Producing spoken sentences: the scope of incremental planning

Chapter · January 2012

1 author:

Linda Wheeldon
Universitetet i Agder

Producing spoken sentences: The scope of incremental planning

LINDA WHEELDON

Abstract: The production of spoken sentences involves the generation of a number of levels of representation: a conceptual representation for the message we wish to convey, a grammatical representation that determines an appropriate word order for that message, and phonological and phonetic representations to guide articulation. In normal speech, these representations can be generated and articulated at rates of several words per second. To account for such processing speed, models of sentence production propose that speech is planned incrementally, so that the articulation of early parts of an utterance occurs in parallel with the planning of upcoming segments. However, exactly how processing at different levels is co-ordinated remains a matter of dispute. In particular, there is disagreement about how much of an utterance must be generated at a particular level of representation before processing at the next level can begin. The focus of this paper is on the timing of early conceptual and grammatical encoding processes. A series of experiments is reviewed, which used reaction-time methodologies to investigate the production of sentences in English and Japanese. The aim of the experiments was to determine the scope of advanced planning prior to sentence onset and to investigate the relationship between conceptual and grammatical processes.

1 Introduction

In order to produce sentences we must translate some non-linguistic idea into an articulated utterance. All models of speech production postulate that this translation process occurs in a number of successive steps (Bock and Levelt, 1994; Garrett, 1980a,b; Levelt, 1989, 1992; Pickering and Branigan, 1998; Chang, 2002; Chang et al., 2006; Ferreira and Slevc, 2007). The starting point is a conceptual structure for an utterance that details the information the speaker wants to convey. This representation is usually called the message (Levelt, 1989). There is very little agreement about how messages are structured. However, the current view is that messages are non-linear and must at least contain conceptual category information and have a thematic structure with concepts assigned to thematic roles.

So for example, the message for the sentence *Anne kissed Peter* would comprise an agent-Anne performing an action-kiss on a patient-Peter. Arguably, the message must contain all the information required by a given language to form a legal utterance e.g., mood, focus, time etc. (Levelt, 1989). The message triggers grammatical encoding processes, which select the appropriate lexical items and generate a syntactic structure to fix their linear order. Finally, the phonological and phonetic structure of the utterance must be generated to guide articulation. This chapter will focus on planning during conceptual and grammatical encoding. The main issue to be addressed concerns the way in which these processes proceed over time, and in particular how much processing occurs prior to the onset of articulation.

1.1 Incremental processing

The speed with which we can articulate complex sentences makes it unlikely that the processing of each component level of representation is completed for the entire sentence before processing at the next begins. Instead, most current models of language production propose that processing at all levels occurs in an incremental fashion (e.g., Kempen and Hoenkamp, 1987; Levelt, 1989, 1992). In Levelt's (1989) blueprint for the speaker, utterances are produced in a piecemeal fashion, with early parts of the utterance being articulated while we plan upcoming parts. For example, if a speaker wished to produce the sentence *Rosa and James danced together last night*, prior to articulation they would have to build the conceptual, grammatical and phonological structures they needed. If they did this incrementally they might first retrieve the information relating to the agents, *Rosa and James*. While the agent information was being grammatically encoded, the speaker might conceptualise the action information danced together, and during the articulation of the phrase *Rosa and James* the speaker could simultaneously grammatically encode the action and conceptualise the time information last night, and so on. In such an incremental system, some conceptual processing must have occurred before grammatical processing can begin, but different pieces of the same utterance can be conceptually and grammatically processed in parallel.

Levelt (1989) argues that an incremental production system is attractive for a number of reasons. It can explain the fluency of speech because it allows the fast release of formulated chunks of the utterance for articulation and spreads the processing load efficiently across several different processors. Incremental processing also reduces the need for the temporary storage of completed chunks of the utterance thereby reducing processing costs. Nevertheless, short-term storage capacity could sometimes be necessary as the order in which the message chunks are released might not be the same as the order required by grammatical encoding processes. For example, the sentence *Rosa and James last night danced together* is not grammatical in English. Therefore, any model of sentence production that outputs language in this piecemeal fashion must also explain how the system determines the correct order of chunks for output. We will return to this issue in section 2.3.

1.2 The scope of advanced planning

Any detailed incremental model of language production must specify, for each processing level, the minimal chunk of information it constructs and delivers as output. Of course planning increments for particular levels of processing will vary in a language dependent way as languages have different backward dependencies that may affect the lower limit of processing scope. For example, in many languages the form of a determiner (e.g., *the*) is dependent on the grammatical gender of the noun it refers to (e.g., *das Auto*, the car/*die Blume*, the flower). This means that a speaker would need to retrieve the noun before they could select the correct determiner. Languages also have colocational dependencies between words that can affect word choice in a backward dependent way e.g., in English it is grammatical to say things like *sink into oblivion* or *fall into disuse* but do not *fall into oblivion* or *sink into disuse* (see Smith, 2000, for a discussion). Such idiomatic phrases generate similar amounts of structural priming as non-idiomatic phrases (Konopka and Bock, 2008) suggesting that their structure is planned on-line during production in the same way, rather than being stored as lexical items.

