Processing Load in Children’s Language Production: A Clinically Oriented Review of Research

Monique Charest
Judith R. Johnston

Abstract
Investigations of the way that processing load variations may affect the form, content and adequacy of children's utterances are surprisingly limited. The purpose of this review article is to identify and synthesize the primary findings from the broader literature on language production, including studies of adults and children with normal and impaired language, that can provide insights into the sources and effects of processing load on children's productions. The evidence in this literature points to three important characteristics of language production processing: (1) The total costs of a speaking situation can exceed a speaker's processing resources, (2) language operations can vary in their costs, and (3) processing costs in one domain of language can affect performance in another. The final sections of this paper discuss the implications of these characteristics of language processing for interpreting children's language productions and supporting their communicative success.

Key words: Language production, processing load, limited capacity, dual task, tradeoffs and familiarity
A number of clinically relevant questions arise from these assumptions. Can the processing load inherent in a particular utterance exceed a child’s mental resources? Do different language operations have measurable costs, and can they vary? If so, can the costs of different operations influence each other? Our review of the wider literature on language processing will attempt to answer these questions. In the final section, we will suggest ways in which this view of processing load can inform clinical practice.

**Question 1: Can the processing load inherent in a particular utterance exceed a child’s mental resources?**

Our first question focuses on the possible role of “capacity limits” in language production. According to capacity perspectives on cognition, although knowledge may be infinite, the resources needed to attend to, activate, manipulate, and hold information in mind are not. As such, there are limits to the amount of cognitive work that can be completed at any given time (Kail & Bisanz, 1982; Kail & Salthouse, 1994). These limits reflect the amount of mental “fuel,” or “resources,” that we bring with us to a task (e.g., Just & Carpenter, 1992), and the speed and efficiency with which those resources are used (Kail & Salthouse, 1994). The fact that mental capacity has limits implies that, even if we have the necessary knowledge, we may not always be able to complete the mental work needed in a given situation. Our performance will depend on the relationship between costs and available resources. When the total amount of work needed to complete some activity – the total processing cost – reaches or exceeds resource limits, effects may be seen in the accuracy or adequacy of performance.

To illustrate with an example from everyday life, consider driving. To get from point A to point B, we need to think about and plan a route to where we are going, take in incoming information (e.g., a car approaching quickly in the next lane, an upcoming traffic light that just turned yellow, a car ahead slowing down), and respond accordingly. Often, we are able to complete all of this work while carrying on a conversation, sipping a coffee, or thinking through a problem from work. This extra activity may be scaled back or put on hold, however, if we suddenly find ourselves negotiating a left-hand turn in heavy traffic. What was previously a manageable amount of work or processing load has now, with the added demands of planning a challenging maneuver, become too much to handle. We stop talking, put down the coffee, and momentarily forget our lesson plans for the day.

As is true for other highly automatized functions, we are rarely conscious of devoting mental effort to constructing sentences. However, researchers have clearly demonstrated that capacity limits in the moment of speaking do exist. Studies of capacity limits are typically designed in two ways: (1) researchers observe what happens to performance when the amount of work – the costs – associated with an activity increase, or (2) they observe what happens to
performance when resources decrease due to the need to perform a second, simultaneous task. Language production researchers have used both approaches to study capacity. They have decreased the resources available to speaking, or increased the costs of speaking.

In one line of investigation, adult speakers have been required to talk while doing something else, such as walking, tapping their fingers, ignoring background noise (Kemper, Herman, & Lian, 2003; Kemper, McDowd, Pohl, Herman, & Jackson, 2006), judging whether printed words are real words or not (Barch & Berenbaum, 1994), and holding in mind other, unrelated words (Hartsuiker & Barkhuysen, 2006). All of these dual-task experiments were designed so that the speakers would have to devote some of their processing resource to the secondary task. And, all of these experiments revealed that language performance suffers in consequence. When adults talked while doing something else, they produced shorter, less complex sentences, were less fluent, and in some cases even produced more grammatical errors (Hartsuiker & Barkhuysen, 2006; Kemper et al., 2006) than they did when no second task was required. Apparently, when the resources available for the language production task were reduced by the introduction of the secondary task, the processing load of some utterance types exceeded available resources. The effects were seen in simplified, and sometimes dysfluent or ungrammatical, utterances.

