Cambridge University Press & Assessment
978-1-009-26455-6 — Grammatical Encoding for Speech Production
Linda Ruth Wheeldon , Agnieszka Konopka
Excerpt
More Information

1 Introduction

Psycholinguistic and linguistic theory agree that sentence production is a generative process involving a separate lexicon and grammar (e.g., Chomsky, 1965; Levelt, 1989). Speakers of a language can retrieve words from their mental lexicon and order them in accordance with their grammar to generate a theoretically infinite number of sentences. This potential for unbounded creativity is at variance with the evidence, to be reviewed in what follows, that spoken language tends toward repetition. Nevertheless, some degree of separation between lexical and syntactic representations and processes is a cornerstone of all current models of grammatical encoding (e.g., Chang, Dell & Bock, 2006; Dell, Oppenheim & Kittredge, 2008; Levelt, Roelofs & Meyer, 1999). Theoretical approaches to the processes of lexical retrieval and syntactic structure building in fluent sentence production are discussed in Section 1. The theoretical framing will focus on the key dichotomy in the field: whether grammatical encoding is driven by lexical (e.g., Bock & Levelt, 1994) or syntactic representations (e.g., Chang et al., 2006; Dell et al., 2008). We will begin with a theoretical overview, which will incorporate a brief discussion of theories of lexical representation and access (e.g., Wheeldon & Konopka, 2018), before turning to how retrieved lexical items are integrated into the unfolding syntax of an utterance.

We then evaluate the evidence for the independence of syntax from lexical representations and the nature of the structural representations generated during grammatical encoding (Section 2). The critical evidence in this area has been largely derived from studies of structural priming. In the early days of this research, the presence of lexically unsupported syntactic priming was taken as evidence of abstract structural processing in sentence production (e.g., Bock, 1986). Further research demonstrated limited involvement of the lexicon in the generation of syntactic structures. Existing rich evidence from within-language and between-language comparisons largely supports the view of the independence of syntax and the lexicon in adult speakers (Branigan & Pickering, 2017; Chang et al., 2006; Mahowald, James, Futrell & Gibson, 2016; Pickering & Ferreira, 2008), but with outstanding questions remaining in developmental psycholinguistics (e.g., Messenger, Branigan & McLean, 2011; Rowland, Chang, Ambridge, Pine & Lieven, 2012). Priming research has also helped to delimit the nature of the syntactic representations generated during sentence production (e.g., Bernolet, Hartsuiker & Pickering, 2007; Branigan, Pickering, McLean & Stewart, 2006; Ferreira, 2003; Fox Tree & Meijer, 1999; Hardy, Wheeldon & Segaert, 2020; Ziegler, Snedeker & Wittenburg, 2017).

In the next section we switch focus to the time-course of grammatical encoding (Section 3). Here, the theoretical debate turns on whether online

Cambridge University Press & Assessment
978-1-009-26455-6 — Grammatical Encoding for Speech Production
Linda Ruth Wheeldon , Agnieszka Konopka
Excerpt
More Information

sentence planning occurs in a lexically incremental fashion (Bock & Levelt, 1994; Griffin, 2001; Meyer, Sleiderink & Levelt, 1998; also see Meyer, Wheeldon, Van der Meulen & Konopka, 2012) or in a structurally driven, hierarchical fashion (Konopka & Meyer, 2014; Lee, Brown-Schmidt & Watson, 2013; Martin, Miller & Vu, 2004; Momma, 2021; Smith & Wheeldon, 1999; Wheeldon, 2013; Wheeldon, Smith & Apperly 2011). The critical evidence for this debate comes from studies of planning scope in picture description paradigms to determine the degree of planning occurring in advance of articulation onset. These paradigms frequently make use of eye tracking, allowing the time-course of planning from the initial uptake of visual information to the onset of speech to be determined (e.g., Konopka, 2019). More recently, cross-linguistic studies have investigated the role of language-specific grammatical constraints on planning (Allum & Wheeldon, 2007, 2009; Hwang & Kaiser, 2014a; Momma, Slevc & Phillips, 2016; Norcliffe, Konopka, Brown & Levinson, 2015; Sauppe, Norcliffe, Konopka, van Valin & Levinson, 2013).

