

U-shaped Trajectories in an L2 context: Evidence from the acquisition of verb morphology

Stefan Williams

Department of Language and Literature
Norwegian University of Science and Technology
stefan.williams@ntnu.no

Pedro Guijarro-Fuentes

Department of Spanish, Modern, and Classical Languages
University of the Balearic Islands
p.guijarro@uib.es

Mila Vulchanova

Department of Language and Literature
Norwegian University of Science and Technology
mila.vulchanova@ntnu.no

Abstract

This study explores U-shaped behaviour in the acquisition of irregular verb morphology across three groups of Norwegian L2 learners of English. This phenomenon is especially interesting due to its significance for the organization and division between the mental lexicon and grammar. We hypothesized that if U-shaped behaviour was observable, then we would find significant differences in participants' performance accuracy levels in conjunction with overregularization errors. We report results on the acquisition of irregular verb morphology, in addition to mean reaction times on different types of responses (accurate responses and overregularized ones). The final analysis includes data from participants from the 8th grade (N=17), 9th grade (N=19), and 10th grade (N=15). We report results of the acquisition of irregular verbs, in addition to reaction latencies. Participants responded to an elicitation task to test performance on 40 items. The results are consistent with the later stages of U-shaped learning. We found an increase in overall accuracy co-occurring with a decrease in overregularization errors. We propose that the existence of U-shaped behaviour in the L2 is indicative of a general underlying input-driven learning pattern, and that this process is an integral part of acquiring knowledge upon exposure to irregularities in a productive paradigm.

Keywords: U-Shaped learning; irregular verb morphology; English, Norwegian; Second Language Acquisition.

Resumen

Este estudio explora el comportamiento en forma de U en la adquisición de morfología de verbos irregulares en tres grupos de estudiantes noruegos L2 de inglés. Este fenómeno es especialmente interesante por su importancia para la organización y división entre el léxico mental y la gramática. Planteamos la hipótesis de que si el comportamiento en forma de U fuera observable, entonces encontraríamos diferencias significativas en los niveles de precisión de los participantes junto con errores de sobrerregularización. Este estudio informa de los resultados sobre la adquisición de morfología verbal irregular, además de tiempos medios de reacción en diferentes tipos de respuestas (respuestas precisas y sobrerregularizadas). El análisis final incluye datos de aprendices del octavo grado (N = 17), noveno grado (N = 19) y décimo grado (N = 15). Informamos de los resultados de la adquisición de verbos irregulares, además de los tiempos de reacción. Los participantes respondieron a una tarea de elicitación compuesta de 40 ítems. Los resultados son consistentes con las últimas etapas del aprendizaje en forma de U. Encontramos un aumento en la precisión general que concurre con una disminución en los errores de sobrerregularización. Proponemos que la existencia de un comportamiento en forma de U en la L2 es indicativo de un patrón de aprendizaje impulsado por modelos subyacentes generales, y que este proceso es una parte integral de la adquisición del conocimiento al exponerse a irregularidades en un paradigma productivo.

Palabras clave: aprendizaje en forma de U; morfología de verbos irregulares; inglés; noruego; adquisición de una segunda lengua.

1. Introduction

Phenomena such as grammatical errors and overregularization have been a topic of interest in the field of language acquisition, especially in relation to verb inflectional morphology. As pointed out by Marcus et al. (1992:1), overregularization of irregular verbs, such as “comed” instead of “came” for instance, has been observed for as long as language development has been a topic of study.

Closer studies of overregularization errors have revealed an interesting developmental curve, first reported by Ervin & Miller (1963; see also Cazden, 1968), and currently referred to as the *U-shaped developmental curve*. Children with English as their L1 produce correct irregular forms of verbs, if they mark the past tense at all, up

until their third year. However, around the age of three, they begin producing errors and overregularize irregular verbs that they previously have been able to mark correctly. Around the same time as they start producing overregularization errors, children begin acquiring the rule for marking regular past tense verbs (Pinker, 1998).

The U-shaped developmental curve has been extensively studied in first language acquisition research. While U-shaped behaviour has also been observed in the domain of second languages, few studies address this phenomenon experimentally in relation to grammar acquisition in an L2 context. This is an exploratory study seeking to attest the existence of U-shaped learning patterns in acquiring English L2 strong verb morphology within three L2 participant groups at different levels of proficiency. In particular, we were interested in the participants' performance accuracy, as well as the extent to which they tend to overregularize irregular verbs. Thus, if we find evidence of U-shaped development during L2 acquisition in the same domain it has been attested in the L1, the questions are: what is the source of this pattern, and what are the underlying mechanisms supporting it in both L1 and L2 learning (Vulchanova, Foyn, Nilsen & Sigmundsson, 2014). Such findings can shed light on how second languages are acquired in comparison to the first language. Moreover, our understanding of language development has consequences for how we approach and assess language learning in a practical sense. In a language learning environment, and for the purposes of instruction, errors might be used as a litmus to indicate what aspects need more attention. Understanding the developmental trends and which cognitive learning mechanisms are at work might help us better assess and evaluate how learners develop their second language, thereby improving the quality of second language instruction.

In exploring U-shaped development during second language grammar acquisition, an elicitation task containing 40 irregular verbs based on Berko's (1958) seminal Wug-test was designed to elicit the learners' knowledge of irregular verb morphology. Data were collected via the online survey platform *SurveyGizmo*. Norwegian learners with English as their L2 were recruited as the focus group for this study. A cross-sectional design was chosen, recruiting learners from three different school grades: the 8th grade (N=20), 9th grade (N=24), and 10th grade (N=23), respectively.

The present study is structured as follows: Section 1 reviews the literature on U-shaped trajectories in language learning and second language acquisition, while also presenting theoretical models to account for the learning curve. Section 2 describes the methodology of the study, whereas Section 3 reports results divided by group, category, and individual. In Section 4, our results are discussed and interpreted. Finally, Section 5 discusses limitations of the study and presents the main conclusions before suggesting possible future research on U-shaped development in second language acquisition.

1.1. U-shaped trajectories during language learning and other domains

Initial studies on U-shaped development began, among others, with Cazden's (1968; see also Ervin & Miller, 1963) longitudinal study of child L1 production. Cazden (1968) observed how children exhibited a period of correct performance when producing the irregular past tense verbs prior to overgeneralization errors. Overregularization of irregular nouns was also observed, whereby the participants applied the productive rule to an irregular stem resulting in forms, such as "feets" instead of "feet". Moreover, there seemed to be individual differences in error rates among participants. The same pattern of performance was later supported by a larger study utilizing data from spontaneous language produced by 83 children (Marcus et al., 1992). In addition to affirming the U-shaped learning curve, this study gave further insight into the acquisition of rule-like behaviour relating to past tense morphology.

U-shaped development during L1 acquisition of verb morphology can be summarized as follows (Pinker, 1991; Pinker & Prince, 1988; Marcus et al., 1992). Children exhibit a period of correct performance when producing irregular verbs, if they mark the past tense at all. Lasting from the age of two and into early school years, errors are produced in the form of overgeneralizing the regular inflection of the past tense, resulting in the overregularization of irregular verbs that were previously correctly inflected. However, with time, learners revert back and start marking the irregular verbs correctly again. This dip in performance (i.e., U-shape) seems to affect most irregular verbs. The phenomenon co-occurs with the acquisition of the rule of regular past tense inflection of verbs. U-shaped trajectories have been documented in other domains of first language acquisition, such as causatives (Bowerman, 1982) and pronouns (Karmiloff-Smith, 1986). This is consistent with the three-phase model of language development proposed by Karmiloff-Smith (1986), as well as with Tomasello's (1992) verb-island hypothesis, and more generally with usage-based models of language development. In these models, children's early grammars are input-driven (exemplars/item-specific), then representation-driven and top-down, and finally combining both aspects, and adult-like.

U-shaped learning is not only limited to language acquisition. Carlucci & Case (2013) report that U-shaped learning occurs in a variety of domains of child development, such as understanding temperature, weight conservation, object permanence, and facial recognition. In questioning whether U-shaped learning is logically necessary or not in relation to some formal learning tasks, Carlucci & Case (2013: 81) conclude that:

[...] the general picture that emerges from the so-far known results [...] is that U-shaped behavior is unavoidable for full learning power in the context of a number of parametrized models of learning featuring a number of cognitively motivated constraints. The results

might be taken as suggestive of the fact that humans might exhibit U-shaped and other nonmonotonic learning patterns [...].

Marcovitch & Lewkowicz (2004) address the question of whether U-shaped developmental trajectories are more than an interesting artefact of developmental processes. They hypothesise U-shaped curves are in fact a hallmark of such processes. Drawing parallels between language acquisition and ontological adaptations, as proposed by Openheim (1981, cited in Marcovitch & Lewkowicz, 2004), they discuss how this kind of behavioural regression is central to the developmental process itself. Thus, the question arises whether U-shaped learning might indeed be a universal underlying, input-driven learning pattern in human cognitive development.

U-shaped development has played an important part in explaining different models of how the language faculty operates. It has especially been a topic of discussion for single-mechanism models of the mental faculty, such as connectionist models, and dual-mechanism models, such as the dual-mechanism hypothesis (Pinker, 1998).

1.2. U-shaped learning and the dual-mechanism hypothesis

U-shaped learning as a phenomenon has been a point of discussion on the controversial topic of the psychological reality of the organization of language knowledge in the brain, how the mental lexicon is organized and how it operates (cf. Rumelhart & McClelland, 1986; Pinker & Prince, 1988; Plunkett & Marchman, 1991; Marcus et al., 1992; Pinker 1998; Pinker & Ullman, 2002; McClelland & Patterson, 2002). The traditional account distinguishes between rote-memory and a productive system (i.e., two distinct psychological systems) (as described by Pinker & Prince, 1991). In contrast, the connectionist account seeks to make memory more powerful (Pinker, 1998) and to model language based on a neural perspective (Jackendoff, 2002). Connectionist theory maintains that there are in fact no rules and that language is the result of pattern associations and spreading activation throughout networks (cf. Rumelhart & McClelland, 1986; Plunkett & Marchman, 1991; Ellis, 2003). Importantly, concepts are acquired through iterated exposure that results in increase/decrease in connection strengths among units. Patterns of association/connection constitute knowledge representation and changes to these patterns constitute learning. As an alternative to both these theories, and combining aspects of both, the dual-mechanism theory maintains that while language relies on two modules, the mental lexicon, as a form of memory, has associative properties (Pinker, 1991).

