On the automaticity of language processing

Robert J. Hartsuiker and Agnes Moors

Abstract

People speak and listen to language all the time. Given this high frequency of use, it is often suggested that at least some aspects of language processing are highly overlearned and therefore occur "automatically". Here we critically examine this suggestion. We first sketch a framework that views automaticity as a set of interrelated features of mental processes and a matter of degree rather than a single feature that is all-or-none. We then apply this framework to language processing. To do so, we carve up the processes involved in language use according to (a) whether language processing takes place in monologue or dialogue, (b) whether the individual is comprehending or producing language, (c) whether the spoken or written modality is used, and (d) the linguistic processing level at which they occur, that is, phonology, the lexicon, syntax, or conceptual processes. This exercise suggests that while conceptual processes are relatively non-automatic (as is usually assumed), there is also considerable evidence that syntactic and lexical lower-level processes are not fully automatic. We close by discussing entrenchment as a set of mechanisms underlying automatization.

1. Introduction

In *The modularity of mind,* Fodor (1983) quoted Merrill Garrett's view on the automaticity of sentence parsing during language comprehension: "What you have to remember about parsing is that basically it's a reflex". Around the same time, Levelt (1989) discussed the automaticity of language production. Whereas some language production processes, such as determining the message one wishes to convey and checking one's speech for errors and other problems were considered to be non-automatic, all the other components were claimed to be largely automatic. According to both authors, some of the core processes in the comprehension and production of language basically run off by themselves, without control and without involvement of working memory. In this chapter we ask to what extent such a view holds.

A priori, language processing seems to be an excellent candidate for a system of automatic processes. Speaking and listening are activities we engage in *routinely* and so are practiced very well. Language processing is difficult to counteract: It is very difficult to listen to someone and deliberately *not* understand what they are saying. Much of language processing is *unconscious*. When we say a word, we retrieve and order its phonemes, but we are usually not aware of the details of phonological encoding. Additionally, language processing proceeds very *quickly and accurately*: we can read about 4 words per second and speak at about half that rate, finding words from a lexicon containing at least 30,000 words, while producing speech errors at a rate of less than 1 per 1000 words (e.g. Levelt, 1989). All of these properties (routinization, specialization, mandatory processing, unconscious processing, and speed) point to a strong degree of automatization.

On the other hand, there are also indications that language processing is not completely automatic. For example, there is a long research tradition investigating the role of working memory in sentence understanding (e.g. Caplan & Waters, 1999; Gibson, 2000; Just & Carpenter, 1992). Although theoretical proposals differ, they have in common that understanding is constrained by a limited memory capacity. Furthermore, Garrett's reflex-like model of parsing breaks down when the reader encounters garden-path sentences such as (1).

(1) The man accepted the prize was not for him.

Arguably, to resolve a temporary structural ambiguity as in (1), processes concerned with error detection and revision need to be involved, which some authors view as non-automatic (e.g. Hahne & Friederici, 1999). Further, second language learners often experience difficulties in the comprehension of complex sentences (e.g. Clahsen & Felser, 2006), which indicates that at least in some stages of language learning sentence comprehension is not fully automatic. To sum up, sentence parsing seems to be constrained by working memory, regularly recruits control processes, and is less automatic for novices (L2 readers) than experts (L1 readers).

The aim of this chapter is to examine the automaticity of language production and comprehension processes, in light of three decades of research in the fields of psycholinguistics and attention and performance since Garrett (1983) and Levelt (1989) made their strong claims. Section 2 specifies what we mean with language processing by carving it up into processes occurring in the individual listener and speaker (i.e., in monologue) as well as processes occurring in dialogue. Within each modality, we divide language processing into a conceptual level, a syntactic level, a lexical level, and a sound level. Section 3 describes the classic view of automaticity and Section 4 describes how the classic view is applied to language processing. Section 5 spells out a contemporary view of automaticity (Moors, 2016) and Section 6 examines the implications of this contemporary view for language processing. We review evidence relevant for the question which components of language processing can be considered to be more or less automatic. In Section 7, we explore the factors (repetition, complexity) and mechanisms (entrenchment, procedure strengthening) underlying automatization (i.e., the evolution of a process towards automaticity). The chapter closes with a discussion in Section 8.

2. Language processing

In looking at automaticity on different levels of language processing, we will rely on the simple framework outlined in Figure 1, which divides language processing according to modality (production or comprehension), linguistic level, and social setting (monologue vs. dialogue).

I. INDIVIDUAL LEVEL

A. COMPREHENSION

1. Interpretation
2. Syntactic level
3. Lexical level
4. Phonological / phonetic level

B. PRODUCTION
1. Conceptualization
2. Formulation: syntactic level
3. Formulation: lexical level
4. Phonological/phonetic/articulatory level

II. SOCIAL LEVEL
1. Conversation

Figure 1. Sketch of representational levels in language processing.

Figure 1 reflects the dominant view in the field that the comprehension and production of language can be thought of as chains of processes through a hierarchy of levels. Thus, production typically begins with the intention to realize a particular communicative goal, and this next leads to the creation or activation of subsequent representations at a conceptual stratum, a lexical stratum, and a phonological stratum (Levelt et al., 1999; but see Pulvermüller, Shtyrov, & Hauk, 2009, for the alternative view that all such representations can be activated in parallel). Note that Figure 1 sketches these processes at a rather rough grain size: If we were to zoom in on a particular representational level, more fine-grained sets of representations and processing steps would appear, but these are outside of the scope of this chapter. It is important to note, finally, that Figure 1 should be taken as a framework to organize our discussion of automaticity of different language processes; we do not commit to any notion that these levels are strictly hierarchically ordered or informationally encapsulated.

3. Classic view of automaticity

Traditionally, cognitive psychology has made a binary dichotomy between automatic and non-automatic processes. For example, several authors (e.g. Schneider & Shiffrin, 1977) characterized automatic processes as ones that are initiated by appropriate stimulus input, without the need for the subject's goal to engage in the process (*unintentional or uncontrolled in the promoting sense*), without demanding attentional capacity (*efficient*) or abundant time (*fast*), and without consciousness (*unconscious*), making it difficult to counteract the process (*uncontrollable in the counteracting sense*). Non-automatic processes are the exact mirror image: They require the subject's goal to engage in the process (*intentional*), they demand attentional capacity (*nonefficient*) and abundant time (*slow*), they are *conscious*, and they are easy to counteract (*controllable in the counteracting sense*). Such a dichotomous view implies that there is coherence among the various features of automatic processes (a process that is unintentional is also efficient, fast, unconscious, and difficult to counteract) as well as among the various features of non-automatic processes (a process that is intentional is also non-efficient, slow, conscious, and easy to counteract). From such a perspective, the job of evaluating the automaticity of language processing boils down to examining for each of the subprocesses of language whether one feature of automaticity applies (e.g. *unconscious*).