In the above studies, it was the primary speaking task that suffered when a secondary task was introduced. The effects of capacity overload can also go the other way – the costs of talking can cause performance in a non-talking activity to decline. In one study, for example, speakers who performed a tracing task made more errors while speaking than when tracing in silence (Power, 1986). Interestingly, errors increased over the first clause of two-clause sentences, then decreased over the second clause. This pattern indicated that language processing costs were particularly high earlier in the sentence when more planning was required. As speakers reached the ends of sentences, less work remained to be done, and there was a decline in language processing costs.

The dual task literature points to the important conclusion that the form and content of speakers’ utterances can be determined not only by what they know, but also by resource limitations. Although the activities in these studies were contrived by the researchers, it is likely that real life situations also often require us to divide our resources between talking and another activity. Some utterances may be more or less achievable within the resources that are available.

If adults can reach capacity limits in speaking, we might expect similar evidence of capacity overload in children’s production. In fact, because children are more capacity-limited than adults (Kail & Salthouse, 1994), we might even expect to see more extensive effects. Dual-task evidence for capacity limit effects on children’s language production, however, is much more difficult to come by. To our knowledge, no research has explicitly set out to investigate changes in children’s sentence production as a function of dual task demands.

However, this gap may be more apparent than real. Although researchers may not have adopted this kind of language to talk about task effects, as clinicians we are certainly used to thinking about whether contexts are more or less facilitative for talking, and a broader reading of the developmental literature provides hints of possible “real life” dual task effects on the expressive language of children. For example, children are more responsive, take more conversational turns, and experience fewer communication failures in familiar than in unfamiliar play routines (Furman & Walden, 1990). Very young children not only talk more, but also demonstrate more varied and advanced lexical and syntactic production in familiar than novel play routines (Farrar, Friend, & Forbes, 1993). Nelson (1986) argues that, when faced with an unfamiliar event or situation, children have to construct an understanding of that event in the moment. The work of constructing that understanding while having to also talk can be thought of as a “secondary” task that requires resources and leaves fewer resources available to the work of producing language. The results may be seen in the reduced complexity or adequacy of the language that is produced.

Other aspects of the play situation may also influence children’s language use. Evans and Craig (1992) reported that school-age children with specific language impairment (SLI) produced shorter sentences in free play with toys than in elicited interview contexts with no toys, and suggested that a contributing factor to this difference may have been that the toys were distracting to the children. While playing with toys, these children were also less likely to produce those syntactic or grammatical forms that were most advanced for them, and showed more between-child variability in their language use. A limited capacity perspective on these findings would suggest that when they only needed to talk, the children were able to devote sufficient processing resources to the construction of longer utterances containing more advanced grammatical forms. However, if mental resources were also devoted to toy play, there were insufficient resources for production of the more advanced forms.

Apparent dual-task effects can also be found in other quarters. For example, typically developing 18- to 30-month-old children show the ability to coordinate joint attention between a play partner and a toy. They are also able to produce language while playing. However, they may not be able to produce language while creating joint attention (Adamson, Bakeman, & Deckner, 2004). Adamson et al. reported that children across this age range were likely to talk at points of “supported child engagement,” when the parent created a joint attending context by focusing on the child and object, but not at points when the child was coordinating attention to both mother and toy. Again, a limited capacity interpretation of these results would suggest that the combined costs of attending to both toys...
and another partner, as well as producing an utterance, exceeded the toddlers’ resource limits.

Further evidence of capacity limitations in children’s language production can be found in studies that look at cost variations within a single type of task. The most typical approach has been to examine production performance when children attempt longer and/or more complex sentences. This line of research has shown that, with increased complexity and/or length, children are more likely to produce “characteristic” child language errors, such as omitting required sentence subjects (Bloom, 1990; Grela, 2003; Valian, 1991) and omitting grammatical morphemes such as bound tense morphemes, auxiliaries and articles (Grela & Leonard, 2000; Namazi, 1996; Owen, 2010; Pizzioli & Schelstraete, 2008). Some studies have also reported increases in children’s phonological and articulatory errors with increased sentence length and complexity (see Masterson, 1997, for a review).