Similar to a traditional rote-&-rule theory (e.g. Chomsky & Halle, 1968), the dual-mechanism approach suggests that there are two modules which interact with

each other to form language. First is the memory component which comprises the mental lexicon, and stores words in the form of either stems or the full irregular form. The second module is the rule-capturing component which governs the productive rules of language, and the regularities underlying grammar/language structure. This component computes the regularities on-line, utilizing the stored stems from the mental lexicon. (1) is a proposal of what such a rule could look like in relation to the productive rule of the past tense:

$$(1) \quad V_{\text{past}} \rightsquigarrow V_{\text{stem}} + d$$

(Pinker, 1998: 5)

In the case of an irregular lexical entry, however, two forms would be stored: the base form and the irregular past tense form. Additionally, there would be grammatical tags which capture the function of the word, as shown in (2).

$$(2) \quad V \quad V$$

Bring ~ Brought_{past}

(Pinker, 1998: 6)

When English is taken into consideration, such an organization seems reasonable. Regular verbs are an open-ended class of verbs inflecting within a regular and productive paradigm. Irregular verbs, on the other hand, comprise a relatively small number of verbs (ca. 180), which inflect in idiosyncratic ways (Pinker, 1998: 5). Attempting to capture semi-regularities observed among irregular verbs, Pinker theorizes that associative memory links features to features. Thus, items which share features are assumed to be superimposed in their memory representations, reinforcing shared patterns. New items, similar to already stored items, will activate the shared features and inherit the already existing learned patterns, ultimately leading to some generalizations (Pinker, 1998: 8). Since memory has connectionist properties, similar to those of the Rumelhart & McClelland's (1986) model, it is capable of capturing the semi-regular patterns among irregular verbs. Such properties would account for findings indicating that people have the ability to create irregular-sounding novel words (Pinker, 1998: 6-9).

Assuming a dual organization of the language faculty, the U-shaped developmental curve in first language acquisition can be explained in the following way. To begin with, children have no prior knowledge of the rules of a specific language. Lexical items are memorized as a lexical entry via exposure, including irregular items. Children begin producing frequent irregular forms correctly quite consistently. Through further input, they extrapolate the regular past tense rule, and thus begin marking the regular

past tense inflection. Co-occurring with the acquisition of the rule, overregularization errors start taking place. Words that have weaker memory traces, due to a lower frequency of exposure are more prone to overregularization. In the absence of a strong memory trace or exposure to a word, the regular past tense rule acts as a default, resulting in the overapplication of the rule. With time, and as the memory traces become stronger, these errors even out. When producing a past tense verb, the user first checks the mental lexicon to see if the entry only contains the stem, or if there is additional information stored, such as in the example (2) above. For irregular verbs, the productive rule is blocked when the irregular form is found in the lexical entry, resulting in the successful production of the irregular form. In cases of regular verbs, on the other hand, there are no irregular entries found which can block the rule, and the successful production of a regularly inflected word can thereby take place.

The extent to which a dual processing account is compatible with data from L1 experimental research is open to debate. Thus, Baayen & colleagues (2002) show that inflected regular verbs are both retrieved from the lexicon and computed on-line in parallel. At the same time, the study by Say & Clahsen (2002) provides support for the dual processing account by demonstrating that overregularized irregular verbs elicit stronger early negativity effects than wrongly inflected regular verbs.

1.3. *U-shaped behaviour in SLA*

Extant research on the plausibility of the dual processing account in L2 suggests frequency effects for irregular items, as reflected in faster and more accurate processing and acquisition, thus supporting a dual processing mechanism (Prasada, Pinker & Snyder, 1990; Seidenberg & Bruck, 1990; but see also critical discussion in Ellis, 2002). Further evidence of the plausibility of the dual processing account in the context of the L2 is provided by the study by Pliatsikas & Marinis (2013). They found a clear distinction between processing regular and irregular past tense forms of English verbs in a group of highly proficient L2 speakers. The authors attribute this distinction to the morphological difference between the two verb types and account for the documented delay in regular verb processing in terms of the activation of the regular rule, which is expected to trigger decomposition of the inflected form. Since irregular past tense forms are stored in the mental lexicon as separate entries, no computational processes are required, thus explaining the apparent facilitation in their RTs in that study.

Of specific interest for the current study is the extent to which L2 learners perform similarly on inflection tasks involving pseudo-verbs with similarity to regular and irregular verbs (Prasada & Pinker, 1993). Murphy (2004) used the design from Prasada & Pinker (1993) to compare L2 learners of English with different L1 backgrounds to

adult native speakers and a group of native English-speaking children (mean age = 8,9 years). The two adult groups performed similarly and replicated the original results obtained in the study by Prasada & Pinker (1993), supporting the model of dual mechanism in verb inflection. However, the child group differed significantly from the adult groups. Child participants inflected fewer items and displayed a similarity effect in both items patterned on regular and irregular verbs, reflecting an otherwise attested difference between children and adults in learning style (Johnson & Newport, 1989). Even though the adult results suggest that a dual mechanism might be common in both L1 and L2, Shirai (2019) is sceptical and observes that similar behavioural results (e.g., accuracy) do not necessarily reflect the same underlying cognitive mechanism in both populations.

The acquisition of morphology has been addressed extensively in the domain of second language, acknowledging its crucial role in L2 proficiency and why it may present problems for the L2 learner (DeKeyser, 2005, among others). According to the Bottle Neck hypothesis (Slabakova, 2008; 2013), morphology is one of the most challenging domains in the process of acquiring a second language. While U-shaped trajectories in acquiring a second language have been stipulated in domains where two competing forms render the same content (e.g., preterite forms of the verb, negation and *do*-support, McLaughlin, 1990), experimental research on U-shaped learning in SLA is still limited. Studies which report changes in accuracy trajectories diverge in both methods of data collection and target phenomena, and often report controversial evidence, which makes between study comparison difficult. Furthermore, a variety of terms have been used to define what might be otherwise defined as U-shaped behaviour in SLA ranging from non-linearity, restructuring, to variation and destabilisation. In addition, the reported evidence of what might look like U-shaped curves in learning a second language may have been caused by external factors, such as erroneous forms in the input provided (e.g., Wode, 1978).

Kellerman (1985) was among the first to identify changes in second language learners' production behaviour as U-shaped behaviour, similar to what has been reported in first language acquisition. His study addressed verb alternation patterns in Dutch learners of English of different ages. U-shaped behaviour was observed in the participants' acceptability judgements of transitive-intransitive alternations in English. The adolescent learner group rejected correct target language structures, while both younger and older participants accepted both transitive and intransitive uses. A possible account for this behaviour can be sought in the relatively lower frequency of the intransitive construction in English. With more exposure and resulting restructuring of the L2 knowledge, the more advanced learner group may no longer be influenced by L1 transfer.

The study by Shirai (1990) focused on lexical acquisition and investigated U-shaped learning among Japanese EFL learners of English of three different proficiency groups, along with one group of American native speakers of English. Shirai (1990) claimed U-shaped behaviour in L2 lexical development differs from U-shaped trajectories in L1 or other U-shaped curves in SLA, as it is driven by a growing L1-independence. No clear pattern emerged from the data in that study. Regarding lexical development, Shirai (1990) reported results indicative of U-shapes in certain contexts. Furthermore, he stated that documenting U-shaped behaviour in L2 lexical acquisition is highly item-specific and unpredictable (Shirai, 1990: 6). This author also proposed a three-stage model, where a transitional period from L1-dependency to L1-independency is intermediated by a restructuring stage, causing a U-shaped behaviour in lexical development.

Restructuring of the language system has been proposed to underlie language acquisition specifically in bilingual learners (Grosjean, 1992) and is often used in SLA research as well. Gass & Selinker (2008) hypothesize that restructuring of knowledge is the main cause of the three stages of U-shaped behaviour and learning. Knowledge restructuring can account for L2 morpheme production data, paralleling the three-phase model originally proposed by Karmiloff-Smith (1986) for child first language acquisition. The initial stages are input-driven, while later stages rely on representations and top-down processes. Further support for this idea comes from findings in other domains, such as article use in English as an L2 (Lu, 2001), *-ing* morphology (Lightbown, 1983), plural inflection for nouns (Wode, 1978).

Closer to the purposes of our study, Lightbown (1983) discussed data from French learners of English in a classroom context. She examined the use of the *-ing* progressive form in English at different proficiency levels (i.e., sixth, seventh, and eighth grade learners) and hypothesized that initially these learners were presented only with the progressive form. In stage one, with no other comparable verb form in their system, L2 learners over-extended the use of the progressive into contexts in which the simple present would have been appropriate. In stage two, when the simple present was introduced, learners not only had to learn this new form, but also readjust their information about the present progressive, and its limits. Because of the confusion and subsequent readjustment and restructuring of the progressive, there was a decline in both use and accuracy. Overtime, learners eventually restructured their L2 knowledge appropriately and were able to use both the progressive and the simple present in target-like ways.