4. Application of the classic view on language processing

In broad strokes, it is often assumed that the "higher" or "central" levels of processing (interpretation, conceptualization) are more capacity demanding than the lower levels concerned with syntactic and lexical processing. In the case of dialogue, further capacity-demanding processes would come into play such as modeling the perspective of one's interlocutor. These claims are based on both theoretical and empirical grounds. Thus, Levelt (1989) views conceptualization as a process that is constrained by working memory on the logical ground that it needs unrestricted access to semantic and episodic memory and representations of the current physical and social context as well as a record of the previous discourse. In contrast, subsequent processes of grammatical encoding and lexical retrieval execute well-defined tasks on the basis of limited and specialized sets of representations (i.e., the grammar and the lexicon). Thus, grammatical and lexical processes can be carried out by autonomous specialist systems that do not tap into central resources (and thus have a high degree of automaticity). In support of capacity-constrained conceptualization, Levelt (1989) cites the distribution of disfluencies ("uhs", "ums", repetitions and the like) in spontaneous speech, which reflects phases of conceptualization (many disfluencies) followed by phases of formulation (few disfluencies).

For language comprehension, a similar argument can be made that lexical and syntactic processes can rely on restricted knowledge about words and rules, whereas processes such as mapping sentence structure on meaning and integrating it with discourse could be seen as more central processes. There is, however, considerable controversy in the literature about the degree of automaticity of syntactic and interpretation processes. For instance, Just and Carpenter (1992) proposed a working memory system that would be used for all verbal tasks. This system would be limited in capacity (with individuals varying in this capacity). Sentences would differ in their demand for capacity, with relatively simple sentences, such as ones with a subject-extracted relative clause (2), demanding less capacity than more complex sentences, such as ones with an object-extracted relative clause (3). If a sentence's demand for capacity would exceed a subject's available capacity, comprehension would break down, leading to increased reading times and errors of comprehension. As support for this theory, Just and Carpenter (1992) cited studies that considered effects of sentence complexity, working memory span, and extrinsic memory load.

(2) The student that scolded the teacher got into trouble.
(3) The student that the teacher scolded got into trouble.

Other authors (Caplan & Waters, 1999) agreed that sentence comprehension makes a demand on computational resources, but they rejected the notion of a general verbal working memory. These authors pointed out conceptual and empirical issues with measures of span and with memory load effects. Crucially, they also cited data from brain-damaged patients with spared sentence comprehension despite severe impairment of short-term memory. Rather than a general verbal working memory, these authors argued for a specific verbal memory for 'sentence interpretation' processes which would comprise all comprehension processes including lexical and syntactic processing, assigning meaning, and integrating with the discourse. Importantly, so-called *post*-interpretive processes, such as using sentence meaning to draw an inference, or revising an initially incorrect interpretation (4)[1], would not have

[1] Many readers initially analyze the sentence as an active transitive with *the defendant* as the agent of *examine*, requiring them to revise their interpretation when encountering the *by*-phrase (Ferreira & Clifton, 1986).

access to this specialized resource. Thus, Caplan and Waters (1999) viewed all of the key processes in sentence comprehension as having a strong degree of automaticity: “Processors are thought to be obligatorily activated when their inputs are attended to. They generally operate unconsciously, and they usually operate quickly and remarkably accurately” (p. 93).

(4) The defendant examined by the lawyer stuck with his story.

One might argue that at the level of dialogue, processes related to audience design (i.e., tailoring one’s utterances for the listener) are constrained by capacity, as this would require some modeling of what the listener already knows. Assuming that language production (and in particular conceptualization) itself already makes a demand on capacity, this predicts that speakers are often unsuccessful in audience design. Indeed, Wardlow Lane, Groisman, and Ferreira (2006) observed that in a referential communication task, speakers sometimes ‘leaked’ information that was relevant only for them and not for their interlocutor. Thus, if the speaker sees a big heart and a small heart and the listener can only see the big heart (because of an object occluding the small heart), an utterance like *the big heart* is overspecified and potentially confusing. Nevertheless, speakers often produced such overspecifications and, when explicitly instructed to conceal the hidden information, were in fact *more* likely to overspecify. The latter suggests that keeping the instruction in mind overloaded the capacity for audience design.

To summarize the classic view, in language production, some processes, such as lexical access or phonological encoding are viewed as automatic, whereas other processes, such as conceptualization and audience design in the context of a dialogue are seen as nonautomatic, as they make a demand on central resources. In language comprehension, most processes are seen as automatic, as many sentences can be understood very quickly, efficiently, and obligatory. Yet the observation that sentence comprehension suffers in the case of structural ambiguity or high complexity, has been taken to indicate that sentence comprehension is constrained by a limited capacity of working memory resources, and thus not so automatic after all.

5. Contemporary view of automaticity

Contradicting the assumptions of the classic all-or-none view of automaticity, authors (Bargh, 1992; Moors, 2016; Moors & De Houwer, 2006) have recently argued that the various automaticity features do not necessarily co-occur and so particular processes should be considered as more or less automatic rather than fully automatic vs. non-automatic. This argument is based on studies that showed that automaticity features are often not necessary and sufficient for the presence of other features. With regard to the relation between goals and attentional capacity, for instance, in tasks involving the search for an object among many irrelevant objects, visual attention is often not solely dependent on the subject’s goals (i.e., their task instructions) but also on visual properties of the display, such as the abrupt onset of a distractor stimulus or the presence of a distractor stimulus that shares a visual feature with the target (Theeuwes, 2010). Thus, having goals is not necessary for the allocation of attention. Similar conclusions can be drawn from studies with a paradigm that is often used in psycholinguistics: the visual world paradigm (e.g. Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). In this paradigm, subjects view a display of objects and listen to speech, while eye-movements are monitored. Subjects typically look at objects that are referred to in speech or objects that share properties with these objects. Importantly, overt visual attention (as indicated by eye movements) to objects that are named, or that are similar to named objects, happen irrespective of whether the subject has the goal to find the object (Huettig, Rommers, & Meyer, 2010). Thus, overt visual attention can be driven by high-level properties of a stimulus in the absence of the goal to direct attention to that stimulus.

In addition to the relation between goals and attention, Moors (2016) also analyzed the relation between attention and consciousness and concluded that attention is neither necessary nor sufficient for consciousness, citing studies showing that subjects become aware of unattended stimuli. Similarly, a stimulus does not need to be conscious in order for a goal-directed action to be applied on it. For example, Van Opstal, Gevers, Osman, and Verguts (2010) found that subjects applied the task (same-different judgments) to primes that were presented below the threshold of awareness.

