Are there core and peripheral syntactic structures?
Experimental evidence from Dutch native speakers with varying literacy levels

Saoradh Favier[1] & Falk Huettig[1,2]

[1]Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

[2]Centre for Language Studies, Radboud University, Nijmegen, Netherlands

***Lingua* (in press)**

Correspondence should be addressed to:

Falk Huettig

Max Planck Institute for Psycholinguistics

P.O. Box 310

6500 AH Nijmegen

The Netherlands

E-mail: falk.huettig@mpi.nl

phone: +31-24-3521374

Abstract

Some theorists posit the existence of a 'core' grammar that virtually all native speakers acquire, and a 'peripheral' grammar that many do not. We investigated the viability of such a categorical distinction in the Dutch language. We first consulted linguists' intuitions as to the 'core' or 'peripheral' status of a wide range of grammatical structures. We then tested a selection of core- and peripheral-rated structures on naïve participants with varying levels of literacy experience, using grammaticality judgment as a proxy for receptive knowledge. Overall, participants demonstrated better knowledge of 'core' structures than 'peripheral' structures, but the considerable variability within these categories was strongly suggestive of a continuum rather than a categorical distinction between them. We also hypothesised that individual differences in the knowledge of core and peripheral structures would reflect participants' literacy experience. This was supported only by a small trend in our data. The results fit best with the notion that more frequent syntactic structures are mastered by more people than infrequent ones and challenge the received sense of a categorical core-periphery distinction.

Key words: core grammar, peripheral structures, grammaticality judgments

Introduction

A strong claim of generative linguistics is that all members of a language community converge on the same internal grammar (Chomsky, 1965; Crain & Lillo-Martin, 1999; Lidz & Williams, 2009; Nowak et al., 2001). Related to this (but not exclusive to the generative tradition) is the notion of ‘core grammar’, an inventory of structures acquired by virtually all native speakers. Properties of the standard language that are absent from this collective body of knowledge are described by some theorists as ‘peripheral’ (e.g., Broekhuis & Keizer, 2012; Broekhuis, 2016). Whereas ‘core’ phenomena are hypothesised to “arise spontaneously in the language-learning child”, those on the ‘periphery’ must be “consciously learned at a later age” (Broekhuis, 2016, p. 298); they are often restricted in register (e.g., formal, written), and may deviate from the rules of the core system (e.g., loan forms). Chomsky (1981; p. 8) for instance argued that core grammar is determined by parameter settings of universal grammar and that there is “a periphery of borrowings, historical residues, inventions, and so on, which we can hardly expect to – and indeed would not want to – incorporate within a principled theory of UG.”

The theoretical core-periphery dichotomy is not without its critics and has been strongly questioned (Culicover, 1999). To our knowledge, however, there have been few attempts to examine ‘core’ and ‘peripheral’ grammatical knowledge empirically in native speaker populations. In a step towards establishing the ‘core’ grammar of Dutch, Hulstijn (2017) collected spoken corpus data showing substantial commonalities in the syntactic patterns produced by a sample of native speakers that was heterogenous in terms of age, education and profession. However, limited conclusions could be drawn about the breadth of participants’ productive knowledge, given the small size of the corpus (Hulstijn, 2017). In the present study we used a different approach to investigate the empirical basis for core and peripheral syntactic structures in Dutch. We first asked Dutch linguists for their intuitions as to the ‘core’ or ‘peripheral’ status of a wide range of grammatical structures in the Dutch language and then asked naïve Dutch participants with varying levels of literacy experience to judge the grammaticality of a selection of core- and peripheral-rated structures, as a proxy for their receptive knowledge.

Also relevant to the question of core and peripheral grammatical knowledge is the growing body of evidence that demonstrates that adult native speakers do not all master the grammar of their language to the same extent. Challenging a core assumption of generative linguistics, considerable individual differences in native syntactic proficiency have been observed across a variety of structures, tasks, and speaker communities (reviewed by Dąbrowska, 2012, and Dąbrowska & Divjak, 2019).

A key determinant of differences in syntactic proficiency appears to be the degree of experience with written language. Montag and MacDonald (2015), for instance, showed that avid readers’ implicit syntactic choices in speech reflected the structural distributions of written language. Wells et al. (2009) found that manipulating written language input to maximise exposure to relative clauses over several weeks boosted processing of the same structure in a subsequent reading task. Dąbrowska (2018) observed a small contribution of print exposure (as measured by Author Recognition; ART) to listeners’ comprehension of basic constructions that occur in everyday spoken language. Langlois and Arnold (2020) reported a positive relationship between print exposure (ART) and the use of syntactic cues to interpret ambiguous pronoun reference. Furthermore, Street and Dąbrowska (2010) observed that auditory comprehension of full passives correlated with self-reported hours of reading in adults matched for educational attainment. Finally, the detection of prescriptive grammatical norm violations in spoken Dutch was robustly associated with literacy experience in a large sample of adult native speakers, even after accounting for general cognitive abilities (Huettig & Favier, under review).