More recently, the study by Geeslin & Guijarro-Fuentes (2006) addressed the acquisition of the Spanish copula contrast *ser* and *estar* (both equivalent to English “to be”, however contrasting aspectually) among a native Spanish-speaking group (N=19)

and a native English-speaking group of second language learners of Spanish (N=7). The age of the participants ranged from 20 to 46 for the native Spanish-speaking group, and 20 to 47 for the native English-speaking group. The two groups were presented with multiple choice contextualized preference tasks. The Spanish-group was only tested once while the English-group was tested on four occasions during a 3-year degree program in Spanish. Using the Spanish-group's responses as a standard for comparison, the responses from the English-group were scored accordingly. Geeslin & Guijarro-Fuentes (2006) found patterns of U-shaped learning in the development of the copula choice for the English-group of second language learners of Spanish. Interestingly, a sharper U-shape curve was evident for a sub-group of the native English-speaking participants who studied abroad in Spain for 4 months.

Kounatidis (2016) reports results from a longitudinal study of intermediate proficiency level students of English. The design elicited the production of present tense forms of copular *be*, preterite forms of verbs and negation tested at 3 time points. The results indicated a U-shaped curve for copular *be*, production of past tense forms of familiar regular verbs, and an increasing slope for both familiar and novel irregular forms over the period studied. While accuracy for the familiar regular forms was close to ceiling at all 3 stages, the production of novel regular verbs declined consistently. This latter result, however, combined with the increase in accuracy rate in irregulars, is not consistent with the classical evidence in support of U-shaped trajectories in past tense acquisition by children in the L1.

Based on a large dataset, Casani (2020a; 2020b) provides relevant evidence of the cross-linguistic universality of U-shaped trends in the acquisition of morphology. In a detailed survey of grammar errors in a learner corpus of Italian as L2, Casani documents an inverted U-shape specifically in the acquisition of verbal (inflections) and nominal morphology (agreement, the article) across four levels of proficiency.

One of the biggest differences between first and second language acquisition is the existence of knowledge of an additional language (the L1) in an L2 learning context. During first language acquisition, one does not have any prior knowledge of any specific language, whereas the opposite is true for L2 acquisition. If we observe a curvature in accuracy and overregularization rates similar to the classic U-shaped curve, this suggests the presence of a more generic underlying learning pattern also in L2.

1.4. The current study

On the backdrop of research presented above, we ask the following research questions:

1. Do Norwegian L2 learners of English exhibit a similar learning curve in relation to the acquisition of irregular verb morphology as (child) L1 learners of English?
2. Can this evidence suggest a common learning pattern for L1 and L2 acquisition?

We expected to find significant differences in accuracy and overregularization rates between three groups of L2 learners at different levels of proficiency. The *U* in the term *U-shaped development* models the dip in accuracy. However, this learning pattern involves two related processes: a dip in accuracy followed by subsequent rise in accuracy accompanied by an increase in rate of overregularization and a subsequent decrease of overregularization. Thus, U-shaped development can be documented by following either the accuracy curve or the overregularization one. Although the latter represents an inverted U-shape, it is an inherent part of the phenomenon referred to as U-shaped development. Observing the entire U-shaped curve in the current study would therefore involve an increase in the rate of overregularization errors for the middle proficiency group, in comparison to the lower and higher proficiency groups. In addition, the overall curvature shape can be broken down into smaller segments reflecting different stages of the process, like observed in L1 grammar inflection learning. Furthermore, accuracy rates and overregularization rates are linked together, and any results should reflect both aspects simultaneously.

2. Methodology

2.1. Participants

The participants in this study were students at a lower-secondary school in Norway. They were recruited from three different grades: 8th grade (N=20), 9th grade (N=24), and 10th grade (N=23), respectively. English skills in Norwegian children and adolescents follow a similar trajectory, and seem to accelerate around lower secondary school (Dahl & Vulchanova, 2014). Therefore, we used school grade as a proxy for English proficiency level. However, in line with the Norwegian school system in general, and, more importantly, in accordance with the European Framework for Languages (<https://www.coe.int/en/web/portfolio/home>), the learner groups were further classified as follows: 20 participants in grade 8 (B1 level of general proficiency in English), 24 participants in grade 9 (B2 level) and 23 participants in grade 10 (C1 level).

Because irregular past tense morphology is learned by rote and using memorization sheets at school (Dahl, 2014), and vocabulary size increases steadily, but slowly in

Norwegian school children (Sivertzen, 2013), we estimated that grades 8-10 would be optimal to test for U-shaped learning curves. Ages 10-15 years mark an increase in metalinguistic awareness, as documented by Evensen (2014). Thus, an additional rationale for using these age groups was the development of L2 metalinguistic skills, which may be necessary for performance on the task. To confirm these assumptions, we interviewed the school-teachers of the participating groups. Several of the teachers reported a decline in overregularization errors during the last year of lower secondary school, consistent with our participant selection. In Norway, grade 8-10 attend the same school. Prior to this, learners receive mostly explicit instruction and memorize irregular past tense verbs through memorization sheets. Later through grades 8-10, however, most instruction relating to irregular verbs is in the form of corrective feedback and exposure to target language texts. L2 learning becomes increasingly driven by exposure to authentic input both at school and outside of school via extensive media use (Brevik & Rindal, 2020).

Participants with known learning deficits or other relevant developmental deficits, such as dyslexia, were excluded. Other exclusionary criteria were balanced bilingualism, having a different L1 than Norwegian, and poor performance on the experimental task (i.e., below 48% accuracy for overall verb production). The final participant count was as follows: 8th grade at B1 level (N=17), 9th grade at B2 level (N=19), and 10th grade at C1 level (N=15).

Participants signed a written informed consent to participate in the study. For participants in the 8th and 9th grade who were younger than 16, consent had to be given by their parents. A questionnaire related to the participant's language background, and potential presence of learning deficits, was also administered to parents and participants, respectively.

Students in Norway begin learning English formally during their first year of attending school (at age 6 years old). After completing the 7th grade and upon entering the lower-secondary level, they will have received 366 hours of education in English, whereas during their time at lower-secondary school, they will receive an additional 228 hours (Utdanningsdirektoratet, 2013). Normally, students in the 8th grade are between 13 and 14 years old, 14 and 15 years old in the 9th grade, and 15 and 16 years old in the 10th grade.

After the research project was approved by both the school and the Norwegian Centre for Research Data (henceforth NSD), a select group of volunteer teachers helped grant access to their classes.

2.2. Stimuli

Participants were given two types of elicitation tasks modelled after Berko's (1958) Wug-test. One pertained to verb inflection knowledge, whereas the other one was related to nominal inflection knowledge (not reported here). To create the items for the verb inflection knowledge task, we began by making an overview of the learning material used by the school for teaching English. We identified three books which the school used primarily for each grade. First, we compiled a list of the irregular verbs found across all books' glossaries. In total, 93 irregular verbs were found. Examples such as "to set", "to come", and "to wake" illustrate a few of the irregular verbs that we identified.

Two conditions based on frequency were created, as previous studies indicate frequency of exposure to be an important factor in relation to U-shaped learning and the production of overregularization errors (Marcus et al., 1992). The two conditions were labelled "easy condition" corresponding to high frequency items, and "hard condition" corresponding to low frequency items, respectively. We checked the frequency ratings for all the verbs in BYU's *iWeb: The intelligent Web-based Corpus*. Because children and adolescents are frequently exposed to the internet (Vulchanova et al., 2017), we chose this web-based corpus based on the assumption that the verb frequencies in the corpus will approximate the frequencies they are exposed to in their daily lives. An open search for verbs, grouped by lemmas, resulted in a list of the top 1000 most frequently occurring verbs in the corpus. Of the 93 irregular verbs found across the books, 20 with a frequency ranking smaller than 50 out of the 1000 hits (RAW FREQ greater than 6056787) were chosen as items for the easy condition (e.g., "to run", "to know", and "to take"). Next, 20 of the irregular verbs with a frequency ranking greater than 300 of the 1000 hits (RAW FREQ smaller than 1013581) were chosen as items for the hard condition (e.g., "to sleep", "to freeze", and "to bite"). The selection process resulted in an elicitation task containing a total of 40 items with irregular inflection. Item presentation followed the same order for all participants. Out of the 40 verbs, 17 were cognate Germanic verbs (e.g., swim, come), and 12 out of these were strong verbs in both the L1 and L2, with 7 out of the 12 inflecting in a similar way. Thus, 7 items out of 40 were cognate, strong and inflected similarly for the preterite (e.g., sing/*synge*; sang/*sang*). Observe, however, that in most of these cases these forms are not equivalent phonologically.

The following example in (3) illustrates a stimulus set up:

(3) STEAL: Paul likes to steal things.

However, during the party, Paul _____ Lisa's heart.

The target word was specified in its non-finite form for the verbs (spelled out in capital letters to highlight that this is the target word) and was also used in the present tense in the introductory sentence. In the second sentence, a blank was left where the participant was asked to fill in the irregular past tense form of the word according to context, which in the above example would have been “stole”. The Wug-test has been documented to reliably elicit target item production in both children (Berko, 1958) and adults (Vulchanova et al., 2012) in various experimental set ups.

2.3. Procedure

Prior to any testing, and before the participants received any information, consent sheets were given to both parents and participating students. Ahead of each testing session, a list of participants was created from the consent sheets collected by the teachers involved.

Two separate online response forms were created to accommodate participants who either needed parental consent or who were able to give consent themselves using *SurveyGizmo* as platform. We also collected reaction times to each question. These measurements provided information on the amount of time each participant spent on answering each item of the elicitation task.

Each testing session took about 60 minutes and was held at the school during teaching hours. Some sessions took place in a mixed classroom environment with other non-participating students present, while others were held in a room separate from the non-participating students. During mixed classroom sessions, non-participants were given individual work by the teacher and encouraged to keep quiet, so as to not disturb the participating students. All testing sessions were administered by the first author of this paper.

2.4. Analysis

For the analysis, grade was used as the independent variable, whereas accuracy and overregularization were used as dependent variables. All responses to items in the elicitation task were first compiled according to three categories: accurate responses, overregularization errors, and other types of errors. Percentages were calculated for each variable. For each variable, the mean, standard deviation, and 95% confidence interval were calculated for each group/grade.