Based on these data and considerations, Moors (2016) proposed an alternative view in which automaticity features are not seen as fixed properties of processes, but rather as factors that jointly influence whether a process will occur or not.[2] The factors included in the automaticity concept (goals, attentional capacity, and time or duration) are just a subset of many factors that can all jointly influence the occurrence of processes, next to factors such as the direction of attention, stimulus intensity, expectations, unexpectedness, novelty, goal relevance, recency, and frequency. Building on the assumption that processes require a minimal input for their occurrence, Moors (2016) proposed that all of the listed factors can influence the quality of representations that form the input of many processes. If this representational quality is high enough it will trigger an (unconscious) process. If the representational quality is even higher, it will trigger a conscious process. Crucially, it is hypothesized that the listed factors influence the representational quality (and hence the occurrence of unconscious and conscious processes) in a compensatory way. For instance, stimulus intensity can trade off with stimulus duration, attention can trade off with stimulus repetition or training, and intention can trade off with recency. Empirical support comes from studies such as that of Tzur and Frost (2007). They manipulated both the luminance of a prime stimulus (i.e., a measure of intensity) and its duration and observed that an increase in luminance can compensate for a decrease in duration with respect to priming and vice versa. Factors that can influence representational quality can be divided in (a) factors related to the current stimulus itself (e.g. duration and intensity, but also un/expectedness, goal in/congruence, and novelty), (b) factors related to prior stimuli (e.g. frequency and recency), and (c) factors related to prior stimulus representations (e.g. goals and expectations). The influences of some of these factors are moderated by attention.

What are the implications of this view for the diagnosis of a process as automatic? Moors and De Houwer (2006) recommended to examine the presence or absence of each automaticity feature separately, and to not draw conclusions about the presence of one feature based on the presence of another. Moreover, given the gradual nature of automaticity, only relative conclusions can be drawn: Rather than classifying a process as automatic or not, a process can be evaluated as being more or less automatic than another one. Moors (2016) went one step further, however,

[2] Note that consciousness is not considered as a primary factor influencing the occurrence of processing, but it is shifted to the output side: the other factors determine whether the process will occur (unconsciously) and whether it will reach consciousness.

arguing that because of the compensatory relation of the factors influencing representational quality (and hence the occurrence of processes), it is not very informative to compare processes with regard to a single feature of automaticity. If a process is found to operate with a smaller amount of attentional capacity than another process, this may either indicate that the former process requires less representational quality than the latter, or that the stimulus input of the first process happened to be more intense, or that their representations were preactivated by an intention, an expectation, or any other type of prior stimulus presentation. Thus, studying the amount of attentional capacity required by a process is not very informative, unless all other factors are kept equal. If that is not possible it is best to map a large network of factors required for processes to operate. In what follows, we adopt Moors' (2016) framework: Thus, we (a) see automaticity as gradual rather than all or none, (b) see automaticity features as a subset of many factors affecting language processes, and (c) assume that such factors can compensate each other, and that it is insightful to consider the joint effect of these many factors on the quality of an input representation.

6. Application of the contemporary view on language processing

In this section, we will recast the issue of automaticity of language processes in terms of the view that whether processes do or do not occur is a function of many driving factors, including but not limited to the factors implied in automaticity features (Moors, 2016). Mapping out the network of these factors can be called a "componential analysis". Rather than attempting to be exhaustive in this analysis, we will present a selected number of examples, with the aim of demonstrating that such an analysis is possible and potentially fruitful. Once again, we divide our discussion according to modality (i.e., comprehension, production, and dialogue).

6.1 Comprehension

In this section we review studies that pertain to the features *efficient* and *uncontrolled*.

6.1.1 Is comprehension efficient?

As sketched above, the theory of a limited capacity for sentence comprehension has had much influence in the field. However, many authors have criticized this view. Thus, tests of verbal working memory capacity, like the reading span test, turn out to be not reliable (Waters & Caplan, 1996). Strikingly, there are patients with severe limitations of working memory capacity with spared sentence comprehension (Caplan & Hildebrandt, 1988). Several authors have further argued that the concept of 'resources' is poorly defined and that a resource theory is impossible to falsify. For instance, Christiansen and McDonald (1999) argued for a connectionist theory in which sentence comprehension is determined by the network architecture and the rate of activation spreading, but not by any working memory system.

On the basis of these arguments, one can question whether a sentence comprehension system that is not capacity-constrained can still account for phenomena that seemingly suggest such constraints. For instance, (5) and (6) (taken from Van Dyke & Johns, 2012) illustrate that sentences are more difficult to understand if there is more material intervening between what Van Dyke and Johns dub a head (*the man*) and its dependents (*was afraid*).

(5) The boy understood the *man was afraid.*
(6) The boy understood the *man* who was swimming near the dock *was afraid.*
(7) It was Tony that Joey liked before the argument began.
(8) It was the dancer that the fireman liked before the argument began.
(9) It was the boat that the guy who lived by the sea sailed / fixed after 2 days.

This is at first glance compatible with a capacity-constrained system, suggesting that there is not enough capacity to hold *the man* active and process the further information, so that the representation of *the man* has decayed by the time *was afraid* needs to be linked to it. However, Van Dyke and Johns (2012) argued that such findings can be explained equally well by the concept of interference. In particular, they argue that the content of an extrinsic memory load affects the quality of sentence processing. For instance, Gordon, Hendrick, & Levine (2002) showed that a memory load consisting of three names (e.g. JOEL – GREG – ANDY) interfered with the processing of a sentence with matched content (7) more strongly than a sentence with mismatched content (8). This suggests that the semantic overlap between memory load and sentence created interference. Van Dyke and McElree (2006) further claimed that interference takes place at the moment when an antecedent is sought for a dependent. These authors presented memory lists such as TABLE – SINK – TRUCK while sentences like (9) were presented, in which the verb (*fixed* or *sailed*) was separated from the antecedent (*boat*) by several words. All items on the memory list can be fixed (as is true of a boat), but in contrast to a boat, none of them can be sailed. If there is interference from memory at the moment an antecedent is sought for the verb, this would predict processing difficulty for *fixed* as compared to *sailed.* This is exactly what was found. On the basis of these and other findings, van Dyke and Johns (2012) proposed an account of sentence comprehension that is *not* constrained by capacity, but that is constrained by stimulus factors (e.g. the presence or absence of multiple potential antecedents for a dependent and cues for retrieval) and person factors (e.g. the ability to exploit retrieval cues).

6.1.2 Is comprehension difficult to control?

Provided we master a language, it seems impossible not to process an attended word or sentence in that language. A striking example is the well-known Stroop effect (e.g. MacLeod, 1991): Speakers are slower to name the ink color of a written word if that word is an incongruent color name (*green* written in red) as compared to a congruent color name (*red* written in red) or a baseline (a series of X's or a non-color word). The Stroop effect demonstrates that the subject comprehends a word even though he/she did not intend to comprehend it and even though comprehension actually hinders the task at hand, which suggests that the subject tried to counteract comprehension (but did not succeed). These arguments plead for the unintentional nature of comprehension processes, as well as for the idea that they are uncontrolled in the counteracting sense, either in the weak sense of being difficult to counteract or in the stronger sense of being impossible to counteract (i.e., obligatory).

