

University of Groningen

Effectiveness of explicit vs. implicit L2 instruction

Rousse-Malpat, Audrey

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2019

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Rousse-Malpat, A. (2019). *Effectiveness of explicit vs. implicit L2 instruction: A longitudinal classroom study on oral and written skills*. [Groningen]: Rijksuniversiteit Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Chapter 4

Effects of a Structure-based vs. a Dynamic usage-based method in oral proficiency⁸

1. Introduction

If the past 50 years of research in L2 instruction has taught us anything, it is that the modern language classroom should be a place where learners are surrounded by meaningful L2 input and engage in motivating communicative activities in order to enhance their general proficiency level (Krashen, 1981; Dörnyei, 2002; Lantolf, Thorne & Poehner, 2015; Verspoor, 2017). However, in the Netherlands and probably many other countries in the world, many L2 teachers find it difficult to create an effective communicative setting. Most teachers resort to standard methods that are based on Structure-Based (SB) approaches, with a focus on grammatical practice as achieving grammatical accuracy is viewed essential (Graus & Coppen, 2018). This is not so surprising as most L2 instructional research also points to the effectiveness of some explicit focus on form (Norris & Ortega, 2000; Spada & Tomita, 2010; Goo et al., 2015). However, the effectiveness is usually measured in short-term experiments and usually only in terms of having an effect on grammatical accuracy and not overall proficiency. However, as Haijma (2013) points out, current methods used in the Netherlands fail to achieve a satisfying level of general L2 proficiency, particularly so for oral proficiency.

Recent developments in the application of Complex Dynamic Systems Theory (CDST) combined with Usage-Based (UB) theories on second language instruction, called a Dynamic usage-based (DUB) approach for short, might offer a good alternative for foreign language teachers who struggle with effective language teaching (see Rouse-Malpat & Verspoor, 2018 and Verspoor, 2017). Teaching methods in line with this approach do not see language as a set of rules but as conventionalized routines, where non-linear learning emerges from the dynamic interaction between input and output. Rather than focusing on grammar, DUB methods focus on “iteration” (Larsen-Freeman, 2012), i.e. frequently repeated exposure to utterances

8 This chapter is based on the following submitted article:
Rouse-Malpat, A., Steinkrauss, R. Koote, L. & Verspoor, M. (submitted). Parlez-vous français? Effects of structure-based vs. Dynamic usage-based approaches on oral proficiency.

in meaningful contexts (Verspoor, 2017) to foster automatization and routinization. In several semester-long experiments, these methods have proven as effective or more effective in a foreign language classroom in countries where language teaching is particularly traditional such as in Vietnam or Sri Lanka (Hong, 2013; Irshad, 2015) on general English proficiency tested by objective tests and by holistic scores on written and oral texts or in Germany for L2 Dutch (Koster, 2015). However, few studies only have looked at the effects of DUB inspired L2 instruction on free-production data over a longer period of time and with both analytical and holistic measures that favor SB as well as DUB approaches, and compared them to the effects of SB teaching that is less traditional than in the studies mentioned above.

In this chapter, the effects of two L2 teaching methods - SB and DUB - on general oral proficiency will be measured after three years of instruction in terms of both holistic and analytical measures. To do so, we interviewed 41 Dutch participants learning French as L2 in secondary school classrooms. Using the SOPA protocol for the interviews, we obtained free-production data that were holistically rated and then transcribed and coded for complexity, accuracy and fluency measures. The aim is to investigate whether there is a difference between the methods in the analytical measures that may favor the SB method and holistic measures that favor the DUB method. After clarifying the fundamental differences between SB and DUB approaches and their implications for instructional methods, we will briefly describe the Dutch instructional context and present the study.

2. Differences between Structure-based and Dynamic usage-based approaches

The fundamental difference between a SB and a DUB approach is the way language itself is viewed. An SB approach assumes language is a complex system in which different autonomous sub-components (such as syntax and lexicon) interact predictably according to “rules”. In contrast, a DUB approach assumes language is a complex dynamic system in which there is no fundamental difference between syntax and lexicon, and that language is used on basis of the individual routines of the speaker. Language learning and use is viewed as a dynamic, non-linear process. These different views have implications for how language should be presented and instructed and the behavior that the learners should aim for.

2.1 Structure-based

The term Structure-Based (SB) is based on Lightbown & Spada (2013). As the name implies, an SB approach assumes language is a highly structured entity, inspired, among others, by structural and generative linguistics, which see language as

highly systematic and driven by rules. Whereas structural linguistics focuses on language as a system of contrasts at different levels, e.g. contrasting morphemes or phonemes, generative linguistics focuses on the simplest, underlying principles that can account for syntactic forms in any language, the so-called Universal Grammar (UG). One of the most important aspects of UG is the concept of creativity, meaning that when a language user has detected the principles applying to their language, creative sentences can be generated, regardless of whether this sentence has appeared in the input before. The rules derived from these principles are general and non-symbolic and operate on lists of symbolic elements such as words and irregular forms (like past tense forms in English). Most of the research done within UG concerns syntactical and morphological structures and how these different and separate subsystems of language are related (Marsden, Mitchell and Myles, 2013). As both structural and generative inspired linguistics take morphological and syntactic forms as the core of language (Cook, 2016), they shape the view that grammar represents the heart of the linguistic system around which the other, more peripheral subcomponents of language (vocabulary, pronunciation, and so on) are built. To learn a language, the learner has to have access to the rules. This view of language is implicitly held by many researchers and educators and has inspired SB views on how languages are learned and taught.

Lightbown and Spada (2013) argue that SB approaches see language learning as rule driven. Generally, it is believed that learning and applying grammatical rules is beneficial in becoming proficient in an L2. There is a high focus on accuracy of the grammatical forms that are presented from “simple” to “complex” and other aspects of language such as vocabulary, formulaic phrases, pronunciation, intonation, pragmatic use and so on are believed to be learned separately. Most of our L2 teaching methods, whether they be audio-lingual, communicative, task-based, or skill theory are Structure-based in that they implicitly or explicitly build on the premise that grammar forms the core of the language to be learnt.

Perhaps inspired by behaviorist theories, the idea of maintaining accuracy from the earliest L2 utterance, has also become the rule. Behaviorist methods teach the grammatical forms with structure drills and pattern practice, and as Cook (2016) points out, such practices are still often evident in contemporary course books. To avoid fossilization, errors are seen as bad habits that have to be prevented or corrected as soon as possible, either through more explanation and practice or through remediation in various forms. Therefore, to be sure learners understand the difficult grammatical concepts, rules are usually explained in the L1 (West and Verspoor, 2016) and even though there are exceptions as we will

see in the current study, most structure-based programs may be considered input-poor environments (Kouraogo, 1993), especially if teachers not only teach grammar in the L1 to be sure the learners understand them but also give instructions and do comprehension checks in the L1.

Many studies have been conducted to investigate the effects of structure-based methods on second language proficiency. Long (1991) made a difference between focus-on-forms approaches, similar to what we mean with SB instruction and focus-on-form approaches, where grammar instruction is provided in a meaningful communicative setting and claim that focus-on-forms instruction is less effective than focus-on-form. Several meta-analyses and studies support this claim and find a beneficial effect for focus on form instruction (Norris & Ortega, 2000; Spada & Tomita, 2010; Doughty, 2003). The effects of different types of instructions are almost exclusively focused on grammatical forms.