The results focus on the relation between grade and rate of accuracy, as well as grade and rate of overregularization errors. A Shapiro-Wilk test revealed that the data

did not follow a normal distribution. Consequently, the Wilcoxon test was adopted, as it does not assume normality. This test was chosen to assess the degree of overlap, or lack thereof, between the distributions to see if they were significantly different or not.

We additionally performed a Pearson-correlation test between the participant's self-rated level of proficiency and accuracy, in addition to overregularization rates, to see whether these variables were related or not.

3. Results

3.1. Results by group and category

Table 1 summarizes the descriptive statistics for each group. A Shapiro-Wilk test for normality was also applied to each variable. Although some of the variables did distribute normally, most did not. Consequently, interpreting the confidence intervals must be met with caution.

Table 1: Descriptive statistics from analysis based on grade

		Mean	SD	Confidence interval 95%		Shapiro-Wilk test
				Min	Max	
Grade:						
Overall accuracy	8 th	75.0	12.0	68.8	81.2	$p = 0.786$
	9 th	77.2	11.9	71.5	83.9	$p = 0.276$
	10 th	86.8	8.4	82.2	91.4	$p = 0.01$
Accuracy, easy condition	8 th	86.5	7.9	82.4	90.5	$p = 0.702$
	9 th	88.4	10.9	83.2	93.7	$p = 0.001$
	10 th	93.0	6.2	89.6	96.4	$p = 0.003$
Accuracy, hard condition	8 th	62.9	17.9	53.8	72.1	$p = 0.411$
	9 th	65.5	15.8	57.9	73.1	$p = 0.393$
	10 th	80.3	12.7	73.3	87.4	$p = 0.096$
Overall overregularization	8 th	7.7	8.1	3.5	11.9	$p = 0.001$
	9 th	9.7	7.1	6.2	13.1	$p = 0.009$
	10 th	5.8	4.5	3.3	8.3	$p = 0.006$
Overregularization, easy cond.	8 th	2.9	3.1	1.4	4.5	$p = 0.0004$
	9 th	3.2	4.2	1.2	5.2	$p = 8.1133e-05$
	10 th	1.0	2.1	-0.1	2.1	$p = 3.481e-06$
Overregularization, hard cond.	8 th	12.1	14.3	4.7	19.4	$p = 0.001$
	9 th	15.8	12.4	9.8	21.8	$p = 0.061$
	10 th	10.0	8.2	5.4	14.6	$p = 0.005$

Note: Numbers are presented in percentages converted from fractions.

The mean for overall accuracy rate between the three grades increases from 75% to 96.8%. In contrast, the rate of overregularization increases between grade 8 and 9, and then drops from 9th to 10th grade.

As seen below, Figure 1 and Figure 2 provide a visualization of the data in boxplot format, while Figures 3 and 4 provide a fitted line plot for participants' accuracy and overregularization rates, necessary to assess the learning curves in our data. A classical U-shaped scenario would see a decline in performance in one domain (i.e., irregular past tense morphology), despite overall increase in general proficiency of the participants. Thus, one would first observe error-free performance, followed by performance with errors, which ends up with error-free performance again, consistent with overall assumptions in the literature (e.g., Kellerman, 1985). Our data, however, show similar accuracy rates for the two lower-level proficiency groups, and an increase in accuracy for the highest level proficiency group (Figure 3). As seen in Figure 1 and Figure 2, the 10th grade group performed at ceiling for frequent irregular verbs (easy condition), but showed variation for the less frequent category (hard condition). However, the overregularization errors display an inverted U-shaped trajectory (Figure 4).

Figure 1. Accuracy scores for easy (turquoise) vs hard (red) condition for verbs, by grade

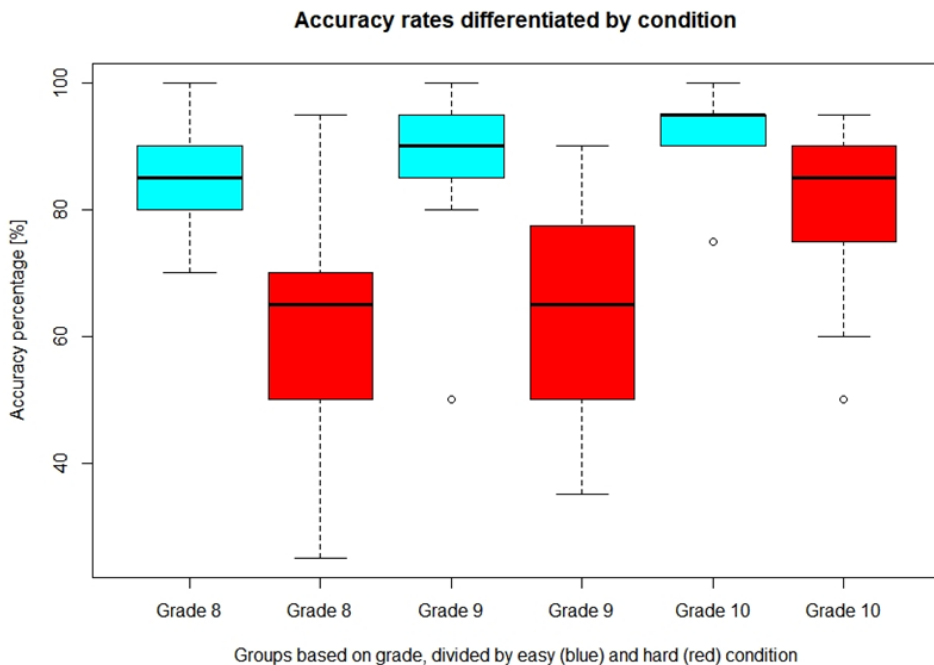


Figure 2. Comparison between easy condition (turquoise) and hard condition (red) for rate of overregularization of verbs, by grade

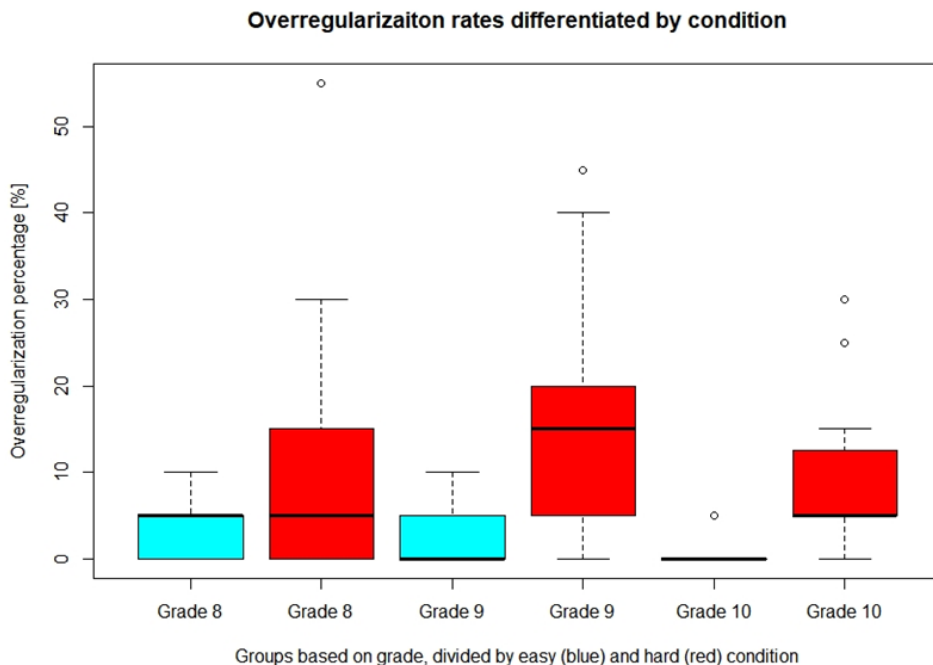


Figure 3. Fitted line plot for overall accuracy scores between grades

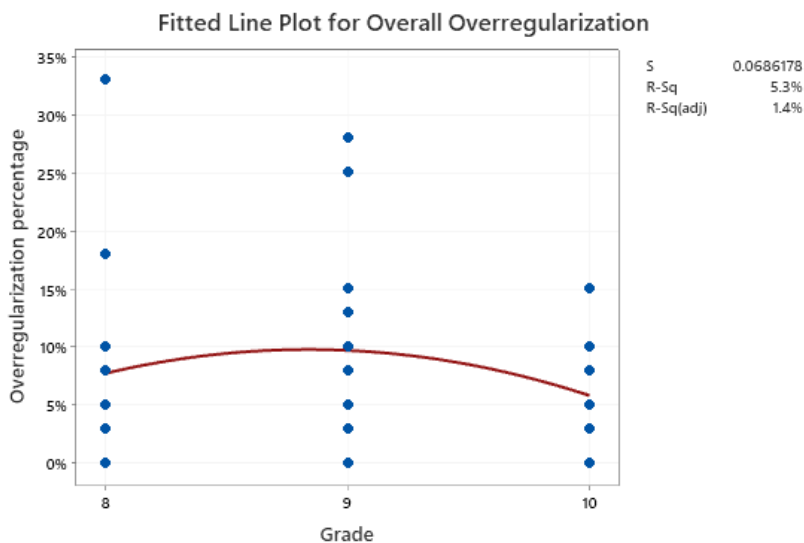
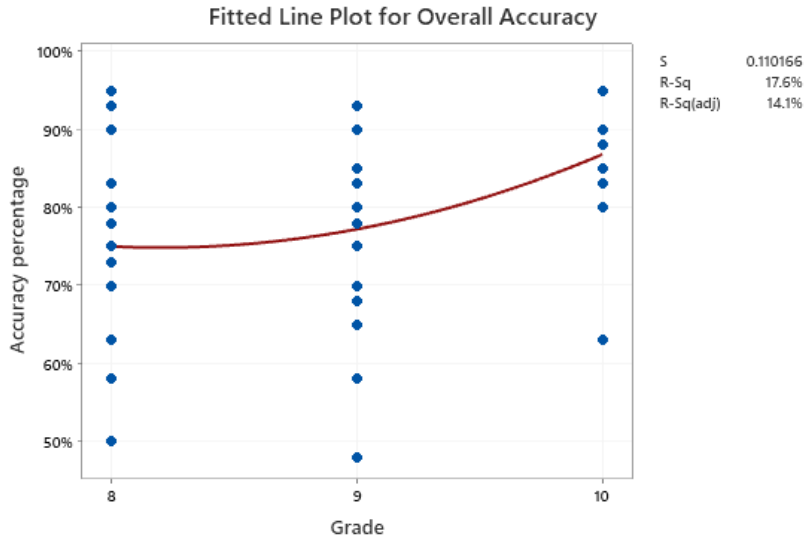


Figure 4. Fitted line plot for overall overregularization rate between grades



There were significant differences between the two conditions within two of the groups (8th and 10th grade), but not for the middle group (9th grade). This was confirmed by a Wilcoxon test which gives an indication of whether two distributions (X and Y) overlap or not, and if they differ significantly, thus testing the null hypothesis for each group compared to each other group. For exact *p*-values, see Table 2. A post hoc Bonferroni correction ($\alpha = 0.016$) for multiple comparisons confirmed the significance effect.