Some of these studies observed children’s spontaneous language production (Bloom, 1990; Namazi, 1996; Valian, 1991). In other studies, the researchers controlled the types of sentences that children would attempt, and in this way could examine the effects of particular sentence types on performance. In two studies by Grela (Grela, 2003; Grela & Leonard, 2000), for example, the experimenter acted out scenes with toys and narrated a short story that went along with the actions. When the final action was produced, the children described what was happening. The experimenter had prepared scenes that invited children to use verbs with differing numbers of objects: zero in the intransitive condition (e.g., “The bear is running”), one in the transit condition (e.g., “The cow is biting the boy”), and two in the ditransitive condition (e.g., “The pig is giving the cup to the mouse”). The results showed that children with SLI as well as younger children with typical language development produced sentences with omitted subjects (Grela, 2003) and with omitted “is” auxiliaries (Grela & Leonard, 2000) most often in the longer, more complex ditransitive sentences (e.g., “The pig __ giving the cup to the mouse”). Findings for particular sentence patterns will, of course, vary according to the language level of the speaker. An utterance that is challenging early in life may not be as challenging later on. However, from a variety of studies we have seen that the costs of speaking can potentially exceed resources throughout the lifespan.

The studies described in this section provide glimpses of the mind at work during language production, sometimes struggling to get the job done. Returning to the question that began this section, the answer is “yes.” The costs inherent in a particular utterance, possibly in conjunction with other demands of the speaking situation, can exceed a child’s mental resources. What children say, and how they say it, can be affected not only by what they know, but also by what they can manage to do in the moment. The next two sections of this review will explore the possibility that processing load and utterance form are affected not only by the total number of operations, but also by the costs of individual operations.

**Question 2: Can we identify costs related to different language operations, and can these costs vary?**

In our discussion of Question 1, we assumed that there are costs associated with producing utterances, and provided evidence that the total processing load can exceed resources. We move now to consider the costs that might be associated with separate components of utterance production, looking at both lexical and syntactic processing in adults and children. The notions of cost and total cost remain relevant but will be treated within a particular framework, one that views language production as a series of differentiated language-processing activities that are organized and unfold in time (Rispoli, Hadley, & Holt, 2008, p. 963). Within this framework (Bock, 1995; Bock & Levelt, 1994; Ferreira & Slevc, 2007; Levelt, 1989), sentence production begins with the construction of a preverbal message, intended to fulfill a particular goal such as providing or receiving information. The message content includes information such as referents (“who” or “what” entities are involved), actions and states, and how the different referents are related (Ferreira & Slevc, 2007). With the message as the starting point, speakers then formulate the lexical and syntactic plan for the sentence. Early in the formulation process, speakers select words to convey the message meaning and assign them to grammatical roles (such as subject of the sentence). These initial lexical representations (called “lemmas”) represent the meaning and grammatical category of the word, but do not carry phonological content. Later in formulation, speakers retrieve the words’ sound form representations, grammatical morphemes, and syntactic plans that specify the order of production. Once words are retrieved and their sentence positions are determined, speakers create phonological plans that then guide articulation.

Much of the support for this differentiated view of production comes from speech error and experimental research showing that different kinds of information become active and are prioritized at different times during production. For example, researchers have noted that whole words can exchange in so-called slips of the tongue, but when they do, grammatical morphemes do not tend to slip with them. They tend to remain in the appropriate sentence position (e.g., “You ordered up ending some…”; _order and end exchange, but the bound morphemes do not_, Bock & Levelt, 1994, p. 948). These “stranding” errors suggest that content words and morphemes are processed separately during production. Moreover, whether and how quickly a particular representation becomes active can be affected differentially by factors that do not seem to influence other representations. For example, early in the time course of accessing a word, the speed with which adults and children process the word is affected by distractor words that are related in meaning, but not in sound. In contrast, at a later stage in processing, speakers are affected by distractors that are related in sound but not meaning (Jerger, Martin, & Damian, 2002; Schriefers, Meyer, & Levelt, 1990). This result demonstrates that...
speakers are “focused” on meaning-based and sound-based lexical planning at different times. In addition, a common phenomenon demonstrates that syntactic processing can be influenced independently of lexical processing. Both adults and children tend to repeat the sentence structures that they have recently produced or heard, even when the current utterance contains different content words, function words or grammatical morphemes than the previous utterance (Bock, 1986; Bock, 1989; Bock & Loebell, 1990; Bock, Dell, Chang, & Onishi, 2007; Huttenlocher, Vasilyeva, & Shimp, 2004; Miller & Deevy, 2006; Pickering & Ferreira, 2008; Shimp, Gámez, Huttenlocher, & Vasilyeva, 2007). A speaker might describe a picture as “the lady is giving candy to the boy” rather than “the lady is giving the boy candy” if he or she has recently said or heard “the pitcher is throwing the ball to the first baseman” or “the girl baked a cake for her friend.” This framework is based on research with adults. However, studies that have been conducted with children, while few in number, find evidence for the same architectures and processes as seen with adults (e.g., Huttenlocher, et al., 2004; Jerger, Martin, & Damian, 2002; Stemberger, 1989). Very young children may prove to be the exception to the rule, but research with this group has barely begun.