Some researchers maintain that exposure is more effective than traditional grammar instruction (Krashen, 1981). Comparing two groups with different amounts of exposure but the same amount of instruction, Martin (1980) found that adult second language learners of English with more exposure scored higher on TOEFL and on other grammar, reading, composition and speaking tests. However, as immersion programs (Harley & Swain, 1984; Day & Shapson, 1991) have shown, this does not mean that exposure alone is sufficient, particularly in the development of accuracy, but that attention to form should be given in classrooms. In line with this idea, Tomasello & Herron (1989) asked their participants to translate sentences containing the verb “to know” from their L1 to their L2 (French) in a study on corrective feedback. In French, “to know” can be translated either by “savoir” or “connaitre”, depending on the grammatical category of the following word: one would choose “savoir” before a verb and “connaitre” before a noun. Results showed that the group who had the explicit corrective feedback of the structure did perform better. On the grammatical teaching of the conditional form in French, Day & Shapson (1991) also found a beneficial effect for grammatical instruction on writing but not for speaking.

In sum, as far as focus-on-forms in an L2 curriculum is concerned, some researchers claim that it is ineffective and thus should be removed from the curriculum; others, such as DeKeyser (1997), are still in favor of including formal structure instruction in teaching because learners need to automatize explicit knowledge. Ellis, R. (2006b) adds that focus on forms approaches may be effective as long as they give the “opportunity for learners to practice behavior in communicative tasks” (p.102).

Summarizing this section, we can say that even though many experts see explicit focus on form (within a meaning-based approach) as more effective than a focus on forms (or Structure-based) approach in Foreign Language teaching, structure-based approaches are still the norm, at least in the Netherlands. Nevertheless, even experts who advocate communicative focus on form approaches still see language as driven by rules that need to be learned, as these are the items that are focused on exclusively, at the expense of general proficiency. It is true that evidence in general points to a beneficial effect of grammar instruction on the acquisition of grammatical rules, but much of the evidence is based on short-term interventions that test mainly for accuracy on the grammatical rules (Norris & Ortega 2000; Spada and Tomita, 2010; Goo et al., 2015). Very few, if any studies, are long-term and have looked at overall proficiency gains, especially in oral production. Finally, as Doughty (2003) points out, in the meta-analyses many issues may have been confounded, and as Andringa and Schultz (2016) have argued, the amount of exposure has not been controlled for adequately.

2.2 Dynamic usage-based

A Usage-Based perspective on L2 learning would predict that language emerges from repetitive exposition to meaningful input and language use (Langacker, 2000; Tomasello, 2003). Linguistic constructions (pairings of form and meaning) are learned through association as they are “heard and used frequently and therefore entrenched, which is the result of habit formation, routinization or automatization” (Verspoor & Schmitt, 2013; p.354). The term Dynamic usage-based (DUB) is inspired by the title of one of Langacker’s articles (2000) in which he argues that a usage-based view is per definition a complex dynamic system theory view. However, in our own use of the term, we would also like to accentuate the fact that language development is per definition non-linear, that some sub-systems need to be learned before others and that variability in the use of structures (which includes making errors) is needed to progress (Verspoor, 2017).

The key difference between a SB and DUB view is that there is no priority for grammar or syntax in language. Language is not driven by rules. Instead, language forms—from concrete morphemes, words, phrases, clauses, sentences, and discourse sequences to abstract lexical categories and morphological and syntactic patterns—are all fundamentally similar as they all bear meaning to different degrees and form a continuum. These forms have dynamically emerged through time through interaction in a speech community and are always subject to change.

Because language is for a great deal conventionalized (N. Ellis, 2012; Wray, 2008; Pawley & Syder, 1983), it is learned through associations, the mechanisms of which are explained by the “Entrenchment and Conventionalized model” by Schmid (2015, 2017). He differentiates four different types of associations: symbolic, syntagmatic, paradigmatic and pragmatic. We can illustrate the differences between those types with the example “Parlez-vous français?” [Do you speak French?].

- The symbolic association is the reciprocal link between form and meaning as the form /parle/ is associated with the meaning ‘articulate, produce meaningful sounds’.
- The *syntagmatic* association is provided by the order of words and constructions. The words in the utterance “Parlez-vous français?” [Do you speak French?] are syntagmatically associated with each other in this order. The inversion of the subject and the verb shows that it is a question and if processed frequently in this order, it will trigger one holistic association between meaning and form as is the case in chunking.
- The *paradigmatic* association is the competing association between the forms and the meanings. “Parler” [to speak] is associated with other synonyms such as “dire” [to say] , “prononcer” [to pronounce] or phrases “entendre parler de” [to hear about]. Paradigmatic associations activate what could have been said or meant instead of what was said or meant.
- And finally, the *pragmatic* association is the link between what is said and the information it conveys in the situational context. The sentence “Parlez-vous français” [Do you speak French] is pragmatically associated with a situation where the linguistic knowledge of the hearer is unclear for the speaker, and where there is a degree of distance between speaker and hearer (cf. Parles-tu français? vs. Parlez-vous français [informal vs. formal ways of saying ‘do you speak French?’]).

Language thus consists of a large array of units of various sizes, which have four types of associations. Verspoor (2017) has called these Form, Use, Meaning Mappings (FUMMs), where Use specifically refers to pragmatic use in a meaningful context.

Because these associations are created gradually, learners need to encounter these linguistic elements in different contexts to reproduce and conventionalize the utterances successfully. Through repetition, the associations are strengthened, routinized and then schematized by the learner (Schmid, 2015). Thus frequency of exposure rather than rule learning is considered as the driving force of language learning (N. Ellis, 2002).

A teaching setting in a DUB perspective would entail that learners are exposed frequently to examples of authentic language from the target language community, that learners discover the forms and the structures of language implicitly or inductively, that the language features are never taken as separated items but rather shown in a communicative context, that speaking, listening, reading and writing are integrated and teach language as a whole, that the target language is learned in interactive communicative activities. In a classroom, this would mean that negotiation of meaning is encouraged, that teaching is learner-centered and appeals to the learner's interest, and that language is used meaningfully with a positive attitude of the teacher towards unpredictability, risk-taking and choice-making (from Verspoor, 2017; see Rouse-Malpat & Verspoor, 2018).

L2 instruction should provide learners with a stream of input, which should be scaffolded for meaning with visuals, gestures, and paraphrases. Learners are asked for output, but early on focusing on meaningful repetitions and guided interactions to minimize the cognitive effort that it takes to make sense from the input and creatively think of output. There is no particular need for grammar explanation, even though some explanations of the grammar in the L2 in an inductive way may help learners learning noticing a certain feature. However, grammar should always be a by-product of the process of language learning and not the primary focus. In fact, this is the original idea behind Long's Focus on Form (1991) as L2 learning occurs during a meaning-based activity where attention to form can be drawn inductively when the learner needs it. It also aligns with the definition of "meaning-focused instruction" (Norris & Ortega, 2003), which provides exposure to rich input and a meaningful use of the target language in context, where grammatical acquisition is not the main goal but the incidental by-product of language acquisition. The main difference between Focus on Form and DUB is that DUB takes a more holistic approach to language and focuses on form-use-meaning mappings (FUMMs) at all levels of language: words, phrases, chunks and conventionalized ways of saying things. "If the language is viewed as a holistic system consisting of an array of smaller and larger conventionalized units called FUMMs, then teaching should be more piecemeal as many FUMMs have to be learned separately and holistically in that the learner is exposed simultaneously to the sounds, words, phrases, and patterns and how they are used in the right context" (Verspoor, 2017; p.150).

