

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*

Jan H. Hulstijn

ACLCLC, University of Amsterdam

*In an attempt to begin to establish the productive knowledge of morpho-syntax (of Dutch), shared by all adult native speakers, this study examined the extent to which several patterns of Dutch syntax are present in the speech of 98 adult native speakers of Dutch, differing in age (18–76) and level of education and profession (EP-Low vs. EP-High), who performed four speaking tasks, differing in formality (formal vs. informal) and discourse type (descriptive v. argumentative). The size of the corpus is 12 hours (80,000 plus word tokens). The study was guided by the following question, derived from Hulstijn's (2015) theory of Basic Language Cognition: Which syntactic patterns (of the ones under examination) are acquired (in all likelihood) by all native speakers? Most of the patterns under investigation were chosen from the perspective of Hawkins' (2004, 2014) efficiency theory. The findings suggest that adult native speakers of Dutch produce subordinate-clause patterns, such as complement clauses, adverbial clauses and relative clauses at their base position in the matrix clause (to the right of the verb, NP, etc.), but that the phenomenon of clause fronting is common only with respect to some adverbial clause types, such as conditional clauses beginning with *als* ('if'). Verb clusters with modal auxiliaries and the passive voice also appear to belong to shared grammatical knowledge. In contrast, the findings suggest that *it*-cleft sentences, *wh*-cleft sentences, the fronting of conjunction-less conditional clauses, fronting of infinitival clauses, and center-embedding a clause within another clause may not belong to shared grammatical cognition. The findings are claimed to be potentially relevant to any theory of language acquisition aiming to explain the following two questions: (i) Why it is that some syntactic patterns are not and other patterns are acquired by all native speakers? and (ii) How can the acquisition of these shared patterns be accounted for in terms of learning mechanisms (nature) and exposure (nurture)?*

* I would like to thank Suzanne Aalberse (University of Amsterdam) and Nel de Jong (VU University Amsterdam) for providing me with valuable feedback on earlier versions of this paper.

1 Introduction

According to authoritative estimates (Comrie, Matthews & Polinsky 2003: 13), around ten percent of the approximately 6,000 languages in the world have existed in both spoken and written modalities for many centuries. Currently, there are around 50 languages spoken by more than 10 million people and around 120 languages spoken by between 1 and 10 million people (Austin 2008). Because of their widely different functions in economically and culturally complex societies, developed over many centuries, many of these languages are currently being used in many genres, in both spoken and written discourse, at home, in educational, professional and leisure-time settings, in private and public fora. More often than not, these multi-functional languages have become standardized as a result of language policy implemented by governments or other institutions in power. Through the descriptive work of many linguists, former and current social and regional varieties of these languages have been documented. As a result of this linguistic work, for many of these multi-functional languages in literate societies, we now have at our disposal dictionaries, grammars and corpora of spoken and written language covering many centuries diachronically and many genres synchronically.

These dictionaries and grammars represent what Miller & Weinert (1998: 376) call “magnavocabulary”, a term borrowed from Ong (1982), and “magnasyntax”. Magnavocabulary “is not the property of any one speaker” (ibid.). Miller and Weinert introduce the term “core” for “the syntax and vocabulary typically used and understood by children at certain ages” (p. 407). The authors argue that “this notion of core is based on frequency as well as simplicity (...). We could say that the core consists of all constructions and vocabulary found in the informal speech of teenage pupils – say age 16 – or of adults who have no higher education”. “(I)t is clear from existing corpora that there is a shared body of syntax whose properties can be specified in terms of degrees of clause embedding, degree of clause combination, proportion of finite subordinate clauses to main clauses, absence of certain constructions” (p. 407).

Given the many functions of languages in modern literate societies and given the uncontested assumption that no singular native speaker has knowledge of the entire vocabulary and grammar to be found in dictionaries and descriptive grammars, the question arises as to which words and which elements of grammar are acquired, and thus shared, by all native speakers. To address the issue of shared and non-shared knowledge of a language, Hulstijn (2011, 2015) proposed the notions of Basic Language Cognition (BLC) and Extended Language Cognition, also referred to as Higher Language Cognition (HLC). Hulstijn (2011: 230f) defines BLC and HLC as follows:

Basic language cognition (BLC) pertains to (1) the largely implicit, unconscious knowledge in the domains of phonetics, prosody, phonology, morphology and syntax, (2) the largely explicit, conscious knowledge in the lexical domain (form-meaning mappings), *in combination with* (3) the automaticity with which these types of knowledge can be processed. BLC is restricted to frequent lexical items and frequent grammatical structures, that is, to lexical items and morphosyntactic structures that may occur in any communicative situation, common to all adult L1-ers¹, regardless of age, literacy, or educational level. [...] *Higher language cognition* (HLC) is the complement or extension of BLC. HLC is identical to BLC, except that, (1) in HLC, utterances that can be understood or produced contain low-frequency lexical items or uncommon morphosyntactic structures, and that (2) HLC utterances pertain to written as well as spoken language. In other words, HLC utterances are lexically and grammatically more complex (and often longer) than BLC utterances and they need not be spoken. HLC discourse pertains to topics other than simple every-day matters, i.e. topics addressed in school and colleges, on the work floor, and in leisure-time activities.

According to Hulstijn (2015: 51), knowing what is included in and what is excluded from BLC – in other words, knowing what a native speaker minimally is in linguistic terms – will serve the empirical study of four fundamental issues in the study of L1 acquisition, L2 acquisition, and bilingualism: (i) explaining commonalities and differences in L1 acquisition in children, (ii) answering the question of whether there is an age-of-onset constraint on attaining so-called ‘native’ control of an L2 (also known as the Critical Period question), (iii) explaining non-trivial individual differences in L2 acquisition, and (iv) making comparisons between non-native and native speakers of a language (e.g., establishing language dominance in bilinguals). BLC Theory, which is proposed as a research agenda containing falsifiable corollaries, aims to bring together views from linguistics (a usage-based account of language use), sociolinguistics (social-psychological attributes factors associated with individual differences in language proficiency), and psycholinguistics (taking processing factors in speech production and speech comprehension into account).

In summary, this paper is based on a combination of two views on language and grammar, somewhat similar to what Chomsky (1986) called external and internal language (E-language and I-language). In one view, language is an *external* object of investigation, the product of spoken and written communication in many different genres of discourse, developed over

¹ A native speaker, also referred to as L1-er, is defined in social terms as “someone who typically acquires the language as a young child (before school age) and maintains the language into adulthood” (Hulstijn 2015: 28). Under this definition, someone can be a native speaker of more than one language. For details, see Hulstijn (2015, ch. 3 & 5) and Hulstijn (submitted).

many centuries of use by sociolinguistically different speech communities. The grammar of such multi-functional languages encompasses many different types of genres (Miller & Weinert's magnasyntax). In the other view, language is a token of an individual's linguistic cognition in which an individual's grammar represents "the cognitive organization of one's experience with language" (Bybee 2013: 2). The question, crucial in the debate with respect to the biological and cognitive make-up of the human species at the current stage of its evolution, is, in the case of modern, multi-functional, spoken and written languages, what the properties are of that part of their magnasyntax that is acquired by all typically developing native speakers, regardless of differences in intellectual abilities or attained levels of literacy and education. To account for the successful acquisition of this subset of magnasyntax, what is it in the mind/brain that newborns need to be equipped with?

In an ideal research world, one would like to create, for a given language, a corpus of oral language, produced by a truly representative sample of native speakers (i.e., people at different ages, different levels of education, different professions, and different ways of spending their leisure time) in all possible situations of oral communication. One would then like to analyze this corpus with respect to vocabulary and morphosyntax and determine the shared grammar. To our knowledge, currently no corpus, representative in this sense, exists for any language. In the current absence of a truly representative corpus of spoken language – in the sense just described – for any of today's multi-functional languages, spoken and written in modern societies, the empirical study reported in this paper is presented as a modest attempt to begin investigating the characteristics of the productive knowledge of syntax shared by all native speakers.

Research question. The study reported in this article examines the extent to which several patterns of Dutch syntax are present in the speech of 98 adult native speakers of Dutch, differing in age and level of education and profession, who performed four speaking tasks. The size of the corpus is 12 hours (80,000 plus word tokens). The study was guided by the following question: Which syntactic patterns (of the ones under examination) are acquired (in all likelihood) by all native speakers?

In the Results section of this paper, a number of syntactic patterns are examined. For each pattern, the question is whether they are produced by speakers at all, and if so, whether associations with speakers' age or level of education or profession are observed. This is important in establishing the grammar shared by all native speakers. First, we examine clause length in general (Section 3.1) and the length of subclauses that precede their matrix clause (Section 3.2). We then examine the incidence of a number of different matrix-preceding subclause types (Sections 3.2.1 through 3.3). This is followed

by examinations of relative clauses (Section 3.4) and heavy pre-nominal adjuncts in NPs (Section 3.5), complex verbs (Section 3.6), and use of the passive voice (Section 3.7).