Studies showing that comprehension breaks down when the stimulus is severely degraded do not support the strong view that comprehension is obligatory, however. This is illustrated by research on the perception of sine wave speech (Remez, Rubin, Pisoni, & Carrell, 1981). Sine wave speech is artificially synthesized speech consisting of a small number of sine waves with frequencies that are based on the formants of a real utterance. When listeners are not informed that they will hear speech, they typically classify this stimulus as “science fiction sounds”, “computer bleeps”, “music”, and, in a minority of cases, as “human speech”. Strikingly, when subjects are informed that they will listen to speech, they indeed perceive the sounds as speech and they can successfully transcribe a sentence. If they are told they will hear speech and are informed about the sentence, they report hearing each word of that sentence very clearly. This fits a view according to which a stimulus of sufficient quality (and which is attended to) will trigger comprehension. But a stimulus of very poor quality, even if it is attended to, further requires that the listener has *an expectation* to hear speech. Thus, expectation can compensate for a lack of stimulus quality with respect to whether comprehension takes place or not. It is interesting to note that Wiese, Wykowska, Zwickel, and Müller (2012) reported an analogous finding in the domain of gaze: Children and adults followed the gaze of people but not of robots, suggesting that an automatic behavior can be contingent on assumptions about the stimulus or situation.

6.1.4 Summary of studies on comprehension

The studies discussed in this section evaluated language comprehension in terms of two automaticity features. With respect to the feature *efficient*, we provided an alternative to the classic, resource-constrained view. With respect to the feature *uncontrolled* we showed that language is not always obligatorily processed: An attended stimulus of poor quality (as in sine wave speech) is not perceived as speech unless the listener *expects* speech.

6.2 Production

Most people find it difficult to give a speech, and many academics struggle with writing papers, suggesting that at least in some situations and especially in the written mode, language production is far from automatic. The question to what extent language production is automatic has not only been studied from a theoretical, but also from an applied perspective. After all, the question of whether talking on the phone (even with a hands-free kit) affects driving has obvious implications for road safety. Bock, Dell, Garnsey, Kramer, and Tubose (2007) had subjects produce and comprehend language while maneuvering a virtual environment in a driving simulator. Content and difficulty of production and comprehension were carefully matched. Both comprehension and production interfered with driving performance to a similar extent, suggesting that neither modality runs off fully automatically.

Theoretical approaches have focused most on verbal working memory and how this should be carved up. Just and Carpenter (1992) reported that working memory span tasks for comprehension (the reading span task) and for production (the speaking span task) correlated only moderately, and therefore suggested a fractionation of verbal working memory into separate resource pools for production and comprehension. We also note that Martin (2005) makes an even more fine-grained division into a semantic short-term memory and a phonological short-term memory. This view is mainly based on studies of patients with dissociable impairments in the retention of semantic and phonological information. For now, we set the issue of fractionating memory systems aside, and engage in a componential analysis of the automaticity of language production, as we did for comprehension. We review research pertaining to the features *efficient* (capacity unconstrained), *unintentional*, and *unconscious*.

6.2.1 Is language production efficient?

Relatively few studies have studied language production while imposing a secondary task. Bock et al. (2007) had people talk and drive but focused on the effect on driving parameters rather than speech production parameters. Kemper, Herman, and Lian (2003) did the reverse: Older and younger adults talked while performing one of several dual tasks (walking, finger tapping, ignoring irrelevant sounds). The two groups differed in their strategies to deal with the dual tasks: Younger speakers simplified and shortened the sentences whereas older speakers spoke more slowly. Both strategies indicate that language production is not fully automatic but at least partly constrained by capacity limitations.

Other studies attempted to isolate specific processes in language production such as syntactic planning and lexical access. One strand of research investigated the production of number agreement between subject and verb in language production. Bock and Miller (1991) had speakers complete sentence fragments like (10) and showed that the speakers sometimes produced agreement errors, as in (11), especially when a plural noun occurred between a singular subject and the verb. Hartsuiker and Barkhuysen (2006) presented this task either with or without a three-word load. They found an interaction between load and (speaking) span, so that by far the most agreement errors were produced by low-span speakers under load. More recently, Allen et al. (2015) presented similar stimuli to diabetics and non-diabetics while their blood glucose level was manipulated with an insulin clamp. Hypoglycemia resulted in a significant reduction of reading span and a significant loss of accuracy on the agreement production task. Finally, Fayol, Largy, and LeMaire (1994) conducted a dictation version of the task in French, exploiting the fact that singular and plural verbs are often homophones in that language, but differ in orthography (e.g. *arrive* vs. *arrivent*, ‘arrive, 1st or 3rd pers. sg.’ vs. ‘arrive, 3rd pers. pl.’). Thus, when the participants hear (12), they need to produce correct agreement on the basis of the subject’s number because the verb’s auditory form provides no information. Fayol et al. (1994) observed that sentences like (12) elicited many agreement errors; these errors were more frequent when there was a secondary task (either click counting or a word load).

(10) The letter from the lawyers …
(11) *The letter from the lawyers were carefully written.
(12) Le chien des voisins arrive. (‘The dog of the neighbors arrives.’)

Summarizing, this line of research suggests that the production of agreement is capacity constrained. Note however that it is not yet clear whether this capacity is needed for the core process of determining the number of the verb, or whether it is needed for a subsequent checking or conflict-resolution process (Hartsuiker & Barkhuysen, 2006).

A further line of research asked whether the scope of planning in production is constrained by capacity. Consider an utterance like (13).

(13) The hat is next to the scooter.

One possibility is that speakers fully plan the utterance before articulation. But it is also possible that they only plan the initial part of this utterance (e.g. *the hat is next*) and then start to articulate it (perhaps while planning the second part). A crucial difference between these accounts is whether or not the second noun (*scooter*) is part of the initial plan. To test this, Meyer (1996) and Wagner, Jescheniak, and Schriefers (2010) had people name two-object scenes with sentences such as (13) while presenting an auditory distractor word that was related to the first noun (e.g. *cap*), the second noun (e.g. *bike*), or not related to either one (e.g. *dog*). Consistent with a broad planning scope, sentence naming onset was delayed both with a distractor related to the first and second noun. However, it has been suggested that there are individual differences in planning scope (Schriefers & Teruel, 1999) and it is possible that these differences are caused by capacity differences. In a follow-up study, Wagner et al. (2010) replicated Meyer's study but now with or without a secondary task. In one experiment, load was induced by having the speakers remember a word list. A probe word was presented after sentence production and speakers indicated whether the probe had occurred in the list. The authors found no load effect on the scope of planning, which seems to argue against a capacity-constrained view of planning. However, when load was increased by increasing the complexity of the sentence (e.g. by adding adjectives) or by adding a further conceptual decision task, the interference effect increased for the first noun but decreased for the second noun, indicating a narrowing of the scope. These findings suggest that planning scope is flexible and dependent on available capacity. Note that the planning scope presumably also depends on the stimuli. Whereas the tested combinations were novel for all subjects, it is conceivable that frequently co-occurring combinations (e.g. *the bow and arrow*) would be planned simultaneously, as assumed by entrenched views.