The model of Levelt and colleagues (Bock, 1995; Bock & Levelt, 1994; Ferreira & Slevc, 2007; Levelt, 1989) is particularly useful for thinking about the architecture of production, that is, the various components, their character and order. To complete our view of language processing, we also need to consider the cognitive mechanisms involved in the representation and deployment of language schemes. Connectionist models provide us with strong interpretive tools for this aspect of production (e.g., Dell, 1986). Goldrick (2007) points to two connectionist principles that can direct our thinking about language processes. First, language representations are patterns of activation within a network of connected units. Second, “processing” is the spreading of activation among connected nodes or units, and the selection or retrieval of particular activated units. Whether and how quickly a given representation is activated and selected depends on its current activation level, the strength of input from connected nodes, and the timing and amount of activation relative to other language forms that could potentially be selected.

Processing models thus invite us to think not only in terms of what speakers know about different domains of language, but also in terms of when, how, and how easily these kinds of knowledge are deployed in the time course from incipient idea to articulated utterance. In previous influential work, researchers have assumed that message creation (i.e., “thinking”) is effortful or resource demanding, but have been less certain that other, specifically linguistic processes are costly (Bock, 1982; Levelt, 1989). However, in recent years, researchers have questioned this view (Cook & Meyer, 2008; Ferreira & Pashler, 2002; Hartsuiker & Barkhuysen, 2006; Smith & Wheelon, 2001). Indeed, there are good reasons to think that the entire production process is constrained by varying costs.

In the realm of lexical processing, for example, we know that words differ in how long they take to be accessed and produced. In studies of picture naming, both adults and children name words faster if they are learned earlier, are more frequent in the language, and have fewer synonyms that can compete for selection (e.g., Alario, Ferrand, Laganaro, New, Frauenfelder, & Segui, 2004; Cycowicz, Friedman, Rothstein, & Snodgrass, 1997; D’Amico, Devescovi, & Bates, 2001; Ellis & Morrison, 1998). They also name words faster if they have recently heard or produced the word itself or a related word (e.g., Lupker, 1988; Pellowski & Conture, 2005; Wheelon & Monsell, 1992).

These findings point to factors that can increase the time requirements of lexical processing. Temporal changes of this sort can be thought of in terms of costs within a limited capacity system. Recent studies with adults have shown that words that require longer processing at the level of meaning or sound (Cook & Meyer, 2008; Ferreira & Pashler, 2002) consume greater resources and thus place greater limits on the performance of other, concurrent activities. Ferreira and Pashler, for example, manipulated whether or not the prior context facilitated activation of the word’s meaning in a naming task. Speakers were faster to name pictures that were preceded by a semantically facilitating sentence stem than a neutral stem (e.g., “He was tired, so he went to _____” / “Here is a picture of a _____” / “Bed”). When the context provided no support, speakers were both slower to name the picture, and demonstrated worse performance on a concurrent tone monitoring task. As with the dual task studies reviewed earlier, the performance decrements in the secondary task tell us that the lexical work consumed resources, and that the more difficult words entailed greater costs.

The costs of syntactic processing are more difficult to identify than the costs of lexical processing since sentence structures, unlike words, cannot be observed in isolation. One naturally occurring language pattern has, however, provided experimental access to abstract syntax. Recall that speakers’ recent language experience somehow facilitates or biases their current production. Speakers tend to repeat sentence structures. By manipulating the patterns that a speaker has just produced or heard, researchers can “prime” the use of a particular structure. The results of syntactic priming studies are particularly interesting from the point of view of processing costs. When speakers re-use a particular sentence structure, they can begin speaking more quickly even if they are not using the same content words. These time savings are observed in studies with both adults (Smith & Wheelon, 2001; Wheelon & Smith, 2003) and children (Anderson & Conture, 2004). Likewise, when speakers produce the same sentence structure many times in a given context, as in an experiment, they come to speak more fluently, with fewer pauses or disruptions (Bock & Loebell, 1990). These results indicate that specific syntactic operations demand time or resources, and that by using primed structures, speakers economize on the work needed to access or build the abstract sentence frame. That
is, the cost of that operation is reduced.

