

(35)a. Paul a décidé [CP de [IP PRO chanter]]

  b. Paul has decided [CP _ [IP PRO to sing]]

  c. Paul a décidé [CP _ [IP PRO de chanter]]

A learner falling into that trap would not easily find positive evidence that '*de*' or other COMPs are in CP, as there is no context in which '*to*' would be ungrammatical in English (causatives, perception verbs, modals, and contraction) while the corresponding '*de*' would be grammatical in French ('*de*' is ungrammatical as well in French in these contexts). I suggest, without developing this point due to lack of space, that one must go beyond word order, and, based on the fourth property identified in Kayne (1981) distinguishing '*de*' from '*to*', introduce considerations of scope with respect to negation to help learners reorganize their L2 grammar.[28]

The overall data on control structures in the corpus seem to point to the conclusion that the structure is mostly acquired. Learners abundantly use control structures and chose linking words most of the time, although not always the correct one. This indicates their awareness that a linking word must be used. However there is no conclusive way, based on word order alone, to decide whether a learner generates the linking word in CP or in IP, as we have seen. This question therefore cannot be easily resolved based on corpus data alone, but the intervening position of clitic pronouns might indicate that '*de*' is generated above IP for at least L12, L20, and L33.

6.2. Data on raising

The important fact is that raising structures are avoided by every learner, even in cases when (negative or positive) transfer of syntactic properties from Standard English would have allowed it to happen. Even though absence of a construction is not definite proof that it is not part of the grammar of the learners, and in any case more research is needed on other structures such as exceptional case marking (ECM) causative and perception verbs, and passivisation, the important point is the contrast between raising and control.

---

[28] Unfortunately, there is no negated infinitive in the corpus that would show the position of '*de*' with respect to negation. More research is needed on this topic.

A possible trigger for the misanalysis of the L2 input by the learners could be based on the observation that, as we have seen, some of the most frequent raising structures in English are ungrammatical in French, and inversely that some control structures in French correspond to raising structures in English (see (3.b) and (14) above, repeated in 36):

(36)a. *I believe [CP PRO to be the most intelligent of all]

   Je crois [CP PRO être le plus intelligent de tous]

   b. *It seems to him/her [CP PRO to be right]

   Il lui semble [CP PRO avoir raison]

In (36.b), we even observe the use of '*sembler*' (seem) as part of an object control structure, with a Dative clitic controller.

These examples are perfect instances of a "subset principle" (Haegeman 1992): learners, when confronted with an absolutely ungrammatical structure in Standard English that appears to be perfectly grammatical in Standard French, will be forced by the language faculty to reorganize their grammar, and, I hypothesize that this contrast is likely to mislead them into creating a UG constrained, yet incorrectly specified grammar of French in which raising would be disallowed. This, I suggest, would account for the fact that raising is not part of the IL grammar of these learners. Another factor might be the influence of Jamaican Creole which does not appear to have raising to subject structures.

6.3. The value and limitations of a corpus analysis

We have seen that some structures are never produced. Although avoidance can sometimes be interpreted as a symptom of transfer (Gass & Selinker 1992), absence of a structure cannot strictly be equated with ungrammaticality in the internal grammar of the learner. Starred sentences are obviously not part of a corpus.[29] Thus, it is impossible to know conclusively, based on corpus data alone, whether these missing structures are part of the grammar of the learner. Pure production data should ideally be complemented by comprehension data and

---

[29] Note that we have some data on learners' self-correction showing that they might be actively trying to modify the current intuitive state of their IL grammar.

grammaticality judgments tasks to have a complete picture of the internalized grammar.

With the use of a relatively small corpus, we can nevertheless draw generalizations, pinpoint specific problems and delineate precise areas of further investigation when the investigation is informed by a comparative approach to the syntax of IL. An advantage of the longitudinal learner corpus is that we are able to analyse each learner at a time the grammar of L2 they are creating. So it is possible that individual differences might emerge showing that some learners would choose one path of acquisition (for instance, in the overuse of '*de*') and other learners another. After this individual analysis, one can then hope to find generalization on the entire population of learners (for instance, with respect to raising). Another benefit of the corpus approach is that the data from our oral corpus of Jamaican learners of French could be compared with data from learners of a different linguistic background (for instance, Australian learners of French) or learners of other languages (for instance, English speaking learners of Spanish, see Valenzuela 2008, Pérez-Tattam 2011 for studies on the acquisition of the Spanish CP system), and with written data. Finally, the learner corpus provides data on students' linguistic behaviour that could serve as the basis for pedagogical implementation.