The notion that literacy experience shapes grammatical knowledge is compatible with usage-based models of language processing, in which acquisition is largely determined by the quality and quantity of the input a language user receives (e.g., Abbot-Smith & Tomasello, 2006; Bybee 2006). In terms of input quality, 'book language' is syntactically more complex and diverse than conversational speech (Kroll, 1977; Roland et al., 2007). Furthermore, skilled readers read more and thus encounter a larger volume of language, which they process at a faster rate than listeners can (260 words per minute for English fiction – approximately twice the typical speech rate; Brysbaert, 2019).

Although there is therefore considerable experimental evidence for literacy-related differences in syntactic proficiency, two issues are noteworthy. First, each of the studies discussed above targeted a small number of structures (between one and ten), making it difficult to draw conclusions about the importance of long-term exposure to 'book language' for syntactic proficiency in general. Second, no previous research has examined literacy-related individual differences in grammatical knowledge with reference to the notion, borrowed from linguistics, of 'core' and 'peripheral' grammar.

Current study

Here, we assessed knowledge of 50 syntactic structures in two groups of non-reading impaired adults, sampled from opposite ends of the literacy experience continuum that exists within a literate society like the Netherlands. In addition to examining literacy-related differences, we aimed to provide a snapshot of the breadth of receptive grammatical knowledge that might be shared by the majority of adult native Dutch speakers. We focused on receptive knowledge of structures that had been designated as either 'core' or 'peripheral' by Dutch linguists during an extensive pre-test of the materials (described in the next section). We assessed participants' knowledge of these core- and peripheral-rated structures using a grammaticality judgment task. Acceptance of a structure as grammatical when presented in two different sentence contexts was taken as a proxy for receptive knowledge of that structure.

For the present study, we predicted that item-level performance would broadly reflect linguists' intuitions as to whether a given structure belonged to 'core' or 'peripheral' grammar. If this is a genuine categorical distinction, we would expect a large discrepancy in accuracy on core- versus peripheral-rated structures. Furthermore, following the usage-based assumption that grammatical knowledge is acquired from the input, we predicted that judgments in general would be subject to considerable individual variation, reflecting individual patterns of experience with language (Kidd et al., 2018). We were specifically interested in written language experience as a determinant of receptive grammatical knowledge. People of varying literacy levels are likely to have gained adequate exposure to core sentence structures. However, we predicted better accuracy on peripheral structures for highly experienced literates, as a function of prior enhanced print exposure.

Method

Participants

The thirty-eight native Dutch speakers who participated in the (Favier, Meyer, & Huettig, under review) study also participated in the current study (mean age = 25.2; 25 females).

These 38 were recruited from a pool of 161 participants with varying degrees of literacy experience who had completed a battery of individual difference measures as part of a large-scale individual differences study (Favier & Huettig, under review). Principal components analysis was performed on six literacy measures (Peabody receptive vocabulary, author recognition, reading habits, spelling, word and pseudoword reading) to derive an underlying construct that explained the maximal amount of variance in the literacy data (Literacy PC1 in Table S1)[1]. For the Favier et al. (under review) and the current study, all participants in the top and bottom quartiles for Literacy PC1 who responded to our invitation were tested. We refer to these groups respectively as high literacy experience (HLE) and low literacy experience (LLE). There was a pronounced group difference in literacy experience, based on Literacy PC1 (t = 8.70, p < 0.001) and ART 2 scores (t = 4.01, p < 0.001). The small difference in non-verbal IQ (Raven's) scores between high and low literacy groups (t = 2.10, p = 0.04) was expected and is in line with previous research (e.g., Olivers et al., 2014). Note that the sample size (N=38) and indeed the participants were thus identical to Favier et al. (under review), a study that observed robust effects of literacy on syntactic processing. Ethical approval was given by Radboud University institutional review board. A descriptive summary of the groups is provided in the Supplementary Materials (Table S1).

Materials

We systematically extracted 180 grammatical structures from a compendium of Dutch grammar (*Algemene Nederlandse Spraakkunst*; Haeseryn et al., 1997). The selection encompassed a broad range of noun-phrase, verb-phrase, and clause-level structures, and reflected a taxonomy of important grammatical phenomena (e.g., mood), their constituent

[1] A detailed description of the tests and the principal components analysis, plus grammaticality judgment stimuli, data, and results of additional analyses are provided in the supplementary materials on https://osf.io/dhqsr/?view_only=5bdefd946fe840bd93b4c1ddb0af716a

categories (e.g., conditional), and sub-types (e.g., with inversion). We generated two semantically distinct but syntactically parallel sentences to exemplify each structure. The lexical content of the examples was kept as simple as possible (sentences adapted from *Algemene Nederlandse Spraakkunst* were often reduced in length and complexity). Sentences (a) and (b) are parallel examples of a conditional construction involving inversion (indicated in italics).

a) *Word ik* ziek, zoek dan een vervanger.
If I become unwell, look for a replacement.

b) *Regent het*, dan gaan we niet naar het strand.
If it rains, we won't go to the beach.

We invited expert informants to complete an online pre-test in which they read a randomised sequence of paired examples, such as (a) and (b). Informants were asked to select the best description for each structure from the following options: Core (known by virtually all adult native Dutch speakers); Peripheral (unknown to many native speakers); Incorrect; Unsure. Written instructions at the start of the survey qualified "known" as relating to receptive knowledge. Space for optional additional comments was provided for all structures. To make the duration manageable, we created three versions of the survey, each comprising a different set of 60 structures (i.e. one third of the long list).