A further dual-task study asked whether the production of single words is constrained by central capacity, and if so, for which stage(s) of word production this is the case. Ferreira and Pashler (2002) had participants simultaneously produce words and discriminate tones. They manipulated the difficulty of word production by manipulating aspects of the task and stimuli that arguably tap into early or late stages of word production. They reasoned that if a particular stage makes a demand on a central capacity, making this stage more difficult should delay performance on the secondary, tone-discrimination task. Manipulations related to lexical access (i.e., picture frequency, predictability, semantic overlap between picture and distractor word) all slowed down performance on the concurrent tone task. In contrast, a manipulation related to phonological encoding (i.e., phonological overlap between picture and distractor word) had no effect on the tone task. Thus, these findings suggest that only early stages of word production are constrained by a central capacity. However, Cook and Meyer (2008) provided some evidence that even the late stage of phoneme selection is capacity constrained. They conducted a version of Ferreira and Pashler's (2002) paradigm in which a picture was named in the presence of either another picture with a phonologically similar or dissimilar name, a phonologically similar or dissimilar word, or a *masked* phonologically similar or dissimilar word. When the distractor was a picture or masked word, phonological relatedness sped up both picture naming and tone discrimination, suggesting that phoneme selection does make a demand on a central capacity. No such effect was found when the word was visible, which the authors attributed to the additive effects of faster phoneme selection but slower self-monitoring processes.

In sum, studies on agreement production, planning scope, and single word production all suggest that language formulation is capacity constrained (i.e., non-efficient) and thus not as automatic as often thought. We now turn to the automaticity features *unintentional* and *unconscious*.

6.2.2 Can parts of language production be unintentional?

The subtitle of Levelt's (1989) book, *From intention to articulation*, suggests commitment to a view of speaking as a process with strong top-down control. The speaker begins with a communicative intention given the situation at hand, and then sets into motion a chain of processes that translate this intention into speech motor activity. Thus, Levelt commits to a non-automatic view, in which information is only expressed when there is a goal to do so (the automaticity feature *intentional;* see Section 3). If language production is indeed fully driven by the speaker's goals, then productions should only contain information that realizes the goal and not unnecessary information that happens to be available. In contrast, the study of Wardlow Lane et al. (2006) discussed above showed that speakers 'leak' privileged information (i.e., that they can see both a small and a big heart) even when it is their explicit goal not to convey this information to the interlocutor. Thus, the authors conclude that "… being part of a communicative intention is not a necessary condition for an accessible conceptual feature to influence grammatical encoding." (p. 276). This conclusion resonates very well with Vigliocco and Hartsuiker's (2002) review of the modularity of language production processes. These authors asked, for each of several interfaces in language production (e.g. between the message and grammatical encoding), whether processing at level n was affected by information at level n + 1 that was not part of its core input. Consistent with Wardlow Lane et al. (2006), they cited numerous studies in which accessible, but irrelevant conceptual features affected grammatical processing. For instance, even though the notional number of the head noun's referent in a distributive sentence like (14) is not relevant for determining the verb's number, numerous studies have shown that agreement errors are much more likely with distributive rather than non-distributive sentences like (15) (e.g. Vigliocco, Hartsuiker, Jarema, & Kolk, 1996). In a similar vein, studies like Cook and Meyer (2008, see above) demonstrate that a picture one does not intend to name still leads to activation of its phonology, and can therefore facilitate the production of another picture that one does intend to name. Together these studies show that conceptual information outside of the speaker's (conscious) goals can still influence production, in contrast to the view that production is always intentional and thus non-automatic in this sense.

(14) The label on the bottles is green.
(15) The baby on the blankets is cute.

However, Strijkers, Holcomb, and Costa (2011) argued that (conscious) intention matters for the time course of the naming process. They presented pictures of common objects with low frequency and high frequency names and measured EEG. In one condition, subjects were instructed to name the pictures. In another condition they performed a non-verbal categorization task on the pictures. Event-related brain potentials were sensitive to the frequency manipulation in both tasks, suggesting that the pictures triggered production processes in both cases. However, in the naming task such frequency effects began already after 152 ms, whereas they began 200 ms later in the categorization task. Thus, although intentions are not necessary to set production into motion, having such an intention speeds up that process considerably.

Of further relevance to the issue of intentionality, Lind, Hall, Breidegard, Balkenius, and Johansson (2014) argued that speakers sometimes have no clear ‘preview’ of what they are about to say and rather infer this on the basis of auditory feedback. They reported a study in which participants engaged in a Stroop task but in which auditory feedback was occasionally manipulated. For instance, when a speaker correctly named the color (grey) of the word *green* printed in grey, auditory feedback would sometimes be replaced, in real time, with a recording of that same speaker saying *green* (the recording was made in an earlier phase of the experiment). On such occasions, speakers often accepted the feedback as if they had produced that speech themselves. From this, the authors concluded that speakers often underspecify their intentions and construct them on the basis of feedback. Therefore, the sense of authorship of one’s speech act would be reconstructive rather than predictive.

6.2.3 Can unconscious stimuli influence language production?

Stimuli do not need to be conscious in order to influence language production. Above, we already gave an example (Cook & Meyer, 2008) of a study in which masked presentation of a word that was phonologically similar to a target picture facilitated naming of that picture. Several studies have also shown semantic effects of masked word primes (Dhooge & Hartsuiker, 2010; Finkbeiner & Caramazza, 2006). In contrast to tasks with visible primes (which typically show semantic interference), masking the distractor leads to semantic facilitation. Finkbeiner and Caramazza (2006) accounted for this in terms of multiple loci at which a semantic relationship plays a role. Masked distractors would lead to facilitation because they would activate their corresponding concept, which would spread activation to the target concept. Visible distractors would do the same, but would further occupy a response output buffer. A semantic relationship would render the distractor difficult to exclude (as it resembles the target) from that buffer, slowing down the process.

6.2.4 Summary of studies on production

Summarizing, the review of production processes focused on three features of automaticity: *efficient*, *unintentional*, and *unconscious*. Studies considering efficiency indicate that language production is constrained by a central capacity. This not only holds for the higher-level processes involved in conceptualization, but also for lower-level processes that deal with syntactic and lexical aspects of formulation. These data contradict the classical view that only higher-level production processes are capacity demanding. With respect to the feature *unintentional*, we have seen that language production is not exclusively driven by (conscious) goals. Speakers are affected by information at the conceptual level that is irrelevant to their current goals and even detrimental to the realization of these goals. At the same time, having the conscious intention to name an object drastically reduces the time required for production processes. This is in line with the hypothesis that factors implied in automaticity (and beyond) can compensate each other in their influence on the occurrence of processes. A final finding is that masked priming studies show that unconscious stimuli can influence language production.

6.3 Dialogue

Most research in psycholinguistics has studied the case of monologue; in typical studies, participants listen to sentences from a tape recording or speak the words labelling objects in pictures into a voice key. This is obviously quite different from the way language is typically used in the real word, namely in dialogues where an interlocutor interacts with another person and alternates between speaking and listening, and where communicative goals arise naturally in the context rather than as the result of an experimenter's instruction. At first glance, dialogue involves complex interactions among individuals. First, dialogue involves the coordination of turn taking. Analyses of the timing of turn taking (Wilson & Wilson, 2005) show that pauses between turns are often shorter than 200 ms and sometimes even 0 ms. This implies that speakers do not wait for the end of the previous speaker's turn to plan their utterance; rather, they must predict when the other speaker will end and plan their own turn so that it seamlessly continues the previous turn.