7. Conclusion

This study illustrates a research combining the UG framework of SLA and the method of learner corpus. The study of IL as a grammatical system of its own is an invaluable source of data for trying to determine the grammar that the L2 learners are building, once limitations of a corpus analysis are recognised. Focussing on embedded interrogatives and control structures, we have hypothesized that the French input data and transfer of properties from the L1s can lead some English/Jamaican speaking learners towards an incorrect analysis of French grammar with respect to the categorial status and choice of phonetic realization of the linking words: '*de*' as a pre-verb (for L14, L17)*,* the possible use of the preposition '*pour*' as COMP (by L14, L18, L31), vs. the correct analysis of '*de*' as COMP (for L12, L20, L33). These factors also account for the contrast

between the productivity of control structures and the absence of raising to subject with typical *seem*-type verbs.

Further topics of interest include research on other non-finite completives (*for/to*, adjectival and nominal complementation, subject and adjunct clauses), a comparison with the acquisition of finite complementation, an investigation of other raising structures (such as passivization), an investigation of the possible semantic distinctions associated with the use of '*de*', '*à*', '*pour*' and a more precise determination of their categorical status, a more precise analysis of the influence of Jamaican Creole, a comparison of Jamaican learners with learners from other linguistic backgrounds (for instance, Australian learners of French), and of course an investigation of the possible pedagogical import of this research.

References

Achard M 2000 'Construal and complementation in French: the perspective dimension' in K Horie (ed) *Complementation* Amsterdam: John Benjamins. Pp.91-120.

Bailey BL 1966 *Jamaican Creole Syntax: a Transformational Approach* Cambridge: Cambridge University Press.

Borsley 2001 'What do prepositional complementizers do?' *Probus* 13:155–171.

Boskovic Z 1997 *The Syntax of Non-finite Complementation: an Economy Approach* Cambridge, MA: MIT Press.

Carnac-Marquis R 1996 'The distribution of à and de in tough-constructions in French' in K Zagona (ed) *Grammatical Theory and Romance Languages* Amsterdam/Philadelphia: John Benjamins. Pp. 35-46.

Chomsky N 1995 *The Minimalist Program* Cambridge, MA: MIT Press.

——— 1999 'Minimalist inquiries: the framework' in R Martin *et al.* (eds) *Step by Step. Essays on Minimalist Syntax in Honour of Howard Lasnik* Cambridge MA: MIT Press. Pp.89-156.

Chomsky N & H Lasnik 1977 'On filters and control' *Linguistic Inquiry* 8:425-504.

Collins P 2009 *Modals and Quasi-Modals in English* Amsterdam/New-York: Rodopi.

Cox TJ 1983 'Teaching the unteachable' *French Review* 57:168-178.

Delattre P 1964 'Le jeu des prépositions dans l'enchainement des verbes en français' *French Review* 38:67-81.

Durrleman-Tame S 2008 *The Syntax of Jamaican Creole: a Cartographic Approach* Amsterdam/Philadelphia: John Benjamins.

Gass S & L Selinker (eds) 1992 Language Transfer in Language Learning Amsterdam/Philadelphia: John Benjamins.

Granger S, J Hung & S Petch-Tyson (eds) 2002 *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* Amsterdam/Philadelphia: John Benjamins.

Gross M 1968 *Grammaire transformationnelle du français* Paris: Larousse.

Haegeman L 1991 *Introduction to Government and Binding Theory* Oxford: Blackwell.

——— 1992 'La théorie générative et l'acquisition des langues secondes: La position des adverbes en français et en anglais' *Cahiers de linguistique française* 13:148-182.

Hawkins R 1989 'Do second language learners acquire restrictive relative clauses on the basis of relational or configurational information? The acquisition of French subject, direct object and genitive restrictive relative clauses by second language learners' *Second Language Research* 5:156-188.

Henry A 1995 *Belfast English and Standard English: Dialect Variation and Parameter Setting* Oxford: Oxford University Press.

Herschensohn J 2007 'L2 functional categories: syntactic development and morphological mapping' in D Ayoun (ed) *French Applied Linguistics* Amsterdam/Philadelphia: John Benjamins. Pp.103-129.

Huot H 1981 *Constructions infinitives du français: le subordonnant 'de'* Genève: Droz.

Jones MA 1996 *Foundations of French Syntax* Cambridge: Cambridge University Press.

Kaye A 2005 'Gemination in English' *English Today* 21:43-55.

Kayne R 1981 'On certain differences between French and English' *Linguistic Inquiry* 12:349-371.

——— 1991 'Romance clitics, verb movement and PRO' *Linguistic Inquiry* 22:647-686.

——— 1999 'Prepositional complementizers as attractors' *Probus* 11:39-73.

Kayne R & I Haik 1980 'De certaines différences entre le français et l'anglais' *Langages* 14:47-64.