2.3 Effectiveness of L2 instruction inspired by DUB theory

There are probably many examples of teaching approaches that are in line with DUB principles, such as TPRS (Total Physical Response and Storytelling), the movie method (Hong, 2013; Irshad, 2015) and AIM (Maxwell, 2001). However, very few have been empirically tested, especially over a longer period of time. A few studies that have been done so far found the DUB inspired methods to be more effective than traditional semi-communicative L2 teaching (Verspoor & Hong, 2017; Rousse-Malpat & Verspoor, 2018). For the AIM method developed in Canada, mixed results were found regarding its effectiveness. Maxwell (2001) and Michels (2008) investigated oral fluency and reported that the AIM learners were better than the non-AIM learners. However, these were very small-scale studies for which no statistics were performed. Nevertheless, studies with more participants and statistical analyses did not reveal any significant differences between AIM and non-AIM either (Mady, et al., 2009). Qualitative findings showed, however, that AIM teachers spoke more French in the classroom and that AIM learners reported to feel more at ease in listening and speaking skills. Bourdages and Vignola (2009) looked at the oral communication skills of two groups of learners (AIM vs. non-AIM) of third graders in Canada and performed an interview. Results showed no significant differences between the groups with regard to proficiency and grammatical accuracy even though AIM students code-switched less with English than the other learners and seemed more willing to communicate. AIM students made more incomplete sentences suggesting that they dared to take more risks but could not maintain sentence-level speech.

In the Netherlands, to date, a few studies have looked at the effects of AIM on oral proficiency using free-production data. Rousse-Malpat & Verspoor (2012) presented evidence that AIM was more effective on oral skills after one year and two years of instruction compared to a communicative explicit method called *Carte Orange*⁹. For specific grammatical constructions such as the present tense and negation, no differences were found but for gender, the non-AIM group was more accurate after the first year but the difference vanished after the second year. Both groups made the same number of errors on all constructions both together and separately. However, there were some qualitative differences. The non-AIM group used more prefabricated chunks for negation whereas the AIM students showed more creativity. The same difference was observed for present tense. Another qualitative difference was the use of Dutch during the interviews. Non-AIM students tended to fall back to Dutch when they did not know a word in French

9 Publisher ThiemeMeulenhoff

or when they wanted to indicate that they did not understand while the AIM students kept speaking French. The same was found on writing skills (Rousse-Malpat, Verspoor and Visser, 2012).

Table 16. Summary Language and Language Learning.

| Structure-based | Dynamic usage-based |
|---|---|
| Modular system of interconnected units: phonemes, morphemes, lexical categories, noun phrases, verb phrases, sentence types | Complex and dynamic network of interconnected constructions: a continuum of form-use-meaning mappings (FUMMs) / constructions from morphemes to sentence patterns |
| Analytic: subsystems of building blocks of linguistic knowledge | Holistic: conventionalized phrases (FUMMs) are sources for linguistic knowledge |
| Rule-driven: Language is built from assembling units, structure comes first. | Meaning-driven: Language is built from usage, structure is emergent. |
| Language is an assemblage of elements that are connected according to predictable rules. | Language has an array of constructions, for many of which similar patterns (schemas) have emerged. |

Table 17. Summary of Language teaching.

| Structure-based | Dynamic usage-based |
|--|--|
| Rules are needed to learn the language | Frequent exposure to meaningful utterances is needed to learn the language |
| Explicit: rules need to be explained, practiced and corrected | Implicit: Patterns are deduced implicitly |
| Rules are learned from simple to complex with explanation and pattern practice | Constructions are used, conventionalized and entrenched with lots of repetitions and revisiting language forms differently |
| Rules need to be practiced to avoid errors | Repeated occurrence of similar utterances is needed to discover recurring patterns |
| Errors need to be eliminated to avoid fossilization | Errors are part and parcel of the road to discovering target-like patterns |
| Use of L1 is needed to explain the rules well | Limited use of L1 as no rules need to be explained |
| Limited exposure to natural L2 use | High exposure to natural L2 use |

Table 16 and 17 summarize our arguments on the difference between SB and DUB for language learning and language teaching. We expect that the different views on language and language teaching have a different effect on language learning. We are particularly interested in the effect of each treatment on the productive oral skills of the learners after three years of instruction. Therefore, we aim at answering the following research question:

Which instructional method (Structure-based or DUB) has more effect on oral proficiency after three years of instruction in terms of general oral proficiency and analytical complexity, accuracy and fluency measures?

3. Method

3.1 The teaching methods

3.1.1 *The Structure-Based method (SB): Grandes Lignes*

The SB method is in our study operationalized as teaching on basis of the use of a text book for L2 French, *Grandes Lignes*¹⁰. It is the best-sold course book for high-school French in the Netherlands. It is a communicative task-based method composed of a textbook and an exercise book. The main goal of the method is that learners master a list of key grammatical rules and the vocabulary list appointed in each chapter by providing them with listening, reading, speaking and writing exercises increasing in terms of difficulty (from reproductive to productive), as described in the previous section on structure-based approaches to language learning and teaching. The chapters are organized around topics that are related to the most common interests of learners of this age such as family, school, sports and holidays.

Learners are first presented with a receptive text containing the target grammatical rule and vocabulary. Then, they are asked to answer questions, usually in the L1, about the meaning of the text. The grammatical rule is explicitly explained by the teacher in the L1 and then practiced in groups of two learners. Vocabulary is given in the form of a word list or a chunk list called “phrases clés” (key sentences) with their translation into Dutch. Those sentences are often literally the expected answer to most productive exercises.

The interaction between the teacher and the learners can be spontaneous in the L2, as it is the case in our study, but the interaction between the learners is thoroughly prepared in advance, if not in the L1. The method fits the description of a deductive focus on form method as described in Long (1991) but in reality there is much focus on forms.

10 Publisher Noordhoff Uitegevers

3.1.2 The Dynamic usage-based method (DUB): AIM

As mentioned in Rousse-Malpat & Verspoor (2018), there are several methods that are inspired by DUB principles and that are used in L2 classrooms. One of them is the Accelerative Integrated Method (AIM) (Maxwell, 2001) for French, English, Spanish, Japanese and Mandarin. In this study, the DUB method has been operationalized by AIM for L2 French in Dutch high schools.

The most important principle of these methods is to provide lots of meaningful L2 exposure and use in the L2 classroom. To achieve that goal, teachers and learners only use the target language in the classroom from day 1 and focus on providing input and scaffolding the output, from pure repetition to use and reuse of chunks present in the input. To do so, a lot of attention is put on oral skills and in a later stage on writing skills.

Teachers use stories to provide learners with meaningful utterances in meaningful contexts. These stories are existing fairy tales for young children (e.g. “les trois petits cochons”- the three little pigs) or created stories involving topics that are relevant to the age group of the learners (e.g. “un garçon populaire”- a popular boy). Each story provides the learners with lots of exemplars scaffolded in terms of complexity.

At first, learners are not asked to create output themselves. They listen to the story told by the teacher and after that, they repeat the story chunked by the teacher imitating the teacher’s words and gestures. In a later phase, they are taught how to re-use whole phrases of the story to answer questions about the meaning of the story. The teacher gives feedback if needed in the form of whole utterances. The goal is to entrench whole phrases that learners will later be able to re-use in other contexts.