1.1 Syntactic efficiency

We chose Hawkins' (2004) grammatical efficiency theory as our guide in deciding which syntactic phenomena in the corpus to examine. Hawkins (2004: 3) proposes a theory aiming at accounting for the design of grammars (of natural languages) on the basis of language performance (language production and comprehension). The theory proposes three main principles that govern syntactic sentence processing: Minimize Domains, Minimize Forms, and Maximize On-line Processing. Of these, the first and third one were relevant for our investigation. We will present these principles briefly (and incompletely) in our own words, focusing on the aspects relevant for our study. For an elaborate presentation, see Hawkins (2004, 2014), which presents an update of the theory.

The principle of Minimize Domains can be illustrated with the following example (Hawkins 2014: 12):

- (1) a. The man [_{VP} looked [_{PP1} for his son] [_{PP2} in the dark and derelict building]]
 b. The man [_{VP} looked [_{PP2} in the dark and derelict building] [_{PP1} for his son]]

The hearer of sentence (1a) can identify the prepositional object of the verb *looked* earlier than the hearer of sentence (1b). In the former case the parsing of the prepositional object constituent (PP1) can be performed immediately after the verb *looked*; in the latter case, the parsing of PP1 can only be performed after the completion of PP2. Thus, in the case of (1a), the processor can construct a syntactic parse of the verb and its prepositional object on the basis of a smaller string of terminal elements than in the case of (1b). The so called Phrasal Combination Domain is smaller in (1a) than in (1b). A Phrasal Combination Domain is the smallest string of terminal elements on the basis of which the processor can construct the Mother node (VP in this example) and its Immediate Constituents (ICs). "The human processor prefers linear orders that minimize Phrasal Combination Domains (by maximizing their IC-to-word ratios), in proportion to the minimization difference between competing orders" (Hawkins 2014: 12). Hawkins calls this the principle of Early Immediate Constituents. It is on the basis of this principle that sentence (1a), in comparison to (1b), makes fewer demands on working memory and on the computational system (Hawkins 2014: 13).

In our wording of it, the principle Maximize On-line Processing states that the human processor prefers sentences in which each subsequent word can be assigned a grammatical function immediately. The processor disprefers sentences where a word leaves room to (i) uncertainty, leading to temporarily unassigned properties, or (ii) ambiguity, leading to misassignment and to backtracking (after subsequent words, that can resolve the ambiguity, have been processed). This can be illustrated by the famous example *The horse raced past the barn fell*. When hearing the word *raced*, the hearer is likely to parse the string *The horse raced* as a subject followed by an intransitive verb. Having heard the remainder of the sentence, the hearer realizes that the initial property assignment of *raced* was wrong and must be reassigned.

Hawkins (2004: 25) summarizes the ideas underlying his efficiency principles with the slogans “Express the most with the least” and “Express it earliest”. It is important to emphasize that the efficiency principles render *probabilistic* tendencies. They are more likely to apply with increasing syntactic complexity (more terminal and non-terminal elements). This is illustrated by the following example (Hawkins 2014: 40). While extraposition of a subject clause to the right is preferred over non-extraposition if the matrix clause is relatively short (compare (2a) with (2b)), this may no longer be the case when the matrix clause contains many constituents and terminal nodes (compare (2c) with (2d)).

- (2)
- a. It is important that he should succeed.
 - b. That he should succeed is important.
 - c. It is important for all the good people who have invested in him that he should succeed
 - d. That he should succeed is important for all the good people who have invested in him.

The complexity (ratio of non-terminal to terminal nodes) of sentence (2a) is bigger than that of (2b) because (2a) contains one terminal node (*it*) more than (2b). However, following the principle of Minimize Domains, in particular the principle of Early Immediate Constituents, (2a) (a structure known as extraposition) can be processed more easily than (2b) (Hawkins 2014: 39). Often, sentence construction by the speaker (and parsing by the hearer) is a matter of dealing with potential trade-offs between formal complexity and processing efficiency. Obviously, the length of a string may play a role. For example, all other things being equal, a sentence beginning with an adverbial clause is syntactically more demanding (putting more demands on short-term memory) for the human processor than a sentence where the adverbial clause assumes its base position, after subject, verb (and object, if present), as predicted by Hawkins’ (2004) principle of Minimize Domains. However, it is not the case

that fronted subclauses are dispreferred altogether. Semantic and pragmatic factors may lead the speaker to put a subclause before its matrix and some syntactically less efficient constructions are preferred when they have become conventionalized (formulaic). This will be illustrated in the Results section.

Although Hawkins (2004) formulated the efficiency principles primarily with the hearer in mind (parsing), the principles seem to apply also for the speaker (speech production). While this matter is debated in the literature (see Hawkins 2014: 50f; McDaniel, McKee, Cowart & Garrett 2015), it does not feature in the current study (see Pickering & Gompel (2006: 480–483) for complexity and memory load in syntactic parsing; Ferreira & Engelhard (2006: 72–79) for complexity in language production; and Pickering & Garrod (2013) for an integrated theory of language production and comprehension). In this study, we focus on speech production and what might be inferred from it about native speakers' mental grammars. The fundamental question of whether two mental grammars have to be postulated, one for speech comprehension and one for speech production, is beyond the scope of this paper (Pickering & Garrod 2014).

1.2 Previous research

According to Macaulay (2013), there are almost no studies in sociolinguistics that investigated differences in syntactic complexity in the speech of people of different social classes. Bernstein (1959, 1962) recorded discussion sessions on the abolition of capital punishment, conducted by 10 middle-class and 14 working-class boys aged 15 to 18, and found the former group produce more subordinate clauses than the latter group but no adequate statistical analyses were conducted. Macaulay (1991) analyzed speech produced by six middle-class and six working-class adults in Scotland and found that the middle-class speakers did in fact “exhibit a greater use of subordination in general and a more frequent use of subordinate noun clauses in particular” (Macaulay 1991: 106).

Hulstijn (2015, chapter 6) reviews the literature of effects of age and level of education on native speakers' language proficiency but no studies were found investigating effects of these factors on syntactic patterns of spontaneous speech. There are a number of studies that looked at the syntax of spontaneous conversations, albeit not with a focus on differences among speakers at different levels of education or social class. The final chapter of the *Longman Grammar of Spoken and Written English* (Biber, Johansson, Leech, Conrad & Finegan 1999; see also Leech 2000) is devoted to the syntax of English as spoken in spontaneous conversations, based on a corpus of around 40 million words.

For Dutch, the language of investigation in the current study, no empirical studies exist comparing syntactic patterns in spoken (or written discourse) of adult native speakers of different ages or of different levels of education or

socio-economic status, with one exception. Jansen (1981) interviewed 40 adult native speakers of Dutch (age range 50–70), differing from one another with respect to gender (20 male, 20 female) and social class (20 low class and 20 middle class). The first part of each interview was conducted in a more formal manner, eliciting formal discourse; in the second part, the interview turned into an informal conversation, eliciting informal discourse. The size of the elicited speech corpus is not mentioned, either in terms of recording time or in terms of the number of word tokens in the transcripts. The aim of the study was to investigate the presence of six features of grammar in a corpus of spoken discourse. To guarantee the validity of the corpus, the researcher controlled for discourse (formal vs. informal), and participants' gender and social class. The six features under investigation are stranding, deletion of sentence-initial element, repetition of the tensed verb, left dislocation with copying by a deictic pronoun, and fronting. These features are all fully acceptable in spoken Dutch but not in (formal) written discourse (with the exception of fronting, which is acceptable also in written discourse). All target structures were used by all four subgroups albeit not to the same extent. However, the aim of the study was not primarily to investigate the association of target-structure production with gender and social class. Thus, the findings of Jansen's study do not speak to our research question (see, however, footnote 4). Unfortunately, the recordings and the transcripts of the interviews are no longer available; otherwise we would have been allowed to use, and indeed had used, this corpus for our study.

Unfortunately, the *Corpus Gesproken Nederlands* (CGN) (The Spoken Dutch Corpus, 2004) – a 9 million-word corpus of modern spoken Dutch – is of little help when it comes to establish the grammar shared by all native speakers. Even though information about speakers' attributes (e.g., gender, age, and level of education) is available in the corpus, the speaking tasks were not (e.g., in the case of informal conversations) or could not (e.g., in the case of teacher talk in schools) be systematically assigned to individuals with different attributes. The advantage of the Mulder & Hulstijn corpus (see Section 2), although much smaller than the CGN, is that we manipulated the speaking tasks (by crossing two two-level factors Formal-Informal and Descriptive-Argumentative) and sampled the participants on the basis of their age and level of education and profession (see next section). This allowed us to do justice to the inherent variationist nature of corpus linguistics (Geeraets 2015). We made use of the CGN occasionally (see Section 3.5).

