

Can linguistic features discriminate between the communicative CEFR-levels? A pilot study of written L2 French

Fanny Forsberg & Inge Bartning
Stockholm University

The study presents results from an ongoing project which tries to match communicative abilities as proposed in the CEFR- scale with the development of linguistic proficiency features. The main purpose is to look for linguistic features of each CEFR-level (A1-C1), in terms of morpho-syntax, discourse organisation and use of formulaic sequences. The selected linguistic phenomena have already been shown to be 'criterial' for acquisitional orders in oral L2 French. To this end, written data have been collected from 42 Swedish university students of L2 French. The method implied included placement of the students on CEFR.scales by DIALANG, production by the students of written argumentative texts and summaries according to CEFR-criteria criteria, raters' judgements and, finally, narrow linguistic analysis of the same productions.

The first results show that

- Measures of morphosyntactic deviances yield significant differences between the CEFR-levels up to B2
- Links can be observed between already established late acquisitional features, like *gérondif*, *dont* and *plus-que-parfait*.

Use of lexical formulaic sequences increases at higher CEFR-levels, but significant differences were found only between A2/B2/C2.

1. Introduction

This chapter presents results from an ongoing project which investigates whether it is possible to find correlations between the development of certain linguistic interlanguage features and the proposed language proficiency levels of the CEFR scale (Council of Europe, 2009). The linguistic phenomena investigated in this study are morpho-syntax (NP and VP morphology), discourse (discourse markers and subjunctions) and the use of formulaic language.

This linguistic analysis has been motivated by the model of six developmental stages of morphosyntax and discourse in oral French as presented in Bartning and Schlyter (2004, p. 282) which was elaborated as an empirically

based bottom-up construct (see Appendix). This study presented results from work using the InterFra corpus, Stockholm University (Forsberg, 2008; Hancock, 2000; Kirchmeyer, 2002) (<http://www.fraitu.su.se.interfra>) and the Lund corpus (see Granfeldt, 2003; Schlyter, 2003). The theoretical underpinnings of the Bartning and Schlyter model (2004, p. 281) include work by Klein and Perdue (1997, the ESF programme with 5 target languages including French L2), studies of grammaticalisation processes (form/function relations, see Bybee & Hopper, 2001) and processability theory (Pienemann, 1998).

The theoretical perspective in this chapter is guided by work on developmental stages (Bardovi-Harlig, 2006; Bartning & Schlyter, 2004; Sharwood-Smith & Truscott, 2005) and formulaic language (Erman & Warren, 2000; Wray, 2008). For earlier studies concerning acquisition and learning routes, esp. in Europe, see the ESF project (Perdue, 1993) and the Pavia project for Italian L2 (Giacalone Ramat, 1992). For studies on accuracy in the tradition of CAF (complexity, accuracy and fluency), see e.g. Van Daele, Housen, Kuiken, Pierrard, and Vedder (2007).

The choice of the three linguistic domains, viz. morphosyntax, discourse phenomena and formulaic language, is motivated by the fact that they have been shown also in recent studies on the InterFra corpus (Bartning, Forsberg, & Hancock, 2009; Bartning & Hancock, in press) to be discriminators in the development of L2 French. The morpho-syntactic areas concern verbal and nominal morphology (see e.g. overviews Ågren, 2008; Granfeldt & Nugues, 2007; Herschensohn, 2006; Véronique, 2009).

Some researchers consider developmental stages as one of the main findings in SLA (see Ellis, 2008, p. 72; Long, 2009), others have recently started to question them (see Hulstijn, this volume; Larsen-Freeman, 2006). This scepticism is thus expressed by Larsen-Freeman (2006) in her article about an emergentist perspective in SLA. She proposes individual profiles permitting great variation with many paths to development suggesting that development of learner language is not discrete and stage-like but more like 'the waxing and waning of patterns', (p. 590). The perspective of interlanguage as a dynamic process is, however, caught by the well-found terms by Bardovi-Harlig (2006, p. 69), 'main routes' and /or 'individual paths' according to different sources of influences.

One of the main aims of SLATE (see Introductory chapter) is to relate communicative development, as expressed by the CEFR-scale, to linguistic development, where the different chapters show examples of various linguistic domains and structures. However, it is important to stress already at this stage that the present study does not make a clear-cut distinction between communicative and linguistic development, since our CEFR-raters use both the communicative criteria stated in general proficiency scales (Finnish National Certificates of Language Proficiency, based on the CEFR general proficiency

scales) and the more language-oriented criteria presented in the manual *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (Council of Europe, 2009). These latter criteria are quite general and vague, however, and do not make reference to language-specific structures, in this case French. One example goes as follows: “Maintains consistent *and highly accurate* grammatical control of *even the most complex language forms*. *Errors are rare and concern rarely used forms* (Council of Europe, 2009, p. 187). Accordingly, our study presents linguistic profiles of written productions that have been rated as belonging to the same general CEFR-levels, based on communicative and/or language-oriented criteria (the two different scales mentioned above). Our aim is thus mainly to map what language-specific features characterize the CEFR-levels in written L2 French.

To our knowledge there has been little earlier work on French L2 concerning the relation between the pragmatic functions described in the CEFR levels and corresponding developmental linguistic features. However, for other L2s such as English, Finnish, Italian, and Dutch there are now several studies as illustrated by this volume (cf. Kuiken, Vedder, & Gilabert, this volume; Martin, Mustonen, Reiman, & Seilonen, this volume; Pallotti, this volume).

Following the reasoning above, our two research questions are:

1. Is it possible to establish linguistic developmental features to match the general communicative CEFR levels? And, if so,
2. How do the proposed linguistic domains of morphosyntax (Bartning & Schlyter, 2004, VP and NP morphology), discourse markers/subjunctions (Hancock, 2000, 2007) and formulaic sequences (Forsberg, 2008) develop along the CEFR-levels?

In this study we focus on written data since there has not been as much research carried out on written French L2 as on spoken French L2, although some studies have worked with written corpora (Ågren, 2008; Bolly, 2008; Granfeldt & Nugues, 2007; Granger, Hung, & Petch-Tyson, 2002) but not in relation to the CEFR scale. In addition, written data is less time-consuming to collect than oral data and it was therefore agreed upon in the SLATE-group that it would be convenient to start off with written production. We stress the point that in order to relate the communicatively defined levels of CEFR to linguistic development, it is appropriate to start with testing students' communicative abilities according to the CEFR levels (cf. Hulstijn, 2006; Martin et al., this volume) and then analyse their productions in terms of linguistic categories.

The general plan of the chapter is as follows: methodological issues are discussed in section 2, the main results from the three domains under investigation, viz. morpho-syntax, discourse markers and formulaic sequences, are shown in sec-

tion 3 followed by the conclusion, section 4. Section 3.1 investigates the non-target-like forms of morpho-syntactic categories drawn from the stages of Bartning and Schlyter and adjusted for written French. The second part of this investigation (section 3.2), which is considered as a qualitative complement to the first part, concentrates on the emergence of a limited number of morpho-syntactic and discursive target features. The third part of the analysis (3.3), focusses on formulaic language and is a quantitative study of the use of a specific category of formulaic language, viz. lexical formulaic sequences, which has been shown to be successful at discriminating between developmental levels (cf. Forsberg, 2008).