Beyond the effects of an immediate repetition of sentence structure, we can see experience-dependent changes in children's syntactic processing across a longer time frame. For example, when children produce sentence structures that are at the upper edge of their developmental ability, they are less fluent. With age, the “comfort zone” of fluently-produced sentence types expands (Rispoli & Hadley, 2001). That is, structures that cause dysfluency in younger children come to be smoothly and easily produced by older children. Rispoli and Hadley proposed that children may attempt any of the sentence structures that are within their knowledge base, but may run into processing difficulty with those structures that have been recently acquired. Since those sentence structures have received little practice, they may be more costly. Wijnen (1990) described a child who showed a sharp rise followed by a sharp decline in dysfluency over a period of several months. During the fluent period, this child relied heavily on a very small number of syntactic patterns. It seemed that the repetitive experience with a limited number of sentence patterns facilitated fluent production. These results support the idea that long-term experience with a specific syntactic structure reduces the processing cost when that structure is produced.

To return to our question, then, the answer seems to be “yes.” The idea of “cost” can apply to separate areas of language operation as well as to the utterance as a whole, and even within a domain, costs can vary. The research on picture naming and syntactic priming shows that a given word or sentence structure can be more or less costly depending upon prior experience, either long-term (e.g., age of acquisition) or from moment to moment (e.g., recent priming). These effects of experience can be interpreted through the connectionist principles of activation and activation spreading. First, once a language form is activated, residual activation over the short-term can provide a “head start” that facilitates subsequent processing. Second, experience results in the strengthening of connections among representations in the production pathway. Stronger connections provide greater activation input and over time facilitate certain activation patterns (Goldrick, 2007). Residual activation and stronger connections can thus result in one form being more likely to be selected than another or to be selected more quickly or easily, possibilities that have been noted for both lexical (e.g., Wheeldon & Monsell, 1992), and syntactic processing (e.g., Pickering & Ferreira, 2008; Smith & Wheeldon, 2001). The resultant reduction in processing cost may free up mental resources for other tasks. We turn now to consider this possibility.

**Question 3: Can the costs of one aspect of language processing influence other aspects?**

We have just seen that the costs associated with different parts of the language production process can vary independently. Now we ask whether the production costs associated with one area can hinder or assist the successful processing of other aspects of an utterance. Levelt and colleagues’ model tells us not only that language production involves multiple activities operating in real time, but also that these activities run in parallel, each working on a different piece of the utterance (Bock, 1995; Bock & Levelt, 1994; Ferreira & Slevc, 2007; Levelt, 1989). A speaker might begin producing the initial part of a sentence as soon as part of the message is formed, while continuing to plan the rest of the message. Or, a speaker will begin to articulate the first words of a sentence before the final words have been retrieved (e.g., Meyer, 1996). This means that speakers concurrently plan messages, activate lexical representations, activate syntactic representations, and speak. They perform multiple language processing activities simultaneously. From a limited capacity perspective, this raises the possibility that a speaker who commits resources to a costly operation (such as accessing a particular word or a particular sentence frame) may be limited in the resources available to other aspects of the sentence. The speaker may sacrifice or scale back processing elsewhere, possibly omitting certain pieces of information, or settling on easier, less costly alternatives (Bock, 1982; 1995; Crystal, 1987; Just & Carpenter, 1992). This is not to suggest that such prioritization decisions occur at a conscious level. As speakers, we are not usually aware of all of the decisions that go into producing a sentence (even though we do sometimes pause to make particular decisions). However, even though these “decisions” may occur without our awareness, within a limited capacity system, decisions in one domain may have implications elsewhere. This phenomenon has often been referred to as a “processing resources tradeoff.” That is, when resources for achieving success in all domains are not sufficient, the speaker “tolerates” lesser performance in one area to allow greater performance elsewhere.

In the realm of lexical costs, young children omit more subjects and produce shorter spontaneous sentences when using more recently acquired verbs compared to familiar verbs (Bloom, Miller, & Hood, 1975). And, when young children imitate sentences containing unfamiliar verbs and/or nouns, they are more likely to omit articles (e.g., “the”; Boyle & Gerken, 1997). These results suggest that when children commit resources to accessing less familiar words, the effort may leave them with insufficient resources to plan and produce other sentence elements. Both of these studies focused on 2-year-old children. Older children, too, may very well demonstrate tradeoffs related to lexical processing costs. Although more research is certainly needed, hints of tradeoff effects can be found. For example, Masterson and Kamhi (1992) reported that school-age children with and without language impairments produced shorter sentences when the sentences contained phonologically complex words. The authors noted that the complex words were also less familiar than simple words, suggesting that phonological planning and/or lexical access costs contributed to decrements or simplifications in children’s syntactic performance.