It is also possible, of course, that the scope of advanced planning might vary in a context or speaker dependent way. There are some findings that suggest we can alter our planning scope to a limited extent under different conditions of time pressure (Ferreira and Swets, 2002) or due to variations in some forms of cognitive load (Wagner et al., 2010) and working memory load (Slevc, 2011). Moreover, individual differences in working memory have been shown to affect processing scope in language comprehension (Swets et al., 2007), and such factors are also likely to influence advanced planning scope in production. The research we describe below does not aim to address these issues. Instead we asked random samples of normal speakers to generate fluent sentences, beginning as quickly as possible. The scope of advanced planning observed under these conditions is arguably optimal for speed and fluency and, as we argue below, provides important constraints on cognitive models of language production.

There is still a great deal of disagreement about exactly how big the planning units are at different levels of processing. It is, of course, not necessary that the scope of processing remains the same at each level. Indeed many models propose that each level has a different processing scope, often with scope decreasing as processing moves closer towards articulation (e.g., Bock and Levelt, 1994; Garrett, 1980a,b; Levelt, 1989). A number of different proposals exist in the literature. Wundt (1900) argued that conceptually a clause must be planned which can then be grammatically encoded in phrasal chunks. In the model proposed by Garrett (1980a,b) and developed by Bock and Levelt (1994), grammatical encoding is lexically driven and requires verb selection. According to this model, there is a clausal scope for lexical access; a speaker retrieves all the open class words for a clause and assigns them to grammatical functions (like subject/or object) before building a syntactic structure. More recently researchers have proposed more tightly incremental word-by-word processing (Griffin, 2001; Levelt and Meyer, 2000; Meyer et al., 1998).

Evidence exists for all of the proposed processing scopes. Planning pauses in speech have been shown to occur more frequently between clauses than clause internally (Butterworth, 1980; Goldman Eisler, 1968). Speech error data have also been used to claim that the words for a whole clause are retrieved prior to speech onset. Word exchange errors such as *this spring has a seat in it* suggest that the exchanging words *spring* and *seat* are retrieved in parallel Garrett (1980a). Similarly, subject-verb agreement errors such as those shown in (1) provide evidence for the clause as a unit of processing. These errors are often caused by the verb agreeing erroneously with a local noun (e.g., *posters are*) rather than the subject noun (e.g., *The slogan is*). Such errors are more likely to occur when the local noun is in the same clause as the verb (1.a) than in a different clause (1.b), again suggesting that elements of a clause are processed in parallel (Bock and Cutting, 1992).

1. a) The slogan on the posters *are* really effective

   b) The boy [that likes the snakes] *are* really happy

It remains unclear, however, at which level of representation the clause functions as a processing unit. The effects described above could occur during conceptual, rather than grammatical processing.
Experimental evidence exists for tighter incremental planning. For example, Schriefers et al. (1998) investigated whether verb retrieval was necessary prior to sentence onset. In their study, German speakers produced descriptions of pictured actions like *The man empties the bucket*. Different grammatical structures were elicited using a sentence completion task, as in (2). A subject-verb-object (SVO) order was elicited in a main clause (2.a). Lead-in phrases were used to elicit different word orders with the same picture; SOV in a subordinate clause (2.b) and VSO following an adverbial (2.c). The critical manipulation was the position of the verb in the sentence.

2. a)		*-Der Mann* ***leert*** *den Eimer* The man empties the bucket
b)	*Auf dem nächsten Bild sieht man wie* On the next picture one sees how	*-der Mann den Eimer* ***leert*** the man the bucket empties
c)	*Und auf dem nächsten Bild* And on the next picture	*-****leert*** *der Mann den Eimer* empties the man the bucket

At picture onset, speakers heard and saw distractor verbs, which were semantically related (e.g., *empty*) or unrelated (e.g., *write*) to the verb in the sentence they would produce. In simple picture naming tasks, a semantically related distractor causes interference and slows naming latency (Schriefers et al., 1990). In this experiment, semantic interference slowed sentence production latencies but only when the verb occurred sentence initially (Example 2.c), indicating that the verb had been retrieved prior to sentence onset. When the verb occurred later in the sentence (2.a and 2.b), onset latencies were unaffected by the semantically related distractors. This finding provides strong evidence that, unless it is in sentence initial position, the verb need not be accessed before speech onset.

Evidence for tightly incremental word-by-word processing has come from studies using eye-tracking technology to record the gaze patterns of speakers while they name pictured objects in noun phrases such as *the hat and fork* (e.g., Griffin, 2001; Meyer et al., 1998; Levelt and Meyer, 2000). The looking patterns derived from these studies are very regular. On the majority of trials speakers fixated on the objects in their order of mention and their shift of gaze to the next picture was closely coordinated with articulation, occurring just before the articulation of the first object's name. The amount of time spent looking at an object was affected by lexical and phonological properties of its name, such as word frequency and word length. Speakers also rarely looked ahead at objects to be named later although some peripheral processing of immediately adjacent objects can occur (e.g., Morgan and Meyer, 2005). These findings suggest that we plan each item to the level of phonological encoding before moving the eyes to the next object to be named, with planning progressing only slightly ahead of articulation. However, these experiments usually involved the production of one sentence structure to one fixed pattern of pictures and it is therefore not clear what they tell us about early conceptual and grammatical planning processes. Eye tracking studies, which have elicited varied sentence structures, report that some aspects of the picture arrays or scenes receive an initial scan prior to gaze returning to the initial object to be named. From then on gaze duration and shift is tightly locked to onset of articulation (Griffin and Bock, 2000). Griffin and Bock (2000) have proposed that during the initial fixation, the conceptual content of the utterance is determined prior to the formulation phase during which words are accessed incrementally and phrase structure is built.