2. Methodological issues

2.1 Participants, data collection and tasks

The participants were recruited among students of French at Stockholm University, from various levels during 2007-2008. Most participants were 1st or 2nd term students of French, which explains the fact that many of the participants are placed at the B1 level and that fewer participants are to be found at the highest CEFR levels according to the DIALANG test (see below).

The students (N= 42) were gathered in the computer room of the language laboratory at Stockholm University, where they were asked to perform three tasks during the course of 2-2.5 hours. Time constraints were thus not entirely rigid, but no one was allowed to spend more than 2.5 hours on all of the tasks. The participants were not allowed to use any aids, such as dictionaries or grammar books, and the spell and grammar check had been deactivated on the computers used for the tasks.

The tasks

- 1) Participants were placed at the CEFR level by the DIALANG test, but only using the sub-test devoted to written production. In order to receive a CEFR level from the DIALANG test, the test taker had to first of all pass the vocabulary placement test, then take the self-assessment test, before finally taking the diagnostic test measuring written production skill. The level of the diagnostic test reported by the student is the level taken into consideration when administering the tasks to the students.
- 2) Participants then performed two written tasks – one which was given to all levels, and one which was specific to their estimated CEFR level.

The written tasks were developed by the authors of this article and were modelled on the tasks in the Cefling project (see Alanen, Huhta, & Tarnanen, this

volume; Martin et al., this volume). They were also developed, as were the Cefling tasks, in order to test communicative abilities as stated in the CEFR descriptors, such as expressing personal views and summarizing. The tasks were thus not designed to trigger some specific linguistic features, but rather, to test communicative abilities. Text length was not indicated, but there was a time limit for the whole set of tasks as indicated above.

Task 1: Given to all levels

Subject: Write a summary of a film that you have seen or a book that you have read recently.

Task 2: Specific for each CEFR level

A2: Write an e-mail to a friend and tell him/her about what you did last weekend.

B1: Argumentative/personal task: Why is it important to learn French?

B2: Argumentative/topic-based task: Can the individual do anything to counter the climate threat?

C1: Genre-specific task: Write a letter of complaint to the “Préfecture de police” regarding permission to stay in France.

C2: Genre-specific task: Write an application letter to a university/school/publisher in France.

(The six CEFR levels are A1, A2, B1, B2, C1 and C2.)

2.2 The rating procedure

The rating procedure selected for this particular study needs to be briefly discussed. Because there is risk of circularity, Hulstijn (this volume) claims that it is not appropriate to use the linguistic scales of the CEFR in the rating procedure if the aim is to relate the *communicative* CEFR levels to *linguistic* features of development. His position is theoretically laudable, but in our view, it entails practical limitations. Which professional language proficiency rater – be he/she trained in the CEFR or not – will not take linguistic form into account at some level, especially in the written modality? Can a person judge a written text, without noticing, for instance, grammar and orthography (cf. Alanen et al., this volume)? For a study aiming at teasing apart communicative adequacy (according to the CEFR) and linguistic complexity (both according to raters’ perception and according to general CAF measures), see Kuiken et al. (this volume).

We have come to the conclusion that it is possible to use raters, such as ours, who take both communicative function and linguistic form into account when rating a task (see criteria below), especially since the trained raters themselves propose this procedure. However, we do not know *which* linguistic features (morphosyntax, discourse and formulaic sequences) seem to discriminate between the

CEFR-levels, determining whether a learner is placed at a B1 level and not a C1 level and this will hopefully constitute the main contribution of the present study.

After the data were collected, the tasks were rated by professional CEFR raters, in order to verify whether the students had, in fact, performed at the CEFR-level for which they had been tested. One main rater of French at the University of Jyväskylä rated all 83 productions. (Unfortunately experienced CEFR raters are rare in Sweden and the Jyväskylä rater was recommended by colleagues.) For some productions, in case of difficult rating decisions, a second rater (also from Finland) was involved to ascertain the main rater's decision (this procedure was suggested by the main rater, as it corresponded to their practices). The raters were asked to use their regular rating practices, which involved making use of the following criteria:

1. *Finnish National Certificates of Language Proficiency* (based on the CEFR levels). In Finland, extensive work has been done at e.g. the University of Jyväskylä to align the national language tests with the CEFR (cf. Alanen et al., this volume). As a result, an evaluation scale, based on the six CEFR levels, is used when testing both Finnish as a second language and modern languages in school and higher education. Given their closeness to the CEFR scale, the raters could thus make use of the criteria stated for each level.

The level is described holistically, taking into account most language skills such as comprehension, writing and speaking. A few sentences are devoted specifically to written production, but they are quite general e.g. "is able to write both private and semi-official texts and to express thoughts as a coherent whole." (http://www.jyu.fi/hum/laitokset/solki/yki/english/about/skill_level/).

2. The other criteria are taken from "*Relating Language Examinations to the CEFR – Written assessment criteria grid*" (http://www.coe.int/T/DG4/Linguistic/Manuel1_EN.asp). Besides the criteria from the Finnish National Certificate, the raters also referred to the manual provided by the Council of Europe (2009), which contains more detailed linguistic criteria, developed in order to facilitate the raters' work.

However, it is precisely these criteria that are regarded as problematic by the SLATE-group, since they are not empirically validated through second language acquisition research. Below follows an example from the written assessment criteria grid, B1 level, overall rating (Council of Europe, 2009, p. 187):

"Can write straightforward connected texts on a range of familiar subjects within his field of interest [1], by linking a series of shorter discrete elements into a linear sequence [2]. The texts are understandable but occasional unclear expressions and/or inconsistencies may cause a break-up in reading."

As indicated by this citation raters thus use both holistic and general criteria [1] as well as criteria that contain more linguistic specifications [2].

However, the more specific criteria do not include any particular linguistic structures to which the raters should pay attention, let alone any structures in French, but they do direct the rater's attention, at least to some extent, to the linguistic form of expression and not only the communicative value of the text.

Following the rating procedure described and discussed above, the 83 productions by the 42 writers (one writer produced only one task), were grouped as shown in Table 1 below. As stated above, most productions were found to be at the B1 level. For statistical comparisons to be made between the levels, only levels A2, B1, B2 and C2 had enough values for the statistical analysis to be performed, hence their marking in bold.

Table 1. The CEFR levels of the productions after the rating

CEFR level (N: participants)	N: productions	N: words	Mean number of words/scriptor
A1 (1)	2	76	38
A2 (6)	12	1831	152
B1 (22)	43	11890	276
B2 (8)	16	5632	352
C1 (2)	4	1399	349
C2 (3)	6	2620	437
Total (42)	83	234 448	

3. Analysis: Linguistic criteria in the written productions

As stated in the introduction, potentially discriminatory features have been selected based on the results of many years of research on French as a second language, e.g. the work of the InterFra-project in Stockholm and the Corpus Lund at Lund University as shown in Bartning and Schlyter (2004). Besides the work on French morphosyntax, the InterFra-project has also been investigating the development of certain discursive features such as connectors (Hancock, 2007) and also the development of lexical competence as manifested in formulaic language (Forsberg, 2008). These different linguistic phenomena, which have to date been shown to be fruitful measures for the description of oral L2 French development, have now been investigated in the present *written corpus*.