Similarly, variations in syntactic costs can influence performance elsewhere. We mentioned earlier that when
young children spontaneously produce utterances with increased syntactic complexity they tend to make more errors in morphology and phonology. But syntactic influences can be facilitative as well. Preschool-aged children with SLI and typical language development produce grammatical morphemes more successfully when they re-use a recently produced syntactic frame (Leonard, Miller, Deevy, Rauf, Gerber, & Charest, 2002; Leonard, Miller, Greila, Holland, Gerber, & Petucci, 2000). Leonard et al. (2000) argue that these performance improvements occurred because the prior activation of the sentence frame decreased the resources needed for syntactic processing, resulting in more resources being available for morphological processing. When children can economize on the work needed to access or build sentence frames, the benefits can be seen not only in time savings (Anderson & Conture, 2004), but also in sufficiency of processing elsewhere. Conversely, these results suggest that when activation support is not available, the costs of syntactic “production operations” (Leonard et al., 2000) can negatively affect performance in other aspects of the sentence.

Similar costs and benefits may be seen from long-term familiarity. Vasileyeva, Huttenlocher, and Waterfall (2006) examined sentence production by children who had heard stories containing many active sentences or many passive sentences over a period of several weeks. In a production task that followed, children’s listening experience influenced both their choices of, and success with, sentence structure. The children who had had relatively little exposure to passive sentence structures were less likely than the other children to choose passive sentence structures. Under certain pragmatic conditions, however, they did produce passives. When they did, compared to their peers they made more grammatical errors, including grammatical morpheme omissions.

Finally, language production models (Bock & Levelt, 1994; Levelt, 1989) explicitly assume that message level planning overlaps with language formulation processes. More costly reasoning, decision-making, or discourse planning might leave fewer resources available for language planning. Indeed, there is some evidence that children produce grammatical morphemes less successfully in speaking situations that are likely to be more challenging at the message level. Thordardottir (2008), for example, reported that school-aged children with SLI produced more grammatical morpheme errors when providing explanations or retelling stories than in conversation. Bound morpheme errors are also more frequent when children retell a story from memory than with visual support (Masterson & Kamhi, 1991).

To return to the question of whether the costs of one aspect of language processing can influence others, the answer appears to be “yes.” A small body of research shows that when children produce language forms that are less familiar, or are not supported by recent activation, they are more likely to produce errors or simplifications elsewhere. And, when the discourse context is more challenging, they are more likely to omit or produce errors in grammatical morphology.

It is important to note, however, that increases in costs in one domain will not always produce the same or predicted effects in all situations or across all speakers. Whether particular processing tradeoffs occur will depend on individual differences in cost and resource allocation “decisions.” For example, Thordadottir (2008) demonstrated that although English-speaking children’s morphology success was influenced by discourse context, the same was not true of children speaking Icelandic. The high degree of inflection in Icelandic apparently leads to greater mastery of these forms and renders them less vulnerable to disruption from the discourse costs. Furthermore, recall that Masterson and Kamhi (1992) found that school-aged children produced simpler sentences when using words that were phonologically complex and likely to be less familiar. These authors examined whether producing these more complex words also resulted in more grammatical morpheme omissions. Surprisingly, they found that the children actually produced grammatical morphemes more successfully with more complex words. Recall, however, that these sentences were also likely to be syntactically simple, and as such the directions of influence are not clear. Several directions of influence are possible. It is possible that the higher grammatical morphology success rates were enabled by the relatively low costs of the simple syntax, despite the high costs of the phonology. These results indicate that we may not always be able to predict what forms will be sufficiently costly to produce overt effects on the utterance, or what sort of resource allocations children will make, particularly in spontaneous speech (Kamhi, Catts, & Davis, 1984; Masterson & Kamhi, 1991). Language production is complex, and the outcomes at any moment will be determined by the speaker’s language experience, knowledge states and priorities. We may not be able to predict the same patterns for all speakers, but the research on processing tradeoffs does indicate the range of effects we can look for in the individual children that we serve.

Summary

The literature reviewed here makes it clear that at all ages, the form and content of speakers’ utterances are shaped not only by what they know, but also by their solutions to the real-time challenge of managing processing costs. Most of the time, production processes and their associated costs are not obvious, occurring as a background to fluent, grammatical, and semantically appropriate speech. However, these processes and their costs can become more visible when the total processing load exceeds resources or when researchers manipulate the outcome or ease of processing along the production pathway (Bock, 1996). These moments serve as a window through which language processing systems can be observed and understood.