Table 2. Within-group differences between easy and hard condition

Variable tested:	Grade:	W:	<i>p</i> -value:
Easy & Hard condition	8th grade	255	0.0001386
	9th grade	326.5	1.849e-05
	10th grade	190.5	0.001033

Accuracy on the frequent group of verbs (easy condition) lies above 85% for all participant groups. The less frequent verbs (hard condition) elicit poorer responses at 62% (8th grade) and 65% (9th grade), with a steep rise to 80% only for grade 10. In

addition, there is much less variation in the easy than the hard condition within each group, as indicated by the interquartile range, the median and the tails in the boxplots. Performance appears to increase in the hard condition across the three groups, but mostly in-between grades 9 and 10. The same is true of the variation in accuracy, with greater variation displayed by the two lower proficiency groups (8th and 9th grade), and much less variation in the highest proficiency group (10th grade). Although the rate of overregularization errors seems stable between 8th and 9th grade within the easy condition of verbs, the medians differ, and this is confirmed by the results from the Shapiro-Wilk test (Figure 2). Furthermore, the 10th grade group exhibits a floor effect for this kind of error for highly frequent verbs. Figure 2 also indicates that the rate of overregularization within the easy condition for the 10th grade is tightly distributed (with almost no variation). Altogether, every grade produced more overregularization errors for less frequent verbs (hard condition) in comparison to frequent ones (easy condition). We also see that the rate of overregularization errors in the hard condition rises in the 9th grade to drop again in the 10th grade, thus forming an inverted U-shape, evident in both Figure 2, and in the curve plot in Figure 4. The fitted line plots in Figure 3 and Figure 4 display opposite trends: an increasing slope in Figure 3 (rise in accuracy) and an inverted U-shape in Figure 4 (drop in overregularization).

Since some of the data did not distribute normally, the non-parametric unpaired version of the Wilcoxon test, which does not assume a normal distribution, was applied in R to see whether the distributions overlapped or not. Table 3 provides the details with group comparisons.

Table 3. Wilcoxon tests

Variable tested:	Grades being compared:	W:	<i>p</i> -value:
Overall accuracy	8th & 9th grade	137	0.445
	9th & 10th grade	66	0.008
	8th & 10th grade	48	0.002
Accuracy, easy	8th & 9th grade	123.5	0.224
	9th & 10th grade	98	0.113
	8th & 10th grade	62	0.011
Accuracy, hard	8th & 9th grade	144.5	0.597
	9th & 10th grade	65	0.007
	8th & 10th grade	53	0.004
Overall Overregularization	8th & 9th grade	127	0.271
	9th & 10th grade	195	0.066
	8th & 10th grade	141	0.614
Overregularization, easy	8th & 9th grade	165	0.916
	9th & 10th grade	180	0.119
	8th & 10th grade	171	0.054
Overregularization, hard	8th & 9th grade	120	0.187
	9th & 10th grade	181.5	0.172
	8th & 10th grade	124	0.907

After running a post hoc Bonferroni correction for multiple comparisons ($\alpha = 0.0083$) for overall accuracy as well as accuracy in the easy and hard condition, the distributions for 8th and 9th grade did not differ significantly. The distributions, when comparing 9th and 10th grade for accuracy in the easy condition, did not differ either. However, significant differences between the 9th and 10th grade were found in both the

overall accuracy in verb production as well as the accuracy rates in the hard condition. These differences were also observed between the 8th and 10th grade performances. For the accuracy on the easy condition, the only significant difference was between the 8th and 10th grade. In sum, the 8th and 9th grade do not differ in accuracy. The 9th and 10th grade differ on overall accuracy and on less frequent (hard) verbs, whereas the 8th and 10th grade differ systematically. Taken together, these results indicate an increase in accuracy, primarily between the two younger (lower proficiency) groups and the older, highest proficiency group.

As for overregularization errors, the distributions overlap and did not differ significantly for the 8th and 9th grade ($p = 0.271$) on overall overregularization rates, and for the 8th and 10th grade ($p = 0.614$), which is to be expected from a U-shaped learning pattern. A trend to significance in overall overregularization errors can be observed when comparing the 9th and 10th grade ($p = 0.066$). A similar trend to significance can be observed for overregularization rates in the easy condition, for the 8th and 10th grade ($p = 0.054$). Note that none of the tests could yield exact p -values, due to ties in the data, which means that all the p -values are approximate computations.

Reaction times were recorded for all responses. Table 4 provides the descriptive statistics for the reaction times on accurate and overregularized responses, whereas Figure 5 and Figure 6 visualize the data.

Table 4. Descriptive statistics for reaction times

	Grade:	Mean	SD	Confidence interval 95%	
				Min	Max
Reaction times for accurate Responses					
	8 th	20.6	7.1	16.9	24.2
	9 th	16.2	7.7	12.5	19.9
	10 th	13.0	4.2	10.6	15.3
Reaction times for overregularized responses					
	8 th	33.2	25.0	20.3	46.0
	9 th	18.7	10.4	13.7	23.7
	10 th	15.7	7.1	11.8	19.6

Note: Time is presented in seconds.

Figure 5. Average time spent on accurate answers with standard deviation error bars. Average time spent on accurate verb responses

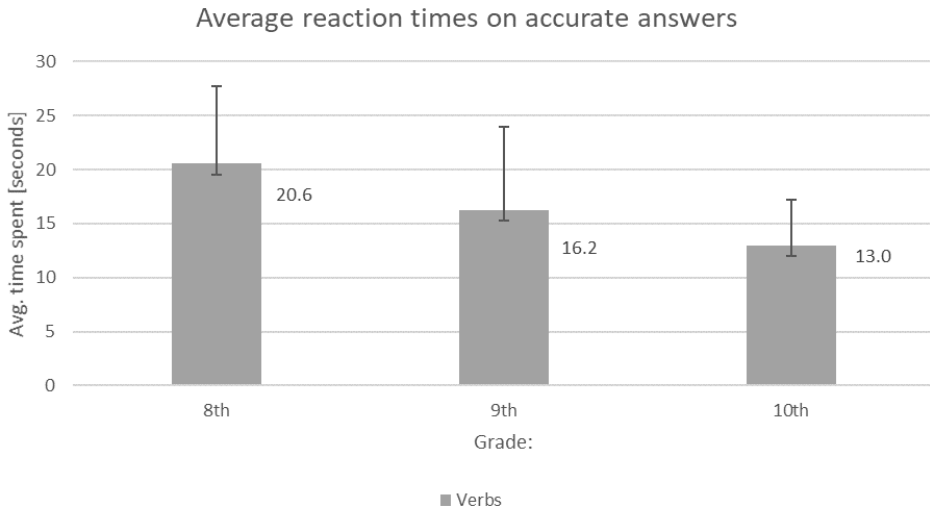
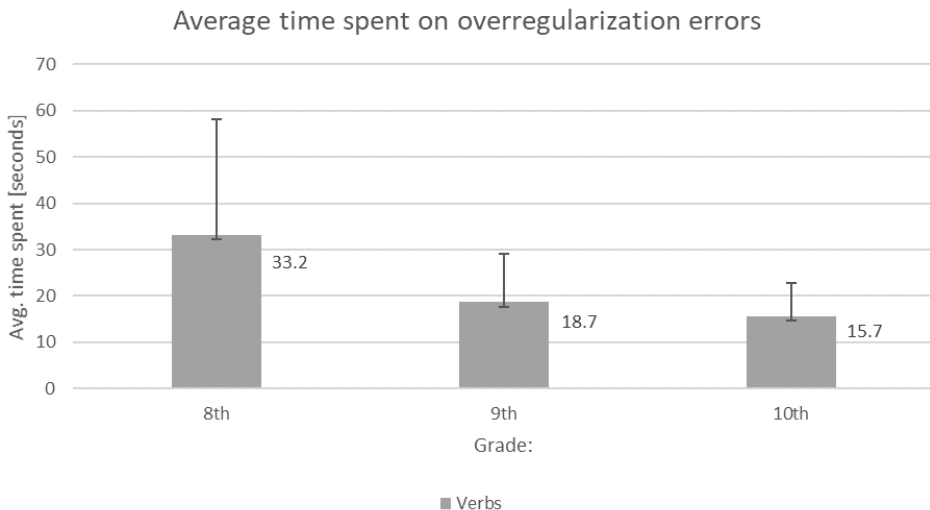


Figure 6. Average time spent on overregularization errors, including error bars for standard deviation. Average time on overregularization errors for verbs



As Figure 5 illustrates, reaction times on accurate responses decrease with grade, suggesting an increase in proficiency and faster lexical retrieval of correctly produced verbs. A similar trend is observable for overregularization errors (Figure 6). This

latter trend is relevant for the current hypothesis. The decrease in time spent on overregularization suggests a process of restructuring the target language knowledge as a result of acquiring the rule. As shown by the standard deviation error bars, the variability also seems to decrease with grade for time spent on verb production in both figures.