Other studies which have independently confirmed the proposed six stages by Bartning and Schlyter (2004) are Housen, Kemps, and Pierrard (2008, oral data), Labeau (2009, written and oral data) and Bolly (2008, written data) for the advanced levels (stages 4-6), and Granfeldt and Nugues (2007) and Ågren (2008) for stages 1-4 in written production. Véronique (2009) also refers to the stages of Bartning and Schlyter (2004) when describing the development of French IL grammar in written and oral data.

It may seem methodologically problematic to apply criteria used in an oral corpus to a written one, but as the studies cited above on written French L2, e.g. Granfeldt and Nugues (2007) and Ågren (2008), have already shown, the criteria proposed in Bartning and Schlyter (2004) work surprisingly well even for written French.

The main criteria for establishing the developmental stages were the following: utterance structure, finiteness, verb morphology, subject – verb agreement, tense, mode and aspect (TMA), negation, noun phrase morphology, gender marking and discourse phenomena. (See Appendix, for a presentation of criteria and stages; and for the InterFra corpus, see <http://www.fraita.su.se/interfra>).

It is important to take into account that the morpho-syntactic stages proposed in Bartning and Schlyter (2004) were not originally meant to be related to the CEFR-levels. It may be tempting, however, to simply map the two scales onto each other, since they both contain six levels or stages, but this cannot be done automatically. The CEFR-scale is developed for all language skills, is supposed to be language independent and is based on teachers' experiences, ranging from the earliest levels to the most advanced level, the levels all being hypothetical. The Bartning and Schlyter stages, on the other hand, are also hypothetical as stages, but at the same time they are empirically based clusters of developmental features / sequences (*itinéraires acquisitionnels*) of two oral corpora of French interlanguage, which obviously offers advantages and disadvantages compared to the CEFR-scale. The obvious advantage is the argument of objectivity; they are based on how actual L2 production develops, without involving personal experiences. The most obvious disadvantage is that they are limited by the corpora and the levels attained by the learners in these corpora. As a consequence, the most advanced level in the Bartning and Schlyter continuum corresponds to the most advanced learners in the corpora at hand, but not necessarily to the most advanced levels of the CEFR-scale (C1-C2).

As already stated, section 3.1 below concentrates on the non-target-like forms of the categories taken from the stages of Bartning and Schlyter (2004) and adjusted for written French. They are here called morpho-syntactic deviances (MSDs). This first analysis of the 83 written productions is a quantitative morpho-syntactic one that takes into account a number of features. The second part of the study (section 3.2) has isolated a limited number of morpho-

syntactic and discursive target-like features, based on the InterFra corpus, and tries to trace the emergence of the features and their use at the different CEFR levels.

The third part (3.3) focusses on formulaic language and offers a quantitative study of the use of a specific category of formulaic language, viz. lexical formulaic sequences, which have been shown to be successful at discriminating between levels, in a number of studies such as Forsberg (2008) and Bartning et al. (2009). These well-documented features discriminating among linguistic levels are now related to the CEFR-based written productions of the study and thus to SLATE work.

3.1 Morpho-syntactic deviances

3.1.1 Presentation of the morpho-syntactic categories

In order to combine the CEFR levels with grammatical development in French L2, we propose the areas of developmental features of NP and VP morphology. In our study we have made a first screening of these features in written L2 French. Space limits do not permit us to problematize the method of identification of these MSDs (morphophonological rules, audible/non audible oppositions, orthography etc.). Nevertheless, we have indicated in the classification below non-target-like/target-like oppositions.

Presentation of the morpho-syntactic categories of deviances (MSDs):

VP morphology

1. Subject-verb agreement, opposition plural/singular in person and number: *ils *sort* (TL (=Target Language): ils sortent), *j'*a* (TL: j'ai); *ils *a* (TL: ils ont),
2. Subject-verb agreement, opposition in person, number: *ils *parle* (TL: ils parlent), *je *peut* (TL: je peux), *il *peux* (TL: il peut), *c'est nous qui *pouvont* (TL: pouvons)
3. Tense, Mode and Aspect (TMA) simplification: the present tense form instead of the *passé composé* (PC), etc. PC instead of *plus-que-parfait* (PQP), non-finite forms instead of finite: *je *donnE*, or the opposite: a finite form instead of a non finite form: *je peux *parle* (TL: parler); *pour *s'occupaient* (TL: s'occupent)
4. TMA subjunctive: *il faut que *j'ai* (TL: j'aie), *que tu *as* (TL: tu aies)
5. TMA auxiliary: *j'*ai tombé* (TL: je suis tombé)
6. Clitic object: *je *lui aide* (TL: je l'aide) (In this study classified under VP since the choice of the object depends on the verb construction, *aider qn*, e.g.)

NP morphology

7. Gender and number on subject personal pronoun: *ils, elles, il/elle* (and their NTL variants)
8. Gender on definite article: **la père* (TL: le)
9. Gender on indefinite article: **un fille* (TL: une)
10. Number on nouns: *les *fille* (TL: les filles)
11. Gender, number on audible attributive adjectives: *une femme *fort* (TL: forte)
12. Gender, number on non-audible attributive adjectives: *une *joli fille* (TL: jolie), *des *joli* (TL: jolies) *filles*
13. Gender, number on audible predicative adjectives: *la fille est *fort* (TL: forte)
14. Gender, number on non-audible predicative adjectives: *la fille est *joli* (TL: jolie)
15. Naked nouns (without obligatory determiners): **liberté* (TL: la liberté)

3.1.2 Morpho-syntactic deviances (MSD) at the different CEFR levels

Table 2 below shows the raw figures of MSDs in the 83 written productions rated at the different CEFR-levels. A distinction has also been made between deviances belonging to the VP and the NP respectively. It becomes quite clear that most of the MSDs are found within the NP even in this written learner corpus, esp. A1-A2, B1 and B2 levels (as in the oral very advanced sub-corpora of InterFra, see Bartning et al., 2009). This tendency of many MSDs in NP morphology in written French also confirms the results of Ågren (2008).

Table 2. Results of morpho-syntactic deviances at six CEFR levels

CEFR-level (N=productions)	VP Total	NP Total	Total MSD	Total No of words
A1-A2 (14)	41 (33%)	83 (67%)	124	1.907
B1 (43)	74 (27%)	205 (73%)	279	11.890
B2 (16)	21 (30%)	48 (70%)	69	5.632
C1 (4)	14 (45%)	17 (55%)	31	1.399
C2 (6)	7 (41%)	10 (59%)	17	2.620
Total (83)				

Whereas table 2 above shows the raw figures of MSDs, table 3 below shows the mean values of MSDs / 100 words at four levels. The figures clearly indicate that the MSDs decrease with higher levels of CEFR, indicating an increase in

students' accuracy. The four selected levels in table 3 (A2, B1, B2, C2) contained sufficiently many productions to be submitted to statistical tests, while productions at the A1 and C1 levels did not¹.