Current research shows that the total processing load for a given utterance is determined both by costs related
to the utterance under construction and by costs incurred from concurrent activities that are not part of the talking task itself. Utterance-specific costs stem from the number and complexity of the operations that are required (e.g., accessing more words, computing more grammatical role relations; Grela & Leonard, 2000), and from the effort needed to complete individual aspects of language processing, such as accessing a less familiar form. Some utterances may be achievable when speakers can fully apply their processing resources to speaking, but become too costly when some resources are committed to other activities (e.g., walking or tapping fingers; Kemper et al., 2003; 2006).

Speakers respond to excessive total costs with changes in form, content, and fluency. Some of these accommodations, such as simplifications to sentence length and complexity, result in a grammatical utterance. Others maintain grammaticality despite temporary disruptions to fluency. And, still others result in non-grammatical utterances, with required elements being omitted or produced in error.

For mature speakers, in the normal course of events, processing load does not pose a significant barrier to communication. Given the complexity of the task, we are remarkably successful at coordinating the many operations required for speaking. Our resources are usually sufficient for the communicative tasks that we attempt, and when they are not, we have flexibility and strategies to adjust to the load while maintaining reasonable communication success. For immature or language-impaired speakers, however, processing load may be more of an issue. Throughout the language learning years, children show language production difficulties that indicate that costs have exceeded resources. For toddlers, one challenge seems to be to speak while maintaining joint attention (Adamson et al., 2004). For preschool-aged children, one challenge is to build more complex sentence structure while producing all of the elements required by the grammar (e.g., Grela & Leonard, 2000). For older children, the challenge is to construct larger units of text without making grammatical errors (e.g., Thordardottir, 2008). The particulars of each study vary, by context, population and effect under investigation. However, the important picture that emerges is that along the path to language maturity, children might be expected to simplify, become dysfluent, and make grammatical errors when processing costs exceed their resources, especially when newly learned or particularly challenging forms are required. There is some indication that for children with language impairments, the pathway from language emergence to maturity is prolonged (e.g., Johnston & Schery, 1976; Rice, Hershberger, & Wexler, 1998). We turn now to consider how the notions of cost and total processing load can influence our understanding of children’s language use and potentially help us accelerate their progress along this path.

Implications for Clinical Practice

The research reviewed above invites us to think specifically about the mental activities that go into creating a single utterance, and the effort that is involved in completing and coordinating these activities. When we think about language production in this detailed fashion we find many points at which a processing load perspective can influence our clinical decisions. In keeping with the subject of this review, we will focus particularly on applications concerning language production. In doing so, there will be occasions where we talk about the nature of our language input. We will not, however, be discussing applications concerning language comprehension, since that task entails a different set of mental activities. Since the applications we will suggest have, by and large, not been clinically investigated, we will present them only as informed deductions about the implications of processing load concepts for our work with children. We will return later to comment on our role in shaping the research needed to validate them.

We begin with three applications of a processing load perspective to clinical language assessment. We are used to thinking about language performance in a binary fashion, that is, as success or failure. The literature we have just reviewed invites us to ask not only whether the child uses a language form, but also to investigate the cost of success and the sources of difficulty.

(1) Successful language production can hide processing costs.

According to the view of language production presented in this paper, an utterance represents the endpoint of a complex series of language processing activities (Bock & Levelt, 1994; Ferreira & Slevc, 2007; Levelt, 1989). The completion of these activities depends on the availability of processing resources that allow us to attend to, activate, transform and hold information in mind. Speakers can usually produce grammatically and semantically appropriate sentences with mental resources to spare for additional tasks. However, for speakers with fewer resources, or for utterances that are more costly, successful utterance production may consume virtually all available resources. When this happens there may be no resources to spare for other mental work, and any additional costs may push the speaker to the point of language breakdown.

The research literature provides examples of language users who achieve success, but end up on the vulnerable margins of capacity. In one study, children with hearing loss achieved normal range performance on a word repetition test, but suffered greater performance decrements than children with typical hearing on a concurrent task that required them to press a button as quickly as possible every time a light appeared (Hicks & Tharpe, 2002). For the children with hearing loss, the costs of listening and speaking seemed to leave fewer resources for performance on this secondary visual–motor task. Further evidence of costly success can be found in a study by Kemper et al. (2006). These researchers reported that older adults who had
recovered from stroke, and performed in the normal range on language assessment tests, nonetheless demonstrated much more language breakdown than healthy peers when they had to concurrently walk, tap their fingers, or ignore noise. They had difficulty coordinating talking with these other activities, and when they did, their utterances often contained grammatical errors. From a processing perspective, these findings suggest that the conclusions drawn from the focused and structured assessment tasks were misleading in that they failed to reveal that the poststroke speakers were operating very close to their resource limits. As soon as the workload increased, communication breakdowns occurred.

