

Promoting incidental vocabulary learning through watching a French Netflix series with glossed captions

Abstract

This paper focuses on the effects of watching an entire season of a French series with the streaming service Netflix in an out-of-classroom context. University Dutch-speaking low- to high-intermediate learners of French were divided over two groups: a control group who only took the tests ($N=37$) and a treatment group ($N=65$). Learners in the treatment group watched six episodes with glossed captions provided by the Chrome extension *Language Learning with Netflix* (i.e., they could access the meaning of the words in the captions whenever they wanted) within a maximum of 21 days. We examined learners' incidental vocabulary learning gains by means of a form and meaning recall test and also analysed them in relation to different variables: word-related factors, the use of glossed captions and learners' vocabulary size. Results revealed that the treatment was effective since participants recalled approximately 35% of the word meanings and 28% of the word forms. Besides, learning gains were positively influenced by the use of the glossed captions as well as vocabulary size scores. Findings also demonstrated that frequency of occurrence positively impacted learning gains, even more when target words appeared more concentrated in one episode rather than across different ones.

Keywords: audio-visual input; glossed captions; vocabulary; frequency; Netflix

1 Introduction

Incidental vocabulary learning, that is vocabulary learned while engaging in a meaning-focused task such as reading or viewing (Webb, 2019), requires massive exposure to second language (=L2) input (Nation, 2013). Yet, classroom instruction alone is not sufficient to learn the huge amount of L2 input. Thanks to the overall accessibility of streaming services (e.g., Netflix), L2 learners are provided with numerous opportunities to engage with authentic L2 input. In that respect, a considerable number of studies has demonstrated that watching captioned audio-visual input results in vocabulary learning gains (e.g., see Montero Perez, Van Den Noortgate, & Desmet, 2013 for a meta-analysis). Yet, most studies have made use of short videos and one-off session treatment, hence leading to small vocabulary gains (except, e.g., Pujadas & Muñoz, 2019; Rodgers & Webb, 2019).

Moreover, the majority of research has focused on adding subtitles (i.e., on-screen text in the L1) or captions (i.e., on-screen text in the L2) to the input in order to stimulate vocabulary gains, although other methods such as glossed captions may also enhance vocabulary gains (Montero Perez, Peters, & Desmet, 2018). In addition, little research has examined the influence of other learner- and word-related factors such as vocabulary size and frequency of occurrence (Rodgers & Webb, 2020). This highlights the need for more studies adopting a multiple-session design, and looking at the impact of glossed captions as well as of learner- and word-related variables on vocabulary learning.

The present paper therefore aims to examine the impact of extensive exposure to French audio-visual input, enhanced with glossed captions, on learners' incidental vocabulary gains. In addition, this study also investigates the influence of learners' vocabulary size, of their use of the glossed captions and of the frequency of occurrence on incidental vocabulary learning.

2 Background

2.1 Audio-visual input and incidental vocabulary learning

So far, most research on vocabulary learning through viewing audio-visual input has focused on the impact of watching a short – from 3 to 45 minutes – video, most often in a one-off learning session (e.g., Peters,

Heynen, & Puimège, 2016). However, recently, extensive viewing has been put forward as an effective out-of-classroom activity for incidental vocabulary learning (Webb, 2015). Rodgers (2013, 2016) argues that films and series are an adequate language learning material because **this type of input meets** Nation's (2007) five conditions for suitable input: (1) available in large quantities, (2) interesting and (3) familiar to the learners, (4) comprehensible and (5) allows learners to gain vocabulary from the input thanks to contextual cues. Moreover, there is growing evidence that out-of-classroom exposure to L2 TV benefits language learning (Lindgren & Muñoz, 2013; Peters, 2018) and it has been shown that the potential for learning vocabulary through viewing may be similar than through written input (Feng & Webb, 2020). Even though extensive viewing has been associated with numerous L2 learning benefits (Vanderplank, 2016), few studies have investigated the impact of longer exposure on L2 learning.

Rodgers and Webb (2019) investigated the impact of watching ten captioned episodes of an American TV series, totalling 420 minutes of exposure over a 10-week period. Participants were evaluated on a tough (i.e., with semantically related distractors) and sensitive (i.e., with distractors from a different part of speech) meaning recognition test. Results indicated that learners gained 23% and 29% of what they could have learned on the tough and sensitive tests respectively. In Authors (XXXX), participants were exposed to 125 minutes of subtitled or captioned videos over a 4-week period and were tested on their meaning recognition and recall. Both groups had relative gains of approximately 14% on the meaning recall test and of 20% on the meaning recognition test.

Similarly, Pujadas and Muñoz (2019) examined the vocabulary gains of young low-intermediate (from Pre A to B1 CEFR levels) learners through watching 515 minutes of a captioned or subtitled TV series over three school terms. Participants were tested on their form and meaning recall gains after each term (i.e., corresponding to eight episodes). No significant difference between captions or subtitles groups was found on any test and both groups had relative gains of approximately 25% for the form and 11% for the meaning recall test. Findings showed an influence of proficiency on scores of the form recall test only. Likewise, in Suárez and Gesa's (2019) study, high school students recalled 29% of the word forms and 16% of the word

meanings after watching 195 minutes of a captioned series. This was significantly more than the control group who did not watch any videos.

In sum, these studies point to a beneficial effect of more sustained exposure to audio-visual input for vocabulary learning for learners of different proficiency. However, the majority of cited studies spread the sessions across a whole school term or year (except Authors, XXXX). There is still limited research adopting a multiple-session design during a shorter period of time. Moreover, all cited studies took place during participants' formal language lessons or in a laboratory context and used L2 English. Because audio-visual input has long been considered as a source of entertainment instead of education (Vanderplank, 2016), there is still little understanding of whether viewing 'entertaining' audio-visual input (e.g., series episodes) in an out-of-classroom context can also lead to similar learning gains. More research is therefore needed on the potential of extensive viewing in an out-of-classroom context (Vanderplank, 2016).

2.2 Factors that influence vocabulary learning through viewing

2.2.1 Glossed captions

The added value of captions versus non-captioning for vocabulary learning is now well established (Montero Perez et al., 2013). On the other hand, the differential effect of subtitles and captions on vocabulary learning is still unclear. The majority of studies examining this issue have shown more positive effects of captions on certain aspects of word knowledge (e.g., Frumuselu, De Maeyer, Donche, & Gutiérrez-Colon Plana, 2015; Peters, 2019; Peters et al., 2016). However, computer-assisted technology now allows to use more advanced types of captioning such as glossed captions. Glossed captions provide learners with the opportunity to access word meaning while watching.