2 Method

The speech produced by adult native speakers of Dutch examined in this study forms part of data collected in a larger study (see Mulder & Hulstijn (2011) for

details). The corpus (henceforth M&H corpus) consists of speech produced in four speaking tasks by 98 native speakers of Dutch, living in a small town not far from Amsterdam. Participants differed in age and level of education/profession (EP), as shown in Table 1. Classification of the latter was based on criteria of Statistics Netherlands (*Centraal Bureau voor de Statistiek* 2001). There were eight subjects in the lowest occupation class (e.g., office cleaner, mail sorter, and housewife) and four subjects in the highest class (university lecturer, researcher, lawyer, and software engineer).

Table 1. Participant numbers by Age and Education-Profession (EP)

| <i>Age group</i> | <i>EP-Low</i> | <i>EP-High</i> | <i>Total</i> |
|------------------|---------------|----------------|--------------|
| Young (18–35) | 18 | 24 | 42 |
| Older (36–50) | 10 | 10 | 20 |
| Senior (51–76) | 24 | 12 | 36 |
| Total | 52 | 46 | 98 |

Participants performed a battery of 12 language-related tasks, consisting of (i) four computer-administered speed tasks, measuring word association, auditory lexical decision, visual lexical decision, and picture naming; (ii) a paper-and-pencil vocabulary knowledge test; (iii) an auditory and a visual word-span task; (iv) four speaking tasks; and (v) a writing task (70 participants). Mulder & Hulstijn (2011) reported on participants' lexical knowledge, lexical fluency, lexical memory as a function of their age and EP. With respect to the four speaking tasks, Mulder & Hulstijn report how well participants performed in terms of communicative success (see below), as well as the length of their responses (number of words and T-units), number of hesitations and number of grammatical errors.

The four computer-administered speaking tasks were designed as cognitively complex monologues, more demanding than most every-day conversations. In terms of the discourse dimensions presented by Biber et al. (2002: 18), the tasks could be considered as relatively (i) informational, (ii) non-narrative, (iii) elaborated, requiring context-independent reference production, and (iv) argumentative (two of the four tasks). The speaking tasks differed along two dimensions, formal vs. informal and descriptive vs. argumentative, in a two-by-two fashion, as described below. For each task, the instruction screens provided a photo picture of the communicative situation and one or several visual-verbal cues concerning the topic. Participants, who were tested individually, were invited to play the role of someone involved in the situation

and to produce a monologue of not more than two minutes. While subjects were speaking, the visual-verbal cues remained accessible on the computer screen.

In the *Unemployment* task (informal, descriptive), the participant tells a friend about the development of unemployment among women and men over the past ten years, using information provided by a graph. In the *Transportation* task (informal, argumentative), the participant discusses, for a friend, the pros and cons of how to solve the problem of traffic congestions by the use of public transportation, bicycles, or automobiles. In the *Hospital* task (formal, descriptive), the participant works at the employment office of a hospital and tells a candidate for a nurse position, on the phone, what the main tasks in the vacant position are. In the *Car Park* task (formal, argumentative), the participant plays the role of the manager of a supermarket, addressing a neighborhood meeting, arguing which one of three alternative plans for building a car park he/she prefers.

Participants' speech was recorded and later transcribed. The audio recordings were played to a panel of three judges (non-linguists) to individually assess the communicative success of participants' speech responses. Grammatical correctness did *not* feature in the rating scale, which was entirely devoted to the quality and amount of information provided. Of the many findings of the Mulder & Hulstijn study (2011), the following two are relevant in the present context. In comparison to EP-Low subjects, EP-High subjects (i) talked significantly longer, producing more words, and (ii) produced responses that are communicatively more successful. It is important to note that these findings provide evidence for the cognitive demand of the speaking tasks: the participants were challenged to show the best of their oral skills, increasing the chances of demonstrating the best of their lexical and grammatical skills.

Mulder & Hulstijn (2011) did not analyze participants' speech in terms of grammatical characteristics. The purpose of the present study is to provide a detailed analysis of several syntactic patterns in the speech data (the corpus). The analyses were conducted four years after the publication of the Mulder & Hulstijn (2011) article. The corpus contains 80,215 word tokens (approximately 12 hours of speech). Although the corpus is extremely small in comparison to current standards in the field of corpus linguistics, its unique features are that it consists of speech produced by speakers selected for age and EP, who all performed exactly the same speaking tasks. In addition, the informational quality of the responses, making up the corpus, was assessed independently.

Missing responses. Not all subjects performed all four speaking tasks. Six subjects failed one task each, because of fatigue and/or of perceived task difficulty: three subjects (all EP-Low) failed the Unemployment task, one subject (EP-Low) failed the Hospital task, two subjects (one EP-Low and one EP-High) failed the Car Par task; the Transportation task was performed by all

subjects. Hence the *n*-sizes for the four tasks are 98 (Transportation), 95 (Unemployment), 97 (Hospital), and 96 (Car Park). The *n*-size of ANOVAs conducted on dependent variables across tasks (Total of words, Total of T-Units, Words per T-unit) was 92. Inferential statistics are only reported when meaningful. In the case of most patterns under investigation, it sufficed to report the number of cases and the number speakers involved (in the EP-Low and EP-High subcorpora).

3 Results

3.1 Clause length

In our search for patterns of Dutch syntax shared by all native speakers and inspired by Hawkins' (2004) notion of processing efficiency, presented in Section 1.1, we directed our focus first on clauses. In coding the clausal units in our data, we followed as much as possible Haeseryn, Romijn, Geerts, De Rooi & Van den Toorn (1997: 111–113) and Klooster (2001). Most clauses in Dutch (and other Germanic languages) contain a finite verb form and a syntactic subject. The following sentence contains two clauses of this type, separated in writing by a comma. Citations from the corpus are followed by the speaker's ID number, EP level, gender, age in years, and self-labeled profession.

- (3) *als ik daar dan ben, pak ik gewoon de fiets naar mijn werk*
 (2, EP-High, male, age 35, assistant professor)
 'When I arrive there, I just take my bike to my work.'

A clause can also consist of a non-finite, infinitival predicate without an overt syntactic subject. Such clauses are called *beknopte bijzinnen* (infinitival clauses) in Dutch grammars (Haeseryn et al. 1997: 111–113); Biber et al. (1999: 198ff) call them "infinitive clauses". Examples are provided in (4) and (5). The vertical slash | indicates the border between the two clauses.

- (4) *u bent hier vandaag bijeen gekomen | om te praten over de nieuwe parkeerplaats*
 (61, EP-Low, male, age 36, digital screen operator)
 'You have gathered here today | in order to talk about the new car park.'
- (5) *je had die files ook kunnen ontwijken | door gewoon je tomtom te gebruiken*
 (61, EP-Low, male, age 36, digital screen operator)
 'You could have avoided those traffic jams | by using your car navigation.'

Our first endeavor was to ascertain any Age- or EP-related differences in the length of clauses, measured in number of words, as a rough index of potential ease of syntactic processing. We acknowledge that a more valid index would also take the number of constituents into account, such that a clause containing few constituents as well as few words would be optimally processable (as far as syntactic parsing is concerned).

Because counting clause length manually is highly time-consuming and given the pilot status of the current report, we decided to select a subsample of the data for computing clause length. A subsample of 30 participants was randomly selected from the 98 participants in this study, matched for Age and EP, such that there were five subjects in each of the six Age (3) by EP (2) combinations of these speaker attributes. These 30 subjects produced 3,143 clauses across the four speaking tasks, with an average of 105 clauses per speaker. The mean length of words per clause was 7.1 (Median = 7.0). This figure was almost identical (but somewhat smaller) to the mean clause-length figure of 7.5 words observed by Biber et al. (1999: 1071) for conversational English, despite the fact that our speaking tasks may have been more “monologic” and hence more prone to elicit longer clauses than the conversation corpus of Biber et al.² Six two-way ANOVAs, with Age and EP as independent factors were run on the six variables of which descriptive statistics are given in Table 2. These variables include the traditional T-unit (Hunt 1965), defined as an independent clause plus any subordinating clauses associated with it.³ None of these ANOVAs produced significant Age or EP effects or interaction effects (all p values < .05). Thus, on average, clause length was the same for all subjects regardless of age or EP.