Each lesson is formed around several blocks of 10 minutes for each activity. The primary focus of the lesson is on the meaning of the story or on the meaning of the interaction between the teacher and the learners or between learners themselves. In every lesson, learners work in small groups and perform tasks together. Depending on their level of proficiency, the learners can be asked to act out the story or to imagine the end of the story.

Focus on form also occurs, albeit inductively, as the method is built to create associations between form and meaning. To do so, teachers use gestures and songs to draw attention to some forms of the language (e.g. present/ future tense) in the meaningful context of the story.

3.1.3 Selection of the sample

The data of this study comes from a larger study involving 229 participants in five high schools. Each participant was followed for three school years and was asked to perform an oral interview at the end of each school year ($n= 3$), which were then scored holistically in terms of general proficiency. For this study, we selected the sub-group for each type of instruction that was most comparable and stable and we analyzed their last oral interview (after 34 months of instruction). The participants in this study had all obtained the highest scores on oral proficiency in their respective method and they had the highest scholastic level. In terms of stability, the classes remained intact throughout the three years. SB group only had one change of teacher at the end of the first year. All three teachers were experienced and rated as good teachers by the learners.

This study includes 41 Dutch pupils without learning problems from two different schools in the Netherlands. A background questionnaire was submitted to ensure that they were absolute beginners in French when the research started in September 2013. They all learnt French in a formal setting and with the same peers at secondary school. They were taught with the same instruction method during the three years of the research.

The two groups were similar with regard to gender, age, and scholastic aptitude. The SB group consisted of 21 students (10 men, 11 women) with a mean age of 14.29 (range 13 to 15). The DUB group consisted of 20 students (9 men, 11 women) with a mean age of 14.45 (range 14 to 15).

All students had a high scholastic aptitude level as measured by the Cito test (a general scholastic aptitude test taken in the last year of primary education), which was shown by Verspoor, De Bot, and Xu (2011) to be a strong predictor of L2 development of English at the Dutch high school level. An independent-samples *t*-test showed that the SB group ($M = 547.35$, $SD = 2.70$) and the DUB group ($M = 547.95$, $SD = 2.54$) did not differ with regard to their scholastic aptitude, $t(38) = -0.72$, $p = 0.474$. Both groups followed a selective six-year high academic, pre-university program of secondary school that usually attracts the highest-level students as shown in Verspoor, De Bot, & Xu (2015).

3.1.4 Hours of instruction and hours of L2 exposure

Accounting for the exact amount of L2 exposure that the participants received is at the least a very difficult task in a classroom study. However, we determined the number of hours of L2 exposure by multiplying the total number of hours of French instruction with the percentage of L2 exposure learners received on average in the classroom.

To do so, we asked teachers to self-evaluate the amount of L2 exposure that the learners received on average and we confirmed their score by timing the actual amount of L2 exposure during the lessons that we observed (one per teacher). Table 18 shows a summary.

Table 18. Overview of total hours of instruction and L2 exposure after three years.

| Method | Hours of instruction | % of L2 exposure | Estimated hours of L2 exposure |
|---------------------------|----------------------|------------------|--------------------------------|
| SB(Grandes Lignes) | 240 | 60 | 144 |
| DUB (AIM) | 216 | 90 | 195 |

The oral test: eliciting free-production data with young beginners (The SOPA test)

In measuring effectiveness as general proficiency, it is often pointed out that both controlled and free production and reception tasks are needed (R. Ellis, 2005). Much work has already been done in measuring proficiency with controlled and receptive tests, but fewer studies involve data of spontaneous L2 production, probably because they are time consuming and relatively difficult to score objectively. In this study, we use a free production test called SOPA developed by the CAL (Thompson , et al., 2002).

The setting is as follows: there are two examiners, an interviewer and a rater, facing two learners of about the same level and personality (the classroom teacher is asked to provide a list of two compatible learners in terms of proficiency and personality). The interviewer interacts with the participants and asks questions that require linguistic proficiency of the participant ranging from receptive to productive, and from simple to complex ones both syntactically and lexically. S/ he rewards the answers with a positive attitude and gives compliments, while the rater takes notes during the interview and discusses the score of the students with the interviewer after the interview has been finished. Rating happens on basis of a rubric developed for SOPA (see Appendix B for a thorough description).

The test is built up in such a way that it tries to elicit the highest actual level of proficiency of the participant at the time by giving the learner increasingly more difficult tasks. This continues until the interviewer senses that the task exceeds the current proficiency level of the learner in which case the difficulty level is again decreased. This leaves the learner feeling positive with regard to their performance.

In total, there are three increasingly difficult activities the participants are asked to perform. The interview starts with a very simple task called “Le Sac Magique” (The magical bag). During this task, the students perform activities with a bag filled with plastic fruits and colored blocks. The interviewer puts the students at ease by starting with simple questions that only require receptive knowledge of French (e.g. “Can you point at the apple?”, “Can you put the yellow block next to the banana?”). Gradually the interviewer starts asking questions eliciting productive knowledge (e.g. “What is your favorite fruit?”, “How often do you eat an apple?”). The aim of this first task is to put the students at ease with activities that they can perform quite well.

The second activity is called “On parle de toi” (All about you) and includes questions about the personal lives of the students. The interviewer asks questions about family, school, and hobbies (e.g. “How old are you?”, “Do you have any brothers or sisters”, “What subjects do you like at school?”). The questions require more productive knowledge from the students, but because the topics are familiar to the students they are able to engage in the conversation. Because the students in this study were already in their third year of French class, we included a few difficult, out-of-the box questions, to check whether the students could understand more complex speech and to elicit longer and more complex answers. An example of such a question is the following: “Imagine that I am the director of this school and I have decided that from now on all students also need to go to school on Saturday. What do you think about that?”. These challenging questions were only discussed with the more proficient students that could handle these questions, which is in line with the SOPA protocol.

The third activity “La Ferme” (The Farm) involves a 3D farm with movable characters and animals. The interviewers use more complex vocabulary and sentence constructions in their questions, which requires more elaborate receptive knowledge. For example, students were asked questions such as “Can you see the red cat in the tree?”. The students were also asked to perform a short dialogue with the characters of a boy and a girl at the farm. This activity provides the students with the opportunity to talk creatively about a topic of their own choice.

4. Analysis

In this study, each interview was rated holistically on general proficiency and analyzed analytically in terms of complexity, accuracy and fluency measures. Each interview was video-recorded and lasted 15 minutes on average ranging from 13 to 20 minutes. Both groups were assessed by the same team of two researchers who had followed a certified SOPA training. During the interview, the rater wrote down what the participants said and whether they understood the questions and commands on a standardized grid with rubrics developed by the Center for Applied Linguistics (<http://www.cal.org/ela/sopaellopa/>), based on the developmental stages of language learners going from fixed formulae, unsuccessful creative language to successful, phrases.

Immediately after the testing session, the interviewer and rater independently gave scores on the different sub-systems (fluency, vocabulary, grammar, comprehension) and then discussed them until they reached consensus. The validity of the scores is ensured by the training of the assessors and the extensive rubrics used to establish a score. The following descriptions of these categories are based on the Rating Guidelines for the SOPA Rating Scale (Thompson, et al., 2006) and the SOPA Rating Scale itself (levels from 1 to 9).