2 Testing processing scope during spoken sentence production

2.1 Evidence for a phrasal processing scope

We turn now to a series of experiments conducted by myself and colleagues, which were designed to investigate the scope of advanced planning during spoken sentence production. In particular, we were interested in how much of a sentence a speaker plans before beginning to speak. The experiments therefore use speech onset latencies as the dependent measure. The first experiment was an extension of a study by Levelt and Maassen (1981), who demonstrated that sentence onset latencies were longer for coordinate noun phrases (e.g., *The circle and the square move up*) than for coordinate sentences (e.g., *The circle moves up and the square moves up*). This finding suggests that speakers did not plan the whole of the utterance prior to speech onset, as the coordinate sentences are more complex than the coordinate noun phrases. However, this study is ambiguous as to the scope of the processing unit, which could be the initial phrase or the initial clause.

Smith and Wheeldon (1999) (Experiment 1) designed an experiment to test between lexical, phrasal and clausal processing scopes for sentence production. In our experiment, speakers described arrays of moving pictures from left-to-right and their sentence production latencies were measured. On the critical trials the visual displays elicited two kinds of single clause sentences (see Figure 1). The sentences in Figure 1 do not differ in their total structural complexity and are perfectly matched for lexical complexity. They differ, however, in the complexity of their initial phrase, with one sentence beginning with a coordinate noun phrase (NP) and the other with a simple NP. We were interested in how quickly speakers started to articulate these sentences. There are a number of possibilities: speakers could retrieve the first content word prior to articulation, in which case production latencies for the two sentence types should not differ. Alternatively they could plan the entire clause prior to articulation and again production latencies should not differ. However, if speakers prefer to process the initial phrase prior to articulation, then they should take longer to begin the sentences with initial coordinate NPs than the sentences with initial simple NPs.

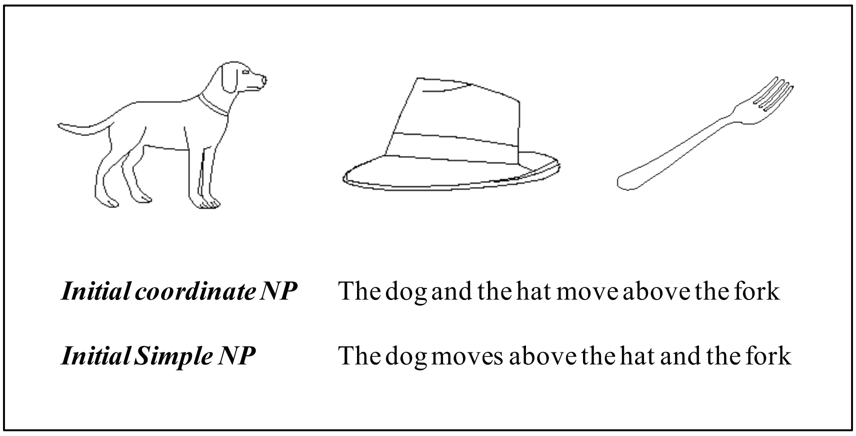

Figure 1: A picture array from Smith and Wheeldon (1999) and examples of the experimental sentences it was designed to elicit

On each trial, the three pictures were presented simultaneously and began to move immediately either up or down. The movement covered 2.5cm of the screen and was completed in 600ms. The pictures were removed from screen 500ms after response completion and there was a two second inter trial interval. A set of 92 different pictured objects were used and many filler trials were included, with different numbers of pictures and sentence structures, to maximise visual, conceptual and syntactic variability from trial to trial.
In this experiment (and the other experiments discussed below), about 30 participants were tested. They were asked to begin to produce the sentences as quickly as possible without making errors, pauses or hesitations. These instructions reduce variation in the data and encourage speakers to generate their preferred minimal amount of structure prior to speech onset. As we were interested in how speakers produce error free, fluent sentences, all trials with errors and disfluencies were excluded from the analyses of the sentence production latencies.
We observed significantly longer latencies (77ms) to sentences beginning with a coordinate NP than to sentences beginning with a simple NP. This finding is not consistent with a lexical processing scope. Speakers could have initiated sentence production having retrieved the first noun in both sentence types. Instead however, the latencies show that they dedicated more processing time to the initial coordinate NPs than the initial simple NPs. For the same reason, this finding is not consistent with a clausal processing scope as more time is dedicated to elements

within the first phrase than to the rest of the clause prior to speech onset.

However, this experiment does not reveal at which level of processing a phrasal scope is operating, as the coordinate NP is visually, conceptually, grammatically and phonologically more complex than the simple NP. Further experiments suggest that visual complexity does not contribute significantly to the effect, as the latency difference is not observed when speakers simply name the moving pictures from left-to-right (Wheeldon and Meyer, 2005). Therefore the visual tracking of single or double picture movements does not contribute to the effect. The translation of a visual scene into a spoken sentence is therefore an essential component of the effect.