Montero Perez et al. (2018) investigated the effects of enhancing a short (i.e., less than five minutes) video with different types of captions, including glossed keyword captions, on the participants' vocabulary gains. The glossed keyword captions group significantly outperformed the other groups on the form recognition and meaning recall tests, and the number of lookups was positively related with the correct recognition of a target word. Hsieh (2019) found similar results when comparing the effects of full captions

with highlighted target words with L1 meaning with other types of captioning (i.e., **full captions**, **no captions**) on form recognition and meaning recognition and recall. Results of these two studies suggest that giving access to the L1 meaning during the viewing activity can enhance vocabulary gains. In both studies, however, only a selection of the words were glossed. Besides, while learners could decide and click on words for which they wanted the translation in Montero Perez et al. (2018), this was not the case in Hsieh (2019) as translations appeared automatically on the screen. There is therefore still little understanding of how learners benefit from and use glossed captions when watching audio-visual input (Gass, Winke, Isbell, & Ahn, 2019).

2.2.2 *Learners' vocabulary size*

Learners' vocabulary size is defined as an estimate of the number of words for which learners know some aspects of word meaning. It is typically measured with a frequency-based vocabulary test (see <https://www.lectutor.ca/tests/> for examples) that includes a selection of items per frequency band. It is strongly assumed that learners' vocabulary size plays a role in incidental vocabulary acquisition through meaning-focused activities (Horst, Cobb, & Meara, 1998; Webb & Chang, 2015).

Previous research **has** demonstrated that, although less consistent than in reading studies (Webb, 2019), learners' vocabulary size positively impacts vocabulary gains through watching short videos, independent of learners' proficiency (Peters et al., 2016; Peters & Webb, 2018). However, its influence on vocabulary gains through more extensive exposure is less clear-cut. Suárez and Gesa (2019) found a relationship between vocabulary size and participants' meaning and form recall scores. Authors (XXXX) looked at the influence of vocabulary size and found only a limited effect on participants' meaning recall and recognition of the target words. Similarly, Rodgers (2013), who researched the impact of watching 420 minutes of a series, did not find any correlation between learners' vocabulary size and learning gains. It thus remains unclear whether and how learners' vocabulary size plays a role in vocabulary uptake in an extensive viewing context.

2.2.3 *Relative frequency of occurrence*

It is assumed that repetitive exposure to a lexical item is necessary in order to acquire its different knowledge aspects (Webb, 2019). Research has shown that the effect of frequency of exposure, though, is influenced by various variables such as the word knowledge aspect tested, learners' individual differences, time of posttest and mode of input – i.e., written text, spoken text or both (Hashimoto & Egbert, 2019; Uchihara, Webb, & Yanagisawa, 2019). Studies have demonstrated that multiple exposure to unknown vocabulary in written and spoken input contributes to vocabulary gains (Horst et al., 1998; Pigada & Schmitt, 2006; Vidal, 2011).

In the same way, frequency of occurrence appears to have a positive impact on vocabulary gains through audio-visual input. Peters et al., (2016) examined the effect of frequency of occurrence through viewing a documentary and found a positive and significant relationship between frequency of occurrence and the four aspects of word knowledge tested (i.e., form and meaning recognition, form and meaning recall). Similarly, Peters and Webb (2018) found that frequency of occurrence had a significant effect on learning gains after watching a full-length documentary.

Nonetheless, the relationship between frequency of occurrence and vocabulary gains through audio-visual input in experiments adopting a multiple-session design seems less substantial. Pujadas (2019) researched the effect of frequency of occurrence (i.e., from two to 20 times) through watching 525 minutes of audio-visual input across a whole school year. Low to intermediate participants' gains were evaluated by means of an aural form recall and written meaning recall tests at the end of each term (i.e., after watching 8 episodes). With increased frequency of 1, the odds for a correct answer on the form and meaning recall tests increased by 5.2% and 6.2%. Similarly, Rodgers and Webb (2019) found a small but positive influence of frequency (i.e., from 5 to 54 times) on their *tough*¹ meaning recognition test only.

¹ Rodgers and Webb (2019) administered two meaning recognition tests: a tough (i.e., with semantically related distractors) and sensitive (i.e., with distractors from a different part of speech) meaning recognition test

When comparing the influence of frequency of occurrence on L2 vocabulary learning from one video and from multiple videos, it appears that the impact of frequency might be diminished by the number of videos watched. Indeed, although studies have demonstrated that intentional decontextualized vocabulary learning occurs more easily through spaced (i.e., items are spaced out over time) than massed learning (i.e., items are encountered massed together) (Nakata, 2015; Nakata & Suzuki, 2019), incidental vocabulary learning seems to be more stimulated by massed learning conditions (Uchihara et al., 2019). Webb (2014) had predicted that the effects of frequency of occurrence “may be greatest when repeated encounters occur within a short span” (p. 2) and a couple of studies have investigated that issue empirically. Rodgers and Webb (2019) found a small positive correlation of frequency of occurrence with their tough test only. They also examined the effect of range of occurrence, that is the number of episodes in which the target words appeared. Results showed it had a negative and small influence on the tough meaning recognition test. Similarly, Pujadas (2019) also found that the word forms and meanings were more likely to be recalled if they appeared in a single episode rather than across different ones.

Importantly, in those studies, the delay between the last encounter with the target item and the posttest (i.e., retention interval) was greater than in studies using a single video. This might have influenced the results and the impact of frequency of occurrence. However, none of these studies examined its potential influence.

3 Rationale and research questions

While the studies reviewed above point to a beneficial role of extensive exposure to audio-visual input for incidental vocabulary learning, more research is still needed on the potential of viewing a French series with glossed captions, in an out-of-classroom context, for vocabulary learning (Vanderplank, 2016). Moreover, this paper will also try to clarify the role of different moderating variables, namely the effect of (1) vocabulary size, (2) of the use of the glossed captions and (3) of word-related factors that might affect incidental vocabulary learning in such a setting. The following research questions guided this study:

- 1) Does watching a series in an out-of-classroom context lead to vocabulary gains?

- 2) To what extent do learner-related variables (i.e., vocabulary size and lookup behaviour) and word-related factors (i.e., retention interval, relative frequency of occurrence) influence vocabulary gains?