4.1 Measuring oral proficiency with holistic measures

4.1.1 Fluency

Oral fluency refers to the size and type of language chunks produced by the students. The ease or hesitancy with which the speech is produced must also be considered when determining the oral fluency rating. For example, a student is assigned to the lowest level (Jr. novice-low, score 1) when s/he “produces only isolated words (i.e. single-word responses) and/or greetings and polite expressions such as good morning and thank you” whereas a student rated as Jr. intermediate-low (score 4) “goes beyond memorized expressions to maintain simple conversations at the sentence level by being creative with the language, although in a restrictive and reactive manner” and “handles a limited number of everyday social and academic interactions”.

4.1.2 Grammar

The holistic score on grammar is primarily determined by the way verbs are used or by the lack of verb use. Other grammatical inaccuracies are considered less crucial than verb inaccuracies when rating grammar. For example, at the Jr. novice-high (score 3) level, a student “creates some sentences with conjugated verbs, but in other attempts to create sentences, verbs may be lacking or are not conjugated”.

4.1.3 Vocabulary

In the rating scale of vocabulary, no specific sets of vocabulary (e.g. colors, numbers, etc.) are included because curricula may introduce certain vocabulary at different moments. Instead, the scores on vocabulary reflect the range of words the participant was able to produce in the target language to express an idea and how much s/he had to fall back on his native language when words in the L2 were missing. For example, a student of the Jr. intermediate-low level “has basic vocabulary for making statements and asking questions to satisfy basic social and academic needs, but not for explaining or elaborating on them” and “use of some native language is common when vocabulary is lacking.”

4.1.4 Comprehension

Rating listening comprehension requires close attention to non-verbal clues and responses that indicate comprehension of the interaction with the interviewer or with the peer. Characteristics of a student at for example the Jr. intermediate-mid level are that the student “understands sentence-level speech in new contexts at a normal rate of speech although slow-downs may be necessary for unfamiliar topics” and “carries out commands without prompting”.

4.2 Measuring oral proficiency with analytical measures

For the analytic analyses all interviews were transcribed and coded. We created one data file per participant. Before the analysis, we removed all the words that were prompted by the interviewer. The samples were then analyzed on the same constructs as the holistic rating: fluency, grammar and vocabulary. Comprehension could not be measured in our corpus; therefore, this component of oral proficiency is not represented in our analytical measures. For most of the measures included in our study we have calculated a ratio, representing a percentage of the total number of French words expressed by the participants or a percentage of the total duration of the interview.

4.2.1 Fluency

We operationalized fluency by calculating the speech ratio and the use of filled pauses. Speech rate was calculated by dividing the total number of words produced in a sample by the total amount of time required for a particular activity or the whole interview (including pause time) expressed in seconds. We based our speech rate on the total duration of the interview because we could not calculate the exact speaking time of each student because they were interviewed in pairs. We considered speech rate to be a valid and objective measure of fluency, as Ortega (1995, as cited in Wolfe-Quintero et al., 1998, p. 14) found in a factor

analysis in her study of the speech of 32 learners that length (words per utterance) was strongly related to rate (syllables per second) and that both “were part of one factor presumed to be fluency”. Filled pauses were calculated by dividing the total number of filled pauses by the total number of French words.

4.2.2 Grammar

Grammar was analyzed with measures of complexity and accuracy. Grammatical complexity was operationalized with the number of complete sentences, which were categorized into four sentence types as defined in Verspoor & Sauter (2000): simple, compound, complex and compound-complex. According to Bardovi-Harlig (1992), the amount of coordination seems to be a more sensitive measure than subordination measures at initial levels of L2 development. Therefore, we decided to analyze the dispersion of sentence types, which allowed us to investigate to what extent the students used different sentence types, including both coordination and subordination.

We followed the classification of sentence types of Verspoor, Schmid and Xu (2012), which is based on Verspoor and Sauter (2000) and allocated each complete sentence in our speech sample to one sentence type: a sentence was either simple, compound, complex or compound-complex. For each sentence type, we calculated the ratio of the number of occurrences of the sentence type (simple, compound, complex or compound-complex) against the total number of sentences. Because these categories are mutually exclusive and therefore complementary, the absolute number of occurrences of each sentence type was converted to percentages. We also calculated a single grammatical complexity variable by taking the ratio of the number of occurrences of all sentence types together against the total number of sentences. Verspoor and Sauter (2000) give the following definitions of these sentence types. Table 19 gives a definition of each type as well as an example from our own corpora.

Table 19. Sentence level based on Verspoor & Sauter (2000)

| Label | Definition | Examples from the data |
|-------------------------|---|--|
| Simple | One main clause | <i>J'ai un cheval</i> (I have a horse) |
| Compound | Two complete main clauses (each with its own subject and finite verb) | <i>Je déteste maths et géographie mais j'aime le dessin</i> (I hate maths and geography but I like drawing) |
| Complex | One main clause and one or more finite dependent clauses | <i>Je préfère manger l'orange parce que c'est délicieux</i> (I prefer eating an orange because it is delicious) |
| Compound-complex | A combination of compound and complex elements | <i>Les vaches doivent manger et je veux que tu m'aide</i> (The cows have to eat and I want you to help me) |

Grammatical accuracy was operationalized as the correct use of three specific grammatical features. This selection is based on Rouse-Malpat and Verspoor (2012) and covers grammatical aspects that have been dealt with in both classes. Because the students included in this study have been instructed for three years, we may expect to find a fair number of occurrences of these constructions in our sample, which allows us to perform statistical analyses on the data.

The first grammatical feature is gender. French has two genders, feminine (articles: la/une) and masculine (l/l'un). Studies on the predictability in French gender attribution and the acquisition of grammatical gender argue that a noun's ending and its grammatical gender are related (Lyster, 2004; Lyster, 2006; Lyster & Izquierdo, 2009). They investigated the effects of form-focused instruction on the acquisition of grammatical gender and found that the form-focused instructed students outperformed the students that received instruction without an explicit focus on form or feedback on grammatical gender. The form-focused instruction was aimed at drawing attention to several noun endings that reliably predict grammatical gender (Lyster, 2004; Lyster & Izquierdo, 2009).

The second construction is negation. In French two negative elements have to be used to form a negative sentence. The two terms *ne* and *pas* need to be placed before and after the conjugated verb, like *Je ne sais pas* 'I don't know'. The *ne* can be left out in spoken French, but the *pas* must always occur after the verb. Errors that occur with the use of negation are therefore mostly that the *pas* is left out or that the negative terms are not in the right place. For Dutch learners, this construction

is particularly difficult because in Dutch there is only one negator (*niet*), which occurs after the verb.

Finally, we investigated the use of present tense, which is the first tense that learners hear in the AIM group and that is explicitly taught in the GL group. We decided to focus on present tense even though other tenses were present in the data because the number of occurrences of the other verb tenses were too low. However, participants also used the past tense (*imparfait* and *passé compose*), the near future (*futur proche*), and the conditional (*conditionnel*) as shown in the table below.

Table 20. Use of verb tenses in each group

| | Total | Present tense | Past tense | Near future | Conditional |
|------------|-------|---------------|------------|-------------|-------------|
| SB | 468 | 437 (93,4%) | 19 (4,1%) | 9 (1,9%) | 3 (0,6%) |
| DUB | 663 | 630 (95%) | 22 (3,3%) | 10 (1,5%) | 1 (0,2%) |

For each of the three features, the accuracy variable was calculated as the ratio of the correct use of the grammatical construction against the total number of occurrences of that construction. We also calculated a single grammatical accuracy variable by taking the sum of the ratio of each feature.