Second, successful dialogue arguably requires each participant to take into account the cognitive state of the other participant. After all, it makes little sense to state something that the other person already knows (e.g. because it was said just before or because it is blatantly obvious given that both dialogue partners share a common perceptual context) or that it is impossible to understand for the other person given that she lacks an essential piece of information. Thus, interlocutors need to have a representation of what is in the *common ground* for both and what information related to the current topic is known or unknown by the other person. However, as the study of Wardlow Lane et al. (2006) discussed above shows, speakers do not always successfully tailor their utterances to the listener's perspective and seem to be worse at this when they are provided with a further task (namely to conceal information from the listener). This suggests that these complex processes are capacity limited, but it may also be the case that interlocutors do not always engage in them.

Finally, dialogue increases demands on monitoring systems (Postma, 2000). This involves self-monitoring for appropriateness and well-formedness but also for communicative success ('Has my interlocutor understood me?'). Furthermore, it involves monitoring the other person's productions ('Have they made a speech error or did they mean that? Have I heard this correctly or have I misperceived this?'). According to an influential view (Levelt, 1989), monitoring is a non-automatic process, although it should be noted that this argument seemed to be solely supported by introspection "… self-corrections are hardly ever made without a touch of awareness" (p. 21).

Given the complex set of demands dialogue seems to make on functions such as memory, inference making, monitoring, and prediction, Garrod and Pickering (2004) asked "Why is conversation so easy?". They proposed a theory that views dialogue as a joint action, in which two actors collaborate to achieve a common goal (like a pair of ballroom dancers or, more mundane, two people carrying away a table). In such joint actions, there are usually no explicit negotiations (e.g. a countdown before lifting up the table); rather, the two actors interactively coordinate by mutually adapting their action to the perception of the other person's actions (i.e., interactive alignment). Similarly, two interlocutors have a joint goal, which Pickering and Garrod (2004) define as aligning their "situation models". These are "multi-dimensional representations containing information about space, time, causality, intentionality, and currently relevant individuals" (p. 8). To converge upon overlapping situation models, interlocutors would not use explicit negotiation nor would they engage much in creating models of each other's mental states. Rather, dialogue partners would align in an interactive way (similar to the example of lifting up the table) and they would do so by mutual priming (i.e., pre-activation) of representations at each linguistic level. Because representations from comprehension would prime representations in production, dialogue partners tend to converge on the same realization of speech sounds, the same words, and the same syntactic choices. Indeed, there is much evidence for priming from comprehension to production at many different levels (see Pickering & Garrod, 2004, for a review).

To summarize, at first glance there seem to be many reasons why dialogue might be non-automatic: It is hard to estimate when to take turns, to model the state of mind of the other person, and to tailor one's utterance to the social and dialogue context. In other words, the processes involved in dialogue are deemed too complex to be automatic. On the other hand, introspection tells us that dialogue is actually really easy especially as compared to holding a monologue. Part of the solution may lie in the consideration that the processes involved in dialogue are less complex than they seem. This aligns with the explanations for automatization discussed in the next section.

7. How to explain automatization?

7.1 Factors influencing automatization

A well-established observation is that processes and behavior become more automatic as a result of practice (also called *repetition*, *training*, *frequency*) and that the complexity of a process is a factor that can impede automatization. We illustrate the role of complexity in the context of sentence comprehension. The question of how much complexity is involved in sentence comprehension in fact has been the topic of a theoretical divide in the field. On the one hand, modular models (Frazier, 1987) hold that the initial analysis of a sentence is determined only by the syntactic categories (e.g. Noun, Verb, Preposition) of the input words. The parsing mechanism builds up a single analysis on the basis of a few syntactic principles (e.g. build a structure that is as simple as possible). Thus, when the listener hears something like *put the apple on the towel* ..., the parsing system will build up an analysis in which *on the towel* is an argument of the verb *put* (it is the target location of the apple) rather than a modifier of *apple*. This is because the latter structure requires a more complex structure (i.e., it would be a reduced relative clause).

On the other hand, constraint-based models would argue that the parsing system uses all potential sources of information right away, including properties of the lexical items and their frequency of co-occurrence with particular structures, information about prior discourse, and knowledge of the world. Such accounts predict that *on the towel* can either be interpreted as an argument of the verb or as a modifier of the noun, depending on the situation. Consistent with this latter view, Tanenhaus et al. (1995) demonstrated that if a display contained an apple on a towel, a box, an empty towel, and an irrelevant object, participants hearing *put the apple on the towel* would first look at the apple and then at the empty towel. When the sentence was then disambiguated (*in the box*) they quickly moved the eyes to the box, the actual target location. However, when the display was changed so that the irrelevant object was replaced by an apple on a napkin, participants no longer misanalyzed the sentence: When hearing *apple* they alternated between both apples, but when hearing *towel* they looked at the apple on the towel and not at the empty towel. Thus, if the parser has a choice between two different analyses, this choice is *not* exclusively driven by only the strictly necessary input, as modularist accounts would have it. Rather, the choice is determined by the input in conjunction with a host of other factors, which even includes the visual context.

7.2 Mechanisms underlying automatization

In addition to asking which factors can influence automatization, researchers have asked which learning mechanisms (situated on a lower level of analysis) could be responsible for the automatization of a process (situated on a higher level of analysis). According to Logan (1988), automatization is explained by a shift from algorithm computation (or multi-step memory retrieval) to direct (or single-step) memory retrieval. Children who learn to add pairs of digits initially count the units of both digits, but once a sufficiently strong association is formed in memory between the pair of digits and their sum, they retrieve this sum from memory, thereby circumventing the counting of units.

As an alternative to the shift to direct memory retrieval, Anderson (1992; see also Tzelgov, Yehene, Kotler, & Alon, 2000) argued that automatization can also be explained by the learning mechanism of procedure strengthening. Repetition of the same procedure (applied to the same or different stimuli) results in the storage of this procedure in procedural memory, which can be retrieved and applied faster and more efficiently thereafter.

Both learning mechanisms of direct memory retrieval and procedure strengthening seem to be covered in the language domain by the notion of entrenchment or chunking[3], a set of mechanisms by which previously serially activated knowledge becomes clustered in a holistic unit or chunk. An equivalent of Logan's direct memory retrieval in the language is provided by Arnon and Snider's (2010) demonstration that speakers produce multi-word phrases (e.g. "don't have to worry") more quickly when these phrases appear more frequently (and that this cannot be reduced to the frequency of the individual words outside of these phrases). This suggests that an association is formed between the words within the phrases ("don't have" and "worry") just like digits are associated with their sum ("1+2" and "3"). An equivalent of procedure strengthening in the language domain is proposed in models of syntactic processing in which repeated exposure to a syntactic structure results in the strengthening of the network connections needed for the production of that structure (Chang, Dell, & Bock, 2006).

[3] The notion of entrenchment has been defined in various ways. Sometimes it is defined as a factor (i.e., repetition). At other times, it is defined as a lower-level process or mechanism (i.e., the creation of a holistic unit or chunk). At still other times, it is defined as an effect (i.e., the holistic unit, or the automaticity of a process). We argue for the clear separation of factors, effects, and mechanisms, and propose using *entrenchment* in the sense of "mechanism".