Twenty-three expert informants participated in the online pre-test. They were professors, assistant professors, and post-doctoral researchers at six Dutch linguistics faculties in the Netherlands and Belgium. Informants were allocated in approximately equal proportions to the three versions of the online survey. Pre-testing stopped when the total number of responses collected for each structure reached either seven or nine (i.e. an odd number).

After aggregating the responses by structure, we discarded those identified as incorrect by more than two informants (five structures discarded in total). 95 structures were judged to be "Core" by all respondents. From this list we selected a representative set of 25 test items, comprising four noun phrase, 11 verb phrase, and 10 clause level structures. Because "Peripheral" judgments showed much less agreement overall, we set a lower criterion for inclusion in this category. A structure was included if more than half of respondents judged it to be peripheral; in other words, if it received at least 4/7 or 5/9 "Peripheral" responses. This resulted in a shortlist of 30 peripheral structures, from which we selected 25 test items (six noun phrase, seven verb phrase, and 12 clause-level structures). We avoided structures that were highlighted as archaic in the comments. Detailed information about the shortlisted structures is provided in the Supplementary Materials.

In addition to the 50 critical items, we created 15 pairs of ungrammatical sentences as foils. These were comparable to the critical sentences in word length and lexical complexity and were designed to increase the difficulty of the test. Each pair of foil sentences contained parallel syntactic anomalies, concerning a noun phrase, verb phrase, or clause, as shown in (c) and (d).

c) *Er wordt geregend.

d) *Er werd gewaaid.

As every core, peripheral, and foil item consisted of two sentences, a total of 130 sentence stimuli were presented in the test. Whereas the pre-test featured a succession of sentence pairs, each corresponding to one structure, the main grammar test presented all sentences individually, resulting in 130 trials. The order of presentation was pseudo-randomised such that examples of the same structure were separated by at least two syntactically unrelated sentences. All participants saw the same pseudo-randomised list.

Procedure

The test was implemented in Frinex, a software packaged developed at the MPI for online experiments. Participants completed the test individually in a quiet room, using a desktop PC and mouse. On each trial, a sentence appeared on the screen followed, after a three-second lag, by two questions and their corresponding response buttons (illustrated in Figure 1). The questions were "Goed Nederlands?" *Good Dutch?* (Ja/Nee response), and "Hoe zeker ben je?" *How certain are you?* (numerical rating scale). The certainty scale was explained in the instructions as follows: 1=geen idee; 2=onzeker; 3=redelijk zeker; 4=zeker. The purpose of the three-second lag was to encourage participants to read the stimulus sentence fully at least once before responding. Only after responding to both questions could they proceed to the next trial, by clicking a button at the bottom on the screen. If no response was recorded, the next trial began automatically after 20 seconds.

Written instructions at the start of the test included five example trials, with Ja/Nee responses completed as appropriate. Two of the examples were foils. The purpose of the example trials was to demonstrate that the question "Goed Nederlands?" entailed a grammaticality judgment, hence calling attention to the syntactic form of the sentences. Participants had the opportunity to seek clarification from the experimenter after reading the instructions. The test took approximately 30 minutes.

Figure 1. Illustration of test interface. Participants had 20 seconds to answer both questions.

Results

137 out of 4940 trials were excluded from the analysis (38 because of a typo in one of the sentences and 99 that timed out before a response was given). The timeout rate was 1.4% in the HLE group and 2.7% in the LLE group. Of the remaining total, 1860 judgments were obtained for core items, 1820 for peripheral items, and 1123 for foil items. Item-level results are summarised below (results for individual structures can be found in the Supplementary Materials). We then consider performance at the participant level and apply inferential statistics to evaluate literacy-related differences in receptive grammatical knowledge.

Item-level analysis

The overall acceptance rate (i.e., rate of “Ja” responses) was 90.3% on core trials, 56.9% on peripheral trials, and 13.2% on foil trials. Mean certainty (rated on a 4-point scale where 4 = certain) was 3.51 for core items, 3.19 for peripheral items and 3.47 for foil items.

Response accuracy was coded as 1 or 0. For core and peripheral trials, “Ja” responses were coded 1 and “Nee” responses 0. For foil trials, the scheme was reversed (i.e., “Nee” = 1). To evaluate consistency within structures, we correlated the two examples of each core and peripheral structure. The strong positive correlation between the proportion of correct responses on example 1 and example 2 ($\tau = .63$) indicates that difficulty *within* core and peripheral structures was largely consistent, allowing us to proceed with structure-level analysis. We calculated structure difficulty by averaging the proportion of correct responses obtained across the two examples. Table 1 presents a descriptive summary of item difficulty by type. The raw data plotted in Figure 2 illustrates the overlap in difficulty between many core and peripheral structures, despite the statistical difference in group means.