Lardière D 1998 'Dissociating syntax from morphology in a divergent L2 endstate grammar' *Second Language Research* 14:359-375.

——— 2000 'Mapping features to forms in second language acquisition' in J Archibald (ed) *Second Language Acquisition and Linguistic Theory* Oxford: Blackwell. Pp.102-129.

——— 2009a 'Some thoughts on the contrastive analysis of features in second language acquisition' *Second Language Research* 25:173–227.

——— 2009b 'Further thoughts on parameters and features in second language acquisition: a reply to peer comments on Lardiere's 'Some thoughts on the contrastive analysis of features in second language acquisition' *Second Language Research* 25(2)' *Second Language Research* 25:409–422.

MacWhinney B 2000 The *CHILDES Project: Tools For Analyzing Talk Mahwah* Hillsdale: Lawrence Erlbaum.

Martineau F 1991 'Clitic climbing in infinitival constructions of Middle French in D Wanner & D Kibbee (eds) *New Analyses in Romance Linguistics* Amsterdam/Philadelphia: John Benjamins. Pp. 235-252.

Martineau F & V Motapanyane 2000 'Infinitive subordinators and verb restructuring in French' in S N Dworkin and D Wanner (eds) *New Approaches to Old Problems: Issues in Romance Historical Linguistics* Amsterdam/Philadelphia: John Benjamins. Pp.217-231.

Myles F 1996 'The acquisition of interrogatives by English learners of French: the role played by structural distance' in K Sajavaara (ed) *Approaches to Second Language Acquisition* Jyväskylä: Jyväskylä Cross-Language Studies. Pp.195-208.

Pérez-Tattam R 2011 *Second Language Competence: The Acquisition of Complex Syntax in Spanish* Newcastle upon Tyne: Cambridge Scholars Publishing.

Peters H 2005 'Development of a longitudinal oral Interlanguage corpus of Jamaican learners of French' *Caribbean Journal of Education* 27:75-96.

——— 2006 'The acquisition of negative construals by Jamaican learners of French: A pedagogical orientation' *Proceedings of the International Conference: Research on Second and Foreign Language Acquisition and Teaching*, Paris, organized by Groupe Langues en contacts et appropriations, DILTEC, Paris III (Sorbonne) (Sept. 6-8, 2006). 18 p. http://www.groupelca.org/h/colloque2006/actes.html)

——— 2009 'Lexical features in oral interlanguage: The acquisition of multifunctional words by adult Jamaican learners of French' The International Journal of the Humanities 7:51-62.

——— Acquisition of French Non-finite Complementation: Transfer in a Diglossic Context. ms. UNSW.

Prévost P 2009 *The Acquisition of French: The Development of Inflectional Morphology and Syntax in L1 Acquisition, Bilingualism, and L2 Acquisition* Amsterdam/Philadelphia: John Benjamins.

Rankin T 2009 'Verb second in advanced L2 English' in M Bowles *et al.* (eds) *Proceedings of the 10th Generative Approaches to Second Language Acquisition Conference (GASLA 2009)* Somerville, MA: Cascadilla Proceedings Project. Pp.46-59.

Rizzi L 1978 'A restructuring rule in Italian syntax' in S J Keyser (ed) *Recent Transformational Studies in European Languages* Cambridge, MA: MIT Press.

——— 1997 'The fine structure of the left periphery' in L Haegeman (ed) *Elements of Grammar* Dordrecht: Kluwer. Pp.281-337.

Rooryck J 1988 'Les verbes de contrôle métaphoriques' *Revue romane* 23:3-8.

——— 1997 'On the interaction between raising and focus in sentential complementation' *Studia Linguistica* 50:1-49.

Rowlett P 2007 *The Syntax of French* Cambridge: Cambridge University Press.

Schwartz B & R Sprouse 1996 'L2 cognitive states and the Full Transfer/ Full Access model' *Second Language Research* 12:44-72.

——— 2007 'Linear sequencing strategies or UG-defined hierarchical structures in L2 acquisition? A reply to Meisel' in S. Karimi, V Samiian & W Wilkins (eds) *Phrasal and Clausal Architecture: Syntactic Derivation and Interpretation* Amsterdam/Philadephia: John Benjamins. Pp.295-318.

Tranel B 1987 *The Sounds of French: an Introduction* Cambridge: Cambridge University Press.

Valenzuela E 2008 'On CP positions in L2 Spanish' in J Liceras, H Zobl & H Goodluck (eds) *The Role of Formal Features in Second Language Acquisition* New-York/London: Lawrence Erlbaum. Pp.534-560.

White L 2003 'On the nature of interlanguage representations: Universal Grammar in the second language' in C Doughty & M Long (eds) *The Handbook of Second Language Acquisition* Oxford: Blackwell. Pp.19-42.