4 Methodology

4.1 Participants

The participants were 111 Dutch-speaking undergraduate students in their first or second year of university. Three participants whose L1 was French and six students who did not attend the sessions in the required period were removed, bringing the total number of participants to 102 (30 males, 72 females). Among them, 65 formed the treatment group who watched the series. The control group consisted of two classes ($N= 37$) and only took the pre- and posttests. Their age ranged from 17 to 20 years and they studied French as L2 ($N= 15$), pedagogical sciences ($N= 28$), law ($N= 43$), economics ($N= 9$), computer science ($N= 2$) or medicine ($N= 7$). According to their scores on the vocabulary size test, students were considered low to high-intermediate learners of French (L2 of this study). Participants of the experimental group received remuneration in exchange for their time, while the control group was not reimbursed.

4.2 Materials

4.2.1 Input

Six episodes from the first season of the series *Dix pour cent* (i.e., *Call my agent!*) were used as input for the experimental group. The participants watched the episodes on the streaming service *Netflix*. The mean length of the episodes was 51 minutes and they varied in length from 47 to 54 minutes, totalling 307 minutes of exposure.

As it is assumed that global comprehension of a series is possible with knowledge of 95% of the input lexical coverage (Webb & Rodgers, 2009), the vocabulary profile of the episodes was calculated using *Lextutor* (www.lextutor.ca/vp/comp). Lextutor matches the words of a text to the words of a (usually corpus based) frequency list. For the first three episodes, 95% of the lexical coverage was attained with the 3,000

Figure 1 – Lexical coverage per episode

most frequent words (including proper nouns), whereas it was reached with the 4,000 most frequent words for the last three episodes (see figure 1).

4.2.2 Glossed captions

Captions were the original captions from *Netflix*. For the glossary, we chose to use the freely available Chrome extension *Language Learning with Netflix*² to give the participants direct access to L1 translations. In order to access the translation, participants had to click on the word. This action paused the episode and a translation with, when applicable, other examples from the word used in the episode, appeared (see figure 2).

Figure 2 – Screen capture of an item in the glossed captions

Since this extension was not developed in the context of this experiment but is a freely available extension and is still in its early development, the French-Dutch dictionary was not always optimal. Therefore, we had to make sure that the translations of the target words were correct (see next section). The participants were informed that the translations might not always be accurate. The participants' screen was recorded as they watched the episodes.

4.2.3 Target items

The target items included in the pre- and posttests were selected based on (1) their frequency as evaluated by Lextutor, on (2) their frequency of occurrence within the six episodes of *Dix pour cent* and on (3) their translation in the glossed captions. First, words that pertained to the 0–2000 frequency lists were not selected because *most of it* was known by the participants. Therefore, the frequency of the target items ranged from the 2000 most frequent to 24000 most frequent words. Second, word families (i.e., *prévenir – prévient; to warn – warns*) that occurred two or more times in the episodes were included in the selection. Finally, we had to make sure the translations of the target words were correct in the Chrome extension *Language*

² <https://languagelearningwithnetflix.com/>

Learning with Netflix. This resulted in a first set of 78 target items. To avoid any test fatigue, we decided not to add any distractors. Targets included 37 nouns, 23 verbs, 9 adverbs and 9 adjectives.

The average frequency of occurrence of the set was 4.72, with a minimum of two and maximum of 20 appearances. The mean range of occurrence (i.e., number of episodes the word appeared in) was 2.50. This gave an average relative frequency (i.e., frequency divided by range of occurrence) of 2.02, ranging from one to 13.

4.3 Instruments

4.3.1 Vocabulary size test

Participants' vocabulary size test in French was measured with the bilingual version of the Vocablab test (Noreillie, 2019). The test contains 150 multiple-choice items, with 30 items per frequency band (0–5000). For each item, learners have to choose the correct meaning of the L2 item between three distractors, the key and an *I don't know* option. The results of this test will be used in the analysis in order to control for learners' differences in French proficiency. The reliability of the test was high ($\alpha = .95$).

4.3.2 Vocabulary tests

Vocabulary gains were evaluated at two different levels of sensitivity: written form recall and aural and written meaning recall. In the form recall test, participants are asked to actively recall the target items by providing the L2 form based on the L1 prompt, whereas the meaning recall test measures participants' passive recall of the target items, by requiring the L1 translation of an L2 prompt. Both tests eliminate guessing and are presumed to better match the needed word knowledge for receptive skills (Schmitt, 2019).

For the form recall pre- and posttest, the first letter, as well as the number of letters, of each target item was given. Reliability for both pre- ($\alpha = .95$) and posttest ($\alpha = .94$) was high. For the meaning recall test, the target words were listed on a sheet of paper and presented twice aurally by a native speaker with 5 seconds between each target. Reliability for both pre- ($\alpha = .91$) and posttest ($\alpha = .93$) was high. The form recall test was administered first to avoid the meaning recall test to facilitate completion (Nation & Webb,

2011). For both form and meaning recall, the pre- and posttests included the same target words, only the order of presentation of the words was changed.

4.3.3 Viewing comprehension tests

Since the viewing activity had to be a meaning-focused one, a viewing comprehension test consisting of 10 multiple-choice and 10 true/false questions was developed for each episode. Questions were in Dutch in order not to interfere with L2 reading skills.

4.3.4 Questionnaire

The questionnaire of the treatment group consisted of Likert-scale questions that focused on: (1) how much contact participants have with French outside of the classroom, (2) the perceived usefulness and the ease of use of the glossed captions, (3) their opinion on the series and on the settings of the experiment (i.e., out-of-classroom), and finally (4) whether they had expected to receive vocabulary posttests. Figure 3 represents examples of items from the questionnaire (see Appendix).

Figure 3 – Examples of items from the questionnaire

The questionnaire of the control group also contained questions about their contact with French outside the classroom, but also on their general contact with French during the 2 weeks between the pre- and posttests. The second part focused on whether they had expected identical posttests and whether they had tried to remember words from the tests or not.

4.4 Procedure

In order to find participants, a call for participation was sent out via e-mail to all students of the campus. The e-mail described the experiment with details on the time needed to complete it, the number of sessions and a general description of the goal of the study (i.e., how we acquire French by watching Netflix). Participants who were interested could sign up via a link.