The Element will also include relevant data from studies of bilingual sentence planning (e.g., Konopka, Meyer & Forest, 2018). This research speaks both to the representation of syntactic structure and to the issue of the effects of cognitive load on planning scope. We will review evidence that grammatical planning scope can be modulated by non-linguistic factors and cognitive limitations, including speed requirements (e.g., Ferreira & Swets, 2002), working memory (e.g., Swets, Jacovina & Gerrig, 2014), and attention (e.g., Jongman, Meyer & Roelofs, 2015; Jongman, Roelofs & Meyer, 2015).

In the final section of the Element (Section 4), we will provide an evaluation of the strengths and weaknesses of the methodological approaches that have been used to date in the field. Finally, we will reassess the theoretical landscape, highlighting gaps and defining the resulting avenues for future research.

1.1 Grammatical Encoding in Speech Production

1.1.1 The Component Processes for Speaking

In this section, we review theories of grammatical encoding for speech production, focusing on the proposed relationship between words and syntax. We begin, however, with setting the process of grammatical encoding in context. All cognitive models of speech production are heavily influenced by Levelt's classic blueprint for the speaker (Levelt, 1989), which in turn built on the seminal work of Garrett (1975). The proposal is that utterances are produced in a number of more-or-less successive processes, and there is also agreement on the broad structure of the processes involved (see Figure 1). The starting

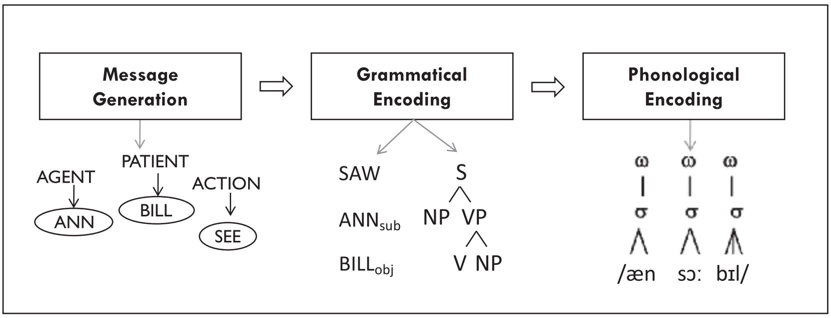

Figure 1 A representation of the key processing stages of spoken sentence production.

point is message generation, which involves the construction of a conceptual representation that details the information that the speaker wants to convey. This representation is usually known as the *message* (Levelt, 1989). The current view is that messages are non-linear and must at least contain conceptual category information. Messages can be very short (e.g., mapping onto utterances like ‘Hi’ or ‘Look there!’) or much longer, including a thematic structure which assigns concepts to thematic roles such as agent or patient (e.g., mapping onto utterances like ‘The politician was amazed by the volume of fan mail’; see Konopka & Brown-Schmidt, 2014, for a review). In addition, messages should contain information that is required to generate a grammatical sentence including time, mood and focus, as well as any language-specific information required by a language for obligatory syntactic or morphological markers (see Levelt, 1989, chapter 3, for a detailed discussion).

The message triggers grammatical encoding processes, which include selecting the appropriate lexical items, assigning grammatical roles and generating a syntactic structure to fix their linear order. The phonological structure of the utterance is constructed in the subsequent phase, where an abstract prosodic representation is generated which forms the input to phonetic and articulatory processes. Grammatical encoding processes therefore form the link between the conceptual structure to be conveyed and the sound structure of the utterance that will convey it. The component processes are *lexical retrieval* and *syntactic structure building*.