Table 5. Pearson-correlation between participant’s self-rated level of proficiency and rates of accuracy and overregularization

Variable:	Grade:	Correlation:	<i>p</i> -value:
Overall accuracy	8 th	0.2761288	0.283
	9 th	0.3016911	0.209
	10 th	0.3591946	0.188
Accuracy, easy cond.	8 th	0.4424952	0.075
	9 th	0.3012573	0.210
	10 th	0.372678	0.171
Accuracy, hard cond.	8 th	0.1908335	0.463
	9 th	0.2343126	0.334
	10 th	0.3027247	0.272
Overall overregularization	8 th	0.001714255	0.994
	9 th	-0.2487454	0.304
	10 th	-0.4861471	0.066
Overregularization, easy cond.	8 th	-0.2362087	0.361
	9 th	-0.3860624	0.102
	10 th	-0.559017	0.030
Overregularization, hard cond.	8 th	0.05121768	0.845
	9 th	-0.1553049	0.525
	10 th	-0.4214636	0.117

An additional Pearson correlation test was performed to see whether the participants' self-rated level of proficiency correlated with accuracy rates, as well as the rate of overregularization errors. A Bonferroni was calculated ($\alpha = 0.0083$) to adjust for multiple comparisons. No significant correlations were found between the participants' self-rated level of proficiency and accuracy rates. As for the overregularization errors, no correlation was found between the participants' self-rated level of proficiency and the overall rate of overregularization errors of verbs (shown in Table 5). However, we did find a significant ($p = 0.03028$) negative correlation ($r = -0.559017$) between the participant's self-rated level of proficiency and rate of overregularization within the easy condition.

In the next section we report the individual results for each of the groups.

3.2. Individual results

Table 6 summarizes individual performance and number of participants at or below chance level (.50), and number who performs above (for both conditions). Further, the table provides details the number of participants who performed above and under the group's average performance.

Table 6. Individual results

		8th grade		9th grade		10th grade	
		Easy cond.	Hard cond.	Easy cond.	Hard cond.	Easy cond.	Hard cond.
Chance level conditions	50% both						
	Above:	17	12	18	13	15	14
	At chance:	0	4	1	4	0	1
	Under:	0	1	0	2	0	0
	Total:	17	17	19	19	15	15
No. of participants							
Under/above mean							
	Above mean:	8	10	13	8	9	8
	Under mean:	9	7	6	11	6	7
	Total:	17	17	19	19	15	15

The results in Table 6 provide evidence of overall good individual performance in all groups with very few participants performing at or below chance (observed only in the hard condition). We also see an increase in performance accuracy, with fewer individuals performing below the mean in both conditions with grade. These results thus complement the individual variation picture presented in the box-plot analyses above.

4. Discussion

Prior to testing, we expected to find significant differences in the levels of accuracy and overregularization errors when comparing the three learner groups in this study, in addition to possibly finding ceiling effects in the most proficient group (i.e., 10th grade at C1 level). We assumed evidence for U-shaped learning would support an underlying input-driven learning pattern shared between L1 and L2 acquisition. We hypothesized a U-shaped trend would be evidenced by an uneven (i.e., curved) rate in performance, with a dip in accuracy and increase in overgeneralization rates in the middle proficiency group, as opposed to a linear increase/decrease trend. With regards to accuracy, we observed a clear increase in performance between the 8th and

the 10th grade, suggesting that knowledge of irregular inflections increases with time, and between the least (B1) proficient and the most (C1) proficient group. Our results are thus consistent with the data in Kounatidis (2016) on past tense forms of irregular verbs.

A significant point of interest in this study was the trend observed at the mid-point (9th grade at B2 level). Indeed, a difference in overall accuracy was observed between the 9th grade and the 10th grade, driven primarily by performance on less frequent (hard) verbs. Concerning overregularization, we observed trends toward significance. In particular, this was found in overall performance between the 9th and the 10th grade, as well as between the 8th and the 10th grade on the frequent (easy) verb condition. These differences suggest the 9th grade was a mid-point in the study (\approx middle proficiency), with a clear difference between that point (level) and the most advanced one. We interpret these findings to be indicative of part of the U-shaped curve, modelling the later stages of the phenomenon, thus providing a parallel to behaviour observed in L1 acquisition (Marcus et al., 1992).

4.1. Evidence for U-shaped learning in the L2

The first question in this study sought to determine whether Norwegian L2 learners of English exhibit U-shaped learning similar to L1 learners of English when acquiring irregular aspects of verb morphology in English. We interpret our results as suggestive evidence in favour of a U-shaped trend. This interpretation is limited to a small interval from the later stages of U-shaped learning during L2 acquisition among Norwegian learners of English. During the later stages of U-shaped learning, the dual-mechanism hypothesis predicts that during U-shaped development, the level of accuracy should steadily increase, while the rate of overregularization errors decrease. Our findings model this trend.

Firstly, overall accuracy rates improve with each grade, rising from 75% correct responses to 86.8% on average. This is confirmed by the systematic difference in accuracy between the 8th and the 10th grade (see Table 1). In addition, the variation in accuracy decreases with grade and proficiency level. This decrease in variability, in conjunction with the increase in performance accuracy, is indicative of an overall increase in proficiency among participants across grades. However, performance does not seem to increase in a linear fashion with grade. Instead, the middle point (i.e., the 9th grade), evidences a rise in overregularization errors combined with no significant difference between accuracy on frequent (easy) verbs and less frequent (hard) verbs (Table 2). This suggests an input-driven restructuring of grammar competence in this group, as reflected in the trend to significance on accuracy on less frequent verbs and

overall overregularization errors between the 9th grade and the 10th grade (Table 3). Importantly, our findings not only document that L2 learners overregularize in the process of the acquisition of L2 morphology, but also demonstrate that this happens at a specific time-point in the course of learning. This shows the process is not linear, thus suggesting an inverted U-shape curve.

These results seem to better reflect the participants' development regarding their proficiency level than their self-reported level of proficiency. A Pearson correlation-test between participants' self-rated level of proficiency and the rate of overregularization errors did not reveal any relationship (Table 5). If anything, the correlation between these two factors seemed to decrease with grade, replicating well-known trivial effects between level of performance and self-assessment (Kruger & Dunning, 1999; but see LeBranç & Painchaud, 1985 for evidence to the contrary). Furthermore, participants' self-rated level of proficiency did not seem to be a reliable measurement in our study. When asked to rate their overall level of proficiency on a Likert-scale ranging from 1-5, participants did not use the whole range, but instead only reported values between 3-5, compromising the correlation analysis results.

Secondly, participants were sensitive to frequency effects. There was a clear frequency effect observed in performance on highly frequent items (easy condition) and the low frequency ones (hard condition), consistent with the results in Murphy (2004). Not only was this effect visible for accuracy (Figure 1), but also for rates of overregularization errors (Figure 2). This observation supports Pinker's (1998) idea of the links between frequency of exposure and memory traces and their role in the lexical retrieval of words. A higher frequency of exposure leads to stronger memory traces. This means that they are less prone to overregularization errors and more likely to be correctly retrieved from the mental lexicon as a first step in the process of producing the target inflection.

Thirdly, there were some observable curves between the grades in overregularization rates in overall verb production (Figure 4), and some interesting differences became evident when differentiating between conditions. In the easy condition, a floor effect was observed for the 10th grade (Figure 2) suggesting a dramatic drop in frequent verb overregularization in the most advanced participant group. In addition, when comparing median values between grades in the hard condition, we saw a rise and drop in overregularization rates resembling an inverted U-shape. This was not confirmed, however, by the Wilcoxon test results.

Finally, reaction times for accurate answers and overregularized items decreased with grade. Not only does the average time spent on these types of responses decrease with grade, but also the variability (see SD and mean values in Table 4). In addition, participants across all grade levels spent more time on overregularized responses than

on accurate responses. This is yet further evidence suggesting an overall increase in proficiency from grade to grade and is consistent with the findings in Pliatsikas & Marinis (2013) of increased reaction times for regular verb forms as a result of an underlying process of decomposition.

We have reason to believe that our data speak in favour of input-driven U-shaped learning. The first indication is connected to the frequency effects observed in all results. Although our participants are L2 learners of English, their responses parallel the frequency effects expected of the forms in the target language English. Thus, the results indicate sensitivity to the target language and to English item frequencies. It should be noted here that the stimuli verbs were neither cognates of Norwegian verbs, nor corresponded in paradigm regularity between the two languages (e.g., the English irregular may correspond to a Norwegian regular verb: “to make” – *å lage*, and “to think” – *å tenke*). This comes to suggest that the results reported here reflect acquiring English as a (target) language in the later stages of the U-shaped curve. As grade and proficiency increase, the variability in all measurements decreases and the participants begin to produce language more uniformly. This parallels the development among L1 speakers from adolescence to adulthood. Although the two languages are typologically similar in their division between a productive past tense rule and sub-regular patterns among irregular verbs, the items used in this study did not have a one-to-one correspondence on inflection paradigm type between the two languages, with only 7 items inflecting similarly in both languages. This similarity, however, did not prevent participants from over-regularizing, as seen in the non-target form “frozed”. The current findings can be taken as support for an overall learning trend rather than cross-linguistic influence (Casani, 2020a, 2020b). Since Casani’s (2020a, 2020b) studies also document a similar trend for basic syntax (word order and negation), as well as for prepositions and conjunctions, the question arises whether U-shaped trajectories can be more universal than originally thought and can be documented across the board for second language acquisition and across languages. Such evidence will necessarily also call for a revision of our theory of the nature of U-shaped trajectories in language acquisition.