Table 3. Mean number of morpho-syntactic deviances/ 100 words

CEFR-level	Mean value
A2	8.6
B1	3.2
B2	1.2
C2	0.4

Table 4 below shows statistically significant differences between five of the CEFR levels concerning morpho-syntactic deviances.

Table 4. Statistical results for the comparison between different CEFR-levels (One way ANOVA, Tukey-Kramer post-hoc test)

CEFR-level	P-Values
A2>B1	0.001
A2>B2	0.001
A2>C2	0.001
B1>B2	0.01
B1>C2	0.01
B2>C2	Not significant

The results are interesting, since statistical differences are found up to the B2 level, but not between B2 and C2 (Table 4). However, there is an important difference in the mean values between B2 and C2, namely that the number of MSDs diminishes (see Table 3), although the difference was not significant.

Furthermore, one interpretation of the results might be that morpho-syntactic development, as manifested in quantity of MSDs, reliably discriminates

¹ The software used for the statistical analysis, GraphPad InStat, did indicate that there were too few values in these given groups for the analysis to be performed.

Table 5. Quantitative results concerning the developmental features in the 83 productions at the six CEFR levels (raw and relative frequencies)

CEFR level	Total words	<i>dont</i>	<i>ce qui</i>	<i>ce que</i>	Gerund	Pluperf	Subj soit, soient	<i>donc</i>	<i>mais</i>	<i>parce que</i>	<i>pourtant</i>	<i>puisque</i>	<i>en effet</i>
A1	76	0	0	0	0	0	0	0	1	0	0	0	0
Token/total									1.32%	0.00%			
A2	1831	1	0	0	0	0	0	1	20	17	0	0	0
		0.05%						0.05%	1.09%	0.93%			
B1	11890	3	11	10	13	1	3	17	106	22	8	2	1
		0.03%	0.09%	0.08%	0.11%	0.01%	0.03%	0.14%	0.89%	0.19%	0.07%	0.02%	0.01%
B2	5632	3	8	5	12	0	4	6	40	7	4	1	1
		0.05%	0.14%	0.09%	0.21%	0.00%	0.07%	0.11%	0.71%	0.12%	0.07%	0.02%	0.02%
C1	1399	1	2	1	5	0	0	2	7	1	0	0	4
		0.07%	0.14%	0.07%	0.36%	0.00%	0.00%	0.14%	0.50%	0.07%	0.00%	0.00%	0.29%
C2	2620	5	3	1	5	4	1	4	16	1	3	3	1
		0.19%	0.11%	0.04%	0.19%	0.15%	0.04%	0.15%	0.61%	0.04%	0.11%	0.11%	0.4%
Total words 234 448													

between CEFR-levels up to level B2. Other linguistic criteria and a larger corpus are probably necessary to characterize the differences between the highest levels (B2-C1-C2).

It is interesting to note that Martin et al. (this volume) also see important differences between A2 and B1, suggesting that the step to B1, The Independent User, is an important one for the learner to take.

With the cautionary statements above in mind, we think that the results are nonetheless thought-provoking because they suggest that morpho-syntax does not develop beyond a certain CEFR-level of written L2 French. One could perhaps propose that problems in morphosyntax are stable and do not altogether disappear. This result seems to concur with recent results proposed by Bartning et al. (2009), even though the results from this last study of late oral French were not matched against the CEFR-scale. However, it did show that three different learner groups, all highly proficient, two being resident in the TL country and one being in a foreign language setting, did not differ significantly with respect to morpho-syntactic deviances. Surprisingly, the MSDs persisted through these high stages (according to the literature near-native speakers' grammar is native-like at very advanced stages, cf. von Stutterheim, 2003).

To sum up, morphosyntax seems to develop along the CEFR-scale up to the level B2, as measured by frequency of grammatical deviances. Our question now is, of course, what other specific grammatical and discursive criteria could be tested against the CEFR-levels. This question will be qualitatively explored in the following section.

3.2 Candidates for indications of tendencies of developmental features

As stated in the introduction, the selection of morpho-syntactic and discursive features (here: discourse markers (e.g. *donc*), and subjunctives (e.g. *puisque*)), is based on features typical of the development of French interlanguage in the SLA literature (for an overview, cf. Bartning, 2009; Herschensohn, 2006). Many of them are presented in the appendix. Among relative pronouns we find: *dont*, *ce que*, *ce qui* (Bartning & Schlyter, 2004; Flament-Boistrancourt, 1984; Hancock & Kirchmeyer, 2005); the TenseModeAspect category is represented by the pluperfect (Bartning, 2009; Howard, 2009) and the subjunctive (Bartning, in press; Howard, 2008): *qu'il soit*, *ils soient*. At higher levels gerund emerges (Kirchmeyer, 2002): *venant*, *en venant*. Finally, we find connectors and subjunctives, as both late and early features, such as *donc*, *pourtant*, *puisque*, *en effet*, *mais* and *parce que* (Hancock, 2000, 2007; Kirchmeyer, 2002). These features have been examined in 83 productions in a pilot study. The results in table 5 below show the raw figures and percentages of the frequency of the use of these different phenomena.

Let us now consider Table 5.

Concerning the relative pronoun *dont* (column 3), according to earlier research on French L2 in *oral* productions *dont* is acquired late. In Bartning and Schlyter (2004) it does not show up until stages 5-6, if at all (see Appendix). It is thus interesting that it appears more frequently from level B1 (even A2) and upwards here. *Dont* is an example of condensed syntax which is expected to turn up late in IL (cf. written L2 French: Flament-Boistrancourt, 1984; oral L2 French: Hancock & Kirchmeyer, 2005).

It is also interesting to see that the other relative pronouns as *ce qui*, *ce que* do not appear until B1, B2. These pronouns refer anaphorically to anterior clauses in utterances and thus presuppose complexity in utterance building, a feature that belongs to higher acquisitional levels (Kirchmeyer, 2002).

The gerund is also a late feature in oral production (Bartning & Schlyter, 2004, Table 3, p. 294). This construction is also a manifestation of complex and condensed syntax which belongs to high levels in oral proficiency. This has been shown by others, e.g. Kirchmeyer (2002). Interestingly, as the data presented in Table 5 show, we see that the gerund turns up in several productions from B1 level and upwards. It is not surprising that it appears here in written productions since it belongs to written genres more than to oral. In any case, it appears to discriminate between the A levels, which show no appearances, and the other levels, which show surprisingly many. In the Bartning and Schlyter (2004) oral corpora the gerund was extremely rare, with only occasional uses at stages 5-6. It would be interesting to investigate the use of the gerund in a larger corpus of spontaneous non-native and native productions, oral and written. A relevant factor which could be revealing is the explicit/implicit dichotomy, as written language invites the learner to reflect on his language and time permits metalinguistic control. This issue will be explored in future studies focusing on the differences between oral/written French interlanguage.