The issue of phonological complexity is more difficult to deal with. Sentences are produced with rhythm and intonation i.e., prosodic structure, and these structures often coincide with syntactic units. Therefore effects of prosodic and syntactic units are difficult to disentangle. However, there is good evidence the scope of phonological encoding prior to speech onset is at most a single phonological word - a unit that can be larger than a lexical word but with a single stressed syllable (Levelt, 1989; Levelt et al., 1999; Wheeldon and Lahiri, 1997, 2002). The initial phonological word is indeed bigger in the coordinate sentences e.g., [the dog and the]$_{\omega}$ than the simple sentences e.g. [the dog]$_{\omega}$, although effects of phonological complexity are usually much smaller (15-20ms) than the effects we have observed (Wheeldon and Lahiri, 1997, 2002). Moreover the effect remains when sentences are matched for phonological structure (Allum and Wheeldon, 2007) (Experiment 1, discussed below).

In addition, it remains possible that the processing scope is not phrasal but lexical and that speakers simply prefer to retrieve the first two content words of a sentence before they begin to speak. The second content word of the simple NP sentences is always the verb "moves" whereas the second content word in the complex NP sentences is a different picture name on each trial. The retrieval of the first two content words of the sentences beginning with the simple NPs would, therefore, be easier than for sentences beginning with complex NPs. To rule out both of these possibilities, Allum and Wheeldon (2007, Experiment1) replicated the experiment in Japanese (see also Martin et al., 2004). We used coloured picture displays to elicit simple and coordinate sentences. Unlike English, the verb final properties of Japanese allow the subject and a complement to take the first two positions in a sentence, as can be seen in the example sentences in (3). Japanese is a topic-comment language and the topic is marked by the particle *wa*. This means that from left-to-right, the Japanese sentences are perfectly matched for lexical and phonological complexity. Nevertheless, Japanese speakers were significantly slower to initiate the coordinate NP sentences than the simple NP sentences (by 32ms). The finding replicates the English experiment described above and provides strong support for the claim that the latency effect is due to the structure of the sentence initial phrase rather than to differences in lexical and/or phonological structure.

3. a) [INU to BOUSHI wa] FOOKU no ue ni arimasu
   [Dog and hat TOP] fork above are
   The dog and the hat are above the fork

   b) [INU wa] BOUSHI to FOOKU no ue ni arimasu
   [Dog TOP] hat and fork above is
   The dog is above the hat and the fork

2.2 Conceptual or grammatical units?

The experiments reviewed so far provide evidence that a phrasal scope operates at some early conceptual or grammatical level of processing during sentence production. However, the sentences tested so far cannot reveal at which level of processing this scope operates. This is because they don't allow us to distinguish between different grammatical and conceptual structures. Both the simple and the coordinate sentence initial NPs in our sentences correspond to two major grammatical units: the subject phrase and the head of the subject phrase. They also correspond to a major conceptual unit: the theme or actor.

In order to distinguish between the grammatical options, Allum and Wheeldon (2007, Experiment 2) tested whether the entire subject phrase is the preferred minimal scope of processing prior to speech onset. In this experiment two sentence types were compared (see Example (4) below). The subject phrase in sentence (4.a) again comprises a coordinate

NP with both nouns acting as hierarchically equal heads. In sentence (4.b) the subject phrase is more complex, comprising a head noun phrase and a modifying prepositional phrase (PP).

4. a) Coordinate NP sentences [The fork and the dog] are blue

   b) PP sentences [The fork above the dog] is blue

If the planning scope is the first complete verb argument phrase (i.e., the whole subject phrase), then speech onsets for the two sentence types should be similar, or perhaps longer for the PP sentences, due to their greater syntactic complexity. If the planning scope can span a unit smaller than a whole verb argument phrase, such as the head of the subject phrase, then latencies to the coordinate NP sentences should be longer than to PP sentences. We found that latencies to coordinate NP sentences were significantly longer (by 102ms) than latencies to the PP sentences. This finding provides evidence that speakers did not encode the whole of a subject phrase prior to articulation but instead planned a smaller unit such as the head of the subject phrase.

However, we still do not know if the planning unit is determined by conceptual or grammatical factors. Is the planning unit equal to a major conceptual unit (e.g. the theme) or simply the first grammatical phrase to be produced? English is a head-initial language, which means that the head of the subject phrase comes first. The head of the subject phrase also plays a central thematic role - as the agent or theme of the message. So for English, conceptual and grammatical units are confounded. In contrast, Japanese is a head-final language, therefore within the subject phrase, subsidiary phrases occur before the subject head. For example, in the Japanese sentence *the fork above the dog is blue* the modifying PP *above the dog* precedes the subject head *the fork*. Allum and Wheeldon (2007) used the head-final characteristics of Japanese to determine whether processing scope corresponds to the sentence initial grammatical phrase or the head of the subject phrase and theme (Experiment 3 in their study).