How well did linguists’ intuitions predict accuracy on peripheral structures? There was a moderate negative correlation between the proportion of “peripheral” ratings a structure received from informants in the pre-test and its performance on the grammar test ($\tau = -.20$). However, the correlation may be interpreted with caution, due to the narrow range in the proportion of peripheral ratings (0.57 – 1.00). The subjunctive (e.g., “Ware hij niet zo rijk geweest, hij had het nooit zo ver gebracht”) was judged “peripheral” by all expert informants. This structure also caused the most difficulty in the test, with an average acceptance rate of 8% (irrespective of literacy group). Relatedly, there was a high rate of false positives (42%) for the ungrammatical foil sentence that resembled a subjunctive (“Ware hij niet zo laat, was alles goed”). Together, these findings suggest that many adult native speakers of Dutch have only partial knowledge of the subjunctive.

Notably, recognition of core-rated structures was not at ceiling. Structures that expert informants unanimously rated “Core (known by virtually all adult native Dutch speakers)” were rejected as incorrect on almost 10% of trials. Of the core-rated structures, comparatives using fronted *zo…als* (e.g., “Zo leuk als we gehoopt hadden is het helaas niet geworden”) caused the most difficulty, with an average acceptance rate of 57%.

Item type	**Mean accuracy (SD)**	**Range**
Core	.90 (.09)	.57 – 1
Peripheral	.57 (.27)	.08 – .99
Foil	.87 (.1)	.65 – .99

Table 1. Item-level performance, summarised by type.

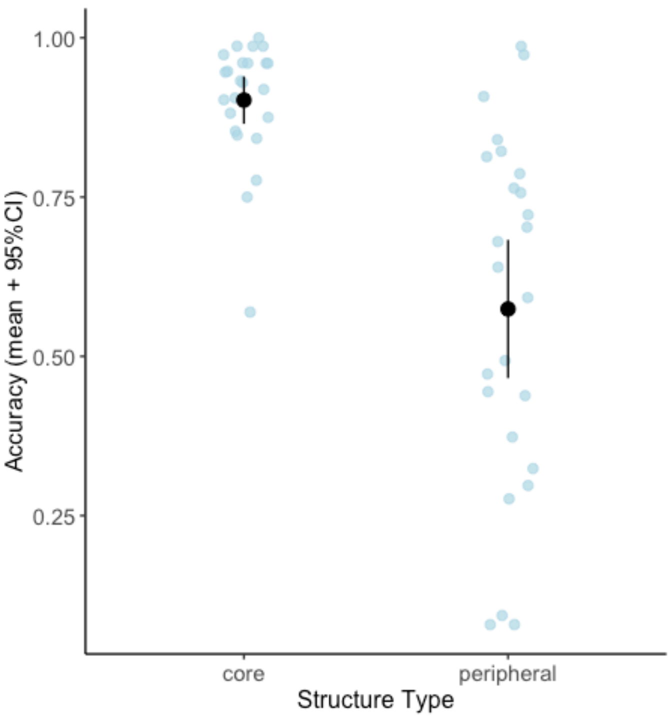

Figure 2. Mean-error-plot showing mean item-level accuracy for core and peripheral structures, plus raw item-level data. Error bars represent 95% confidence intervals.

Participant-level analysis

We aggregated each participant's proportion of correct responses by item type. We also calculated individual d-prime (d') scores, a measure of overall test performance that controls for potential response bias (see Supplementary Materials for discussion). Table 2 summarises the results by literacy experience group. The groups show a similar pattern of performance across the three items types, e.g., they were least accurate on peripheral items. The HLE group was numerically more accurate than the LLE group in accepting both core and peripheral structures, and also showed a small advantage in mean d' scores. Group-level performance on individual structures is reported in the Supplementary Material. The number of structures that performed at ceiling (100% accuracy) differed between groups, with 14 in the HLE group (12 core) compared to only two structures at ceiling in the LLE group (both core). Interestingly, the correct rejection of ungrammatical foils appears unrelated to literacy experience. Figure 2 plots accuracy by literacy group on 'core' and 'peripheral' structures.

Group	Core	Peripheral	Foil	d'
HLE	.92 (.07)	.59 (.13)	.87 (.08)	1.82 (.30)
LLE	.88 (.09)	.54 (.11)	.87 (.11)	1.73 (.36)

Table 2. Summary of group-level performance. Mean proportion correct aggregated by literacy group and item type. D-prime scores (d') aggregated by group. Standard deviations given in brackets. HLE = High Literacy Experience; LLE = Low Literacy Experience.