The less frequent feature of our illustrations is the use of the pluperfect and, if it is used at all, it occurs at the highest levels. This pattern follows earlier studies of French interlanguage, e.g. works by Howard (2009) and Bartning (2009).

The subjunctive verbal forms *soit*, *soient* appear also at B1. This is an acknowledged late feature in oral French according to Bartning and Schlyter (2004) and Howard (2008).

The selection of discourse markers and subordinations has been made in order to be independent of text genres, such as narrative and argumentative texts. As table 5 (above) also shows, *mais* and *parce que* turn up already as connectors at A1-A2 as in early oral French L2 (cf. Hancock, 2000). At levels A2-B1 there is a remarkable increase of *mais* and then a decrease at B2-C2. This result also reflects tendencies already found in oral French IL: the well-known overuse of *mais* in the rather limited repertoire of connectors in early IL. The

counts for *parce que* reveal an increase at A2 with very few occurrences at C1-C2, the hypothesis being that *parce que* is taken over by the use of other causal connectors such as *puisque* etc. Table 5 also clearly shows that the repertoire of different connectors (*donc*, *pourtant*, *puisque* and *en effet*) grows at B1 and the writers have access to more than just two connectors *mais* and *parce que*.

Table 5 also informs us that the grammatical structures *dont*, *ce que*, *ce qui*, the gerund, the pluperfect and the subjunctive, as well as the remaining connectors *donc*, *pourtant*, *puisque* and *en effet*, appear at all levels from B1. None of these phenomena, grammatical or discursive, turns up at the earlier CEFR levels A1-A2 (with the exception of two occurrences of *dont* and *donc*). They all belong to more elaborated language.

We now turn to the third domain of linguistic features to be investigated in this chapter in our search for potential candidates of developmental measures along the CEFR levels, namely, the use of formulaic language. In the conclusion, results from the investigations of the MSDs, discursive phenomena and formulaic language will be viewed together and be related to the CEFR levels.

3.3 Formulaic language: Lexical formulaic sequences in relation to the CEFR-scale

Formulaic language, such as collocations, idiomatic expressions and social routines, are known to be a stumbling block for second language learners and users (Schmitt, Grandage, & Adolphs, 2004; Wray, 2008). However, 'formulaic language' or 'formulaic sequences' is sometimes also used in the SLA literature to refer to unanalyzed sequences that help beginner learners to communicate before they master creative rules. Sometimes these formulaic sequences correspond to target-like formulaic sequences, such as 'You're welcome', but they can also be sequences which are only unanalyzed in the learner's production such as *Monique *j'habite* (Monique *I lives), in Myles, Hooper, and Mitchell's (1998) study. In the present study, these latter sequences will not be treated since the study only takes into account target-like, idiomatic sequences. Furthermore, "Idiomatic expressions and colloquialisms" are also mentioned in the Written Assessment Criteria grid of the CEFR (Council of Europe, 2009, p. 186) as a feature which is not mastered before the C2 level (the most advanced CEFR-level). This encouraged us even more to test this criterion on written L2 development.

Erman and Warren (2000) presented a taxonomy of formulaic sequences, where they were divided into Lexical, Grammatical and Discursive prefabs, based on their main function in language use. Forsberg (2008) and Lewis (2008) applied this taxonomy to second language data, French and English respectively, and both found that the Lexical prefabs were the sequences causing difficulties for second language learners. They were thus shown to be an efficient yardstick of second language proficiency, especially when distinguishing

between advanced formal learners and very advanced second language users residing in the TL community. Consequently, we decided to see how this category is used at the different CEFR levels.

3.3.1 Identification and classification of FSs

The two most commonly used ways of identifying formulaic language in corpora are the statistical method and the phraseological method (cf. Granger & Pacquot, 2008). The first method identifies recurrent sequences based on different statistical measures such as *log likelihood* and *Mutual information score* (MI). Words that occur together more often than predicted by chance are labelled as collocations (a category of formulaic language). This method is automatic and involves no human judgement. Within the phraseological methodology, on the other hand, the researcher identifies potentially formulaic/conventional sequences based on linguistic criteria related to syntactic, semantic and pragmatic restrictions.

In view of the small size of our corpus and the fact that the statistical method requires a large corpus, it was decided that a phraseological method would be more appropriate for this study.

The present study makes use of Erman and Warren's (2000) original categorisation of prefabs (their term) which was slightly modified in Erman, Forsberg, and Fant (2008). Sequences can thus be categorized into Lexical, Grammatical and Discursive types.

Only the Lexical FSs category will be presented in more detail here. For an overall presentation of the categorisation, see Forsberg (2008) or Forsberg (2010).

Lexical FSs:

Clausal:

je vous en prie ('you're welcome')

c'est pas grave ('that's OK')

métro, boulot, dodo (no equivalent)

Phrasal:

agréablement surpris ('positively surprised')

faire du sport ('practice a sport')

poser une question ('pose a question')

Lexical FSs incorporate at least one content word. They are used for extralinguistic reference (as opposed to grammatical and discursive FSs) and denote actions (such as *faire la fête* 'to party'), states (*avoir peur* 'to be scared'), objects (*pomme de terre* 'potato') and so on (Forsberg, 2008, p. 96). They are sub-clas-

sified into clausal and phrasal sequences. Clausal sequences are full clausal, propositional language-specific sequences, often with pragmatic connotations among which conversational routines are probably the best known whereas phrasal sequences are primarily used for their denotative meanings, and as a rule constitute phrases, sometimes with open slots, such as *X tenir X au courant de X* ('keep X posted on X').

When it comes to the practical identification of these sequences, Erman and Warren (2000) make use of the criterion *restricted exchangeability*. In order for a sequence to qualify as *conventional* (a prefab in their terminology), an exchange of one of the words for a synonymous word must always result in a change of meaning or a loss of idiomaticity (Erman & Warren, 2000, p. 32).

The first step in identification is to find the Lexical FSs that meet the *restricted exchangeability* criterion (Erman & Warren, 2000). This is then complemented by internet searches using Google.fr. The Google tests are carried out following a specific procedure. To test the extent to which restricted exchangeability applies to a sequence, an analogous sequence, which has been subject to one of the modifications listed below, is constructed. The modifications were established based on Erman and Warren (2000) and on empirical evidence, i.e. some of the modifications were found to be decisive through work with the data.

1. One of the words is exchanged for a synonymous word
2. One of the words is exchanged for an antonymous word (for example *ça marche mal* 'it works bad' instead of *ça marche bien* 'it works well')
3. Change of article (from definite to indefinite or absence of article)
4. Change of number (from plural to singular or vice versa)
5. Change in word order (for example *égalité femmes/hommes* 'equality women/men' instead of *égalité hommes/femmes* 'equality men/women')

For a sequence to be counted as formulaic, it has to appear at least twice as frequently on Google as any of the modified versions. To sum up, the methodology is based on linguistic criteria and the researcher's intuition, which is complemented by searches on Google.fr, in order to ascertain the researcher's intuitions.