1.1.2 Lexical Retrieval Processes

Lexical retrieval refers to the activation and retrieval of words from the mental lexicon. During production, activation at the conceptual level triggers a lexical search. Psycholinguistic models largely agree that lexical representations exist independently of semantics, at the *lemma* and *lexeme* levels (Kempen & Huijbers, 1983). *Lemmas* are abstract, modality-general and language-specific lexical entries that are activated by information at the conceptual level. In turn, lemma selection activates *lexemes*, that is, representations that include word-form information (see Caramazza & Miozzo, 1997, vs. Roelofs, Meyer & Levelt, 1998), and then phonological encoding processes. For example, a speaker wishing to convey information about one person (a woman) transferring something (a book) to another person (a man) will generate a message-level representation consisting of conceptual nodes that correspond to the nominal concepts *woman*, *man* and *book*, as well as the action of transferring X to Y, and this information may activate the *lemma* nodes for the nouns ‘woman’, ‘man’, ‘book’ and the verbs ‘give’ and ‘donate’. Lemmas include item-specific syntactic information, such as grammatical gender for nouns and restrictions on syntactic alternations for verbs (e.g., the verb ‘give’ can be used with both prepositional-object [PO] and double-object [DO] syntax, while the verb ‘donate’ can only be used with PO syntax).

The majority of models describing lexical access focus on retrieval of individual words – most often nouns (e.g., ‘woman’, ‘man’, ‘book’) – or production of short sequences of words in simple or complex noun phrases (NPs) (e.g., ‘the woman’, ‘the woman and the man’). The likelihood of selecting a lemma and the speed of selecting one noun lemma over another vary as a function of (a) word-specific variables (e.g., lexical frequency, age of acquisition, name agreement), (b) properties of the words’ lexical neighbours (e.g., neighbourhood density, recent activation of neighbouring lexical nodes, the degree to which relationships between words are taxonomic or thematic), and (c) the proposed architecture of the production system (e.g., the direction of information flow between the conceptual, lexical and phonological levels). Two classes of models, Levelt and colleagues’ serial model (Levelt et al., 1999; also see Roelofs, 1992) and Dell and colleagues’ interactive models of lexical access (Dell, 1986; Dell, Schwartz, Martin, Saffran & Gagnon, 1997), have led the theorising in the field. In both models, the concepts or lexical nodes that are most strongly activated are selected for production, but the models differ in the degree to which they allow activation from lower levels to influence selection: serial models assume a feedforward flow of activation from concepts to lemmas and to phonological encoding, while interactive models allow for feedback from lower levels.

Lexical retrieval models also differ in their assumptions about the selection process at the lemma level, specifically the degree to which lemmas do or do not compete for selection (Levelt et al., 1999 vs. Mahon, Costa, Peterson, Vargas & Caramazza, 2007; see Abdel-Rahman & Melinger, 2009, for a review). The predictions of these models are often tested with the picture–word interference paradigm, where speakers name individual pictured objects while ignoring superimposed printed words. Retrieval times normally increase in the presence of semantic competitors, such as when trying to name the picture of a cat while seeing the printed word 'dog', and decrease in the presence of phonological neighbours, such as when trying to name the picture of a cat while seeing the printed word 'cap'. Debates concerning the size and direction of these effects often hinge on determining the joint effects of multiple individual processes: conceptual priming (semantically related words prime each other), lexical interference (taxonomically related words compete against each other for selection), lexical facilitation (thematically related word prime each other) and phonological facilitation (phonologically related words prime each other). Production of a *sequence* of words, either in phrases (e.g., 'the cat and the dog') or without a phrasal context ('cat dog'), naturally multiplies the number of processes to be completed and adds an additional parameter: retrieval of each word (word *n*) can be influenced by anticipatory activation of word *n+1*, and likewise, retrieval of word *n+1* is influenced by production of word *n*. As in most picture-word interference paradigms, retrieval of word *n* is slower when word *n+1* is a semantic competitor, but retrieval of word *n+1* is also slower when word *n* is a semantic competitor (an effect known as cumulative semantic interference).