In this experiment, we again used colour coded pictures to elicit the sentences we wanted. In these sentences we progressively extended the initial PP relative to the head phrase, while keeping the overall subject phrase length the same (see Example 5).

5. a) KANI no ue no YAKAN to TSUKUE to PANDA wa aka DESU
   [crab above] kettle and desk and panda TOP red ARE
   The kettle and the desk and the panda above the crab are red

   b) KANI to YAKAN no ue no TSUKUE to PANDA wa aka DESU
   [crab and kettle above] desk and panda TOP red ARE
   The desk and the panda above the crab and the kettle are red

   c) KANI to YAKAN to TSUKUE no ue no PANDA wa aka DESU
   [crab and kettle and desk above] panda TOP red IS
   The panda above the crab and the kettle and the desk is red

If Japanese speakers can initiate articulation having processed the first grammatical phrase then sentence production latencies should increase as the length of the PP increases (from 5.a to 5.c). However, if they have to process up to the head of the subject phrase, a major thematic unit, then latencies to all sentence types should not differ as the subject phrases in all sentences are matched for overall length and complexity. We found that sentence onset latencies increased by approximately 50ms with each noun added to the sentence initial PP. This result suggests that the initial syntactic phrase is more thoroughly processed at some stage than the whole subject phrase.

The experiments described so far allow us to draw a number of conclusions. First, the scope of the process we are tapping into is not the whole subject phrase or theme. Instead, speakers can initiate sentence production having processed a sentence initial phrase. Moreover, sentence onset latency is determined by the size of this initial phrase irrespective of the thematic or grammatical role it plays in the sentence to be produced; similar effects are observed in Japanese and English, despite differences between the two languages in the ordering of head and subsidiary phrases. Critically, these ordering differences are determined by *grammatical convention* alone and do not mirror the conceptual saliency of the information. It would be odd to suggest that the information in a subsidiary preposition phrase was conceptually more salient to Japanese speakers than the topic of the sentence. These findings therefore suggest that the locus of the effect is in the process of *grammatical* encoding.

2.3 The scope of lexical access

We turn now to the question of the scope of lexical access. Does the phrasal scope we have shown so far determine the scope of lexical access prior to articulation? This is an important question as the answer has consequences for the modeling of grammatical encoding processes. Critically, if the scope of grammatical encoding is a phrase and this scope controls lexical access in advance of articulation then how does a non-linear conceptual message ensure the correct order of lexical access for output?

In traditional, lexically-driven models, the order of lexical activation is driven by the conceptual weighting of lexical concepts (Bock and Levelt, 1994; Levelt, 1989, 1992; Pickering and Branigan, 1998). So for example, in a given thematic structure, the agent or theme might be the most salient concept. This would cause its associated lexical item to become most highly activated and therefore be assigned to the subject role in the sentence to be produced. Once this grammatical function assignment has occurred, syntactic structure can be built in an incremental fashion. For example consider again the sentence *Anne kissed Peter*. If the agent *Anne* was conceptually most salient then the lexical representation of *Anne* would become highly active and be assigned the subject role, resulting in an active sentence. However, if for some reason the patient *Peter* was conceptually most salient, then its lexical representation would receive the most activation causing it to be assigned the subject role, resulting in the passive sentence *Peter was kissed by Anne*.

However, conceptual weighting cannot explain the ordering of activation of lexical items for a sentence initial modifying phrase (e.g., *above the dog*). In this situation, no clear conceptual weighting factor such as salience, animacy or focus is in operation. The required order of activation is based on a purely grammatical convention. So if, prior to utterance onset, lexical access occurs only for an initial subsidiary phrase, then it must be possible for syntactic processes to interact directly with conceptual processes to ensure the correct order of activation of lexical items. Data showing a phrasal scope of lexical access can, therefore, provide evidence against a lexically mediated approach to the generation of syntax and for an approach that allows word ordering prior to lexical access.

Recently models have been proposed that allow syntactic structure to be generated based on thematic structure without reference to lexical content (Chang, 2002; Chang et al., 2000, 2006). In these models, the process of lexical selection is kept completely separate from syntactic sequencing decisions. The selected lexical concepts activate their associated lexical items but this activation process is blind to the thematic role of the lexical concept. This means that the word for the concept *Anne* will be activated in the same way regardless of the thematic role it plays in the message. The most appropriate word order is determined by the thematic structure and the speakers acquired syntactic knowledge. In this model, therefore, syntactic structure can be generated prior to, and independently of, lexical access.
Allum and Wheeldon (2009) argued that if the scope of grammatical encoding defines the scope of lexical access, then in the PP subject phrases only the first noun should be retrieved prior to speech onset for both Japanese and English speakers - irrespective of the role it plays in the sentence. We tested this claim using a picture preview technique. The sequence of events on each trial is shown in Figure 2.