² Cross-language comparisons should be interpreted with caution. For instance, Dutch allows more compounding than English, which may result in shorter texts in Dutch than in their English translations.

³ For the purposes of this study, the T-unit (Hunt 1965) or any variant of it is too crude a measure. We therefore do not delve into the T-unit literature (e.g., Biber et al. 1999: 1069; Chaudron 1988; Foster, Tonkin & Wigglesworth 2000).

Table 2. Descriptive statistics of Clauses, T-units, and Words across the four speaking tasks (N = 30)

| | Sum | Minimum | Maximum | Mean | SD |
|---------------------|--------|---------|---------|------|-----|
| Number of Clauses | 3,143 | 53 | 204 | 105 | 34 |
| Number of T-units | 1,992 | 38 | 120 | 66 | 21 |
| Number of Words | 26,358 | 471 | 1534 | 879 | 265 |
| Words per T-units | | 7.6 | 18.5 | 13.5 | 2.5 |
| Clauses per T-units | | 0.9 | 2.2 | 1.6 | 0.3 |
| Words per Clause | | 2 | 21 | 7.1 | 2.9 |

Note: The Word count includes both clauses and non-clausal units. The coding of T-units also included non-clausal units. Thus, the product of 1.6 (clauses per T-unit) times 7.1 (words per clause) does not equal 13.5 (words per T-unit); the product is smaller (11.4).

3.2 Subclauses (*adverbial or wh-cleft clauses*) preceding matrix clauses

The first place in sentences of SOV and SVO languages can be taken by a constituent other than the syntactic subject. This fronting (also called extraposition or left dislocation) is normally pragmatically driven (e.g., topicalization). In Dutch (and German), fronting leads to what is often referred to as subject-verb inversion, i.e., the first constituent after the fronted constituent is a finite-verb form, followed by the subject. When the fronted constituent consists of a clause (adverbial clause, subject clause, object clause), the speaker has to wait delivering the verb and the subject until the subclause has been completed. Similarly, the hearer, when hearing that a sentence begins with a subclause, has to postpone the parsing of the finite verb and the subject. Thus, all other things being equal, a sentence beginning with an adverbial clause is syntactically more demanding (putting more demands on short-term memory) for the human processor than a sentence where the adverbial clause assumes its base position, after subject, verb and object, as predicted by Hawkins' (2004) principle of Minimize Domains. When listening to a fronted *that* clause (*dat* in Dutch, *dass* in German), the hearer cannot yet assign its syntactic property with certainty because the *that* clause may either be a subject- or an object clause. This is dispreferred by the processor because of Hawkins' principle of Maximize On-line Processing. In spoken, but not in written language, fronted subject- or object clauses are often followed by the resumptive pronoun *dat* which helps the processor parse the matrix clause, as if the fronted constituent consisted of only one word (see also footnote 4).

In Dutch and German, generally speaking, almost all types of subclauses containing a finite verb observe a verb-final (S)OV pattern, whereas almost all types of declarative main clauses observe a verb-second pattern, with the finite

verb in second position. There is one subclause type that is characterized by the absence of a clause initial conjunction and by a finite-verb-first VSO word order (see Section 3.2.4). Main clauses can also be preceded by infinitival clauses. Generally then, listeners, while hearing an utterance-initial subclause, know, on the basis of one or more overt, formal subclause characteristics, that they are not yet hearing a main clause. Only when the subclause has been completed, can the listener round off its parse and store the information of the clause as a proposition that, in one way or another, is semantically related to the proposition of the matrix clause.

Clause length (number of constituents or number of words) is arguably related to processing cost. In Section 3.1, we did already observe that clause length in the M&H corpus of spoken Dutch is rather short, even when one considers that participants fulfilled cognitively demanding speaking tasks, requiring delivering relatively much information in descriptive or argumentative monologues. We will now present the findings of subclause length.

As expected, the 30 subjects in the subsample mentioned above, produced many more subclauses (including infinitival clauses) following their matrix clause ($M = 24$; $SD = 12$) than preceding it ($M = 5$; $SD = 3$). This big difference was reliable in a repeated-measures ANOVA ($\eta^2_p = .69$; $F [1,29] = 64.856$; $p < .001$). Note that the fronted clauses were only of the adverbial type, while the subclauses following subject and verb also included *dat* clauses, such as complement clauses and subject clauses. (Not surprisingly, all 30 subjects produced *dat* clauses, following the verb in the matrix clause.). None of the 30 subjects produced sentence-initial *dat* clauses.⁴ No effects of Age or EP were obtained in the number of matrix-preceding or matrix-following subclauses. All together, these 30 subjects produced 155 pre-matrix clauses, with a mean word length of 6.8 words ($SD = 2.4$). Thus, on average, matrix-preceding clauses were rather short, in line with Hawkins' principle of Minimize Domains.

3.2.1 Conditional clauses preceding a matrix clause

By their semantic nature, conditional clauses are highly likely to be produced before rather than after their matrix clause, as illustrated in (6) and (7):

⁴ Jansen (1981: 190f), in the study mentioned in Section 1.1, found 84 instances of fronted *dat* clauses followed by the resumptive pronoun *dat*. Of these, 57 and 27 instances were produced, respectively, by middle class and lower class participants, while 55 and 29 instances were produced by male and female participants.

- (6) *en als het u wat lijkt, kunnen we misschien een afspraak maken*
(155, EP-Low, female, age 66, office cleaner)
'And if you like this, we could arrange a meeting.'
- (7) *als je te lui bent om te werken, dan wat heb ik dan aan je*
(157, EP-High, male, age 74, artist)⁵
'If you are too lazy to work, then, what use are you to me?'

In the entire corpus (98 speakers), 210 matrix-preceding conditional clauses were found, of which 209 started with the conjunction *als* and one with its more formal synonym *indien*. Of these 210 instances, 154 were produced in the two argumentative tasks and 56 in the two descriptive tasks. An ANOVA produced no significant effects of Age or EP or a significant interaction. The relative frequency of matrix-preceding *als* clauses (divided by number of words or divided by number of T-units) was not significantly affected by Age or EP either. There were 14 EP-Low subjects and 10 EP-High subjects who produced no matrix-preceding *als* clauses at all.

3.2.2 Other sentence-initial adverbial clauses

In total, 81 adverbial clauses (not starting with *als*) preceding a main clause were produced, 53 in the argumentative and 28 in the descriptive tasks, by 23 EP-Low and 22 EP-High subjects. Examples are provided in (8) to (10).

- (8) *dus wat dat aangaat, schiet je er weinig mee op*
(7, EP-High, female, age 23, elementary-school teacher)
'Thus, as far as that is concerned, that will bring you little further.'
- (9) *zoals u kunt zien, is de tweede optie de goedkoopste*
(23, EP-High, female, age 25, university student)
'As you can see, the second option is the cheapest.'
- (10) *ook omdat u op verschillende afdelingen bezig bent, zal dat wel een groot gedeelte ook zijn*
(108, EP-High, female, age 42, leader of a daycare center)
'also, because you will work in different departments, that will be a large part (of the job).'

3.2.3 Sentence-initial infinitival clauses

Only 13 instances of sentence-initial infinitival clauses (*om* + infinitive), preceding a main clause were found. They were produced by 4 EP-Low speakers and 8 EP-High speakers (9 instances). See (11)–(13) for examples.

⁵ The word *dan* ('then') functions as a resumptive pronoun, so that the matrix clause can be optimally processed under the principle of Minimize Domains.

- (11) *en om nou naar Den Haag op de fiets te gaan, dat zie ik gewoon niet zo zitten*
 (24, EP-High, female, age 25, project assistant at an institute of higher education)
 ‘To go by bike to The Hague, that doesn't really work for me.’
- (12) *om meteen te beginnen met de eerste manier die ziet u hier achter [sic]⁶ u op het scherm*
 (25, EP-High, male, age 28, product controller environmental management)
 ‘To start immediately with the first solution, that is what you see on the screen behind you’
- (13) *om beter u van dienst te kunnen zijn, kunnen we misschien beter een afspraak maken*
 (71, EP-Low, male, age 25, butcher)
 ‘To serve you better, we should perhaps make an appointment.’

3.2.4 Sentence-initial clauses without an overt conjunction

In both English and Dutch (and German), it is possible to omit an overt conjunction in a conditional subclause. In English, this omission leads to subject-verb inversion. The same is true in Dutch (and German), meaning that this type of subclause has no verb-final word order. With respect to syntactic parsing, this type of conditional clause creates more ambiguity (competition with yes-no questions) than when the clause begins with an overt conjunction. In terms of the Competition Model of Bates & MacWhinney (1989), there is initially (i.e., after hearing the finite verb and the subject) more cue competition in this conjunction-less clause type, which creates the danger of misassignment or non-assignment, thus violating Hawkins' principle of Maximize On-line Processing.