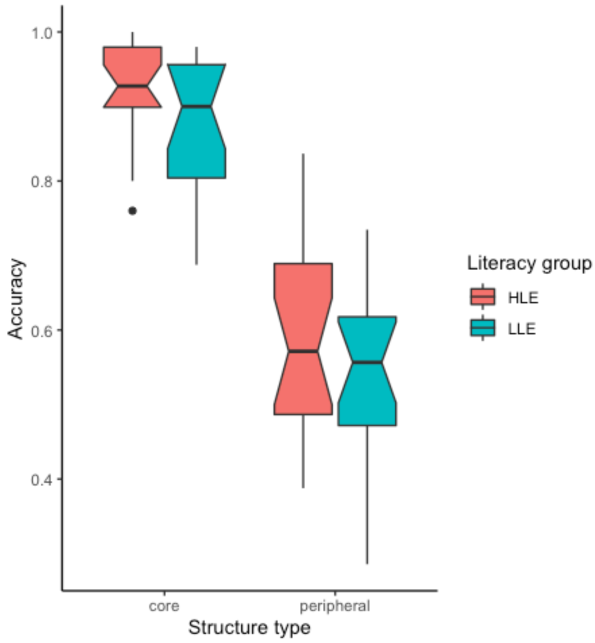

Figure 3. Notched box-and-whisker plot showing the distribution of participant-level accuracy by literacy group, aggregated by structure type. HLE = High Literacy Experience; LLE = Low Literacy Experience. Each coloured box represents the interquartile range (IQR, i.e., 25th – 75th percentile); the 'notches' correspond to 95% confidence intervals for the median (marked in black). The 'whiskers' extend from minimum to maximum (respectively defined as Q1-1.5*IQR and Q3+1.5*IQR). The single outlier is shown as a black point.

We used the *lme4* package in R (version 1.0.153; Bates et al., 2014) to fit a mixed logit model to the accuracy data. For simplicity, we analysed core and peripheral trials only (thus excluding data from all foil trials). The binomial dependent variable was correct ('1') or incorrect ('0'). Structure type (Core/Peripheral) was a fixed factor in the model, with Core taken as the reference level. We included our index of literacy experience (Literacy PC1) as a continuous predictor, as well as its interaction with Structure Type.[2] Raven's Matrices and Backward Digit Span scores from Time 1 were added to the model as covariates, to account for the potential contribution of non-verbal IQ and verbal working memory respectively. All

[2] Model comparison revealed a better fit (log likelihood) with the continuous predictor Literacy PC1 as opposed to the categorical predictor literacy experience group.

continuous predictors were mean centred. We included random intercepts for participant, sentence, and structure level (noun phrase/verb phrase/clause).

Predictor	Coefficient	SE	*z* value	95% CI
Intercept	3.07	0.38	7.99	2.21, 3.82
Literacy PC1	0.13	0.07	1.71	-0.02, 0.27
Type: Peripheral	-2.53	0.33	-7.78	-3.17, -1.89
Non-verbal IQ	0.03	0.02	1.58	-0.01, 0.07
Verbal WM	0.04	0.05	0.71	-0.06, 0.13
Literacy PC1 x Type: Peripheral	-0.03	0.05	-0.50	-0.13, 0.08

Table 3. Summary of fixed effects in the mixed logit model (N = 3680; Log likelihood = -1439.2). Intercept represents the log-odds of a correct response on a core trial for a participant with average literacy experience (Literacy PC1), non-verbal IQ, and verbal working memory (WM).

Table 3 presents the fixed effects in the mixed logit model of response accuracy (for the variance captured in the random part of the model see Table S4, Supplementary Material). The large positive coefficient for the intercept reflects the high average accuracy on core trials. There was only a very small effect of literacy experience on the log odds of responding correctly on core trials (the positive coefficient for Literacy PC1). As predicted, there was a robust effect of structure type on accuracy, such that peripheral structures were far less likely than core structures to be recognised as grammatical. There was a weak positive relationship between accuracy and non-verbal IQ, although this was not statistically robust, and there appeared to be no association with verbal working memory. Finally, we did not find evidence for an interaction between literacy experience and structure type, indicating that the small advantage associated with increased literacy experience did not differ between core and peripheral structures.

Discussion

In order to investigate the empirical basis for ‘core’ and ‘peripheral’ syntactic structures in the Dutch language, we collected grammaticality judgments from adult native Dutch speakers with varying levels of literacy experience. Half of the target structures had previously been classified by a panel of linguist informants as “Core (known by virtually all adult native Dutch speakers)”, and half as “Peripheral (unknown to many native speakers)”.

‘Core’ structures

Consistent with the intuitions of linguist informants, there was a large discrepancy in overall performance on core and peripheral structures. On average, core structures were over 30% more likely than peripheral structures to be accepted as correct Dutch. For example, all participants demonstrated knowledge of the *aan het* continuous construction (e.g., “De schilder was verf aan het mengen”). Unsurprisingly, our results broadly support the notion that the majority of adult native speakers share at least some grammatical knowledge (i.e., ‘core’ grammar). The limited convergence we observed amongst participants on core-rated structures, however, does not fit easily with the categorical definition of ‘core’ grammar espoused in generative linguistics (e.g., Broekhuis & Keizer, 2012, Chomsky, 1981). Several structures unanimously classified as ‘core’ by linguists performed well below ceiling in our educationally diverse sample (e.g., clause-level ellipsis; “Theo is vaak weg, maar ik bijna nooit”). This echoes previous findings of substantial individual variation in the comprehension of supposedly ‘core’ constructions amongst native English speakers (e.g., universal quantifiers; Street & Dąbrowska, 2010; Dąbrowska, 2018).