Each participant of the treatment group had to take part in 7 sessions of approximately one hour. After signing up for the experiment, they received an e-mail with details on how to plan their sessions on a

schedule created for the experiment. There were a couple of conditions to meet: (1) the first session, during which they completed the pre-tests and vocabulary size test, had to take place the week before the first session of viewing and (2) the rest of the sessions had to be spread over a maximum of two weeks. The posttests and questionnaire were administered right after the last episode so that the students did not have to come back an eighth time. This means they had a maximum of 21 days to complete the whole experiment. The mean number of days to complete the experiment was 16 days, with a minimum of 11 and a maximum of 21 days.

During their first session, participants first received the form recall test, then the vocabulary size test and finally the meaning recall test. This session usually took them 50 minutes to complete. The six following sessions were dedicated to viewing the episodes and completing the viewing comprehension tests. Before the first episode, explanations were given on the use of the glossed captions. Participants were told they were not obligated to use them and that their screen would be recorded. They watched the series on an individual laptop with a headset. In order to accentuate the out-of-classroom, more ‘informal’ context, we decided to use tools (i.e., Netflix and *Language Learning with Netflix*) that learners could use if they were trying to learn a language by themselves, outside of the curriculum. Besides, they were offered free drinks during each session. During the last session, participants completed the form recall, questionnaire and meaning recall posttests after watching the last episode and completing the viewing comprehension test.

Participants from the control group took part in 3 sessions: one for the pre-tests, one for the vocabulary size test and one for the posttests. Because of time constraints, the vocabulary size test and pre-tests could not be administered during the same session. The form recall and meaning recall posttests were separated by a short questionnaire. They completed the vocabulary size test during the first session, two weeks later they received the pre-tests and two weeks later the posttests and questionnaire. Each session took approximately 20 to 30 minutes.

4.5 Scoring and analyses

The viewing comprehension and vocabulary tests were scored dichotomously. In the form recall test, spelling mistakes were considered incorrect if they involved a change in pronunciation of the word (e.g., *monstrieux* instead of *monstrueux*). However, if a spelling mistake did not change the pronunciation (e.g., *susette* instead of *sucette*), the answer was considered correct. In the meaning recall test, a good definition, synonym or translation was considered correct. In case of polysemous words, only the meaning shown in the episode and used in the glossed captions was accepted as a correct answer. All tests were scored by the first author. In case of hesitation, the answer was discussed with a Dutch speaker highly proficient in French. In order to check lookup behaviour, the first author watched all screen recordings. Each time a target word was clicked on was considered as a lookup.

In order to account for the variation in participants' scores in the pre-tests, a word was considered 'known' when it was answered correctly in the pre- and posttests and 'learned' if it was unknown in the pre-test but known in the posttest (Horst et al., 1998). Relative gains are then calculated for each subject by dividing the *learned words* by the number of target words still available for learning in the posttest (i.e., $(\text{learned words} / (\text{Number of target words} - \text{known words})) * 100$). These relative gains were used to answer both research questions.

We performed two independent-samples *t*-tests with learners' relative gains on the form and meaning recall tests as dependent variables and group (i.e., control or treatment) as independent variable to answer the first research question. In order to answer the second research question, Generalised estimating equations (GEE) were conducted to perform a repeated measures logistic regression. The assumptions of logistic regression (e.g., dichotomous response variable, multi-collinearity) were all met. In the analyses, the repeated measures within participants were the 78 target words. The predictors were the vocabulary size test results, the number of times a target word was looked up, the relative frequency and retention interval. Retention interval was calculated for each target word and participant individually ($M = 2.87$, $SD = 3.36$, $Min = 0$, $Max = 10.03$).

5 Results

5.1 Vocabulary size

The descriptive statistics of the vocabulary size test are displayed in Table 1. An independent-samples *t*-test indicated there was no significant difference between the treatment and control groups' scores ($t(100) = .012, p = .991$).

Table 1 – Vocabulary size test results

| Group | 0-1K Mean (SD) | 1K-2K Mean (SD) | 2K-3K Mean (SD) | 3K-4K Mean (SD) | 4K-5K Mean (SD) | TOTAL (out of 150) Mean (SD) |
|-----------------------------|-------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|---|
| <i>Treatment (N=65)</i> | 25.89 (3.16) | 22.31 (4.70) | 17.95 (4.97) | 16.04 (5.52) | 13.07 (5.05) | 95.28 (21.59) |
| <i>Control (N=37)</i> | 26.00 (2.21) | 22.35 (3.87) | 17.75 (3.92) | 16.51 (4.61) | 12.70 (4.04) | 95.32 (16.26) |

5.2 Viewing comprehension tests and questionnaire

The results of the comprehension tests indicated that all participants had a good comprehension of the episodes, with a mean of 84%.

Participants reported that they do not often watch French TV with or without captions (on a scale from 1 to 5 where 1= disagree, 5= totally agree, $M = 1.56, SD = 0.92$), and also do not read in French (same scale, $M = 1.38, SD = .8$). When asked whether they assumed there were going to be vocabulary posttests, 85% participants of the treatment group agreed. Different reasons as to why they expected the posttests were given: 37 stated they guessed because of the pre-tests, 10 because there were glossed captions, seven said that it was described in the title of the research (which was 'Learn French with Netflix') and six gave no explanation. In the control group, 40% thought they would have to take the same tests.

On a Likert-Scale from 1 to 5 (1=disagree, 5= totally agree), participants from the treatment group indicated that the series was enjoyable ($M = 4.62, SD = .49$). When asked whether the vocabulary of the

episodes was too difficult, they mostly disagreed ($M = 1.81$, $SD = .79$), although 66% said they would not have understood the series without captions. The vast majority (95.7%) reported having watched the series as if they were at home but were still attentive when watching the episodes (on a scale from 1 to 10, $M = 8.68$, $SD = .87$).

Regarding the use and perceived usefulness of the glossed captions, the results showed that only 17% of the participants would have preferred Dutch subtitles and half of the participants stated they would use the glossed captions if watching a series at home. Table 2 lists the descriptive statistics for the five other items on use and perceived usefulness of the glossed captions.