Of this type of conditional clauses, preceding a main clause, 35 instances were produced, 10 by EP-Low and 25 by EP-High subjects. As in the case of the *als* clauses (Section 3.2.1), there were more instances in the two argumentative tasks (27 instances) than in the two descriptive tasks (8 instances). Examples of this clause type are given in (14)–(16).

- (14) *mocht u nog meer willen weten, dan horen wij dat graag van u*
 (120, EP-High, male, age 52, head of administration unit)
 ‘Should you want to know more, then we will be pleased to hear.’

⁶ The screen was not behind the audience but behind the speaker.

- (15) *willen wij het onder de grond gaan doen, dan zullen de kosten voor ons als supermarkt bijna één miljoen euro bedragen*
 (75, EP-High, male, age 52, outlet manager)
 ‘Should we want to build it underground, then the costs for us as supermarket will amount to one million Euro.’
- (16) *werk je dichtbij dan ga je lekker met de fiets*
 (58, EP-High, male, age 30, account executive)
 ‘(If) you work nearby, then you can travel by bike.’

Of the 35 conjunction-less conditional clauses, 9 contained a past-tense form of the verb *moeten* or *zullen*, typical of this type of conditionals, often of the formulaic type, as in (14).

In Dutch, a conjunction-less fronted subclause need not express a condition. It can also, together with the following main clause, express a contrast or comparison between two propositions. In English, the subclause in such sentences begins with the overt conjunction *while*, as in the fictitious example (17).

- (17) *Heeft de auto het voordeel dat je voor de deur in kunt stappen, de trein heeft het voordeel dat je niet in een file hoeft te staan.*
 ‘Whereas the car has the advantage that you can step into it at your front door, the train has the advantage that you need not be in a traffic jam.’

Sentences of this type were not found in the corpus although the two argumentative speaking tasks (the Transportation task and the Car Park task) gave room for statements, comparing two or more alternative state of affairs.

3.2.5 *Wh-cleft sentences*

In Germanic and Romance languages, *it-cleft* sentences and *wh-cleft* sentences allow the speaker to highlight a constituent from a simple, undivided sentence (e.g., 18a, 19a, 20a) by pulling it out, splitting the sentence into a biclausal sentence (18b, 19b, 20b). In the case of *it-clefts*, this focusing process results in a matrix clause (containing the highlighted element) followed by a subordinate complement clause (18b). In the case of *wh-clefts* (in the literature also called pseudo-cleft sentences), the first clause is a relative clause and the second clause is the matrix clause beginning with a copula (19b, 20b).

- (18) a. *The teacher spoke first*
 b. *It was the teacher who spoke first*
- (19) a. *I asked patience from him*
 b. *What I asked from him was patience*

- (20) a. *On the day after the operation I asked patience from him*
 b. *What I asked from him on the day after the operation, was patience*

This fronting of information in the form of extraposition comes at the price of increased grammatical complexity (ratio of non-terminal to terminal nodes). Given the speaker's decision to highlight certain information by clefting a sentence and putting the focused constituent in the *wh* clause, the resulting two-clause cleft construction can nevertheless be relatively efficiently processed by speaker and hearer if the matrix clause is short, as in (18b) and (19b), conforming to the principle of Minimize Domains. Processing efficiency of the matrix clause decreases when the number of its constituents increases (compare (19b) with (20b)). Thus, there is a tension resulting from a trade-off between structural complexity and ease of processing, typical of verbal information processing (Hawkins 2004: 25).

In the M&H corpus, only one instance was found of an *it*-cleft sentence (21), produced by an EP-High speaker; 18 instances of *wh*-cleft sentences were found, produced by 4 EP-Low subjects and 13 EP-High subjects (14 instances). Examples (22) and (23) have rather short *wh* clauses.

- (21) *het zijn vooral mannen die werkeloos zijn*
 (30, EP-High, male, age 24, student at a college of higher education)
 'It is mostly men who are unemployed.'
- (22) *maar waar het voornamelijk om gaat, is dat het aantal parkeerplaatsen, die daarbij vrijkomen op zich relatief weinig is*
 (12, EP-High, male, age 31, lawyer)
 'But what the main point is, is that the number of parking spaces which will thereby become available, is relatively small.'
- (23) *wat me opviel, was dat vooral de [//] het percentage wat van de vrouwen die werkloos zijn een stuk hoger is*
 (21, EP-High, male, age 29, technical draughtsman)⁷
 'What struck me, was that in particular the [//] the percentage of women who are unemployed is a lot higher.'

The *wh*-cleft sentences were rather short, in line with Hawkins' principle of Minimize Domains. Some appeared to be of a rather formulaic type containing common collocations, as in (22) and (23). Only the following two *wh*-cleft sentences were rather long, violating the Minimize Domains principle.

⁷ The speaker pauses briefly after *de* and then self-corrects herself, replacing *de* by *het* (which matches the gender of the noun *parkeergarage*).

- (24) *maar wat toch een hele doorslaggevende factor blijft voor de gemeente en zowel ook voor ons, is dat met name de kosten op het dak veel lager zijn*
(73, EP-High, female, age 48, teacher)
‘But what remains a very decisive factor for the city and also for us, is that in particular the costs [of a parking lot] on the roof will be a lot lower.’
- (25) *wat ook niet geheel onbelangrijk is en toch wel dertig procent van de tijd in beslag gaat nemen, dat is dat u ingezet zult gaan worden op de eerste hulp*
(120, EP-High, male, age 52, team leader salary administration)
‘What is also not entirely unimportant and yet will take up over thirty percent of the time, that is that you will be deployed at the First Aid unit.’

3.3 Center-embedded clauses

Embedding a clause within another clause (i.e., after the beginning⁸ and before the end of the matrix clause) is generally considered a factor increasing the processing load of the matrix clause, violating the principle of Minimize Domains. In the M&H corpus, 35 instances of embedded clauses of this type were found (17 EP-Low, 18 EP-High), produced by 11 EP-Low and 10 EP-High subjects. The word length of these embedded clauses was generally rather short (Mean = 5.9, Mode = 4). In some cases, the speaker repeated the beginning of the matrix clause (or a part of it) after completion of the embedded clause. This is likely to help both speaker and listener in processing the matrix clause. Examples involving center-embedding are shown in (26)–(28).

- (26) *en ja het vervelende is dat je, wat het nu is, dat je echt afhankelijk bent*
(52, EP-Low, female, age 26, pharmacy assistant)
‘And yes, the unpleasant thing is that you, whatever it is, that you remain dependent.’
- (27) *ik denk dat de oplossing is, dat de mensen maar gewoon moeten accepteren, dat ze, als ze in de stad gaan werken, dat ze in de file staan en dat er geen wegen moeten worden gebouwd, dat de mensen uiteindelijk moeten besluiten om met het openbaar vervoer te gaan*
(64, EP-High, male, age 35, coach of disabled people)
‘I think that the solution is that people simply have to accept that they, if they start working in the city, that they will be in traffic jams and that no roads should be constructed, that eventually people have to decide to use public transportation.’

⁸ In this coding, instances were excluded when the beginning of the matrix clause consisted only of a subordinate conjunction, without any other constituent, and if this conjunction was repeated.

- (28) *dus kan u op uw gemak, als u even hiernaartoe komt, op uw gemak met ons praten wat de tijden zijn van het [I] van het werk*
 (155, EP-Low, female, age 66, office cleaner)
 ‘Thus, you can at your leisure, if you come to this place (hospital), at your leisure talk with us about working hours.’

Two-word clauses embedded within another clause, such as *vind ik* (‘I think’; $n = 1$), *geloof ik* (‘I believe’; $n = 0$), *denk ik* (‘I think’; $n = 21$), *wie weet* (‘who knows’; $n = 0$), and *zoals gezegd* (‘as [I] said’; $n = 1$) were excluded from this count. These expressions appeared more frequently at the end of a sentence. Nine instances of embedded clauses started with *als* (‘if’). Five instances started with *zoals* (‘like/as if’). Two instances were infinitival clauses. Some instances had a main-clause structure.

3.4 Relative clauses

Relative clauses in Dutch begin with *die*, *dat*, *waar* as well as some less common relative conjunctions. The most frequent relative conjunction is *die*, which can function as a subject argument or object argument of the predicate. All instances of subject relative clauses and object relative clauses with *die* were located and their clause length counted, including *die* itself. Note that the minimal word length of a subject clause is 2 (e.g., *de kinderen die binnenkomen*; ‘the children that enter’), while the minimal word length of object relative clauses is 3 (e.g., *de opties die we hebben*; ‘the options that we have’). Table 3 provides the breakdown of these subject and object sentences by EP. The proportion of object clauses to subject clauses was 43% and 42%, for the EP-Low and EP-High subcorpora respectively; the Chi Square test produced no significant result.