'Peripheral' structures

Interestingly, there was not much unanimity in the pre-test classification of peripheral structures by our expert informants (linguists), perhaps because the delineation of the 'periphery' is not straightforward for Dutch (Los, 2016). The shortlisted structures also varied considerably in their performance on the test, correlating weakly with the proportion of peripheral ratings received in the pre-test. Somewhat contrary to linguists' intuitions, eight peripheral structures obtained an average acceptance rate of over 75%. In particular, the continuous construction with past and present participles (e.g., "De deur op slot gedaan hebbende, verliet hij het huis") performed unexpectedly well, given the consensus among informants that it would be unknown to many native speakers. In contrast, as many as half of the peripheral structures scored below chance level in our sample, highlighting the disparity between descriptive grammars ("magnasyntax") and the knowledge that most native speakers actually acquire (Miller & Weinert, 1998). The low prevalence of these structures in the general population might be explained by their restricted usage (e.g., highly formal registers), combined in many cases with irregularity (Broekhuis, 2016). For example, the comparative construction within an exclamative (as in "Ze moeten toch altijd doen als wisten ze alles!"), which was rejected in 92% of trials, deviates from canonical subordinate clause word order. The wide range in accuracy on peripheral-rated structures provides further support for a continuous distribution of prevalence, and casts doubt on the viability of a categorical distinction between 'core' and 'peripheral' grammar.

Measuring grammatical knowledge

The present data on adult native Dutch speakers' receptive syntactic knowledge complements Hulstijn's description of syntactic *production* in a similar population (2017). That pilot study was intended as a first step towards establishing the productive inventory of syntactic patterns shared by (virtually) all adult native Dutch speakers (with the broader aim of defining what a native speaker minimally constitutes in linguistic terms). When comparing the present findings to Hulstijn (2017), it is important to bear in mind that differing task demands can give rise to asymmetries in performance across comprehension and production (McCauley & Christiansen, 2013). Because accurate grammaticality judgment can be achieved with only a "shallow parse" of the sentence, supported by semantic knowledge, language users may accept structures as grammatical without having the mastery needed to use them in production. For example, pre-nominal participle phrases were accepted with 85% accuracy on our test but were almost entirely absent from the 80,000-word spoken corpus described by Hulstijn (2017). On the other hand, fronted conjunction-less clauses expressing contrast were rejected as ungrammatical on almost 50% of trials in our test *and* did not feature at all in the corpus, suggesting that this property of Dutch may be truly 'peripheral'. Similarly, several 'core' structures that were used by the majority of speakers in Hulstijn (2017) were also amongst the best performing in our data (e.g., relative clauses and fronted conditional clauses with *als,* both of which obtained at least 90% accuracy).

To gain a more comprehensive and nuanced picture of potential 'core' properties of Dutch, further work should systematically evaluate grammatical knowledge by measuring both comprehension and production in a large, heterogenous sample of adult native speakers. This relates to the fundamental issue that grammatical knowledge cannot be directly observed; we cannot target abstract syntactic structures independently of their lexical realisation. In our grammaticality judgment test, for example, the specific lexical content of each sentence token likely introduced experimental noise. To mitigate any extraneous influence of lexis on participants' judgments, we presented two sentence tokens for each structure. The lexical overlap between tokens was minimal, yet we obtained a strong correlation between judgments of token 1 and token 2, suggesting that participants' responses were not contingent on specific lexical content. Although this result is promising, future research could minimise experimental noise further by using a larger number of lexically distinct tokens per item. If strong within-item consistency was again observed, we could conclude with greater confidence that the task taps knowledge of abstract syntactic structures.

More generally, it will be crucial for future research to provide *converging evidence* by measuring grammatical knowledge in different ways. We regard such an approach as important to reduce the confounds inherent in any single measure. The measure we used here (grammaticality judgment) certainly has its limitations as a proxy for grammatical knowledge. For example, as a metalinguistic task, grammaticality judgment arguably lacks ecological validity in comparison to measures involving more 'typical' language use (e.g., comprehension questions). However, parallels can be drawn with the lexical decision task, which is certainly 'atypical' in terms of language use, but nevertheless is regarded as a useful psycholinguistic tool by many psycholinguists (who believe that such as task, even when taking its limitations into account, can tell us something about an individual's lexical knowledge). Moreover, while previous studies have assessed native grammatical knowledge using sentence-picture matching, comprehension questions, etc. (e.g., Dąbrowska, 2018; Ferreira, 2003), these methods are often applicable to only a limited range of constructions (e.g., passives). Using grammaticality judgment, we were able to target 50 different constructions, providing us with a broad-based and efficient probe of adult native speakers' grammatical intuitions/knowledge.

The current study constitutes an initial step in a line of research that did not previously exist. Future work could build on this foundation. Magnitude estimation approaches could be used to increase the sensitivity of grammaticality judgment measures (Bard, Robertson, & Sorace, 1996). An important step will be to develop a measurement tool that targets a similarly broad (or even broader) range of structures but addresses some of the limitations of the traditional, binary-response grammaticality judgment task. For example, we used certainty ratings to reflect the probabilistic dimension of grammaticality judgment, but future studies could explore alternative approaches to elicit probabilistic judgments (e.g., asking to what degree a sentence “sounds natural”; Featherston, 2005). From a usage-based perspective, such a design could be more sensitive to individual differences in grammatical knowledge related to (written) language experience (Kidd et al., 2018). Another possibility to explore is whether the tendency to *predict* upcoming syntactic continuations (Favier et al., under review; cf. Huettig & Pickering, 2019) provides a sensitive measure of individuals’ grammatical knowledge.