5. Conclusion

This study set out to explore whether Norwegian L2 learners of English exhibit a similar learning curve in the acquisition of irregular and regular aspects of verb morphology as L1 learners of English do. Our evidence suggests that our group of Norwegian L2 learners of English do in fact display later stages of U-shaped learning behaviour in the acquisition of verbs. Accuracy rates increased with grade, whereas overregularization rates decreased. Furthermore, 10th grade exhibited a floor effect for the overregularization rate of highly frequent verbs. A Wilcoxon test found a trend to significance in rate of overregularization of highly frequent verb items between 8th

and 10th grade. This result can likely be ascribed to the small sample size, and we speculate that we could have found a clearer result and more power from a bigger sample. Our results speak in favour of the U-shaped trajectory as an underlying general pattern characteristic of language learning and related to the operation of storage (memory) mechanisms and the gradual acquisition of rules and pattern extraction to construction, as a result to exposure to input, and consistent with usage-based frameworks of language acquisition (Ellis, 2002).

Although we have suggestive evidence for an interval of a U-shaped curve, this interpretation must be met with caution. Our limited data (i.e., sample size, age groups) cannot confirm nor disprove the existence of classical U-shaped learning in the acquisition of L2 morphology. Consequently, this reduced our choice of statistical tests for our analysis. Another limiting aspect to our study is the amount of variability in many of the measurements and variables within each group. Separating and organizing the participants according to levels of proficiency may have yielded an even clearer picture of the developmental curvature that we are interested in. However, grouping participants according to proficiency level is not unproblematic either, since we lose the benefit of variability in the data. A final limitation is that we did not control for variation between British and American Englishes. Due to the diverse input which Norwegian learners of English receive and the resources we had, it was not feasible to control for this.

However, further studies are indeed necessary to confirm this. In particular, within-participant longitudinal designs, similar to Geeslin & Guijarro-Fuentes (2006), would be highly relevant to address this phenomenon in further detail. The current study should also be replicated including data from earlier grades at lower proficiency level to see if the entire U-shaped curve can be observed. In addition, testing English L2 learners, but with a typologically dissimilar L1, such as Mandarin Chinese, and analysing data based on L2 learner corpora from other second languages of the type reported in Casani (2020a, 2020b) would be important to establish the universality of the phenomenon. We also believe a study which compares Norwegian L2 learners with child Norwegian L1 speakers would provide important evidence of the nature of U-shaped trajectories and the underlying factors that impact on this pattern.

6. Acknowledgements

The authors would like to thank the participants and their teachers for their time and participation in the test. We would also like to thank Giosuè Baggio for his help with our statistics, Olav Landmark Pedersen for his expertise using R, and Brittany Arnold for her valuable feedback during our final revision.

7. References

Baayen, H., Schreuder, R., De Jong N., & Krott, A. (2002). Dutch Inflection: The Rules that Prove the Exception. In S. Nooteboom, F. Weerman, & F. Wijnen (Eds.). *Storage and Computation in the Language Faculty. Studies in Theoretical Psycholinguistics* (Vol 30, pp. 61-92). Dordrecht: Springer.

Berko, J. (1958). The Child's Learning of English Morphology. *WORD*, 14(2-3), 150-177.

Bowerman, M. (1982). Starting to talk worse: Clues to language acquisition from children's late speech errors. In S. Strauss (Ed.). *U-shaped behavioral growth* (pp.101-145). New York: Academic Press.

Brevik, L. M., & Rindal, U. (2020). *Teaching English in Norwegian classrooms: From research to practice*. Oslo: Universitetsforlaget.

Carlucci, L., & Case, J. (2013). On the Necessity of U-Shaped Learning. *Topics in Cognitive Sciences*, 5, 56-88.

Casani, E. (2020a). Valutare la competenza morfosintattica in italiano L2. Una validazione corpus-based dei livelli del QCER. In E. Nuzzo, E. Santoro, & I. Vedder (Eds.). *Valutazione e misurazione delle produzioni orali e scritte in italiano lingua seconda* (pp.15-26). Florence: Franco Cesati Editore.

Casani, E. (2020b). Corpus d'apprenants et recherche linguistique. Une étude sur l'italien. Actes du Colloque CEDIL'18 «Qu'a changé le numérique? Perspectives linguistiques et didactiques». Université Grenoble Alpes, 30 mai - 1 juin 2018. <https://hal.univ-grenoble-alpes.fr/hal-02641560/document>

Cazden, C. B. (1968). The Acquisition of Noun and Verb Inflections. *Child Development*, 39(2), 433-448.

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper & Row.

Dahl, A. (2014). *Young Second Language Learners: The acquisition of English in Norwegian first-grade classrooms*. PhD thesis. Department of Language and Literature, Norwegian University of Science and Technology.

Dahl, A., & Vulchanova, M. (2014). Naturalistic acquisition in an early language classroom. *Frontiers in Psychology*, 5, 1-9.

DeKeyser, R. M. (2005). What Makes Learning Second-Language Grammar Difficult? A Review of Issues. *Language Learning*, 55, 1-25.

Ellis, N. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143-188.

Ellis, N. C. (2003). Constructions, Chunking, and Connectionism: The Emergence of Second Language Structure. In C. J. Doughty, & M. H. Long (Eds.). *The Handbook of Second Language Acquisition* (pp.33-68). Williston: Blackwell Publishing Ltd.

Ervin, S. M., & Miller, W. R. (1963). Language development. In H. W. Stevenson (Ed.) & J. Kagan, C. Spiker (Collaborators) & N. B. Henry, H. G. Richey (Eds.). *Child psychology: The sixty-second yearbook of the National Society for the Study of Education, Part 1* (pp. 108-143). Chicago: University of Chicago Press.

Evensen, M. W. (2014). *Searching for patterns in English as a second language: a Norwegian study of 10 and 15 year old learners*. Unpublished MA thesis. Department of Language and Literature, Norwegian University of Science and Technology.

Gass, S. M., & Selinker, L. (2008). *Second Language Acquisition. An introductory course* (3rd ed.). New York: Routledge.

Geeslin, K. L., & Guijarro-Fuentes, P. (2006). A Longitudinal Study of Copula Choice: Following Development in Variable Structures. In N. Sagarra, & A. J. Toribio (Eds.). *Selected Proceedings of the 9th Hispanic Linguistic Symposium* (pp. 144-156). Somerville, MA: Cascadilla Proceedings Project.

Grosjean, F. (1992). Another View of Bilingualism. In R. J. Harris (Ed.), *Advances in psychology* (Vol. 83). *Cognitive Processing in Bilinguals* (pp. 51-62). Elsevier Science Publishing Company Inc.

Jackendoff, R. (2002). *Foundations of Language. Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

Johnson, J.S., & Newport, E.L. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60-99.

Karmiloff-Smith, A. (1986). Stage/structure versus phase/process in modeling linguistic and cognitive development. In I. Levin (Ed.), *Stage and Structure: Reopening the Debate* (pp. 164- 190). Norwood, N.J.: Ablex.

Kellerman, E. (1985). If at first you do succeed... In S. M. Gass, & C. G. Madden (Eds.). *Input in Second Language Acquisition* (pp. 345-353). Rowley: Newbury.

Kounatidis, J. (2016). Destabilisation, IL variation and restructuring in foreign language learning. *Yearbook of the German Cognitive Linguistics Association*, 4(1), 261-288.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121-1134.

LeBlanc, R., & Painchaud, G. (1985). Self-Assessment as a Second Language Placement Instrument. *TESOL Quarterly*, 19(4), 673-687.

Lightbown, P. M. (1983). Exploring relationships between developmental and instructional sequences in L2 acquisition. In H. W. Seliger, & M. H. Long (Eds.). *Classroom Oriented Research in Second language Acquisition* (pp. 217-245). Rowley: Newbury.

Lu, C., F.-C. (2001). The acquisition of English articles by Chinese learners. *Second Language Studies*, 20(1), 43-78.

Marcovitch, S., & Lewkowicz, D. J. (2004). U-Shaped Functions: Artifact or Hallmark of Development? *Journal of Cognition and Development*, 5(1), 113-118.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F., & Clahsen, H. (1992). Overregularization in Language Acquisition. *Monographs of the Society for Research in Child Development*, 57(4), i-178.

McClelland, J. L., & Patterson, K. (2002). Rules or connections in past-tense inflections: what does the evidence rule out? *TRENDS in Cognitive Sciences*, 6(11), 465-472.

McLaughlin, B. (1990). Restructuring. *Applied Linguistics*, 11(2), 113-128.

Murphy, V. A. (2004). Dissociable systems in second language inflectional morphology. *Studies in Second Language Acquisition*, 26(3), 433-459.

Pinker, S. (1991). Rules of Language. *Science*, 253, 530-535.

Pinker, S. (1998). Words and Rules. *Lingua*, 106, 219-242.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28, 73-193.

Pinker, S., & Prince, A. (1991). Regular and Irregular Morphology and the Psychological Status of Rules of Grammar. *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistic Society: General Session and Parasession of The Grammar of Event Structure* (pp. 230-251). Berkeley: Berkeley Linguistic Society.

Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *TRENDS in Cognitive Sciences*, 6(11), 456-463.

Pliatsikas, C., & Marinis, T. (2013). Processing empty categories in a second language: When naturalistic exposure fills the (intermediate) gap. *Bilingualism: Language and Cognition*, 16, 167-182.

Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition*, 38, 43-102.

Prasada, S., & Pinker, S. (1993). Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes*, 8, 156.

Prasada, S., Pinker, S., & Snyder, W. (1990). Some evidence that irregular forms are retrieved from memory but regular forms are rule generated. *Bulletin of the Psychonomic Society*, 28(6), 519-519.