An ANOVA was conducted on the number of relative clauses (beginning with *die*, *dat*, or *waar*). A main effect of EP was obtained ($\eta^2_p = .122$; $F [1,92] = 12.767$; $p = .001$). The mean numbers of relative clauses were 3.4 (EP-Low) and 6.1 (EP-High). No Age effect was obtained and the Age x EP interaction was not significant either. There were 8 EP-Low subjects and there was 1 EP-High subject who produced no relative clauses at all.

The figures in Table 3 show that the relative clauses were short. Of all 243 relative clauses beginning with *die*, only 12 had a length of 10 words or more. Of all 105 relative clauses beginning with *waar*, 11 had a length of 10 words or more. Only 23 instances of relative *dat* were observed (12 EP-High and 11 EP-Low; 15 subject clauses and 8 object clauses). The mean length of these 23 clauses was 4.8 words ($SD = 1.8$).

Table 3. Subject- and Object-Relative Clauses (RC) by EP

| Type of RC | EP-Low (<i>n</i> = 52) | | | EP-High (<i>n</i> = 46) | | |
|---|---------------------------------|---------------------------|----------------------------------|---------------------------------|---------------------------|----------------------------------|
| | Subcorpus of 36,512 word tokens | | | Subcorpus of 38,902 word tokens | | |
| | No. of speakers | No. of instances observed | Clause length in words Mean (SD) | No. of speakers | No. of instances observed | Clause length in words Mean (SD) |
| Subj. clauses with <i>die</i> | 28 | 65 | 5.5 (2.2) | 46 | 106 | 5.4 (2.3) |
| Obj. clauses with <i>die</i> | 16 | 28 | 5.4 (2.1) | 25 | 44 | 5.0 (2.2) |
| Clauses with loc. <i>waar</i> | 18 | 22 | 5.4 (2.8) | 23 | 36 | 6.7 (3.1) |
| Clauses with <i>waar</i> + non-adj. prep. | 17 | 22 | 6.2 (2.4) | 17 | 25 | 5.8 (1.7) |
| Clauses with <i>waar</i> + adj. preposition | 20 | 30 | 7.3 (4.2) | 27 | 58 | 7.3 (3.3) |
| Total | | 167 | 5.9 (2.8) | | 269 | 5.9 (2.7) |

Note: loc. = locative; adj. prep. = adjacent preposition

Thus, although experimental studies have shown that subject relative clauses are easier to parse for both speaker and listener than object relative clauses (McDaniel et al. 2015), both types occurred equally frequently in the M&H corpus, possibly because of their short word-length.

3.5 Noun phrases with pre-nominal participle phrases

Noun phrases (NPs) are constituents headed by a noun (N). While hearing an NP, it is important for the listener to establish as soon as possible which word is the noun, heading the NP. In Germanic languages, an article followed by an adjective (e.g., *a little*) provide high cue validity (Bates & MacWhinney 1989) to the hearer that these two words form the beginning of a NP and that the next word might well be the head noun (e.g., *a little boy*). In Dutch, as in most other Germanic languages, the head noun can be preceded by a phrase headed by a present or past participle or by an infinitival constituent. Particle phrases of this type make it more difficult for the hearer to parse the NP as a whole because of the distance between DET and N (Hawkins' principle of Minimize Domains). This typically occurs in information-dense formal documents or in literary fiction, as the following examples illustrate. (In English, such adjuncts, when consisting of more than two words, are usually placed after the head noun, as the

translations show, which make them easier to parse.) Arguably, none of the examples from fiction given here (short stories written by Simon Carmiggelt, 2013) are of the formulaic type. They strike the reader as novel, creative expressions, sometimes with unusual combinations, as in (32).

- (29) *een door honger uitgeteerde man* (p. 247)
a by hunger wasted man
‘a man wasted away by hunger’
- (30) *een in het familiealbum geplakte foto* (p.240)
a in the family.album glued photo
‘a picture glued into the family album’
- (31) *een blijkbaar tot zijn gevolg behorende jongen* (p. 248)
a evidently to his retinue belonging boy
‘a boy evidently belonging to his retinue’
- (32) *een tot het uiterste bedwongen das van reine zijde* (p. 262)
a tot the utmost controlled scarf of pure silk
‘a scarf of pure silk, controlled to the limit’

Pre-nominal participle phrases are also quite normal in official documents. The following two examples are taken from the Renkema corpus, a 48,000-word corpus of correspondence between government and parliament 1975-1976.⁹ Example (34) even contains a present-participle phrase embedded in the higher pre-nominal past-participle phrase.

- (33) *de in voorbereiding zijnde wettelijke regeling* (line 182)
the in preparation being legal regulation
‘the statutory regulation in preparation’
- (34) *de door de tot de KVP-fractie horende leden gedane suggestie*
the by the to the KVP-party belonging members made suggestion
‘the suggestion made by the members belonging to the KVP party’ (l. 188)

The 98 speakers in the M&H corpus of spoken Dutch produced only three instances of (rather short) pre-nominal participle phrases, presented in (35)–(37).

⁹ This small corpus has been added to the so-called Eindhoven Corpus (1976), and is currently available from <http://tst-centrale.org/nl>.

- (35) *het*¹⁰ *nieuw aan te leggen supermarkt*
 the new to construct supermarket
 (55, EP-Low, female, age 29, activity coach)
 ‘the soon to-be-built supermarket’
- (36) *theoretisch gerichte opleidingen*
 theoretically oriented programs
 (57, EP-Low, female, age 31, customer-support employee)
 ‘theoretically-oriented programs’
- (37) *een goed geoliede machine*
 a well oiled machine
 (79, EP-Low, male, age 38, restorer)
 ‘a well-oiled machine’

Westendorp (2016) conducted an investigation into the use of pre-nominal present or past participles in the speech of members of parliament, contained in the Corpus of Spoken Dutch (CGN). She randomly selected 20 sessions of recorded speech in parliament, looking only at speech produced by members of parliament with a higher level of education (88 speakers, 40,000 words). Only the following four instances of prenominal adjuncts with present or past participles were found, one of which (39) being in fact ungrammatical.

- (38) *een redelijk goed uitvoerende, goed werkende wet*¹¹ (CGN fn 000163)
 ‘a fairly well executing, well-functioning law’
- (39) *de breed door veel partijen ingediende motie* (CGN fn 000167)
 ‘the motion brought forward widely by many parties’
- (40) *de via de Flexwet toegenomen mogelijkheden voor een tijdelijk dienstverband* (CGN fn 000242)
 ‘the possibilities, increased by the Flex law, for temporary employment’
- (41) *de door de opsteller van het amendement gevreesde vermenging*
 ‘the fusion, feared by the author of the amendment’ (CGN fn 000212)

Thus, even highly educated members of parliament, being exposed daily to heavy pre-nominal constructions in official documents, such as draft texts of bills and letters from the government, seem to rarely use these constructions in oral production, while participating in debates in parliament.

¹⁰ This expression contains an article/gender error. The definite article for *supermarkt* is *de*, not *het*.

¹¹ This utterance probably contains a self-correction: *goed uitvoerende* (‘well executing’) is replaced by *goed werkende* (‘well-functioning’).

3.5.1 Present participles without modifiers

If we exclude present participles that function separately as normal adjectives (e.g., *een uitdagende baan*, ‘a challenging job’), we found only 15 instances of present participles in the M&H corpus (4 EP-Low, 11 EP-High), 10 of which in pre-nominal position (1 instance produced by 1 EP-Low speaker; 9 instances produced by 7 EP-High speakers). Examples are provided in (42) to (44).

- (42) *de binnenkomende kinderen*
(4, EP-High, male, age 27, first officer on a ship)
‘the entering children’
- (43) *deze groeiende woonkern*
(14, EP-High, male, age 31, ERP technical application administrator)
‘this growing residential area’
- (44) *slaande portieren*
(172, EP-High, male, age 60, unit head in the police forces)
‘slamming doors’

3.6 Complex verb phrases

Wondering whether EP-Low speakers produced sentences with verb clusters, we examined the incidence of verb clusters consisting of one finite verb form and two or three non-finite verb forms. In the three-verb clusters, 77 different combinations of a finite Aux form with a non-finite AUX (including passive auxiliaries *worden* and *zijn*) were obtained. The most frequent combination (109 instances) consisted of *moeten* + [past participle] + *worden* (‘should be + [past participle]’). Table 4 provides the total numbers of verb clusters observed in the EP-Low and EP-High subject groups. ANOVAs did not produce Age or EP effects in any of the four subtypes nor on all clusters together. On average, EP-Low and EP-High subjects produced 3.3 (SD = 3.2) and 4.8 (SD = 3.5) verb clusters, respectively. There were 4 EP-Low subjects, and there was 1 EP-High subject who produced no verb clusters of the types just mentioned at all.