4.2.3 Vocabulary

Vocabulary was analyzed in terms of sophistication and diversification. For lexical sophistication, the average word length was calculated. For lexical diversity, the Guiraud's index (the number of types divided by the square root of two times the number of tokens) seemed the most appropriate measure because the length of the speech samples varied considerably. Finally, we looked at lexical access, operationalized by the total number of French words and the number non-French words. Lexical access shows a certain degree of automaticity and entrenchment and which is likely to vary according to the level of proficiency (Segalowitz, 2010 in Ankerstein, 2014).

5. Results

5.1 The holistic analysis

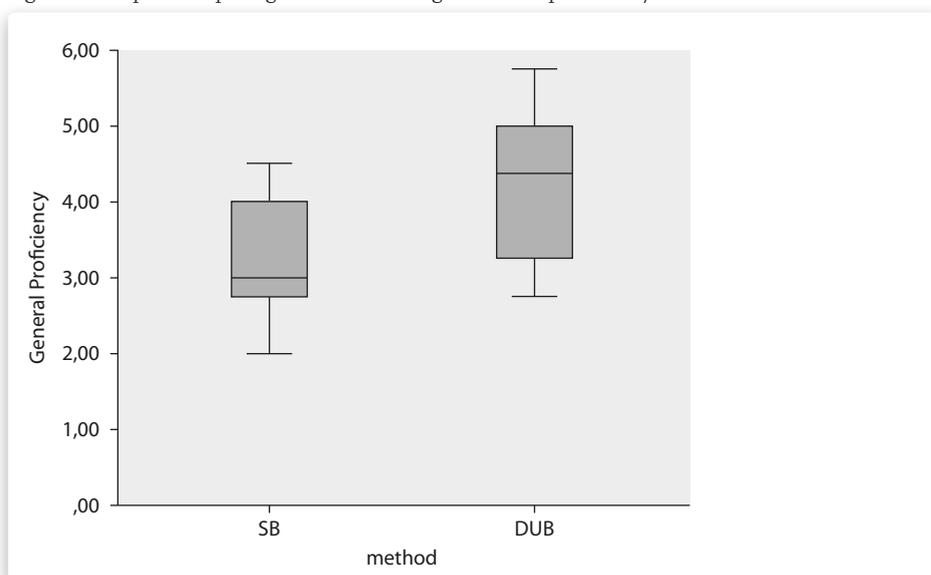
On the holistic measures of oral proficiency, our participants received a grade from 1 to 9 for fluency, vocabulary, grammar and comprehension. To evaluate whether any of our variables were measuring the same construct, we performed a correlation analysis. We performed a two-tailed Spearman's Rho correlation ($\alpha=.05$), which revealed that our four variables were highly and significantly correlated with each other.

Table 21. Correlations between variables.

| | Fluency | Grammar | Vocabulary | Comprehension |
|---------------|---------|---------|------------|---------------|
| Fluency | - | | | |
| Grammar | ,894** | - | | |
| Vocabulary | ,921** | ,946** | - | |
| Comprehension | ,853** | ,828** | ,841** | - |

Because our variables are significantly strongly related, we averaged them into one variable representing general oral proficiency. A Kolmogorov-Smirnov test and a Levene's test revealed that our data for this new variable was normally distributed and that we could assume equality of variance. Therefore, we compared the DUB group with the SB group on their general oral proficiency using an independent t-test ($\alpha=.05$).

Figure 4. Boxplot comparing SB and DUB on general oral proficiency.



As Figure 4 shows, the SB group scored lower ($M = 3.24$, $SD = 0.71$) on general oral proficiency compared to the DUB group ($M = 4.29$, $SD = 0.91$). An independent-samples t-test indicated that this difference was significant ($t(39) = -3.72$, $p = .001$) with a large effect size (Cohen's $d = 1.16$).

To find out which analytical measures contributed to our perception of general proficiency, a two-tailed Pearson's r correlation analysis ($\alpha = .05$) was performed to assess the relationship between the holistic and the analytic scores.

Our intention is to compare our two groups only on those analytical measures that correlate significantly with general proficiency measured holistically and may therefore be assumed to tap into the learner's proficiency.

Table 22. Correlations between general proficiency and analytical measures.

| | General Proficiency |
|---|----------------------------|
| Fluency. Speech rate | ,829** |
| Fluency. Filled pauses | -,489** |
| Grammatical complexity | ,720** |
| Grammatical accuracy | ,370* |
| Vocabulary. Lexical sophistication | ,022 |
| Vocabulary. Guiraud | ,483** |
| Vocabulary. French words | ,822** |
| Vocabulary. Non-French words | -,355* |

(* = $p < .05$; ** = $p < .01$)

As Table 22 shows, general oral proficiency correlated significantly with almost all the analytical measures. Strong positive correlations were found between general oral proficiency and speech rate ($r = .829$, $p < .001$), general oral proficiency and syntactic complexity ($r = .720$, $p < .001$) and general oral proficiency and the number of French words ($r = .822$, $p < .001$). A moderate positive correlation was found between general oral proficiency and lexical diversity measured by Guiraud's index ($r = .483$, $p = .001$). A weak positive correlation was found between general oral proficiency and grammatical accuracy ($r = .370$, $p < .017$). A moderate negative correlation was found between general oral proficiency and the number of filled pauses ($r = -.489$, $p = .001$) and a weak negative correlation was found between general oral proficiency and the amount of non-French vocabulary ($r = -.355$, $p = .023$). There was no correlation between general proficiency and lexical sophistication, which can be due to the fact that both groups used basic words of approximately 4 letters (SB group: $M = 4.08$, $SD = 0.34$; DUB group: $M = 4.07$, $SD = 0.17$).

In the next paragraphs, the results of the analytical measures representing fluency, grammar and vocabulary will be presented.

5.2 The analytic analysis

5.2.1 Fluency

For every activity of the SOPA interview taken separately, the speech rate of the DUB group was significantly higher than the speech rate of the SB group, indicating that the DUB group was more fluent. The same applied to the entire interview. See Table 23 for details.

Table 23. Results of the fluency analysis per game.

| | SB | | DUB | | | | | |
|---------------|------|------|------|------|-------|-------|-------|------|
| | Mean | SD | Mean | SD | t | df | p | d |
| Game 1 | 0.15 | 0.06 | 0.21 | 0.07 | -2.75 | 39 | 0.009 | 0.86 |
| Game 2 | 0.22 | 0.08 | 0.33 | 0.14 | -2.96 | 29.31 | 0.006 | 0.93 |
| Game 3 | 0.25 | 0.07 | 0.37 | 0.12 | -3.79 | 29.87 | 0.001 | 1.19 |
| Total | 0.20 | 0.06 | 0.29 | 0.10 | -3.52 | 39 | 0.001 | 1.09 |

As table 23 shows, for each game and also for the entire interview, the speech rate of the DUB group was higher than the speech rate of the SB group. Besides the speech rate, we also calculated a ratio of filled pauses. There was no significant difference between the DUB group ($M = 0.16$, $SD = 0.07$) and the SB group ($M = 0.18$, $SD = 0.06$) with regard to the number of filled pauses, $t(39) = 0.99$, $p = .333$.

5.2.2 Grammar

Results show that both groups used mostly simple sentences but that other types of sentences were present in the corpus as well as we can see in table 24.