In a recent meta-analysis, Bürki, Elbuy, Madec and Vasishth (2020) concluded that existing research does not adjudicate between models assuming competitive and non-competitive lexical access. Oppenheim and Nozari (2021) also showed that behavioural indexes such as the presence of semantic interference or facilitation cannot be used to conclusively distinguish between competitive and non-competitive lexical access, as competitive and non-competitive selection rules can produce similar behavioural outcomes. A more promising approach is to track context-specific changes in retrieval speed in order to model experience-driven changes in activation levels and connections between the conceptual level and word level (see Dell & Jacobs, 2016; Dell, Nozari & Oppenheim, 2014; Oppenheim, Dell & Schwartz, 2010, and Oppenheim & Nozari, 2021, for more detail with supporting empirical evidence and simulations). For example, the degree to which both taxonomically and thematically related distractors interfere with production of a target word depends on the way these relationships are represented in the model, rather than depending on selection rules.

The models of lexical retrieval reviewed in the preceding text are concerned with the nature of lexical representations for content words (mostly nouns) and thus do not make explicit claims about processes responsible for integrating sequences of lexical items into longer utterances. In the rest of the Element, we focus primarily on a different long-standing debate in psycholinguistics – namely, the contribution of the lexicon to grammatical encoding (see Bock, 1982, 1987, for early reviews). This area of research focuses on production of longer, multi-word utterances with complex syntactic structures and, critically, utterances requiring retrieval of verbs.

1.1.3 The Need for Syntax

Producing grammatically correct multi-word utterances requires that words be produced in a specific order, that is, that they be sequenced according to language-specific word-order rules. This sequencing is referred to as linearisation. Interestingly, while it is clear that linguistic utterances *are* structured, the nature of the structural representations generated to output grammatically correct word sequences is debated. This puzzle concerns the degree to which the lexicon is involved in the generation of sentence structure.

Broadly speaking, the generation of sentence structure has been described, in different accounts, as a by-product of lexical retrieval processes or as the outcome of processes operating outside of the lexicon (e.g., see Bock, 1987, for a review). Lexicalist (or functional) accounts propose that there is no strict separation between the lexicon and grammar: speakers retrieve lexical items as required by the preverbal message they want to communicate, and it is the lexical retrieval process that initiates the building of a syntactic structure. In other words, the building of a linguistic structure is dependent on lexical activation. By implication then, syntax is largely epiphenomenal. However, the linearisation of a longer, complex message that requires activation of multiple content words poses a problem for this account, as lemma activation can be responsible for the activation of 'local' syntactic information but is less likely to be responsible for the building of larger syntactic frames (also see Section 3 for a discussion of planning scope in multi-word utterances). Abstract structural accounts are better able to account for linearisation in longer utterances, as they propose that larger *structures* (or *frames*) are built by abstract syntactic procedures independently of the lexical items that will be slotted into them. These procedures are sensitive to word-specific syntactic requirements, but they are not, crucially, triggered by activation of individual lemmas.

The viability of the lexical account, and thus the origins of the debate between lexical and abstract accounts, has historical roots. Language research has been largely skewed in favour of comprehension rather than production, and comprehension studies show strong reliance on the lexicon during parsing. In comprehension, listeners receive a linguistic signal that comes in word by word over time and they must integrate this information to decode the speaker's message. Naturally, given that listeners process incoming information as soon as it becomes available, the processor may give more weight to new lexical information (which can be quickly integrated with those parts of the utterance that have already been heard) than to structural information (as the structural representation of a spoken utterance is built up or inferred from a *string* of words rather than from individual words). Listeners do generate predictions about upcoming words, but evidence of prediction based on the semantic or lexical content of a sentence (be it coarse-grained, i.e., involving entire words, or finer-grained, i.e., involving sublexical units) is currently more plentiful than evidence of prediction of structure based on grammatical markers or parts of speech (see Huettig, Rommers & Meyer, 2011, for a review). Thus, the demands of comprehension for structural processing may be less stringent than in production and may effectively ‘hide’ potential effects of abstract structural processes. Levels of engagement during comprehension can also vary, such that ‘good enough’ processing (i.e., the build-up of underspecified representations) may be sufficient for successful comprehension in many contexts (Karimi & Ferreira, 2016). Indeed, finding evidence of the involvement of abstract structural processes in comprehension requires development of more sensitive measurement tools or ensuring greater engagement on the listener’s part (see Tooley & Bock, 2014).