Table 2 – Descriptive statistics on use and perceived usefulness of glossed captions

| Items on a Likert-scale from 1 to 5 where 1= disagree, 3= no opinion, 5= totally agree | Mean | SD |
|--|-------------|-----------|
| <i>I found the glossed captions useful.</i> | 3.06 | 1.30 |
| <i>I found the glossed captions confusing, because the translations were not always correct.</i> | 3.40 | 1.30 |
| <i>I found the glossed captions user-friendly.</i> | 3.52 | 1.37 |
| <i>I used the glossed captions because I assumed vocabulary tests would follow.</i> | 1.86 | 1.13 |
| <i>The glossed captions helped me answer the vocabulary posttests.</i> | 2.66 | 1.30 |

5.3 RQ1: Does watching a TV series in an out-of-classroom context lead to vocabulary gains?

To determine whether the mean gains on the vocabulary tests were significant, two paired samples t -tests were carried out. The analyses for the experimental group showed that the gains on the form ($t(64) = 11.48$, $p < .000$, $d = .83$) and meaning recall tests ($t(64) = 14.26$, $p < .001$, $d = .98$) were significant. The gains of the control group on the form ($t(36) = 7.34$, $p < .001$, $d = .73$) and on the meaning recall ($t(36) = 3.38$, $p = .002$, $d = .35$) tests were also significant.

As can be observed in Table 3, there were large differences between the mean absolute gains of both groups on the form and meaning recall tests. Independent-samples t -tests showed the difference

between both groups' means to be significant on form recall ($t(100) = 7.56, p < .001, d = 1.56$) and meaning recall ($t(100) = 7.87, p < .001, d = 1.62$) tests.

Table 3 – Descriptive absolute and relative gains vocabulary tests

| Group | Vocabulary test | Known words in % (<i>SD</i>) | Mean absolute gain (<i>SD</i>) | Mean relative gain (<i>SD</i>) | Minimum absolute gain | Maximum absolute gain |
|-----------------------------|-----------------|-----------------------------------|-------------------------------------|-------------------------------------|-----------------------|-----------------------|
| Treatment group (N = 65) | Form recall | 29.8 (18.38) | 14.40 (6.75) | 27.49% (13.61) | 4 | 31 |
| | Meaning recall | 49.62 (15.42) | 13.49 (5.99) | 36.47% (15.86) | 3 | 33 |
| Control group (N = 37) | Form recall | 20.23 (8.81) | 7.92 (3.23) | 12.95% (5.60) | 2 | 16 |
| | Meaning recall | 43.90 (14.03) | 7.46 (3.69) | 17.25% (8.79) | 0 | 16 |

5.4 RQ2: To what extent do learner-related variables (i.e., vocabulary size and lookup behaviour) and word-related factors (i.e., retention interval, relative frequency of occurrence) influence vocabulary gains?

First, the use of the glossed captions is discussed. On average, participants looked up 6 target words out of 78 ($SD = 6.75$). Participants used the glossed captions to a limited extent: 21% did not look up any target word, 40% looked up one to five target words, and the rest looked up 11 to 30 words with only five participants looking up more than 20 items. Nine target words were not looked up at all, while the 69 other target words were looked up by on average 5.88 participants out of 65 ($SD = 7.11$) with a minimum of 1 and maximum of 43.

As stated before, two Generalised Estimating Equations were conducted. Meaning and form recall gains were entered as dependent variable and relative frequency and retention interval, as well as vocabulary size results and lookup behaviour were the covariates. Through a backward stepwise selection, any

predictors or interactions that were not significant were removed and we reran the model with only the significant main effects.

Table 4 – Number of observations per vocabulary test GEE

| | Correct answers | Incorrect answers | Total observations |
|-----------------------|------------------------|--------------------------|---------------------------|
| Form recall | 916 (25.7%) | 2655 (74.3%) | 3571 |
| Meaning recall | 877 (34.3%) | 1678 (65.7%) | 2555 |

For form recall, the analysis was performed with 3751 observations (i.e., number of words still available for learning in the post-test, see Table 4). As can be observed in Table 5, all variables except retention interval had a significant impact on form recall gains. For each additional time a word was looked up, the odds for recalling the form were 1.4 times higher. When the relative frequency of the target word increased by one, the odds for a correct answer on the form recall test were 1.35 times higher. Regarding vocabulary size, there were only 2% more odds to answer correctly on the form recall test with each additional known word on the vocabulary size test.

Table 5 – GEE results for form recall

| | B | χ^2 | df | p | Exp(B) |
|--------------------|----------|----------------------------|-----------|----------|---------------|
| (Intercept) | -3.830 | 139.559 | 1 | .000 | .022 |
| Relative frequency | .301 | 73.303 | 1 | .000 | 1.351 |
| Vocabulary size | .023 | 55.028 | 1 | .000 | 1.023 |
| Lookups | .334 | 9.307 | 1 | .002 | 1.397 |

Regarding meaning recall results, the analysis was performed with 2555 observations (see Table 4) and all variables were related to learning (see Table 6). Retention interval had a limited impact on the learning gains. With one additional day separating the last encounter with a target word from the test, the odds of recalling the word meaning decreased by less than 1%. The odds for recalling the meaning of a word increased by 1.62 each time it was looked up. When the relative frequency of the target word increased by

one, the odds for a correct answer on the meaning recall test were 1.36 times higher. Finally, with each additional word known on the vocabulary size test, the odds for a correct answer increased by 3%.

Table 6 – GEE results for meaning recall

| | B | χ^2 | df | <i>p</i> | Exp(B) |
|--------------------|----------|----------------------------|-----------|-----------------|---------------|
| (Intercept) | -2.622 | 37.635 | 1 | .000 | .073 |
| Relative frequency | .306 | 51.950 | 1 | .000 | 1.357 |
| Retention interval | -.062 | 10.685 | 1 | .001 | .939 |
| Vocabulary size | .025 | 59.949 | 1 | .000 | 1.026 |
| Lookups | .486 | 17.253 | 1 | .000 | 1.625 |

6 Discussion

This paper expanded on the literature on incidental vocabulary learning through viewing in several ways:

(1) it is one of the first studies to have a multiple-session design and to be conducted outside a classroom context, (2) it examined the influence of glossed captions via participants' lookup behaviour and (3) explored the influence of learner- and word-related factors.

6.1 RQ1: Vocabulary gains through watching a TV series

In answer to the first research question, the results indicated that the treatment was beneficial, since participants recalled, on average, the form of 14 words and 13 word meanings. This represents relative gains of 27% and 36% respectively. Our results support those found in Suárez and Gesa (2019) and Pujadas and Muñoz (2019). In the former, learners scored 29% on the form recall and 16% on the meaning recall tests after 195 minutes of exposure, which corresponded to 22 word forms and 15 word meanings. Compared to the results of the present study, Suárez and Gesa's (2019) findings might seem higher. However, it should be noted that all students in their study were pre-taught the set of target words, hence improving their scores. Similarly, in Pujadas and Muñoz (2019) students who watched the series with captions recalled 9 word forms and 6 word meanings after watching 525 minutes of videos, corresponding to 13% and 6% respectively. The reason why learners in the present paper scored considerably higher on the meaning recall

test might partly be explained by Pujadas and Muñoz's test format, since recalling meaning was dependent on the correct recalling of the form.