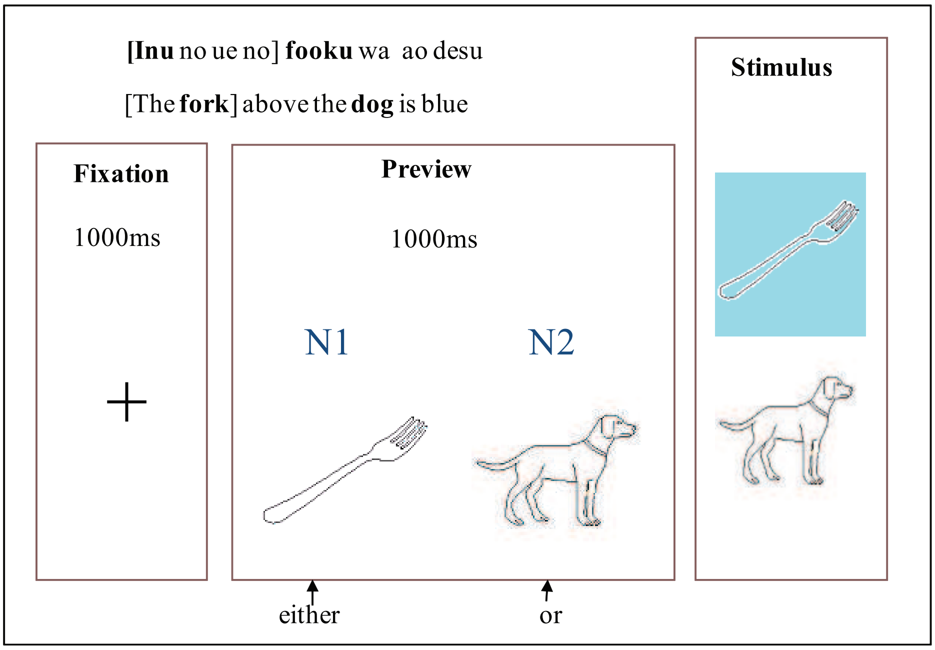

Figure 2: Sequence of events on a picture preview trial (Allum and Wheeldon, 2009).

On each trial, speakers first saw a fixation cross. This was replaced for one second by either a blank screen or by one of the pictures subjects would have to use in the sentence they were about to produce. Finally subjects saw the visual display they should describe and their speech onset latencies were recorded. Pictures were previewed on one third of all trials and speakers knew that the previewed picture would always occur in the upcoming sentence to be produced. Again a large number of filler trials were used to vary the syntactic structures produced as well as the visual and sentence positions of the previewed pictures. We reasoned that, if during the production of the target sentences, the preview picture name was retrieved prior to speech onset, then we should see a preview benefit in the sentence production latencies. In other words, if you need to retrieve the word *fork* before you can start your sentence, then having already retrieved its name during the preview period should speed you up. The target sentences included a simple subject phrase head and a subsidiary PP (see Figure 2). Pictures were previewed in the first or the second phrase of the target sentences and the same experiment was run in both English and Japanese. The results are summarised in Table 1. As can be seen, sentence production latencies were significantly faster when the sentence initial picture was previewed. However, there was no significant facilitation due to the preview of a picture that occurred in the second phrase. This pattern of results was the same for English and Japanese speakers.

Table 1: The results of Allum and Wheeldon (2009), Experiments 4 and 5: Picture preview in prepositional phrase sentences in English and Japanese. Mean RTs and percentage error rates (in parentheses) are shown for the three Picture preview conditions, as is the difference between unpreviewed and previewed conditions. Significant effects are marked with an asterisk.

	JAPANESE		**ENGLISH**	
Condition	RT (%err)	RT Diff	RT (%err)	RT Diff
No Preview	916 (5.2)		1080 (8.3)	
Preview (N1)	740 (5.5)	176*	992 (6.7)	88*
Preview (N2)	908 (7.0)	8	1084 (8.7)	-4

This pattern of results mirrors that of the scope experiments reviewed in the previous section, in that it is consistent with the claim that only the lexical items for the first phrase of a sentence need to be retrieved prior

to speech onset. However, it is possible that this is a purely methodological effect. Perhaps previewing non-initial information simply has the effect of confusing speakers thereby slowing them down and masking any preview benefit for the second noun. Alternatively, if the second noun to be produced is highly active due to preview, it might compete with the first noun to be produced for selection, again slowing sentence onset latencies.
However, if the scope of lexical access is really determined by phrase structure then we should see a different pattern of preview effects for coordinate NP sentences (Example 6), which should show a facilitatory effect of preview for both first and second nouns.

6. a) [Inu to fooku wa] ao desu

   b) [The dog and the fork] are blue

The results of the coordinate NP experiments are shown in Table 2. The pattern of results is very different to that for the prepositional phrase sentences. Now there is a significant facilitation of sentence production latencies due to preview of both first and second nouns and once again the results for Japanese and English speakers are very similar. The preview effect for initial nouns is significantly larger than that for second nouns, showing an added benefit of preview for words to be produced first. Critically, however, the second noun also shows a preview benefit when it occurs in the first coordinate NP of the sentence. Together, the preview experiments provide evidence that speakers retrieve the nouns for the first phrase of a sentence before they begin to produce it.
The final experiment I want to describe was designed to rule out an alternative explanation for the findings of the preview experiments described above. In these experiments, the required sentence structures were elicited using displays in which the phrase constituency of the pictured objects was determined by colour grouping. Colour is a very salient perceptual feature. Is it possible therefore, that the preview effect was determined by colour grouping rather than by grammatical grouping? We were able to test this possibility in Japanese because this language has two different forms of coordination. The experiments described so far made use of *To..wa* coordination, as in Example 6 above.