In contrast, the distinction between lexical sources of structure and abstract structural processes is more salient and thus more relevant in production. The processing demands of language production on the speaker are arguably higher than the demands of comprehension on the listener. To produce an utterance, speakers must first decide what they want to say (albeit not necessarily in large, sentence-sized chunks) and must then begin generating the linguistic material they will need to communicate their message from scratch. This involves both structural and lexical processing, so stronger reliance on lexical than structural information may not be as viable in production as it is in comprehension: producing a sequence of words cannot bypass structural processing and rely exclusively on lexically specific syntactic information. An empirical challenge in the field of language production is therefore the need to delineate the boundary between lexically driven and lexically free influences on word order, and to explain when and how these processes interact.

1.1.4 Models of Grammatical Encoding: The Relationship between Words and Syntax

Models of grammatical encoding differ in the relationship they propose between words and structure. There are different claims about which level of representation encodes links between lexical and structural information, with some models encoding explicit links between lexical concepts and thematic

Cambridge University Press & Assessment
978-1-009-26455-6 — Grammatical Encoding for Speech Production
Linda Ruth Wheeldon , Agnieszka Konopka
Excerpt
More Information

roles at the conceptual level (e.g., Chang, 2002; Chang et al., 2006), and others in grammatical representations between lemmas and syntactic information, allowing lexical retrieval and structure building to interact during grammatical encoding (e.g., Bock & Levelt, 1994; Cleland & Pickering, 2003, 2006; Ferreira, 2000; Ferreira, Morgan & Slevc, 2018; Levelt, 1989; Levelt et al., 1999; Momma, 2021; Pickering & Branigan, 1998). Models also diverge in the degree to which lexical or structural information guide grammatical encoding.

The earliest models of grammatical encoding were lexically driven and accorded a central role to lemma representations, which comprised semantic and syntactic-lexical information (e.g., Bock & Levelt, 1994; for reviews, see Bock & Ferreira, 2014; Ferreira & Slevc, 2007; Ferreira et al., 2018). Later versions of this approach limited lemmas to encoding aspects of lexical syntax, including grammatical category (e.g., noun, verb, adjective) as well as syntactic features (e.g., tense, number, grammatical gender; e.g., Levelt et al., 1999, see also Roelofs & Ferreira, 2019). These models also assume a discrete flow of information, with lemma selection occurring during grammatical encoding prior to the activation of phonological form (see Section 1.1.2). Two distinct stages are proposed for structure building. In the initial stage, termed *functional encoding*, the lemmas which best match the conceptual representation in the message are retrieved and assigned to grammatical functions appropriate for the thematic structure (e.g., agent → subject, patient → object, for a transitive active sentence such as ‘Anne saw Bill’). Following function assignment, an appropriate phrase structure is generated to which the lemmas are attached. The process for generating phrase structure was elaborated in a model proposed by Pickering and Branigan (1998), which also incorporated links from lemma representations to nodes specifying the possible phrase structures in which they can occur. These ‘combinatorial nodes’ were initially linked only to verbs and encoded subcategorisation information. Later versions of the model extended the approach to nouns (Cleland & Pickering, 2003, 2006). Following function assignment, the selection of phrase structures in the model is driven by activation spreading from the lemmas with the most highly activated combinatorial node being selected (*constituent assembly*). Due to the direct links between lemmas and syntactic structures, this approach provides a clear mechanism through which lexical and syntactic representations can interact to determine the structure of the sentence produced.

Another approach which encodes explicit links between lemmas and syntactic structures employs tree-adjoining grammar (TAG; Ferreira, 2000; Frank, 2002; Momma, 2021, 2022). Momma (2021, 2022) proposes a TAG-based grammatical encoding model in which the syntactic structure for an utterance is constructed based on elementary trees. Elementary trees are complex