Pickering and Garrod (2004) argued for entrenchment as an important explanation for the increasing fluency or automaticity of dialogue. They explicitly challenged the view that during each production, the speaker moves through all levels of processing as in Figure 1. Whenever they can, speakers will skip production levels, specifically when they produce a series of words that has become a *routine*. Indeed, casual speech contains many chunks of words that are recycled from the past, including, but certainly not limited to, idioms and stock phrases. Thus, when a speaker produces "thank you very much" in a conversation, he might retrieve this in one go rather than going through the "normal" stages of conceptually driven word retrieval and combination. Such a view nicely fits with Arnon and Snider's (2010) finding that the frequency of multiword utterances predicts their production latencies. But what is more, Pickering and Garrod argue that routinization does not only happen at the level of the language, but also in the much more restricted situation of single conversations. Thus, one can easily imagine a conversation in which a speaker introduces a novel multiword utterance that is repeated several times (because it is the topic of conversation and because of massive priming). As a result, this utterance would become entrenched (or routinized) on the fly, allowing the conversation partners to re-use it in one go, just like routines that are more fixed in the language. Garrod and Pickering (2004) argue that an architecture in which there is massive priming and routinization on the fly solves many of the "problems" of dialogue. For instance, because interlocutors would, in the course of dialogue, converge upon a shared set of aligned representations (an "implicit common ground") there is no need to infer each other's state of mind. And because of entrenchment taking place within the time frame of a single dialogue, speakers can short-circuit some of the complex production processes and instead retrieve longer utterances directly.

8. Discussion

Is parsing basically a reflex? Does language formulation proceed largely automatically? There are no simple yes-no answers to such questions, because automaticity features do not form coherent sets. Thus, a process can be more automatic with respect to one feature, while being less automatic with respect to another. In our discussion of language *comprehension*, we examined the features *efficient* and *intentional*. Regarding efficiency, there is an extensive literature that has argued for a capacity-constrained (less automatic) language comprehension system, while at the same time debating the nature of this capacity. But it is possible that effects that seem to imply a limited capacity can be explained by other factors, such as interference during retrieval operations. Another automaticity feature is *unintentional*: whether the goal to engage in the process of comprehending causes it. Certain stimuli, even when the person has the goal to comprehend them do not trigger comprehension by themselves, but only when the perceiver explicitly expects speech.

Language production can similarly be seen as a collection of more and less automatic processes depending on the specific automaticity features considered. Regarding the feature *efficient*, language production, like comprehension, often has been viewed as constrained by a capacity. In contrast to the classic view of automatic lower-level processes in production, there is evidence supporting capacity limitations for lexical access and grammatical encoding. However, consideration of the feature *intentional* shows that language production is not a fully non-automatic process either. Even though the (conscious) intention to name an object speeds up production processes, it may not be needed to trigger it. Further, if feedback is distorted, speakers sometimes claim agency for the production of words they did not intend to say and in fact did not say. Finally, with respect to the feature *unconscious*, we have seen that information does not need not to be conscious in order to affect production.

Dialogue, at first glance, appears to make a demand on highly non-automatic processes such as making predictions about when it would be a good moment to take turns or reasoning about the mental states of the other person. In contrast, the interactive alignment theory of Pickering and Garrod (2004) claims that dialogue relies on less complex processes than is usually assumed and hence that less demand is needed than sometimes thought.

It thus appears that none of the language processes we have considered can easily be characterized as fully (or even largely) automatic vs. fully (largely) non-automatic. This is consistent with Moors' (2016) argument that automaticity features do not form coherent sets. Instead, it may be more fruitful to consider automaticity features as a subset of many mutually compensatory factors that jointly influence whether and how a particular process will be carried out. Thus, an attended sentence will elicit comprehension processes, unless its quality is sub-threshold. But poor stimulus quality can be compensated by expectations. Similarly, an intention to speak is not necessary to trigger production processes but does affect the speed of such processes. Such a framework, in which automaticity features are only a subset of a legion of factors driving (or blocking) cognitive processes, and in which these features do not necessarily co-vary, does raise the question of whether the automaticity concept still adds anything to the analysis of cognitive processes. Navon (1984) famously compared the concept of resources to a "soup stone.[4]" Perhaps automaticity is a soup stone too.

With regard to the explanation of automatization, we pointed at factors such as repetition, recency (or priming), and complexity. We also discussed entrenchment as a set of learning mechanism (such as direct memory retrieval and procedure strengthening) that can render a production or comprehension process more automatic. Entrenchment plays a major role in Pickering and Garrod's (2004) mechanistic theory of dialogue, both in the sense of having entrenched, long-term memory representations of units in one's language (e.g. a commonly used four-word combination) and in the sense of creating entrenched units on the fly during a dialogue (e.g. a particular referring expression that one of the interlocutors introduces in the dialogue and which is then repeated). Entrenchment (in both of these senses) creates subroutines that alleviate the burden of processing and thereby increase automaticity.

We end this chapter with a discussion of two theoretical issues that deserve further scrutiny in the future. First, one can make a theoretical distinction between core language processes and secondary language processes. Blurring of this distinction can have repercussions for the assignment of automaticity features to these core processes. With core processes, we mean processes that translate a linguistic representation at level n in a hierarchically organized sequence of processes to a representation at level $n+1$: for instance, the process of retrieving a word on the basis of a concept. Secondary processes are processes that oversee core processes and intervene. For instance, if the core process of word retrieval makes a mistake (e.g. retrieving *dog* instead of *cat*), a self-monitoring system may detect this and set into

---

[4] Navon described a Jewish legend about a magical stone that was claimed to result in soup when added to boiling water. However, the soup would be better when vegetables were also added and even better if you put in meat as well…

motion processes of interruption and correction (e.g. Hartsuiker & Kolk, 2001). Similarly, if the parser has produced an initial analysis that turns out to be incorrect, a revision system might abandon the initial analysis and replace it with an alternative analysis. Detailed assessment of automaticity needs to find a way of distinguishing whether a given automaticity feature applies to the relevant core process itself, or to a secondary process operating on it. For instance, Hartsuiker and Barkhuysen (2006) argued that effects of extrinsic memory load on agreement production did not affect the core processes that determine the number of the verb, but rather a checking process, which, given enough capacity, would weed out incorrect number markings.

Second, the concept of automaticity is separate from but related to the issue of modularity, as cognitive modules would have many of the features of automatic processes. The issue of modularity has had a profound influence on the literature for many decades. Many studies on the modularity of language processes have focused on information encapsulation. Thus, if a language processing level is a module, it responds only to its strictly necessary input and does not take into account other information. Much work in language comprehension and production has provided evidence against information encapsulation: Thus, parsing decisions are affected by information in the visual context, and language production decisions are affected by accessible but irrelevant conceptual features.

This chapter has reviewed a subset of the language processing literature and interpreted the findings in terms of a componential view of automaticity. An obvious caveat is that our selection of the literature is not exhaustive but exemplary. Thus, a more ambitious goal would be to try to map out the interrelations between many more factors (related to automaticity and beyond) for the complete set of core cognitive processes involved in production and comprehension, in monologue and dialogue, while taking into account the possible influences of secondary processes like checking and revising.