Our results are also in line with other studies with a long treatment. Authors (XXXX) tested the meaning recall of their participants after 125 minutes of exposure to educational videos, results showed that an average of 5 word meanings were recalled after the treatment (i.e., 13% in relative gains). In Rodgers and Webb (2019) participants recognised 23% and 29% of the target words after 420 minutes of exposure.

Learning gains found on the test might, however, underestimate the vocabulary uptake that actually occurred. First, words from the 0-2000 frequency bands were not tested, while the results of the vocabulary size test demonstrated that some learners did not master these frequency bands. Second, words that were not correctly translated in the glossed captions were not selected as potential targets but learners might have acquired them. This is confirmed by answers on the questionnaire, on which learners could list words and other aspects of language (e.g., grammar, pronunciation) they thought they acquired during the treatment. Answers comprised some single or multiword items that were not tested such as *metteur en scène*, *cocotte*, *fait chier*, *n'importe quoi*, *deuil*, *bourré*, *enceinte*, *menacer*. Participants also indicated that they thought they improved their listening comprehension, the pronunciation of certain words, the verb time, sentence construction and fluency. This again demonstrates that watching audio-visual input is a beneficial activity for language learning.

6.2 Influence of learner-related variables

6.2.1 Vocabulary size

Both form and meaning recall gains were positively influenced by an increase in the results of the vocabulary size test. That means that participants who scored higher on the vocabulary size test learned more target words than the others. This trend corroborates Gesa and Suárez's (2019) results, which also pointed to a positive influence of vocabulary size on both form and meaning recall gains. However, it contradicts findings of Rodgers (2013) who did not find any correlation between tests' results and vocabulary size, as

well as Authors (XXXX), in which the influence of vocabulary size on learners' meaning recall gains was very limited.

The greater influence in our study might be explained by different variables. First, there was more variation in the vocabulary size test results than in Authors (XXXX). Second, it can be hypothesised that the series was easier to follow and more adapted than the input in Authors (XXXX) and the series used in Rodgers (2013). Indeed, Noreillie, Kestemont, Heylen, Desmet, and Peters (2018) argued that knowledge of only 1000 or 2000 lemmas, or 86% and 93% of lexical coverage, is enough for adequate listening comprehension of French. It might be assumed that, for viewing, these numbers may even be lower since imagery can support comprehension and word-meaning mapping (Rodgers, 2018). In the present study, 86% and 93% lexical coverage were attained with the 1000 and 2000 levels, which 70% of participants mastered. In Rodgers (2013), on the other hand, only 1.06% of the participants mastered the frequency bands needed to attain 95% coverage of the ten episodes, and in Authors (XXXX) only one participant out of 50 attained the 93% coverage for only 5 videos out of 15. We therefore argue that, thanks to their greater linguistic knowledge, our participants encountered fewer unknown words, could understand the input better, hence were more able to guess, interpret and acquire new word meanings.

6.2.2 Look up behaviour

Although not all target words were looked up by the participants (69 out of 78) and that not all participants (52 out of 65) looked up target words meaning in the glossed captions, lookup behaviour still had an impact on recalling the form and meaning of the words. The influence was slightly greater on the meaning than on the form recall gains. This is not surprising, as the glossed captions offer the L1 translation of the L2 word, which is exactly what is required on a meaning recall test. It is indeed crucial for the form-meaning mapping to occur that learners see the form and meaning together (Nation, 2013). When clicking on a word in the captions, learners are looking to get a translation and might not allocate as much attention to the L2 word form, hence probably remember the L1 translation better than the L2 form. Our results corroborate those of Montero Perez et al. (2018), although the real impact of lookup behaviour on vocabulary learning is difficult

to interpret, as they looked at its impact on the form recognition. There is therefore still more research needed on this subject.

The fact that the glossed captions were not accessed much is not surprising in view of the answers given on the questionnaire. Indeed, learners acknowledged that they mostly tried to infer the meanings based on the context and/or on the imagery. More importantly, some learners also stated that they only consulted the translations of words they considered as essential for comprehension of the story. This phenomenon has also been found in a reading study conducted by Hulstijn (1993). However, in Hulstijn's study, lookup behaviour seemed to be influenced by learners' vocabulary knowledge, which was not the case in the present paper. It might be a consequence of the speed of the captions, since participants also declared that they were often not quick enough to click on the words. Participants then realised they did not need to understand each word to understand the story and therefore stopped trying to get the translations.

6.3 Influence of relative frequency of occurrence

Relative frequency of occurrence influenced both form and meaning recall results. In both cases there was a positive relationship, meaning that words were more likely to be recalled if the relative frequency was higher. Relative frequency is higher when, for example, a word is encountered multiple times in the same episode than across several ones. As a reminder, the mean relative frequency of the target words was of 2.02, ranging from 1 to 13.

The present results are hard to compare with previous findings as this is one of the first studies with several experimental sessions to take relative frequency into account. However, Rodgers and Webb's (2019) results corroborate our findings since they found a small positive correlation between word relative gains on a meaning recognition test and relative frequency. Similarly, Pujadas (2019) have also found that words appearing several times in a single video were more likely learned than the ones occurring in multiple one. Further comparison with viewing studies examining the frequency of occurrence in one-off learning sessions seems to confirm the more beneficial effect of *massed* learning compared to *spaced* learning (Uchihara et al., 2019). Indeed, Peters et al. (2016) and Peters and Webb (2018) found a considerably greater

influence of frequency of occurrence (e.g., 25% more odds to recall word meanings for each additional occurrence). Nonetheless, retention interval might have played a role in our results. It was therefore checked whether the interaction between relative frequency and retention interval significantly influenced learning gains, but it was not the case. This means that relative frequency was not modulated by retention interval.

Retention interval on its own, though, impacted meaning recall in a very limited way, since the odds for recalling the word meaning decreased with fewer than 1% with each day separating the encounter with the word and the posttest.