Table 4. Number of instances (Matches) of 3-verb and 4-verb clusters observed in the speech of EP-Low and EP-High speakers and number of speakers producing these instances

| | EP-Low ($n = 52$) | | EP-High ($n = 46$) | | All subjects ($N = 98$) |
|--------------------------------------|---------------------|--------------------|----------------------|--------------------|------------------------------|
| | No. of matches | No. of subjects | No. of matches | No. of subjects | |
| 3-verb clusters | | | | | |
| AUX + AUX + INF | 79 | 35 | 117 | 37 | 196 |
| AUX + PP + (<i>te</i>) <i>zijn</i> | 2 | 2 | 5 | 4 | 7 |
| AUX + PP + <i>worden</i> | 85 | 37 | 84 | 38 | 169 |
| 4-verb clusters | | | | | |
| | 7 | 5 | 15 | 10 | 22 |
| Total verb clusters | 173 | | 221 | | 394 |

Note: AUX = auxiliary verb; INF = infinitive; PP = past participle

3.7 *Passive voice*

We counted all predicates rendered in the passive voice. No significant Age or EP differences were obtained but there was a trend ($p = .07$) for EP-High subjects ($M = 3.4$; $SD = 1.8$) to produce more passives than EP-Low subjects ($M = 2.5$; $SD = 2.1$). Seven subjects (all EP-Low) did not use any passives at all. The stimulus information in the Car Park task provided subjects with the verbal cue *bomen kappen* ('to cut trees'), as a feature of one of the car park scenarios. The corpus contains 170 utterances expressing the need to cut trees. Of these 170 cases, a majority of 102 (60%) were rendered in the passive voice, in 29 cases (17%) an active form was used, in 17 cases (10%) an intransitive form was used, and in 22 cases (13%) the stimulus *bomen kappen* was repeated. The differential use of these four means of expression was not significantly mediated by EP. The following examples illustrate each of these four ways of formulation: passive (45), active (46), intransitive (47), and infinitival construction (48).

- (45) *er moeten bomen gekapt worden*
(166, EP-High, female, age 51, activity coach)
'Trees have to be cut.'
- (46) *dan moeten we veel bomen kappen*
(20, EP-High, female, age 25, elementary-school teacher)
'Then we have to cut many trees.'
- (47) *dan zullen de bomen weg moeten*
(72, EP-Low, female, age 20, administration assistant)
'Then the trees have to go away.'

- (48) *en daarnaast, ja, bomen kappen*
 (6, EP-High, male, age 29, bookkeeper)
 ‘And in addition, yes indeed, cut trees.’

4 Discussion

In an attempt to begin to establish the productive knowledge of morpho-syntax (of Dutch), shared by all adult native speakers, this pilot study examined the extent to which several patterns of Dutch syntax are present in the speech of 98 adult native speakers of Dutch, differing in age and level of education and profession (EP), who performed four speaking tasks. The study was guided by the following question: Which syntactic patterns (of the ones under examination) are acquired (in all likelihood) by all native speakers? Most of the patterns under investigation were chosen and examined from the perspective of Hawkins’ (2004, 2014) efficiency theory, supplemented, where necessary, with the notion of cue competition in Bates & MacWhinney’s (1989) Competition Model. The findings can be summarized as follows.

Clause length. In a subcorpus (30 participants), it was observed that average clause length (including main clauses) was rather short (7.1 words; Section 3.1). Fronted subclauses were slightly shorter (6.8 words; Section 3.2). Embedded clauses (Section 3.3) and relative clauses (Section 3.4) were even shorter (5.9 each). Clause length was not reliably associated with speakers’ age or EP. Thus, clause length was rather short, given the argumentative and descriptive nature of the tasks (Biber et al., 2002).

Patterns used by a large majority of both EP-Low and EP-High speakers. Relative clauses were produced by 43 of the 52 EP-Low subjects and by 45 of the 46 EP-High subjects (Section 3.4). Fronted conditional clauses with *als* were produced by 38 EP-Low subjects and by 36 EP-High subjects (Section 3.2.1). Verb clusters consisting of at least three verbs were produced by 48 EP-Low speakers and by 45 EP-High speakers. The passive voice was used by all EP-High subjects and by 45 of the 52 EP-Low subjects.

Patterns used by a minority of both EP-Low and EP-High speakers. Fronted infinitival clauses were produced by only 4 EP-Low and 8 EP-High subjects (Section 3.2.3), while *wh*-cleft sentences were produced by 4 EP-Low and 13 EP-High subjects (Section 3.2.5). Center-embedded clauses (containing a finite verb) were produced by 11 EP-Low and 10 EP-High subjects (Section 3.3).

Patterns used by other proportions of EP-Low and EP-High speakers. Fronted conjunction-less conditional clauses were produced by 10 EP-Low and 25 EP-High subjects (Section 3.2.4). Adverbial clauses (not starting with *als*)

preceding a main clause were produced by 23 EP-Low and 22 EP-High subjects (Section 3.2.2).

Patterns almost or entirely absent. Pre-nominal participle phrases were produced only by 3 EP-Low subjects (Section 3.5), while pre-nominal present participles without modifiers were produced by 1 EP-Low subject and 7 EP-High subjects (Section 3.5.1). One instance was found of an *it*-cleft sentence (Section 3.2.5). No instances were found of fronted subject or object clauses. No examples were found of fronted conjunction-less clauses that express a comparison or contrast (Section 3.2.4).

The findings suggest that adult native speakers of Dutch produce subordinate-clause patterns, such as complement clauses, adverbial clauses, and relative clauses at their base position in the matrix clause (to the right of the verb, NP, etc.), but that the phenomenon of clause fronting is common only with respect to some adverbial clause types, such as conditional clauses beginning with *als* ('if'). Verb clusters with modal auxiliaries and the passive voice also appear to belong to shared grammatical knowledge. In contrast, the findings suggest that *it*-cleft sentences, *wh*-cleft sentences, the fronting of conjunction-less conditional clauses, fronting of infinitival clauses, and embedding a clause within another clause may not belong to shared grammatical knowledge. Further empirical research is needed, looking at grammar and lexicon simultaneously, as suggested by Gries & Ellis (2015). Obviously, we propose our suggestions concerning the contents of shared linguistic knowledge with caution, given the limitations of our study, a major limitation being that non-occurrence of a structure in a small corpus does not necessarily mean that the speakers would not be able to produce the structure. In Hulstijn (submitted), I address the fundamental conceptual-methodological issues underlying this fact.

Furthermore, the findings appear to support Hawkins' (2004, 2014) efficiency principles of Minimize Domains and Maximize On-line Processing: less efficient patterns, such as clause fronting and clause embedding, occur, if at all, when the fronted or center embedded clauses contain only a few words and/or are of the formulaic type, thus hardly burdening the speaker's processing memory.

Age and EP. Age and EP did not, or hardly, matter with respect to the syntactic patterns under investigation. However, Mulder & Hulstijn (2011) had observed pronounced Age and EP associations of the same 98 individuals in other domains: Participants of older age performed more poorly than subjects of younger age on all measures of lexical fluency and lexical memory, but better on the vocabulary test. EP-High subjects scored significantly higher than the EP-Low subjects on the vocabulary measures and on the auditory word-span task. In the four speaking tasks, they talked longer (producing more words), their responses were communicatively more successful, and they made fewer

grammatical errors per T-unit in the two formal tasks. The association between EP and grammatical patterns in speech production deserves further investigation.

Discourse type effects. The four speaking tasks differed in formality and discourse type, in a 2 x 2 fashion. With respect to the syntactic patterns under investigation, no associations with formality were observed. In comparison to the descriptive subcorpus, the argumentative subcorpus contained more instances of fronted adverbial clauses (mostly conditional propositions, typical in constructing an argumentation). This illustrates how discourse type may bias (positively or negatively) the production of certain patterns. For instance, had participants been asked to perform a narrative task also, we might have observed instances of predicative adjuncts, characterizing the (temporal) state in which a protagonist finds herself/himself, as in *Vrolijk lopen Kikker en Eend door het bos* ('Cheerful, Frog and Duck are walking through the woods') (Velthuijs 2010). The production of such patterns might be mediated by literacy (e.g., listening to the reading aloud of children's books).