Table 24. Number of occurrences according to sentence type

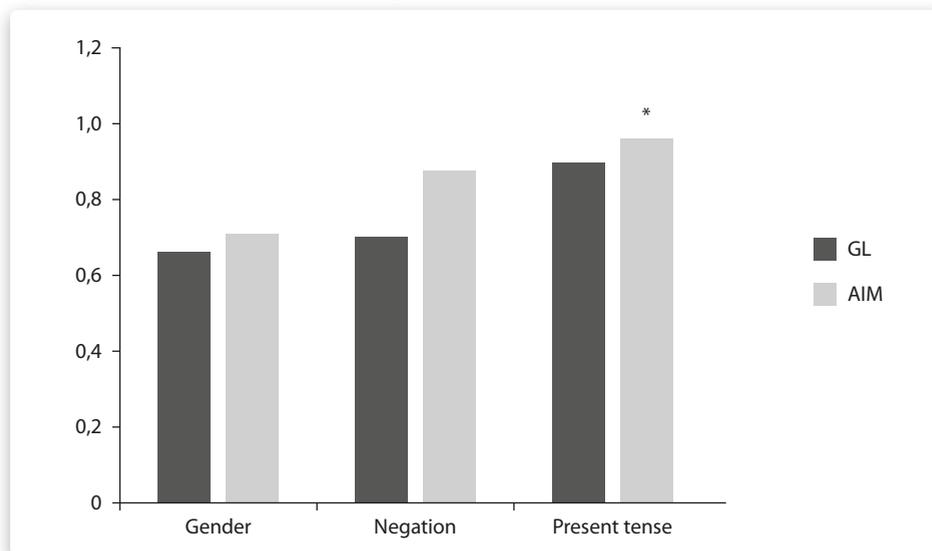
| | Total | Simple | Compound | Complex | Compound-complex |
|-----|-------|-----------|-----------|---------|------------------|
| SB | 408 | 375 (92%) | 24 (5,8%) | 8 (2%) | 1 (0,2%) |
| DUB | 417 | 307 (74%) | 43 (10%) | 39 (9%) | 28 (7%) |

An independent samples t-test revealed that the DUB group ($M = 0.25$, $SD = 0.16$) used significantly more non-simple sentence constructions than the SB group ($M = 0,07$, $SD = 0.07$), $t(24.77) = -4.58$, $p < .001$ with a large effect according to Cohen's d ($d = 1.47$).

For grammatical accuracy, an independent samples t-test showed that there was only a significant difference between the groups in the correct use of present tense with the DUB group outperforming the SB group (SB ($M = 0.90$, $SD = 0.09$),

DUB ($M = 0.96$, $SD = 0.03$; $t(23.96) = -2.99$, $p = .006$). The effect size of this difference was large ($d = .92$). Gender and negation were found to be used equally accurately by the two groups (see Figure 5).

Figure 5. Ratio of correct occurrences of present tense



A more detailed analysis of the type of errors made by both groups in the present tense revealed that they produced similar errors; usually they used the verb in the infinitive form. However, small differences were found in the type of error made by each group, as the SB group produced more semantic and tense errors than the DUB group, and the DUB group produced more subject-verb disagreement errors than the SB group. Table 25 shows the different error types and how often they occurred in both groups (Note that due to rounding not all measures add up to 100%).

Table 25. Error analysis of the present tense

| Error Type | Example and frequency SB | Example and frequency DUB |
|----------------------------------|---|--|
| Infinitive form | *Il regarder les poulets (He looks at the chickens) (51%) | *Je n'apprendre beaucoup (I don't to learn a lot) (48%) |
| Semantic error | *J'ai à le petit école (I have in the primary school) (14%) | *Je suis quinze (I am fifhteen) (4%) |
| Subject-verb disagreement | *Les objets a le couleur (The objects has the couleur) (10%) | *On n'avons pas le musique (We don't have music) (22%) |
| Incorrect infinitive | *Tu aimes habite à la ferme? (You like living on the farm?) (6%) | *Je dois me reveille (I have to wake up) (13%) |
| Verb omission | *Je ne pas les devoirs (I don't homework) (4%) | *Je quatorze ans (I fourteen years) (9%) |
| Tense error | *J'ai vois des oiseaux (I have seen the birds) (8%) | - |
| Word order | - | *Le tomate fait me malade (A tomato makes me ill) (4%) |
| No liaison | *Je oublie (I forget) (2%) | - |
| Subject omission | *Prend de petit-déjeuner (To take breakfast) (2%) | - |

5.2.3 Vocabulary

Results show that lexical diversity as measured with Guiraud's index was not significantly different ($t(39) = -1.35, p = .186$) between the DUB group ($M = 6.73, SD = 0.50$) and the SB group ($M = 6.51, SD = 0.58$).

A word count analysis of all interviews showed that the DUB group used 5035 French words ($M = 251.75, SD = 88.01$) and the SB group used a total of 3951 French words ($M = 188.14, SD = 50.41$). An independent samples t-test showed that this difference was significant, $t(39) = -2.85, p = .007$.

We also analyzed in detail to what extent the students used other foreign languages than French and in which situations these languages were used. Table 26 shows the results of this analysis.

Table 26. Use of other foreign languages

| | Dutch | English | German | Spanish |
|------------|-------|---------|--------|---------|
| SB | 109 | 42 | 2 | 2 |
| DUB | 20 | 16 | 3 | 1 |

According to table 26, participants mainly fall back to Dutch and English during the interview when they cannot use French. They usually use single words when they lack vocabulary (e.g. “party”, “homework”, “spinazie” (spinach) and “wandelen” (to walk)) or code-switch French and Dutch in one sentence (e.g. “c’est le mama de petite vogel” (It’s the mother of the small bird)). Sometimes they also use entire Dutch sentences (e.g. “Ik weet het niet” (I don’t know’’)).

6. Discussion

In this chapter, we aimed to answer the question whether a Structure-based or Dynamic usage-based instructional method had more effect after three years on oral proficiency. Oral proficiency was measured in terms of holistic scores and analytical scores. The structure-based approach was based on the implicit view that language is a set of rules and inspires an analytic, grammar-driven approach. The DUB approach, in contrast, views language as a network of conventionalized routines, which inspires the use of linguistic routines at the phrase and clause level. This particular study used the oral free production data from 41 Dutch participants learning French as L2 in secondary school classrooms.

The findings show that the methods had the same effects on the following analytical measures: filled pauses, most measures of grammatical accuracy, and vocabulary. These particular measures correlated only moderately with the holistic general oral proficiency scores. These findings show that a method that does not particularly focus on accuracy can have effects, or even more effects, on some aspects of grammar than a method that focuses mainly on grammar. They also show that at the beginning stages of acquisition, fluency and complexity are at least as important as accuracy in sounding proficient.

However, there were also some clear differences between the two teaching methods. The most striking finding is that the DUB learners were better at general proficiency and at speech rate, grammatical complexity, L2 use and avoidance of non-target language use, which all correlated very strongly with each other. This means that the more French words were used, the faster the participants spoke, and the more complex sentence types were used, the more proficient the participants sounded overall. Previous research investigating CAF measures also showed such strong relationships between general proficiency and speech rate (e.g. Lennon, 1990; Riggenbach, 1991; Towell, Hawkins, & Bazergui, 1996), and holistic ratings and syntactic complexity (Verspoor et al., 2012).