3.3.2 Results Lexical FSs in relation to the CEFR levels

A quantitative study was carried out which calculated the number of Lexical FSs per 100 words in all of the groups. As observed earlier in this article, only groups A2, B1, B2 and C2 have enough productions to pass the statistical tests. The mean numbers of Lexical prefabs /100 words are shown in the table below.

Table 6. Mean value of Lexical CS at the CEFR levels

Level	Mean no Lexical FSs / 100 words
A2	1.01
B1	2.03
B2	3.07
C2	4.18
A2<B2	p<0.001
A2<C2	p<0.001
B2<C2	P<0.01

The numbers suggest that the higher the CEFR level, the higher the number of Lexical FSs. However, a statistical analysis using One-Way ANOVA with Tukey-Kramer post-hoc test showed that significant differences were only found between the following groups: A2 – B2 – C2. Consequently, table 6 above shows that there are differences between low levels and high levels of the CEFR-scale and also between intermediate and high levels, but it does not show significant differences between ‘neighbouring’ levels such as A2/B1 and B1/B2.

One possible reason for this lack of significance, which applies to at least B1/B2, is that we have very few participants at the B2 level. More participants would probably yield a statistical significance between the B1 and the B2 level. Longer texts would probably also yield more robust results, since lexis, due to its lower frequency in interlanguage, requires longer texts.

If the results obtained for Lexical FSs are compared to those regarding MSDs, two interesting aspects are found: MSDs are better at discriminating between each level of the CEFR up to B2 level, whereas Lexical FSs do not seem to render differences that are fine-grained enough to separate between e.g. A2 and B1. On the other hand, Lexical FSs do the job that MSDs do not succeed in doing, i.e. discriminate between the higher levels B2 and C2. Possibly, this is due to the fact that Lexical FSs develop modestly up to a certain level and that development continues even at the highest levels, possibly never ending, since we are dealing with the growth of lexis, which is constant even in the L1.

4. Summing up

This chapter thus investigated the linguistic development of grammatical, discursive and formulaic structures in productions made by informants placed at different CEFR levels.

Our research questions were:

1. Is it possible to establish linguistic developmental features to match the communicative characteristics of the CEFR levels? and, if so,
2. How do the proposed linguistic domains of morphosyntax (VP and NP morphology), discourse markers/subjunctions and formulaic sequences develop along the CEFR-levels?

As an answer to the first question on the basis of the investigation of 83 productions, the results above show that there is a decrease of MSDs across the levels with significant differences between the levels (see Tables 3 and 4). Thus it seems that linguistic features do discriminate between CEFR levels and more specifically the features chosen for this study. Measures of morpho-syntactic deviances thus yield significant differences between the CEFR levels up to B2. These results of late MSDs in highly proficient learners/users concur with findings in Bartning et al. (2009).

An answer to the second question, as shown in Table 5 above, is that there is development of a selection of grammatical and discursive features across the levels. These developmental features are represented by *dont*, *ce que*, *ce qui*, gerund, pluperfect and some connectors. Our study also presented results concerning the lexicon: it was shown that the use of lexical formulaic sequences increases at higher CEFR levels, but significant differences were only found between A2/B2/C2 (Table 6). This is yet another linguistic feature that can be used as a measure of progression in IL development and in the CEFR scale.

In the future, when the written CEFR corpus has been enlarged, the MSDs and their TL equivalents, as well as the developmental features in Table 5 above, will be examined in order to work out a developmental continuum of written French IL linked to the CEFR scale.

It thus seems, according to this study, that a relationship is to be found between linguistic development (the three different measures) and communicative development as expressed in the CEFR scales, at least as regards written production in L2 French. Possibly the three measures, i.e. morpho-syntactic deviances (accuracy), emergence and use of discourse/grammatical markers and, finally, the rate of lexical formulaic sequences could be proposed as constituting ingredients of a global index of interlanguage development. However, the measures and the rating procedures need to be further refined before drawing any further conclusions.

Acknowledgements

We thank Hugues Engel, Department of French, Italian and Classical languages, Stockholm University, for his valuable help with some data collection and relevant comments on earlier drafts of this paper and Victorine Hancock for her precious remarks. We also warmly thank our two raters from the University of Jyväskylä, Finland. Finally, we thank the anonymous reviewers of this volume for their valuable comments.

References

- Ågren, M. (2008). *À la recherche de la morphologie silencieuse. Sur le développement du pluriel en français L2 écrit* (Doctoral dissertation). Lund University, Sweden.
- Bardovi-Harlig, K. (2006). Interlanguage development. Main routes and individual paths. *AILA Review*, 19, 69–82.
- Bartning, I. (2009). The advanced learner variety: 10 years later. In E. Labeau & F. Myles (Eds.), *The advanced learner variety: The case of French* (pp. 11–40). Bern: Peter Lang.
- Bartning, I. (in press). Late morpho-syntactic and discourse features in advanced and very advanced L2 French – a view towards the end state. In S. Haberzettel (Ed.), *The end state of L2 acquisition*. Berlin: Mouton de Gruyter.
- Bartning, I., Forsberg, F., & Hancock, V. (2009). Resources and obstacles in very advanced L2 French. Formulaic language, information structure and morphosyntax. *EUROSLA Yearbook*, 9, 185–211. Amsterdam: Benjamins.
- Bartning, I., & Hancock, V. (in press). Morphosyntax and discourse at high levels of second language acquisition. In K. Hyltenstam (Ed.), *High level proficiency in second language use*. Berlin: Mouton de Gruyter.
- Bartning, I., & Schlyter, S. (2004). Itinéraires acquisitionnels et stades de développement en français L2. *Journal of French Language Studies*, 14(3), 281–299.
- Bolly, C. (2008). *Les unités phraséologiques: Un phénomène linguistique complexe?* (Unpublished doctoral dissertation). Université catholique de Louvain, Belgium.
- Bybee, J., & Hopper, P. (2001). *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins.
- Council of Europe. (2009). *Manual for relating language examinations to the Common European Framework of Reference for Languages* [Online version]. Retrieved from http://www.coe.int/t/dg4/linguistic/manuel1_EN.asp
- Ellis, R. (2008). *The study of second language acquisition*. Oxford: Oxford University Press.
- Erman, B., Forsberg, F., & Fant, L. (2008, October). *Nativelike selection in high-level L2 use*. Paper presented at High-level proficiency in a second language, Stockholm, Sweden.
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*, 20(1), 29–62.