7 Limitations

While the present study demonstrates that incidental vocabulary learning through extensive viewing with glossed captions is possible, the study has some limitations. A first limitation concerns the fact that the control group also made significant gains by only taking the tests (Peters & Webb, 2018; Rodgers, 2013). However, as previously argued, the participants in the treatment group might have learned words that were not included in the tests. Moreover, the difference in gains between the treatment and control groups was still significant with a considerable effect size. Another limitation is the type of glossary chosen in the first place. For reasons of ecological validity, it was decided to work with a tool that is available to any learner who wishes to learn languages with streaming services. However, while very useful, the development of the French-Dutch dictionary was suboptimal. Some translations were therefore incorrect, or even not translated at all. Another drawback was that multiword items such as idioms or collocations were not considered as a whole but always as separate unique word items, altering the L1 translation. Yet, since we only selected words that were correctly translated, this did not influence the results.

8 Conclusion and implications

The present research points to the potential of extensive viewing for L2 incidental vocabulary learning. It demonstrated that watching a whole series with glossed captions in an out-of-classroom context leads to substantial form-meaning knowledge gains. Moreover, the findings support the hypothesis that learners with a higher vocabulary size will learn more words, but that acquiring new word forms and meanings is still

possible at a lower proficiency. In the same way, looking up the meaning of words during viewing seems to stimulate its learning, even though learning also happened when guessing from context. Finally, the present paper was one of the first viewing studies that made use of a French programme and a multiple-session design. It is therefore also one of the first studies investigating the effect of relative frequency of occurrence. The results seem to indicate that the impact of frequency of occurrence on word learning is moderated by range of occurrence, in other words, words encountered more concentrated in a single episode appear to be more likely learned than those encountered once across multiple ones.

The findings of the present study have pedagogical implications. Results confirm that viewing as an out-of-classroom activity can lead to incidental word learning in the same way as extensive reading (Feng & Webb, 2020). Teachers should therefore encourage their students to engage in (extensive) viewing for L2 vocabulary learning as an out-of-classroom activity, even at a low proficiency level. Furthermore, findings point to the effectiveness of glossed captions for vocabulary learning. Teachers should therefore consider using this type of on-screen text, by for example, using new extensions such as *Language Learning with Netflix*, to support learners in their learning process.

References

- Besnehard, D. (2015). *Dix pour cent [Television series]*. Paris, France: France 2.
- Feng, Y., & Webb, S. (2020). Learning vocabulary through reading, listening, and viewing. *Studies in Second Language Acquisition*, 42(3), 499–523. <https://doi.org/10.1017/S0272263119000494>
- Frumuselu, A. D., De Maeyer, S., Donche, V., & Gutiérrez-Colon Plana, M. (2015). Television series inside the EFL classroom: Bridging the gap between teaching and learning informal language through subtitles. *Linguistics and Education*, 32(Part B), 107–117. <https://doi.org/10.1016/j.linged.2015.10.001>
- Gass, S., Winke, P., Isbell, D. R., & Ahn, J. (2019). How captions help people learn languages: A working-memory, eye-tracking study. *Language Learning & Technology*, 23(2), 84–104.

- Hashimoto, B. J., & Egbert, J. (2019). More Than Frequency? Exploring Predictors of Word Difficulty for Second Language Learners. *Language Learning*, 69(4), 839–872. <https://doi.org/10.1111/lang.12353>
- Horst, M., Cobb, T., & Meara, P. (1998). Beyond a Clockwork Orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, 11(2), 207–223.
- Hsieh, Y. (2020). Effects of video captioning on EFL vocabulary learning and listening comprehension. *Computer Assisted Language Learning*, 33(5–6), 567–589. <https://doi.org/10.1080/09588221.2019.1577898>
- Hulstijn, J. H. (1993). When Do Foreign-Language Readers Look up the Meaning of Unfamiliar Words? The Influence of Task and Learner Variables. *The Modern Language Journal*, 77(2), 139–147.
- Lindgren, E., & Muñoz, C. (2013). The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *International Journal of Multilingualism*, 10(1), 105–129. <https://doi.org/10.1080/14790718.2012.679275>
- Montero Perez, M., Peters, E., & Desmet, P. (2018). Vocabulary learning through viewing video: The effect of two enhancement techniques. *Computer Assisted Language Learning*, 31(1–2), 1–26. <https://doi.org/10.1080/09588221.2017.1375960>
- Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, 41(3), 720–739.
- Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning: Does Gradually Increasing Spacing Increase Vocabulary Learning. *Studies in Second Language Acquisition*, 37(4), 677–711. <https://doi.org/10.1017/S0272263114000825>
- Nakata, T., & Suzuki, Y. (2019). Effects of massing and spacing on the learning of semantically related and unrelated words. *Studies in Second Language Acquisition*, 41(2), 287–311. <https://doi.org/10.1017/S0272263118000219>

- Nation, P. (2007). The Four Strands. *Innovation in Language Learning and Teaching*, 1(1), 1–13.
<https://doi.org/10.2167/illt039.0>
- Nation, P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Noreillie, A. (2019). *It's all about words: Three empirical studies into the role of lexical knowledge and use in French listening and speaking tasks*. (Unpublished dissertation) Catholic University of Leuven, Belgium.
- Noreillie, A., Kestemont, B., Heylen, K., Desmet, P., & Peters, E. (2018). Vocabulary knowledge and listening comprehension at an intermediate level in English and French as foreign languages. *ITL - International Journal of Applied Linguistics*, 169(1), 212–231. <https://doi.org/10.1075/itl.00013.nor>
- Peters, E. (2018). The effect of out-of-class exposure to English language media on learners' vocabulary knowledge. *ITL - International Journal of Applied Linguistics*, 169(1), 142–168.
<https://doi.org/10.1075/itl.00010.pet>
- Peters, E. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from audiovisual input. *TESOL Quarterly*, 53(4), 1008–1032. <https://doi.org/10.1002/tesq.531>
- Peters, E., Heynen, E., & Puimège, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System*, 63, 134–148.
<https://doi.org/10.1016/j.system.2016.10.002>
- Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, 40(3), 551–577.
<https://doi.org/10.1017/S0272263117000407>
- Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, 18(1), 1–28.