References

Allen, K. V., Pickering, M. J., Zammitt, N. N., Hartsuiker, R. J., Traxler, M. J., Frier, B. M., & Deary, I. J. (2015). Effects of acute hypoglycemia on working memory and language processing in adults with and without type 1 diabetes. *Diabetes Care, 38*(6), 1108-1115.

Anderson, J. R. (1992). Automaticity and the ACT* theory. *American Journal of Psychology, 105(2),* 165-180.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language, 62*(1), 67-82.

Bargh, J. A. (1992). The ecology of automaticity: Toward establishing the conditions needed to produce automatic processing effects. *American Journal of Psychology, 105*(2), 181-199.

Bock, K., Dell, G. S., Garnsey, S., Kramer, A. F., & Kubose, T. T. (2007). Car talk, car listen. In A. Meyer, L., Wheeldon, & A. Krott (Eds.), *Automaticity and control in language production* (pp. 21-42). Hove, UK: Psychology Press.

Bock, K., & Miller, C. A. (1991). Broken Agreement. *Cognitive Psychology, 23*, 45-93.

Caplan, D. (1992). *Language: Structure, processing and disorders. Issues in the biology of language and cognition*. Cambridge, MA: MIT Press.

Caplan, D., & Hildebrandt, N. (1988). *Disorders of syntactic comprehension*. Cambridge, MA: MIT Press.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences, 22*(1), 77-126.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113*(2), 234-272.

Christiansen, M. H., & MacDonald, M. C. (1999). Fractionated working memory: Even in pebbles, it's still a soup stone. *Behavioral and Brain Sciences, 22*(1), 97-98.

Clahsen, H., & Felser, C. (2006). How native-like is non-native language processing? *Trends in Cognitive Sciences, 10*(12), 564-570.

Cook, A. E., & Meyer, A. S. (2008). Capacity demands of phoneme selection in word production: New evidence from dual-task experiments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(4), 886-899.

Dhooge, E., & Hartsuiker, R. J. (2010). The distractor frequency effect in picture-word interference: evidence for response exclusion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(4), 878-891.

Fayol, M., Largy, P., Lemaire, P. (1994). When cognitive overload enhances subject-verb agreement errors. A study written in French language. *Quarterly Journal of Experimental Psychology, 47*(2), 437-464.

Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28*(6), 1187-1199.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25(3),* 348-368.

Finkbeiner, M., & Caramazza, A. (2006). Now you see it, now you don't: On turning semantic interference into facilitation in a Stroop-like task. *Cortex, 42*(6), 790-796.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance 12: The psychology of reading* (pp. 559-586). Hillsdale, NJ: Erlbaum.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences, 8*(1), 8-11.

Gibson, E. (2000). The dependency locality theory: a distance-based theory of linguistic complexity. In Y. Miyashita, A. Marantz, & W. O'Neil (Eds.), *Image, language, brain* (pp. 95-126). Cambridge, MA: MIT Press.

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological Science, 13*(5), 425-430.

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience, 11*(2), 194-205.

Hartsuiker, R. J., & Barkhuysen, P. N. (2006). Language production and working memory: The case of subject-verb agreement. *Language and Cognitive Processes, 21*(1-3), 181-204.

Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology, 42*(2), 113-157.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica, 137*(2), 151-171.

Kemper, S., Herman, R. E., Lian, C. H. T. (2003). The costs of doing two things at once for young and older adults: Talking while walking, finger tapping, and ignoring speech or noise. *Psychology and Aging, 18*, 181-192.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122-149.

Levelt, W. J. M. (1989). *Speaking: From articulation to intention*. Cambridge, MA: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1-38.

Lind, A., Hall, L., Breidegard, B., Balkenius, C., & Johansson, P. (2014). Speakers' acceptance of real-time speech exchange indicates that we use auditory feedback to specify the meaning of what we say. *Psychological Science, 25*, 1198-1205.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95(4)*, 492-527.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological Bulletin, 109*(2), 163-203.

Martin, R. C. (2005). Components of short-term memory and their relation to language processing. Evidence from neuropsychology and neuroimaging. *Current Directions in Psychological Science, 14*, 204-208.

Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language, 35*(4), 477-496.

Moors, A. (2016). Automaticity: Componential, causal, and mechanistic explanations. *Annual Review of Psychology, 67*, 263-287.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132*(2), 297-326.

Navon, D. (1984). Resources. A theoretical soup stone? *Psychological Review, 91*(2), 216-234.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169-190.

Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition, 77(2),* 97-131.

Pulvermüller, F., Shtyrov, Y., & Hauk, O. (2009). Understanding in an instant: Neurophysiological evidence for mechanistic language circuits in the brain. *Brain and Language, 110*(2), 81-94.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212*, 947-950.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*(1), 1-66.

Schriefers, H., & Teruel, E. (1999). Phonological facilitation in the production of two-word utterances. *European Journal of Cognitive Psychology, 11*(1), 17-50.

Strijkers, K., Holcomb, P. J., & Costa, A. (2011). Conscious intention to speak proactively facilitates lexical access during overt object naming. *Journal of Memory and Language, 65*(4)*, 345-362.*

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632-1634.

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica, 135*(2), 77-99.

Tzelgov, J., Yehene, V., Kotler, L., & Alon, A. (2000). Automatic comparisons of artificial digits never compared: Learning linear ordering relations. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26(1), 103-120.*

Tzur, B., & Frost, R. (2007). SOA does not reveal the absolute time course of cognitive processing in fast priming experiments. *Journal of Memory and Language, 56*(3), 321-335.

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language, 55*(2)*,* 157-166.

Van Dyke, J. A., & Johns, C. L. (2012). Memory interference as a determinant of language comprehension. *Language & Linguistics Compass, 6*(4), 193-211.

Van Opstal, F., Gevers, W., Osman, M., & Verguts, T. (2010). Unconscious task application. *Consciousness and Cognition, 19*(4)*,* 999-1006.

Vigliocco, G., & Hartsuiker, R. J. (2002). The interplay of meaning, sound, and syntax in sentence production. *Psychological Bulletin*, *128*(3), 442-472.

Vigliocco, G., Hartsuiker, R. J., Jarema, G., & Kolk, H. H. J. (1996). One or more labels on the bottles? Notional concord in Dutch and French. *Language and Cognitive Processes, 11*(4), 407-442.

Wardlow, Lane, L., Groisman, M., & Ferreira, V. S. (2006). Don't talk about pink elephants. Speakers' control over leaking private information during language production. *Psychological Science, 17*(4)*,* 273-277.

Wagner, V., Jescheniak, J. D., & Schriefers, H. J. (2010). On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(2), 423-440.

Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology, 49*A, 51-74.

Wiese, E., Wykowska, A., Zwickel, J., & Müller, H. J. (2012). I see what you mean: How attentional selection is shaped by ascribing intentions to others. *PLoS ONE, 7*(9), e45391.

Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin and Review, 12*(6)*,* 957-968.