- Flament-Boistrancourt, D. (1984). *La pratique des relatifs français chez les néerlandophones* (Unpublished doctoral dissertation). University of Lille III, France.
- Forsberg, F. (2008). *Le langage préfabriqué – formes, fonctions et fréquences en français parlé L2 et L1*. Bern: Peter Lang.
- Forsberg, F. (2010). Using conventional sequences in L2 French. *International Review of Applied Linguistics in Language Teaching*, 48, 25–51.
- Giacalone Ramat, A. (1992). Grammaticalisation processes in the area of temporal and modal relations. *Studies in Second Language Acquisition*, 14, 297–322.
- Granfeldt, J. (2003). *L'acquisition des catégories fonctionnelles. Étude comparative du développement du DP français chez des enfants et des apprenants adultes* (Doctoral dissertation). Lund University, Sweden.
- Granfeldt, J., & Nugues, P. (2007, June). *Évaluation des stades de développement en français langue étrangère*. Paper presented at TALN 2007, Toulouse, France.
- Granger, S., Hung, J., & Petch-Tyson, S. (Eds.). (2002). *Language Learning & Language Teaching: Vol. 6. Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam: John Benjamins.
- Granger, S., & Pacquot, M. (2008). Disentangling the phraseological web. In S. Granger & F. Meunier (Eds.), *Phraseology. An interdisciplinary perspective* (pp. 27–49). Amsterdam: John Benjamins.
- Hancock, V. (2000). *Quelques connecteurs et modalisateurs dans le français parlé d'apprenants avancés. Étude comparée entre suédophones et locuteurs natifs* (Doctoral dissertation). Cahiers de la Recherche 16. Stockholm University, Sweden.
- Hancock, V. (2007). Quelques éléments modaux dissociés dans le paragraphe oral dans des interviews en français L2 et L1. *Journal of French Language Studies*, 17, 21–47.
- Hancock, V., & Kirchmeyer, N. (2005). Discourse structuring in advanced L2 French: The relative clause. In J.-M. Dewaele (Ed.), *Focus on French as a foreign language* (pp. 17–35). Clevedon: Multilingual Matters.
- Herschensohn, J. (2006). Français langue seconde: From functional categories to functionalist variation. *Second Language Research*, 22, 95–113.
- Housen, A. Kemps, N., & Pierrard, M. (2008). The use of verb morphology of advanced L2 learners and native speakers of French. In E. Labeau & F. Myles (Eds.), *The advanced learner varieties: The case of French* (pp. 41–61). Bern: Peter Lang.
- Howard, M. (2008). Morpho-syntactic development in the expression of modality: The subjunctive in French L2 acquisition. *Canadian Journal of Applied Linguistics*, 11, 171–192.
- Howard, M. (2009). Short- versus long-term effects of naturalistic exposure on the advanced instructed learner's L2 development: A case study. In E. Labeau & F. Myles (Eds.), *The advanced learner varieties: The case of French* (pp. 93–123). Bern: Peter Lang.
- Hulstijn, J. (2006, December). *Linking meaning (function) and form (lexis, grammar, prosody)*. Paper presented at the SLATE Workshop, Amsterdam, Netherlands.

- Kirchmeyer, N. (2002). *Étude de la compétence textuelle des lectes d'apprenants avancés. Aspects structurels, fonctionnels et informationnels* (Doctoral dissertation). Cahiers de la Recherche 6. Stockholm University, Sweden.
- Klein, W., & Perdue, C. (1997). The basic variety (or: Couldn't natural languages be much simpler?). *Second Language Research*, 13, 301–347.
- Labeau, E. (2009). An imperfect mastery: The acquisition of functions of imparfait by anglophone learners. In E. Labeau & F. Myles (Eds.), *The advanced learner variety: The case of French* (pp. 63–90). Bern: Peter Lang.
- Larsen-Freeman, D. (2006). The emergence of complexity, fluency and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, 27, 590–619.
- Lewis, M. (2008). *The idiom principle in L2 English* (Doctoral dissertation). Stockholm University, Sweden.
- Long, M. (2009, September). *Second language acquisition and language teaching*. Paper presented at Stockholm University, Sweden.
- Myles, F., Hooper, J., & Mitchell, R. (1998). Rote or rule? Exploring the role of formulaic language in the foreign language classroom. *Language Learning*, 48, 323–364.
- Perdue, C. (1993). *Adult language acquisition: Cross-linguistic perspectives* (Vol. 2). New York: Cambridge University Press.
- Pienemann, M. (1998). *Language processing and second language development. Processability theory*. Amsterdam: Benjamins.
- Schlyter, S. (2003). Development of verb morphology and finiteness in children and adults acquiring French. In C. Dimroth & M. Starren (Eds.), *Information structure, linguistic structure and dynamics of learner language* (pp. 15–44). Amsterdam: Benjamins.
- Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived clusters psycholinguistically valid? In N. Schmitt (Ed.), *Formulaic sequences* (pp. 127–151). Amsterdam: Benjamins.
- Sharwood-Smith, M., & Truscott, J. (2005). Stages or continua in SLA: A MOGUL solution. *Applied Linguistics*, 26, 219–240.
- Van Daele, S., Housen, A., Kuiken, F., Pierrard, M., & Vedder, I. (Eds.). (2007). *Complexity, accuracy and fluency in second language use, learning and teaching*. Brussels: KVAB.
- Véronique, D. (Ed.). (2009). *L'acquisition de la grammaire du français, langue étrangère*. Paris: Didier.
- Von Stutterheim, C. (2003). Linguistic structure and information organisation. The case of very advanced learners. *EUROSLA Yearbook*, 3, 183–206. Amsterdam: Benjamins.
- Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford: Oxford University Press.

APPENDIX

Table 1. Overview of the six developmental stages proposed by Bartning and Schlyter (2004, p. 293)

STAGE	MORPHO-SYNTACTIC AND DISCURSIVE FEATURES
<p>Stage 1 – Initial stage</p> <p>This stage brings to mind the prebasic and basic varieties (see Bartning & Schlyter, 2004, p. 295; Klein & Perdue, 1997).</p>	<p>Utterance structure</p> <ul style="list-style-type: none"> - Nominal and non-finite utterance structure (<i>je // mes amis</i> 'I meet friends'); some bare nouns ; - formulaic expressions (<i>je ne sais pas</i> 'I don't know', <i>je voudrais</i> 'I would like') <p>Connectors</p> <ul style="list-style-type: none"> - Emergence of the connectors <i>et</i> ('and'), <i>mais</i> ('but') and <i>puis</i> ('then') <p>Negation</p> <ul style="list-style-type: none"> - Negation of the type Neg X (<i>non grand lit</i> 'no double bed') and preverbal negation <p>TMA</p> <ul style="list-style-type: none"> - Emergence and rare use of <i>passé composé</i> (very few contexts are marked for past tense) <p>Verb morphology</p> <ul style="list-style-type: none"> - Mostly non-finite verb forms but some finite verb forms; - no opposition between personal verb forms; <p>Gender</p> <ul style="list-style-type: none"> - Default value on the value of masculine/feminine on determiners <p>Pronouns</p> <ul style="list-style-type: none"> - Use of first person pronouns <i>je</i> without elision
<p>Stage 2 – Post-initial stage</p>	<p>Utterance structure</p> <ul style="list-style-type: none"> - Continued use of non-finite utterance structure but also some finite forms; - frequent use of the 'passe-partout' formulas <i>c'est</i> ('it is'), and some <i>il y a</i> ('there is'); - subordination (see below connectors) <p>Connectors</p> <ul style="list-style-type: none"> - Simple subordination with temporal, causal and relative clauses <p>Negation</p> <ul style="list-style-type: none"> - Use of the preverbal negation <i>ne</i> (without <i>pas</i>) along with the first uses of the TL form <i>ne ... pas</i>; <p>TMA</p> <ul style="list-style-type: none"> - Emergence of modal auxiliaries <i>pouvoir</i> 'be able to', <i>vouloir</i> 'to want to'; - increase of <i>passé composé</i> forms; - use of <i>imparfait</i> with <i>être</i> ('be') and <i>avoir</i> ('have'); - first uses of periphrastic future; <p>Verb morphology</p> <ul style="list-style-type: none"> - Still frequent use of non-finite forms; - emergence of forms of irregular verbs in the present plural 3rd person as <i>ils *prend</i> 'they take'; - subject – verb agreement between 1st and 2nd person singular of non-thematic verbs (<i>être, avoir</i>); - alternative forms between the verb forms of the 1st person present plural <i>nous V-ons</i> and the short form <i>*nous V</i> (<i>*nous parle</i> 'we speak') <p>Gender</p> <ul style="list-style-type: none"> - Continued default value on determiners - Some adjectival agreement <p>Pronouns</p> <ul style="list-style-type: none"> - Object pronouns in postposition of the verb