An interesting question to be addressed by future studies concerns the nature of the continuum observed in the present grammaticality judgment data, and to what extent it might approximate a Zipfian distribution (Zipf, 1949; cf. Lestrade, 2017). Further research is required to address this because i) the structures in our study were sampled from a compendium of Dutch grammar rather than natural language corpora (for which Zipfian curves have previously been observed), and ii) we did not present a random sample of syntactic structures but a selection based on pre-defined criteria. In other words, sampling only the most and least known structures (according to linguists) likely distorted the distribution of the data, thus precluding any meaningful speculation about the actual shape of such a distribution in Dutch. The broader question however is an intriguing one, which future research could shed light on. Ideally, the distribution could be characterised by a large-scale prevalence study in which the syntactic structures of Dutch were more comprehensively targeted (perhaps equivalent to the Dutch Lexicon Project; Brysbaert et al., 2016).

Examining the distribution of the top 100 most frequent structures in a large corpus of naturalistic speech would be another way to begin to approach this question. It would also be interesting to compare any resulting Zipfian curve to the distribution of grammaticality ratings for the same structures (although a very large amount of data would likely be needed to detect any underlying Zipfian distribution).

Influence of literacy experience

We predicted that participants with more literacy experience would perform better overall in recognising the structures as correct Dutch, in line with the usage-based assumption that grammatical knowledge is shaped by input (e.g., Bybee, 2006) and that highly experienced literates get higher quality and quantity input. Interestingly, we observed only a small numerical difference in accuracy (approximately 5%) in favour of the HLE group. Modelling analysis that accounted for differences in general cognitive abilities revealed the independent contribution of literacy experience to be statistically marginal. In addition, there was no evidence for an interaction between literacy experience and structure type, indicating that there was no additional benefit of literacy experience for recognising peripheral structures. This is surprising, since given the characteristic low frequency and restricted usage of these structures, we had predicted that highly experienced literates would be the most likely to have encountered them before. Further, specific prior exposure was expected to benefit peripheral structures in particular because of their complexity and/or irregularity (MacDonald & Christiansen, 2002).

What then may explain the absence of a robust literacy effect in the present study? One may argue that the sample size (N=38) was simply too small. This alternative explanation is unlikely to account for the absence of a (literacy) effect found here because the same participants (pre-selected for their literacy differences from a pool of 161 individuals) had also participated in another study on the same day, which observed robust effects of literacy on syntactic processing (Favier et al., under review). While it is conceivable that a future study with a very large sample size may find a statistically significant difference, it would therefore likely find only a small effect at best. We believe that a more likely explanation for the similar performance of HLE and LLE groups in the present study is that native speakers' grammatical knowledge as assessed by grammaticality judgments *over a large range of structures* (50 structures in the present study) overall is fairly good. Many 'grammatically legal' structures in Dutch may simply be 'too peripheral' for almost all native speakers, occurring so infrequently that most participants had never (or very rarely) encountered them before, regardless of literacy experience. Corpus analyses of contemporary Dutch texts could be used to evaluate the empirical basis for this. The present study suggests that although there are certainly some syntactic structures that people with low literacy experience are less familiar with (e.g., prescriptive usage of *als/dan, mij/ik, hun/ze, die/dat*, Favier & Huettig, under review), these are comparatively few. The notion that literacy-related differences emerge only for some structures is supported by our item-level data. Although the level of accuracy across groups was generally high, the item-level data reveal that the HLE group was six times more likely than the LLE group to perform at ceiling on *some* core structures. If we take 100% acceptance as the criterion for inclusion, the body of grammatical knowledge shared by the HLE participants was relatively large (comprising about half of the core structures tested). In contrast, LLE participants unanimously converged on only two core structures (given these strict inclusion criteria).

Conclusions

We observed systematic differences in the grammaticality judgments of adult native Dutch speakers that broadly corresponded to Dutch linguists' intuitions regarding 'core' and 'peripheral' grammatical knowledge. Importantly however, within these categories, there was substantial variability in participants' judgments, which suggests that a categorical distinction between a 'core' grammar and a 'periphery' may not be tenable. Contrary to our expectation, individual differences in literacy experience only explained a small amount of the variance in grammatical judgements of 'core' and 'peripheral' syntactic structures. Thus, overall, the present findings appear to fit best with usage-based views that there is a continuum of syntactic knowledge and that more frequent syntactic structures are mastered better (and by more people) than infrequent ones.

Acknowledgments

We thank Antje Meyer for her advice throughout this project and Brigitte Bauer, who advised extensively on the selection of examples for each structure. We also thank Jan Hulstijn and Ray Jackendoff for their helpful comments on an earlier version of this paper.

References

Abbot-Smith, K., & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review*, *23*(3), 275–290. https://doi.org/10.1515/TLR.2006.011

Bard, E. G., Robertson, D., & Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language*, 32-68.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. *ArXiv:1406.5823 [Stat]*. http://arxiv.org/abs/1406.5823

Broekhuis, H. (2016). Syntax of Dutch: The data set. *Nederlandse Taalkunde*, *21*(2), 297–327. https://doi.org/10.5117/NEDTAA2016.2.BROE

Broekhuis, H., & Keizer, M. E. (2012). *Syntax of Dutch. Nouns and Noun Phrases Volume 1*. Amsterdam University Press.

Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language*, *109*, 104047.

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(3), 441.

Bybee, J. L. (2006). From Usage to Grammar: The Mind's Response to Repetition. *Language*, *82*(4), 711–733. https://doi.org/10.1353/lan.2006.0186

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Chomsky, N. (1993). *Lectures on government and binding: The Pisa lectures* (No. 9). Walter de Gruyter.

Crain, S., & Lillo-Martin, D. C. (1999). *An introduction to linguistic theory and language acquisition*. Oxford: Blackwell Publishers.

Culicover, P. W. (1999). *Syntactic nuts: Hard cases, syntactic theory, and language acquisition* (Vol. 1). Oxford University Press on Demand.

Dąbrowska, E. (2012). Different speakers, different grammars: Individual differences in native language attainment. *Linguistic Approaches to Bilingualism*, *2*(3), 219-253.

Dąbrowska, E. (2018). Experience, aptitude and individual differences in native language ultimate attainment. *Cognition*, *178*, 222–235. https://doi.org/10.1016/j.cognition.2018.05.018

Dąbrowska, E., & Divjak, D. (2019). *Cognitive Linguistics—Key Topics*. Walter de Gruyter GmbH & Co KG.

Featherston, S. (2005). Magnitude estimation and what it can do for your syntax: Some wh-constraints in German. *Lingua*, *115*(11), 1525-1550.

Favier, S., & Huettig, F. (under review). Long-term written language experience affects syntactic awareness and usage but not syntactic priming of spoken sentences.

Favier, S., Meyer, A. S., & Huettig, F. (under review). Literacy enhances syntactic prediction in spoken language processing.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, 47(2), 164-203.

Haeseryn, W. J. M., Romijn, K., Geerts, G., Rooij, J. D., & Van den Toorn, M. C. (1997). *Algemene Nederlandse Spraakkunst* [2 banden].

Huettig, F., & Pickering, M. J. (2019). Literacy advantages beyond reading: Prediction of spoken language. *Trends in Cognitive Sciences*, *23*(6), 464-475.

Hulstijn, J. H. (2015). *Language Proficiency in Native and Non-native Speakers: Theory and research*. John Benjamins Publishing Company.

Hulstijn, J. H. (2017). Syntactic patterns in the speech of native speakers of Dutch, differing in age and level of education and profession: A pilot study exploring the size of shared language knowledge. *Linguistics in Amsterdam*, 10(1), 60-90.

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual Differences in Language Acquisition and Processing. *Trends in Cognitive Sciences*, *22*(2), 154–169. https://doi.org/10.1016/j.tics.2017.11.006

Kroll, B. (1977). Combining ideas in written and spoken English: Discourse across time and space. In E. O. Keenan & T. L. Bennett (Eds.), *Discourse across Time and Space* (pp. 69-108). Los Angeles, CA: University of Southern California.

Langlois, V. J., & Arnold, J. E. (2020). Print exposure explains individual differences in using syntactic but not semantic cues for pronoun comprehension. *Cognition*, *197*, 104155. https://doi.org/10.1016/j.cognition.2019.104155

Lestrade, S. (2017). Unzipping Zipf's law. *PloS One*, *12*(8), e0181987.

Lidz, J., & Williams, A. (2009). Constructions on holiday. *Cognitive linguistics*, *20*(1), 177-189.

Los, B. (2016). The rise of verb-second. *Nederlandse Taalkunde*, *21*(2), 265–273. https://doi.org/10.5117/NEDTAA2016.2.BLOS

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, *109*(1), 35–54. https://doi.org/10.1037//0033-295X.109.1.35

McCauley, S. M., & Christiansen, M. H. (2013). Toward a unified account of comprehension and production in language development. *Behavioral and Brain Sciences*, 36(4), 366-367.

Miller, J. E., Miller, J., & Weinert, R. (1998). *Spontaneous spoken language: Syntax and discourse*. Oxford University Press on Demand.

Montag, J. L., & MacDonald, M. C. (2015). Text exposure predicts spoken production of complex sentences in 8- and 12-year-old children and adults. *Journal of Experimental Psychology: General*, *144*(2), 447–468. https://doi.org/10.1037/xge0000054

Nowak, M. A., Komarova, N. L., & Niyogi, P. (2001). Evolution of universal grammar. *Science*, *291*(5501), 114-118.

Olivers, C. N. L., Huettig, F., Singh, J. P., & Mishra, R. K. (2014). The influence of literacy on visual search. *Visual Cognition*, *22*(1), 74-101.

Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, *57*(3), 348–379. https://doi.org/10.1016/j.jml.2007.03.002

Street, J. A., & Dąbrowska, E. (2010). More individual differences in language attainment: How much do adult native speakers of English know about passives and quantifiers? *Lingua*, *120*(8), 2080–2094. https://doi.org/10.1016/j.lingua.2010.01.004

Wells, J., Christiansen, M., Race, D., Acheson, D., & Macdonald, M. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, *58*(2), 250–271. https://doi.org/10.1016/j.cogpsych.2008.08.002

Zipf, G. K. (1949). *Human behavior and the principle of least effort. An introduction to human ecology*. New York and London: Hafner publishing company.