We assume that this head start in fluency and complexity is the result of how the DUB method made L2 exposure meaningful. Presenting the learners with input made of conventionalized routines from stories and pushing them to produce output from the very beginning gave them the experience they needed in the L2 to perform well during the interview. The meaningful exposure, with enough repetition, helped the learners to form all types of linguistic associations (symbolic, syntagmatic, paradigmatic, and pragmatic) and be flexible in their L2 use. The use of gestures is a very important component of the method as it, on the one hand, helps remembering the words, and creates a strong link between the phonological form of the word and its meaning. All this time spent on drilling routines and producing as much language as possible with implicit or inductive focus on form made the participants speak faster and use more types of sentences. We can illustrate this with an example from the interviews. In all interviews, we have asked the question “Qu’est-ce que tu aimes manger?” (What do you like to eat?). Less proficient students often reply in single words “La pomme” (An apple), while more proficient student use complete sentences to answer and add more information on their own initiative, like “J’aime la pomme et j’ai un pomme dans mon sac” (I love an apple and I have an apple in my bag). The DUB learners were encouraged to speak in complete sentences in class, which probably helped them to entrench sentence frames, which can easily be accessed and reused in different situations. In line with Langacker (2000), we would argue that this limits conscious processing and allow the students to continue speaking supporting both fluency and creativity.

Finding that L2 learners with an implicit type of method with high L2 exposure are more fluent or complex is not a novelty. Studies in immersion programs in Canada already pointed out that these types of methods worked well on these aspects. It is more striking to find no difference or even an advantage for the DUB method on grammatical accuracy measures, especially considering the fact that the SB method spent a significant amount of time explicitly explaining and practicing the grammatical rules. We had expected the SB learners to outperform the DUB learners in this respect. The lack of such a finding could stem from the fact that the three grammatical constructions analyzed for this study may be categorized as “simple” constructions and are therefore perfectly suitable for learning through hearing and producing many examples. It would be interesting to analyze more complex grammatical structures, which were unfortunately not present in our data at this stage. Our results on the accuracy of gender contradict the findings of Lyster (2004) and Lyster and Izquierdo (2009), who found that students who were explicitly instructed about grammatical gender outperformed the students

who were instructed without an explicit focus on grammatical gender. However, their studies involved groups that may have been similar in the amount of L2 exposure and differed only in terms of awareness raising in one of the groups. In the DUB group however, there was also awareness raising of gender by means of specific gestures for feminine and masculine determiners. Apparently, the gesture associated with a word was as effective as learning the article together with the noun from a vocabulary list as in the SB condition.

The study shows that as far as vocabulary was concerned there were not many differences. The SB learners learned lists of vocabulary whereas the DUB learners were asked to repeat words accompanied by a gesture. Both techniques seem to be effective; however the gesture technique has the advantage of avoiding the L1 while entrenching the vocabulary, which could partly explain why the DUB learners used less of the L1 in their interviews. The SB learners used significantly more non-French vocabulary and did so in more different situations. However, we must add that both SB and DUB learners used some Dutch words when they lacked the vocabulary, but the SB learners also used a few complete Dutch utterances, to verify if they understood the question of the interviewer correctly or to explain why they could not answer in French. The DUB learners usually communicated in French and none of them used complete Dutch utterances in their speech. Some DUB learners used certain strategies to avoid the use of Dutch, like nonverbal communication (gestures, facial expressions) and circumlocution. Our findings with regard to the use of non-French vocabulary are comparable to the findings of AIM research in Canada and in the Netherlands. Bourdages and Vignola (2009) found that AIM students used significantly fewer bilingual and English utterances than the non-AIM students. In the Netherlands, Rousse-Malpat and Verspoor (2012) also found that the non-AIM students used more Dutch during the interview and in more situations than the AIM students after one and two years of instruction.

As far as fluency is concerned, the groups were similar in their use of filled pauses. Our results suggest that counting filled pauses at the beginning stages of acquisition might not be a valid way to distinguish fluent and non-fluent speakers. The findings in fluency research to date were also ambiguous with regard to the utility of filled pauses as a measure of fluency (e.g. Kormos & Dénes, 2004; Rekhart & Dunkel, 1992; Van Gelderen, 1994). It might be more interesting to investigate the distribution of pauses of foreign language learners, as it is suggested that their pausing pattern is influenced by their L1 (Raupach, 1980).

There are a number of limitations to this study that need to be acknowledged and addressed in further research. First of all, we were limited in controlling for some known confounding variables in classroom-based research, such as teachers and attitudes and motivation of the learners. Both groups differed in the number of hours of French lessons and of L2 exposure, even though they were the two most comparable groups from our larger study. Furthermore, the generalizability of our findings is limited, as the scope of our study concerned two groups of learners with a high scholastic aptitude in Dutch secondary education. The findings can thus not be generalized to all learners.

With regard to the measures used in this study, our investigation of analytical measures was limited to frequency counts and three grammatical constructions. Although our choice was inspired by the existing literature on CAF, we are perfectly aware that our measures are limited in terms of range. Complexity, accuracy, and fluency are each complex subsystems with multiple components, which show nonlinear and dynamic development. Therefore, the development of CAF in L2 oral performance should be traced over time and assessed with multiple measures in future research. We also felt that some aspects of the linguistic system were not well represented by our measures; that is to say the pragmatic aspect of oral language use. We tried to grasp this pragmatic aspect by measuring L2 use, which is a limited aspect of pragmatics. We see from the analysis of the semantic errors that there is a difference between the groups in using verbs in the appropriate context, so we feel that functional adequacy of the language could also be part of the analysis.

Finally, we have found that the DUB method has more effect than the SB method, but we do not know exactly why. We cannot say whether it is the frequency of exposure, the learning environment, the amount of authentic input, the enhanced output or the scaffolding in the form of gestures. We have suggested that the differences between the two groups could be explained by the underlying assumptions each method has of how languages are learned but we cannot point at single variables that influence particular elements of L2 development because many variables dynamically interact over time and are impossible to tease apart. However, follow-up studies could yield more insight into the specific aspects that contribute to the effectiveness of foreign language instruction in classroom settings.

7. Conclusion

In this chapter, we investigated the effects of two different teaching methods based on two different views on how languages are learned: one Structure-based, viewing language as a set of rules, and one Dynamic usage-based, viewing language as a set of conventionalized routines. Beginners of L2 French were tested on their oral proficiency after three years of instruction. Their oral output was scored holistically and analytically. Results showed beneficial effects of the DUB method on general proficiency, fluency, grammatical complexity, accuracy of the present tense and L2 use. There were also several aspects where both methods had the same effects: filled pauses, accuracy of negation and gender and lexical diversity. There was no advantage of the SB group on any measure. We discussed that this difference in effects can be explained by the underlying assumptions each method takes to explain how second languages are learned.

Our findings suggest that a teaching method inspired by DUB principles offers a good alternative for foreign language teachers who struggle with effective language teaching. Teaching methods inspired by this theory do not see language as a set of rules but as conventionalized routines, where learning emerges from the dynamic interaction between input and output. Our findings also suggest that grammar instruction is not a prerequisite for grammatical accuracy at the beginning phase of L2 acquisition. On the contrary, learners without grammar instruction were even more grammatically accurate on one aspect than the explicitly instructed students. However, the lack of grammar instruction does not mean that there was no attention to grammar at all in the DUB classes. The DUB method uses an inductive approach to teaching form-use-meaning mappings, including grammar, in which learners are stimulated to detect or notice patterns in the input and work out the rule for themselves. These findings suggest that teaching approaches should be focusing more on providing the learners with high exposure to the target language and enhanced input, in which grammar is related to meaning and use.

The chapter presented the results of a detailed study on the effects of both programs on speaking. The next chapter investigates the effects of both programs, using the same data set, on writing with a special focus on writing complexity.