- Pujadas, G. (2019). *Language learning through extensive TV viewing: A study with adolescent EFL learners*. (Unpublished dissertation) University of Barcelona, Spain.
- Pujadas, G., & Muñoz, C. (2019). Extensive viewing of captioned and subtitled TV series: A study of L2 vocabulary learning by adolescents. *The Language Learning Journal*, 47(4), 479–496. <https://doi.org/10.1080/09571736.2019.1616806>
- Rodgers, M. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions*. (Unpublished dissertation). Victoria University of Wellington.
- Rodgers, M. (2016). Extensive listening and viewing: the benefits of audiobooks and television. *The European Journal of Applied Linguistics and TEFL*, 5(2), 43–57.
- Rodgers, M. (2018). The images in television programs and the potential for learning unknown words. *ITL - International Journal of Applied Linguistics*, 169(1), 191–211. <https://doi.org/10.1075/itl.00012.rod>
- Rodgers, M., & Webb, S. (2020). Incidental vocabulary learning through viewing television. *ITL - International Journal of Applied Linguistics*, 171(2), 191–220. <https://doi.org/10.1075/itl.18034.rod>
- Schmitt, N. (2019). Understanding vocabulary acquisition, instruction, and assessment: A research agenda. *Language Teaching*, 52(2), 261–274. <https://doi.org/10.1017/S0261444819000053>
- Suárez, M. M., & Gesa, F. (2019). Learning vocabulary with the support of sustained exposure to captioned video: do proficiency and aptitude make a difference? *The Language Learning Journal*, 47(4), 497–517. <https://doi.org/10.1080/09571736.2019.1617768>
- Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The Effects of Repetition on Incidental Vocabulary Learning: A Meta-Analysis of Correlational Studies. *Language Learning*, 69(3), 559–599. <https://doi.org/10.1111/lang.12343>
- Vanderplank, R. (2016). ‘Effects of’ and ‘effects with’ captions: How exactly does watching a TV

programme with same-language subtitles make a difference to language learners? *Language Teaching*, 49(2), 235–250. <https://doi.org/10.1017/S0261444813000207>

Vidal, K. (2011). A Comparison of the Effects of Reading and Listening on Incidental Vocabulary Acquisition. *Language Learning*, 61(1), 219–258. <https://doi.org/10.1111/j.1467-9922.2010.00593.x>

Webb, S. (2014). Repetition in Incidental Vocabulary Learning. In C. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 1–6). <https://doi.org/10.1002/9781405198431.wbeal1425>

Webb, S. (2015). Extensive viewing: language learning through watching television. In D. Nunan & J. C. Richards (Eds.), *Language learning beyond the classroom* (pp. 159–168). New York, NY: Routledge.

Webb, S. (2019). *The Routledge Handbook of Vocabulary Studies*. New York, NY: Routledge.

Webb, S., & Chang, A. C.-S. (2015). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *Language Teaching Research*, 19(6), 667–686. <https://doi.org/10.1177/1362168814559800>

Webb, S., & Rodgers, M. (2009). Vocabulary Demands of Television Programs. *Language Learning*, 59(2), 335–366. <https://doi.org/10.1111/j.1467-9922.2009.00509.x>

Appendix

Opdracht: Answer to the questions as honestly as possible.

- Name and surname:
- Age: year old
- Language(s) that are used at home:
- With my family (with grandmother, cousin, uncle...), I usually speak:
.....
- With my friends, I usually speak
- In my high school, most lessons were taught in(language).
- Outside of the school environment, I often get in contact with French:
 - Yes
 - No
 - If yes, please give a number of examples (e.g., at work, 2x/week)
.....
- How do you evaluate your reading skills for French? (1 = very bad, 10 = outstanding)

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|----|

- How do you evaluate your listening skills for French? (1 = very bad, 10 = outstanding)

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|----|

- How do you evaluate your general language skills for French? (1 = very bad, 10 = outstanding)

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|----|

| | Totally agree | Agree | Not really agree | Disagree | No opinion |
|--|---------------|-------|------------------|----------|------------|
| I often watch French movies, series, videos, TV programs,... To what extent? | | | | | |
| I often watch subtitled French movies, series, videos, TV programs,... To what extent? | | | | | |
| I often read French books. To what extent? | | | | | |
| I often read French press (e.g., journal, magazine). To what extent? | | | | | |
| I expected to receive tests at the end of the experiment. | | | | | |
| I found the serie nice to watch. | | | | | |
| I found the subtitles distracting, I would rather not activate them. | | | | | |
| I would rather have Dutch subtitles. | | | | | |
| The vocabulary of the episodes was difficult. | | | | | |
| I did not understand the episodes because there were too many unknown words. | | | | | |

| | Totally agree | Agree | Not really agree | Disagree | No opinion |
|---|---------------|-------|------------------|----------|------------|
| I found the glossed captions useful. | | | | | |
| I found the glossed captions confusing, because the translations were not always correct. | | | | | |
| I used the glossed captions because I assumed vocabulary tests would follow. | | | | | |
| I used the glossed captions mainly for words I did not know. | | | | | |
| The glossed captions helped me answer the vocabulary posttests. | | | | | |
| I used the glossed captions mainly because of the comprehension tests given after each episode. | | | | | |
| The glossed captions helped me answer the comprehension tests. | | | | | |

| | | | | | |
|---|--|--|--|--|--|
| I found the glossed captions user-friendly. | | | | | |
| I have mainly used the glossed captions in the first episode. | | | | | |
| Would you use the glossed captions if you watched a French series at home? | | | | | |
| Did you use the glossed captions in each episode? <input type="radio"/> Yes <input type="radio"/> No Why? | | | | | |
| When did you consult the glossed captions? Can you explain the main reason(s) why you used the glossed captions? | | | | | |

- Did you find the glossed captions user-friendly?
 - ☐ Yes
 - ☐ No
- I had a suspicion that the main goal of the experiment was to learn vocabulary:
 - ☐ Yes, why?
 - ☐ No (why?):

- I was attentive while watching the episodes

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|----|

- I have watched the series as if I was at home (e.g., I did not feel obligated to use the glossed captions)

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|----|

- I was attentive while completing the tests

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|----|

- I have mainly focused on understanding the content of the episodes when I was watching

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|----|

- I have often tried to remember/learn new words during the experiment:
 - Yes
 - No
- If yes, can you give a number of examples of words you have learned thanks to this experiment?
 -
 -
- Have you learned other linguistic aspects through watching the series? If yes, can you explain this?
 -
 -
 -
 -
- Did you feel that your listening skills improved throughout the episodes?
 - Yes
 - No
- Would you have understood the series without French subtitles?
 - Yes
 - No