Table 2: The results of Allum and Wheeldon (2009), Experiments 1 and 3: Picture preview in coordinate NP sentences in English and Japanese. Mean RTs and percentage error rates (in parentheses) are shown for the three Picture preview conditions, as is the difference between unpreviewed and previewed conditions. Significant effects are marked with an asterisk.

	JAPANESE		**ENGLISH**	
Condition	RT (%err)	RT Diff	RT (%err)	RT Diff
No Preview	1084 (8.0)		955 (11.2)	
Preview (N1)	962 (7.0)	122*	835 (6.7)	120*
Preview (N2)	1036 (7.5)	48*	895 (6.9)	60*

To typically binds the two items closely as a set and *wa* often functions contrastively. The coordinated items are therefore bound as a unit and contrasted with other items. Alternatively, it is possible in Japanese to use *mo..mo* or a 'listing' coordination as in Example 7 below. This form of coordination has a looser binding both conceptually and syntactically and is not contrastive in function.

7. a) [Inu mo fooku mo] ao desu
   The dog and the fork are blue

The items in the *mo..mo* coordination play the same thematic role as in the *to..wa* coordination but the concept is different. In the sentence *Keio University and Waseda University are excellent* the *to..wa* coordination would imply that these universities are both excellent in contrast to others. The *mo..mo* coordination would imply that these two Universities are excellent as well as other universities. The two kinds of coordination also have different effects on the scope of application of adjectives. In the sentence *The red chair and the desk are made of wood* the *to..wa* coordination would extent the colour adjective to apply to the second noun whereas the *mo..mo* coordination would not. These differences led us to think that the two forms of coordination might also differ in their scope of lexical access with the two constituent simple NPs in the *mo..mo* coordination behaving as separate units. We therefore reran the Japanese coordinate NP preview experiment but simply asked speakers to use *mo..mo* coordination. Importantly, however the same colour coded visual display was used to elicit the *mo..mo* coordination as had been used to elicit the *to..wa* coordination. The results can be seen in Table 3.

Table 3: The results of Allum and Wheeldon (2009), Experiment 2: Picture preview in mo…mo coordinate NP sentences in Japanese. Mean RTs and percentage error rates (in parentheses) are shown for the three Picture preview conditions, as is the difference between unpreviewed and previewed conditions. Significant effects are marked with an asterisk.

Condition	RT (%err)	RT Diff
No Preview	1006 (4.7)	
Preview (N1)	934 (3.9)	72*
Preview (N2)	1036 (2.3)	-30

Mo..mo coordination yielded a very different pattern of results to *to..wa* coordination despite the use of visually identical displays to elicit the sentences. First noun preview again significantly speeded sentence production latencies compared to no preview. However, second noun preview slowed sentence production latencies, although not significantly. This pattern of results provides strong support for our claim that syntactic rather than visual grouping determines the scope of lexical access during sentence production.

3 Conclusion

The series of experiments I have reviewed provide evidence in support of a number of claims about the scope of advanced planning during spoken sentence production. First, they provide strong evidence that language production is incremental. All of the experiments show that more processing time is devoted to the beginning of an utterance prior to speech onset, than to the utterance as a whole. These experiments also provide information about the nature of the minimal unit speakers prefer to plan before beginning their utterances. This unit proved to be a phrasal chunk, which did not correspond to a minimal syntactic phrase such as the first noun phrase of a coordinate NP. The experiments show repeatedly, that speakers preferred to process the whole of a coordinate NP prior to speech onset, as latencies to sentences beginning with coordinate NPs were consistently slower than latencies to sentences beginning with simple NPs. The data also show that the processing unit does not necessarily correspond to a major grammatical unit such as the subject phrase or its head element. Nor does it correspond to a major thematic unit such as agent or theme. The Japanese experiments demonstrate that speakers can initiate their utterance having processed a sentence initial prepositional phrase, which plays a subordinate role both thematically and grammatically. It is not necessary to access the lemma for the head of the subject phrase prior to speech onset. Crucially, the sentence initial placement of the preposition phrase in Japanese is a *grammatical* convention determined by the head-final syntax of the language. We therefore argued that the phrasal scope of processing we observed operates at the level of grammatical encoding. Nevertheless, it remains problematic to find a satisfactory definition for the processing unit our experiments have identified (see Allum and Wheeldon, 2007, for a discussion). Although the evidence suggests that the unit operates at the level of grammatical encoding, its definition seems to require some reference to thematic structure as there is no clear syntactic definition. The unit is not a minimal syntactic phrase which would be a simple NP (e.g., *The dog*) in all of the sentences we have tested. Instead, it seems to be a phrasal unit corresponding to a thematic unit in the message - not necessarily a verb-argument role but perhaps minimal thematic unit such as a modifier. The experiments testing the different forms of coordination in Japanese manipulated the degree of both conceptual and syntactic binding. It remains unclear how each of these levels of structure contribute to different effects observed for the two forms.