Literacy. One might argue that the study conducted here does not really address the question of how small or large BLC might be because BLC is defined in terms of listening and speaking while the participants in this study all had learned to read and write in school and thus had been exposed to written input. Indeed, it should be acknowledged that all participants were literate, although not to the same extent. It is virtually impossible, however, to find adult illiterate native speakers of magnasyntax languages (see Section 1) because of compulsory primary-school education in practically all countries concerned (cf. Huettig & Mishra 2014). However, it is not unrealistic to assume that there exist many EP-low adults who, after their school years, need not read much, avoid reading books, and avoid listening to the speech of EP-high people (e.g., on television). Thus, for instance, in the case of a 60-year old EP-low man who has been carpenter during adulthood, most of his cumulative language input may have been in the oral modality and only a small proportion in the written modality.

Concluding remark. As said in the Introduction, BLC Theory is not a theory of language acquisition, but it is presented as a tool to help make empirical some truly fundamental questions of language acquisition and language use, by proposing that some elements of the grammar of E-language ('magnasyntax', Section 1) are acquired (and somehow represented in I-language) by all native speakers (Basic Language Cognition), while other elements are not (Extended Language Cognition). Grammatical constructs, such as NP, VP, PP, N, V, P, subject/object, agent/patient, and patterns such as adverbial clauses, relative clauses, *wh*-cleft sentences, etc. are abstract. This

implies that their presence or absence in people's mental grammars¹² cannot be directly observed (see Hulstijn (submitted) for an extensive discussion of this fundamental issue). Their claimed or assumed existence in I-language has to be inferred from observable speech-production data. The present pilot study made a modest first attempt to delimit the borderlines of syntactic knowledge acquired by all native speakers.

5 References

- Austin, P.K. 2008. *1000 Languages: The Worldwide History of Living and Lost Tongues*. London: Thames & Hudson.
- Bates, E. & B. MacWhinney. 1989. Functionalism and the competition model. In B. MacWhinney & E. Bates (eds.), *The Crosslinguistic Study of Sentence Processing*. Cambridge: Cambridge University Press, 3–73.
- Bernstein, B. 1959. A public language: Some sociological implications of a linguistic form. *British Journal of Sociology* 10: 271–276.
- Bernstein, B. 1962. Social class, linguistic codes, and grammatical elements. *Language and Speech* 5: 221–240.
- Biber, D., S. Conrad, R. Reppen, P. Byrd & M. Helt. 2002. Speaking and writing in the university: A multidimensional comparison. *TESOL Quarterly* 36: 9–48.
- Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow, UK: Pearson Education.
- Bybee, J.L. 2013. Usage-based theory and exemplar representations of constructions. In T. Hoffmann & G. Trousdale (eds.), *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press, 1–24.
- Carmiggelt, S. 2013. *Gedrukt: Verhalen*. (Selection of the most successful columns of Simon Carmiggelt, 1913–1987). Amsterdam: Van Oorschot.
- Centraal Bureau voor de Statistiek [Statistics Netherlands]. 2001. *Standaard Beroepenclassificatie* [Standard classification of professions]. Voorburg/Heerlen: Centraal Bureau voor de Statistiek.

¹² The term ‘mental grammar’ is loosely used here to refer to the largely implicit knowledge which a native speaker has of (i) the grammatical patterns and (ii) the frequency of occurrence of these patterns in her/his language, in so far she/he has been exposed to it (i.e., someone's individual input). Note that the definition of BLC, given in the Introduction, does not separate knowledge from use. For that reason, the more general term ‘cognition’ is used in the theory's main constructs BLC and HLC. For a theory of language acquisition and use, there is no a priori need to postulate rules (that operate on members of abstract categories) in the mental grammar of native speakers. In other words, with the term ‘mental grammar’, we do *not* refer to a tacit, underlying “system of principles and rules” (Haegeman 1991: 9), as in the generative tradition.

- Chomsky, N. 1986. *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger.
- Chaudron, G. 1988. *Second Language Classrooms: Research on Teaching and Learning*. Cambridge: Cambridge University Press.
- Comrie, B., S. Matthews & M. Polinsky. 2003. *The Atlas of Languages* (revised edition). New York: Facts On Line, Inc.
- Corpus Gesproken Nederlands (Spoken Dutch Corpus). 2004.
<http://lands.let.ru.nl/cgn/ehome.htm>
- Ferreira, F. & P.E. Engelhardt. 2006. Syntax and production. In M.J. Traxler & M.A. Gernsbacher (eds.), *Handbook of Psycholinguistics* (2nd edition). London: Academic Press, 62–91.
- Foster, P., A. Tonkyn & G. Wigglesworth. 2000. Measuring spoken language: A unit for all reasons. *Applied Linguistics* 21: 354–375.
- Geeraerts, D. 2015. *Cognitive Sociolinguistics and Modularity*. Plenary lecture at the 6th Conference of the Association Française de Linguistique Cognitive (AFLICO), Université de Grenoble, 26 May 2015.
- Gries, S.T. & N.C. Ellis. 2015. Statistical measures for Usage-based linguistics. *Language Learning* 65: 228–255.
- Haegeman, L. 1991. *Introduction to Government and Binding Theory*. Oxford: Blackwell.
- Haeseryn, W., K. Romijn, G. Geerts, J. de Rooij & M.C. van den Toorn. 1997. *Algemene Nederlandse Spraakkunst* (2nd edition). Groningen: Martinus Nijhoff Publishers.
- Hawkins, J.A. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
- Hawkins, J.A. 2014. *Cross-linguistic Variation and Efficiency*. Oxford: Oxford University Press.
- Huetting, F. & R.K. Mishra. 2014. How literacy acquisition affects the illiterate mind – A critical examination of theories and evidence. *Language and Linguistics Compass* 8: 401–427.
- Hulstijn, J.H. 2011. Language proficiency in native and nonnative speakers: An agenda for research and suggestions for second-language assessment. *Language Assessment Quarterly* 8: 229–249.
- Hulstijn, J.H. 2015. *Language Proficiency in Native and Non-native Speakers: Theory and Research*. Amsterdam: John Benjamins.
- Hulstijn, J.H. Submitted. A usage-based, probabilistic view on the mental grammars of adult native speakers and end-state nonnative speakers: Conceptual and methodological challenges for measuring language proficiency. Submitted to *Language Learning*.
- Hunt, K. 1965. *Grammatical Structures written at Three Grade Levels*. Champaign, IL: National Council of Teachers of English.

- Jansen, F. 1981. *Syntaktische constructies in gesproken taal*. (Syntactic constructions in spoken language). PhD dissertation, Leiden University. Amsterdam: Huis aan de Drie Grachten.
- Klooster, W.G. 2001. *Grammatica van het hedendaags Nederlands: Een volledig overzicht*. Den Haag: SDU uitgevers.
- Leech, G. 2000. Grammars of spoken English: New outcomes of corpus-oriented research. *Language Learning* 50: 675–724.
- Macaulay, R.K.S. 1991. *Locating Dialect in Discourse: The Language of Honest Men and Bonnie Lassies in Ayr*. New York: Oxford University Press.
- Macaulay, R.K.S. 2013. Grammar and social class. In C.A. Chapelle (ed.), *The Encyclopedia of Applied Linguistics*. Malden, MA: Blackwell.
- McDaniel, D., C. McKee, W. Cowart & M.F. Garrett. 2015. The role of the language production system in shaping grammars. *Language* 91: 415–441.
- Miller, J. & R. Weinert. 1998. *Spontaneous Spoken language: Syntax and Discourse*. Oxford: Clarendon Press.
- Mulder, K. & J.H. Hulstijn. 2011. Linguistic skills of adult native speakers, as a function of age and level of education. *Applied Linguistics* 32: 475–494.
- Ong, W. 1982. *Orality and Literacy: The Technologizing of the World*. London: Methuen.
- Pickering, M.J. & S. Garrod. 2013. An integrated theory of language production and comprehension. *Behavioral and Brain Sciences* 36: 329–392.
- Pickering, M.J. & S. Garrod. 2014. Neural integration of language production and comprehension. *Proceedings of the National Academy of Sciences* 111(43): 15291–15292.
- Pickering, M.J. & R.P.G. van Gompel. 2006. Syntactic parsing. In M.J. Traxler & M.A. Gernsbacher (eds.), *Handbook of Psycholinguistics* (2nd edition). London: Academic Press, 455–503.
- Velthuijs, M. 2010. *Kikker en z'n vriendjes* (Frog and his friends). Amsterdam: Leopold.
- Westendorp, M. 2016. *Noun Phrase Complexity in Dutch Parliamentary Debates*. Unpublished paper, MA Linguistics program, University of Amsterdam.

Jan H. Hulstijn
Department of Linguistics
University of Amsterdam
Spuistraat 134
1012 VB Amsterdam
The Netherlands