STAGE	MORPHO-SYNTACTIC AND DISCURSIVE FEATURES
<p>Stage 3 – Intermediate stage</p> <p>A regular and systematic interlanguage with overgeneralisations and analogies that make the system sometimes not TL (see Bartning & Schlyter, 2004, p. 295).</p>	<p>Utterance structure</p> <ul style="list-style-type: none"> - More finite utterance structures; - Subordination (see below connectors) <p>Connectors</p> <ul style="list-style-type: none"> - Use of causal, temporal, relative, interrogative clauses, and the subordinator <i>que</i> ('that'); overuse of <i>mais, parce que</i> <p>Negation</p> <ul style="list-style-type: none"> - TL use of the negation <i>ne ... pas</i> on the finite forms of the verbs, - Emergence of <i>ne ... rien</i> <p>TMA</p> <ul style="list-style-type: none"> - Use of <i>passé composé</i> with marked past contexts; - use of the periphrastic future; - use of the <i>imparfait</i> on lexical verbs - first use of isolated cases of <i>futur simple</i>; - first use of the subjunctive; <p>Verb morphology</p> <ul style="list-style-type: none"> - Still some non-finite verb forms; - the 1st person present plural is now mostly correct; - subject – verb agreement with opposition between the 3rd persons of singular and plural begins to be established with <i>avoir</i> and <i>être (al ont, est/ sont)</i>; - alternation between <i>ils *prendre</i> and <i>ils *prend</i> (with some cases of the correct <i>ils prennent</i>); <p>Gender</p> <ul style="list-style-type: none"> - More TL use on the definite determiner than on the indefinite; - overuse of the masculine (determiners, adjectives); - problems with adjectival agreement in preposition and attributive sentences <p>Pronouns</p> <ul style="list-style-type: none"> - Object pronouns placed before the lexical verb in simple and complex tenses (often incorrectly after the auxiliaries <i>est/ a</i>)
<p>Stage 4 – Low advanced stage</p>	<p>Utterance structure</p> <ul style="list-style-type: none"> - Multipropositional syntactic structures; discourse complexity that demands temporal and conditional expressions (see below TMA): the emerging forms of tenses, mode and aspect; these contexts are not always systematically marked by relevant forms; the form/function relations are not yet TL <p>Connectors</p> <ul style="list-style-type: none"> - Use of <i>alors, après, finalement, mais, parce que</i> and temporal <i>puis</i>; significant overuse of the polyfunctional markers <i>mais</i> and <i>parce que</i> <p>Negation</p> <ul style="list-style-type: none"> - Complex negation with <i>ne ... personne, jamais, rien</i> <p>TMA</p> <ul style="list-style-type: none"> - The <i>passé composé</i> and the <i>imparfait</i> are more and more TL; - Emergence of the typical French use of the conditional, the pluperfect, the subjunctive <p>Verb morphology</p> <ul style="list-style-type: none"> - The non-finite forms in finite contexts <i>je donne</i> 'I give' disappear, except for verb forms in /-r/ like <i>je lire</i> ('I read'); - the forms <i>ils ont, sont, vont, font</i> dominate over <i>ils * a/est/valfait</i>; - continued use of <i>ils *prend</i> but in competition with <i>ils prennent</i> <p>Gender</p> <ul style="list-style-type: none"> - Overuse of the masculine on determiners and adjectives <p>Pronouns</p> <ul style="list-style-type: none"> - Object clitic pronouns before verbs; - subject clitic pronouns with elision

STAGE	MORPHO-SYNTACTIC AND DISCURSIVE FEATURES
<p>Stage 5 – Intermediate advanced stage</p>	<p>Utterance structure</p> <ul style="list-style-type: none"> - Multipropositional utterances, some infinitives and the gerund ; - inflectional morphology becomes functional <p>Connectors</p> <ul style="list-style-type: none"> - Appearance of <i>donc</i>; native-like uses of <i>parce que</i> <p>Negation</p> <ul style="list-style-type: none"> - Use of <i>personne ... ne, rien ... ne</i> ; - TL use except for the omission of <i>ne</i> <p>TMA</p> <ul style="list-style-type: none"> - Development of inflectional morphology (subjunctive, conditional, pluperfect) - The pluperfect is not used in all obligatory contexts, nor is the conditional; <p>Verb morphology</p> <ul style="list-style-type: none"> - Fragile zones of morphology in multipropositional utterances <p>Gender</p> <ul style="list-style-type: none"> - Still overuse of the masculine in determiners; - problems in agreement with preposed adjectives (feminine) <p>Pronouns</p> <ul style="list-style-type: none"> - More or less TL use, even the relative <i>dont</i>
<p>Stage 6 – High advanced stage</p>	<p>Utterance structure</p> <ul style="list-style-type: none"> - High degree of embedding and of integrated propositions ; - capacity of keeping several information levels simultaneously in the same utterance; - discourse structuring according to L1 (fewer constituents than NS in the pre-frontfield, 'préambule') <p>Connectors</p> <ul style="list-style-type: none"> - TL use of connectors; use of <i>donc, enfin</i>; macro-syntactic relatives <p>Negation</p> <ul style="list-style-type: none"> - TL use; variation on the omission of <i>ne</i> <p>TMA</p> <ul style="list-style-type: none"> - Use of the pluperfect (sometimes still replaced by the <i>passé composé</i>); - the conditional in most contexts <p>Verbal Morphology</p> <ul style="list-style-type: none"> - Stabilised inflection which becomes functional ; some rare <i>ils *prend</i> may turn up in complex syntax/discourse ; <p>Gender</p> <ul style="list-style-type: none"> - Same as stage 5. Adjectival agreement TL but some form of the type <i>la *petit ville</i> may turn up; problems with gender, the indefinite determiner before feminine nouns starting with vowels <p>Pronouns</p> <ul style="list-style-type: none"> - TL in form and position