Our experiments also provide information about the relationship between conceptual and grammatical encoding. The picture preview experiment tested the scope of lexical access prior to sentence onset. These experiments provided evidence to suggest that the phrasal scope, identified in previous experiments, determines the extent of lexical access prior to articulation. This finding is critical for the modelling of sentence production as it supports a model in which syntactic linearization processes precede lexical access processes, with the former guiding the latter.

References

Allum, P. and Wheeldon, L. R. (2007). Planning scope in spoken sentence production: The role of grammatical units. *Journal of Experimental Psychology: Learning Memory and Cognition*, 33(4):791–810.

Allum, P. and Wheeldon, L. R. (2009). Scope of lexical access in spoken sentence production: Implications for the conceptual-syntactic interface. *Journal of Experimental Psychology: Learning Memory and Cognition*, 35(5):1240–1255.

Bock, J. K. and Cutting, J. C. (1992). Regulating mental energy: Performance units in language production. *Memory and Language*, 31:99–127.

Bock, J. K. and Levelt, W. J. M. (1994). Language production: Grammatical encoding. In Gernsbacher, M., editor, *Handbook of Psycholinguistics*, pages 945–984. Academic Press, New York.

Butterworth, B. (1980). Evidence from pauses in speech. In Butterworth, B., editor, *Language Production: Vol.1. Speech and Talk*. Academic Press, London.

Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science*, 26:609–651.

Chang, F., Dell, G. S., Bock, J. K., and Griffin, Z. M. (2000). Structural priming as implicit learning: A comparison of models of sentence production. *Journal of Psycholinguistic Research*, 29:217–229.

Chang, F., Dell, G. S., and Bock, K. (2006). Becoming syntactic. *Psychological Review*, 113:234–272.

Ferreira, F. and Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, 46:57–84.

Ferreira, V. S. and Slevc, L. R. (2007). Grammatical encoding. In Gaskell, M. G., editor, *The Oxford Handbook of Psycholinguistics*, pages 453–470. Oxford University Press.

Garrett, M. F. (1980a). Levels of processing in sentence production. In Butterworth, B., editor, *Language Production: Vol.1. Speech and Talk*. Academic Press, London.

Garrett, M. F. (1980b). The limits of accommodation: Arguments for independent processing levels in sentence production. In Fromkin, V. A., editor, *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen and Hand*. Academic Press, New York.

Goldman Eisler, F. (1968). *Psycholinguistics: Experiments in Spontaneous Speech.* Academic Press, New York.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, 82:B1–B14.

Griffin, Z. M. and Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, 11:274–279.

Kempen, G. and Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, 11:201–258.

Konopka, A. E. and Bock, J. K. (2008). Lexical or syntactic control of sentence formulation? Structural generalizations from idiom production. *Cognitive Psychology*, 58:68–101.

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

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, 42(1-3):1–22.

Levelt, W. J. M. and Maassen, B. (1981). Lexical search and order of mention in sentence production. In Klein, W. and Levelt, W. J. M., editors, *Crossing the Boundaries in Linguistics*, pages 221–252. Reidel, Dortrecht.

Levelt, W. J. M. and Meyer, A. S. (2000). Word for word: Multiple lexical access in speech production. *The European Journal of Cognitive Psychology*, 12(4):433–452.

Levelt, W. J. M., Roelofs, A., and Meyer, A. (1999). A theory of lexical access in speech production. *Behavioural and Brain Sciences*, 22:1–75.

Martin, R. C., Miller, M., and Vu, H. (2004). Lexical-semantic retention and speech production: Further evidence from normal and brain-damaged participants for a phrasal scope of planning. *Cognitive Neuropsychology*, 21(6):625–644.

Meyer, A. S., Sleiderink, A. M., and Levelt, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, 66:B25–B33.

Morgan, J. L. and Meyer, A. S. (2005). Processing of extrafoveal objects during multiple-object naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31:428–442.

Pickering, M. J. and Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic persistence in language production. *Journal of Memory and Language*, 39:633–651.

Schriefers, H., Meyer, A. S., and Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, 29:86–102.

Schriefers, H., Teruel, E., and Meinshausen, R. M. (1998). Producing simple sentences: Results from picture-word interference experiments. *Journal of Memory and Language*, 39:609–632.

Slevc, L. R. (2011). Saying what's on your mind: Working memory effects on sentence production. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 37(6):1503–1514.

Smith, M. and Wheeldon, L. R. (1999). High level processing scope in spoken sentence production. *Cognition*, 73:205–246.

Smith, M. C. (2000). Conceptual structures in language production. In Wheeldon, L., editor, *Aspects of Language Production*, pages 331–374. Psychology Press, Hove.

Swets, B., Desmet, T., Hambrick, D. Z., and Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General*, 136:64–81.

Wagner, V., Jescheniak, J. D., and Schriefers, H. (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*, 36:423–440.

Wheeldon, L. R. and Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, 37:356–381.

Wheeldon, L. R. and Lahiri, A. (2002). The minimal unit of phonological encoding: prosodic or lexical. *Cognition*, 85:B31–B41.

Wheeldon, L. R. and Meyer, A. (2005). Planning sentence structure: Speech latency and gaze patterns during the production of word lists and sentence. In *Paper presented at the 4th Workshop on language production. Maastricht*.

Wundt, W. (1900). *Die Sprache.* Kroner, Leipzig.