Rumelhart, D., & McClelland, J. L. (1986). On learning the past tense of English verbs. Implicit rules or parallel distributed processing? In J. L. McClelland, D. Rumelhart, & the PDP Research Group (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge: MIT Press.

Say, T., & Clahsen, H. (2002) Words, Rules and Stems in the Italian Mental Lexicon. In Nooteboom S., Weerman F., Wijnen F. (Eds.), *Storage and Computation in the Language Faculty. Studies in Theoretical Psycholinguistics* (Vol 30). Dordrecht: Springer.

Seidenberg, M., & Bruck, M. (1990). *Consistency Effects in the Generation of Past tense Morphology*. Paper presented at the thirty-first meeting of the Psychonomic Society, New Orleans.

Shirai, Y. (1990). *U-shaped Behavior in L2 Acquisition*. In H. Burmeister, & P. L. Rounds (Eds.), *variability in Second Language Acquisition: Proceedings of the Tenth Meeting of the Second Language Research Forum* (Vol. 2, pp. 685-700). Eugene: Department of Linguistics, University of Oregon.

Shirai, Y. (2019). *Connectionism and Second Language Acquisition*. New York: Routledge.

Sivertzen, T. G. (2013). *The role of input in early second language acquisition: A study of the long-term effects of initial extra English input on Norwegian 4th graders' receptive vocabulary*. Unpublished MA thesis. Department of Language and Literature, Norwegian University of Science and Technology.

Slabakova, R. (2008). *Meaning in the Second Language*. Berlin: Mouton de Gruyter.

Slabakova, R. (2013). What is easy and what is hard to acquire in a second language: A generative perspective. In: García Mayo, M, Gutiérrez Mangado, & MJ, Martínez-Adrián, M. (Eds.), *Contemporary approaches to second language acquisition* (pp.5-28). Amsterdam/Philadelphia, PA: John Benjamins.

Tomasello, M. (1992). *First Verbs: A Case Study of Early Grammatical Development*. Cambridge: Cambridge University Press.

Utdanningsdirektoratet. (2013). *Læreplan i engelsk (ENG1-03)*. Retrieved September 17, 2019 from <https://www.udir.no/kl06/ENG1-03>.

Vulchanova, M., Foy, C. H., Nilsen, R. A., & Sigmundsson, H. (2014). Links between phonological memory, first language competence and second language competence in 10-year-old children. *Learning and Individual Differences*, 35, 87-95.

Vulchanova, M., Talcott, J. B., Vulchanov, V., Stankova, M., & Eshuis, H. (2012). Morphology in autism spectrum disorders: Local processing bias and language. *Cognitive Neuropsychology*, 29(7-8), 584-600.

Vulchanova, M., Baggio, G., Cangelosi, A., & Smith, L. (2017). Editorial: Language Development in the Digital Age. *Frontiers in Human Neuroscience*, 11, 447. DOI=10.3389/fnhum.2017.00447

Wode, H. (1978). Developmental sequences in naturalistic L2 acquisition. In E. Hatch (Ed.), *Second Language Acquisition* (pp.101-117). Rowley: Newbury.

Appendix

Part 1: Background and participant information

1. Please submit your participant ID for this survey (For instance: ID1498):

Answer: _____

2. When were you born?

Answer: _____

3. Please specify your gender

- Male
- Female

4. What is your native language (morsmål)?

- Norwegian
- English
- Other

Please specify your other native language(s):

Answer: _____

5. What other language(s) do you speak in addition to your native language(s)? Please specify language(s) and competence (high – medium – low) in the textbox below.

If you do not speak any additional languages, please write 'none'.

Examples of how to answer: English (medium) Arabic (high) Mandarin/Chinese (low)

Answer: _____

6. Do you have a diagnosis that could potentially affect your language learning (e.g. dyslexia, impaired hearing, etc.)? If yes, please specify. If no, simply write : 'no'

Answer: _____

7. Have you lived in an English speaking country for more than 2 months? If so, where and for how long? If no, simply write: 'no'.

Answer: _____

8. Do you have close family and/or friends that speak English?

Answer: _____

9. Approximately how many hours do you use English throughout a day? (Including at school, and other activities such as reading English online media etc.)

Answer: _____

10. When do you use English in everyday life?

- While watching TV/movies/series with subtitles
- While watching TV/movies/series without subtitles
- When talking to friends/family
- At school
- When reading/writing English blogs/websites
- While reading English books (not including what you read at school)
- When writing English (not including assignments at school)
- While playing games (PC-games, X-box, Playstation etc.)

11. Are you involved in one or more activities on your free time that involves use of English? (e.g. theatre/role playing, gaming, etc.)

- Yes, please specify below
- No

Please specify which activities: _____

12. How would you rate your own English competence?

- I can understand English text well
- I can understand English speech well
- I can write English well

13. Please rate how well you know English on a scale from 1-5. 1 = Poor, 5 = Very good

	1	2	3	4	5
English:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Part 2: Tasks

Task 1:

SET: Every year, this guy sets out on a new journey.

What did he do last year?

He_____ out on a journey.

Task 2:

RUN: Joey likes to run 10km every day.

What did Joey do yesterday?

He_____10km.

Task 3:

DO: Joey does something stupid every week.

What did Joey most likely do last week?

He_____something stupid.

Task 4:

MAKE: Betty makes good food when Joey comes over for dinner.

Yesterday, Joey came over for dinner. What did Betty do?

She_____good food for Joey.

Task 5:

GET: Every year, Joey gets Betty something nice for her birthday.

This year, Joey_____Betty a nice present.

Task 6:

SAY: Whenever Joey and Betty argue, they say mean things to each other.

Last time they argued, both_____mean things to each other.

Task 7:

GO: Paul likes to go to the arcade to play games.

Paul_____last Friday to the arcade at 12 o'clock.

Task 8:

TAKE: Paul takes Betty's sister Lisa to go for a hike in the weekends.

Last week, Paul_____Lisa for a hike around a lake for a change.

Task 9:

SEE: Paul loves to see movies on his spare time.

Last night, Paul_____Mad Max: Fury Road at home.

Task 10:

KNOW: Since Paul has seen the movie before, he knows what will happen.

One could say that Paul_____what would happen at the end of the movie.

Task 11:

FIND: Betty's friend Anne likes to find nice stones along the beach.

Once, walking along the beach, Anne_____a green stone!

Task 12:

COME: Anne usually comes over to Joey's birthday parties.

This time, Anne_____late to the party.

Task 13:

THINK: Joey thinks that begin late is rude.

During yesterday's party, when Anna was late, Joey_____that Anne was rude.

Task 14:

GIVE: Anna usually does not give presents.

At the party though, Anna_____Joey a gift!

Task 15:

KEEP: Joey like to keep his presents hidden away from people after parties.

After yesterday's party, Joey_____his presents hidden.

Task 16:

SHOW: Betty like to show pictures of Joey at parties.

At yesterday's party, Betty had_____all her pictures of Joey!

Task 17:

FEEL: Usually, Joey feels happy when Betty shows people pictures of him.

However, at the party, Joey_____embarrassed by some of the pictures.

Task 18:

LEAVE: Most of the time, Lisa leaves early when going to parties.

At yesterday's party, Lisa_____early.

Task 19:

TELL: Paul likes to tell jokes at parties.

At yesterday's party, Paul_____a funny joke that made everybody laugh!

Task 20:

PUT: Betty likes to put berries on top of cakes.

Betty_____berries on top of the cake she baked last year.

Task 21:

HIDE: During hide and seek, you should hide from the others who are playing the game.

Lisa_____an hour ago, and the others have still not been able to find her!

Task 22:

SLEEP: Sleeping is important

However, when you have_____for 15 hours, that's too long!

Task 23:

SING: Betty likes to sing.

Betty_____on TV during last year's American Idol!

Task 24:

BLOW: Joey thinks it is hard to blow out all the candles on the cake.

This year though, Joey_____out all the lights on the cake!

Task 25:

STEAL: Paul likes to steal things.

However, during the party, Paul_____Lisa's heart.

Task 26:

HURT: Paul hates to get hurt when playing sport.

Last time Paul played American football, Paul_____his leg.

Task 27:

WAKE: Joey wakes up 7 o'clock every morning.

However, Joey_____up at 1 o'clock last Sunday!

Task 28:

SHUT: When Paul enters his home, he usually shuts the door when he goes inside.

Yesterday, when Paul came home, he_____the door after himself.

Task 29:

SHAKE: Joey likes to shake his milkshakes before he drinks them.

Yesterday, Joey_____his milkshake more than usual!

Task 30:

SLIDE: It's easy to slide on an icy road and fall over.

Last year, Anna_____on the ice and broke her leg!

Task 31:

SWIM: Betty likes to swim in the weekends.

Last weekend, Betty_____5km in a pool!

Task 32:

TEAR: Every Monday, Paul has to tear down all the posters on the poster-wall.

Last Monday, Paul_____down 100 posters!

Task 33:

FREEZE: Brian likes to prepare all his food a week in advance, and freeze it afterwards.

Last night, after preparing his food, Brian_____his food.

Task 34:

SHINE: Brian likes it when the sun shines on him.

Yesterday, the sun_____through the clouds on Brian.

Task 35:

BITE: Brian is afraid of getting bitten by zombies.

During a nightmare, Brian got_____by a zombie!

Task 36:

BET: Brian has a gambling problem and likes to bet way too much money on poker.

Yesterday, Brian_____10 000kr on a poker game and lost!

Task 37:

FLY: In his spare time, Brian likes to fly.

Yesterday at the party, Brian_____through the window.

Task 38:

LEND: People usually lend Brian money after he loses a poker game.

Joey_____Brian 10 000kr after his last poker game.

Task 39:

SWING: Joey likes to play baseball and swing the baseball bat.

Yesterday, Joey_____the bat and hit a homerun!

Task 40:

CREEP: Kevin likes to creep around the forest.

Last year, Kevin_____around the forest and got in trouble afterwards.