These studies, in conjunction with the other research reviewed in this paper, underscore the fact that successful language performance may be achieved at varying costs. The same utterance may be produced by one speaker easily and automatically and by another speaker only with great mental effort. As clinicians, we might ask ourselves how often our clients achieve communication success at the limits of their capacity. In the study by Hicks and Tharpe (2002), the secondary mental activity was an experimental task, where poor performance carried no real consequences. In real life, the “secondary tasks” that suffer may carry greater consequences. The preschooler who is busy explaining to his teacher what has just happened in the yard may not notice he is walking through another child’s play space and stepping on the toys. Or, the child who is slow, albeit successful, in determining the meaning of an utterance may not be able to simultaneously attend to the new forms it contains.

The level of competence that leads to success in a focused clinical context may not be adequate for the challenges of “real world” classrooms or playgrounds. We can investigate this possibility by including different speaking situations in our assessment. If a child’s language production deteriorates as internal or external costs rise, it would suggest that prior successes had been achieved at a high cost. From this perspective, variation is not a barrier to accurately describing a child’s knowledge, but is the phenomenon of interest that provides a window into the costs of using that knowledge.

Consider the example of a school-aged child with a history of language impairment who now scores in the average range in standardized testing and whose conversational language consists of utterances that are almost always grammatical. However, the tests that were used primarily required single word and single sentence productions, and furthermore the child’s utterances in conversation are simple and spoken within short turns. These latter facts should lead us to question whether this child has truly “outgrown” his earlier language difficulties. If, in a narrative task, this same child produces disjointed or ungrammatical fragments, omits words and grammatical morphemes and makes frequent lexical errors, these lapses could indicate that language processing remains costly and that the child’s system cannot meet any increase in demands. In such a situation, despite test scores and often grammatical production, a conclusion that this child has achieved “normal” language ability would be premature.

(2) When utterances breakdown, “what you see” may not be “what you’ve got.”

As reviewed earlier, costly processing in one aspect of language can produce performance decrements in other aspects. The surface form of an utterance represents the speaker’s solution to the problem of managing the total speaking costs, and the final product may not tell the whole story of the challenges that occurred along the way. A child who says, “I _ giving doggie a bath,” may do so not only because the auxiliary is a challenging form, but because syntactic processing or lexical processing was also costly, and the child “prioritized” processing in those domains to the detriment of morphological processing. That is, from a limited capacity perspective, a morphological error, an omission, or any other error may actually be a symptom of problems elsewhere.

As a second example, consider the challenges faced by school age children in classroom discussions. Those who have difficulty with language comprehension may not have the resources to listen, determine meaning, track topic changes and also plan their own contribution. The result might be off-topic comments if the child allocates resources to the production task or lack of visible participation if the child invests in comprehension. From a processing point of view, however, the child’s problem may lie neither in topic maintenance nor in volubility, but in meeting the total cost of simultaneously comprehending one utterance and planning another.

Fortunately, the language processing literature not only indicates the importance of thinking about the costs of production success and the sources of production difficulties, it also suggests specific observation strategies that can help us do so.

(3) We can manipulate complexity, recent activation and familiarity to discover where production is costly.

Difficulty in a particular area of production is not always a sign of trouble elsewhere, but tradeoff relationships are possible and should be explored. As clinicians, if we suspect that a client’s performance in one domain is being hindered by costs incurred elsewhere, we can borrow strategies from the research literature to investigate this possibility:

- Activate specific words or sentence patterns by providing opportunities to hear and produce them repeatedly in a series of speaking turns, then watch for improved performance in other domains as pre-activation reduces the cost of the repeated form. Streim and Chapman (1987), for example, found that 6- and 8-year-old children described pictures using sentences that were longer when they had they had recently named one of the characters in the picture.
- Create activities that invite the speaker to use both early-acquired, familiar forms and later-acquired, less
familiar forms, then watch to see whether performance improves in one domain when other domains require only familiar forms.

As one example, we could use the second strategy to investigate the possibility that lexical retrieval costs are at the root of poor morphological performance. With an older preschool aged child, the activity might involve the pretense of phoning the market to order assorted foods. The therapist could prepare a shopping list with pictures and then, pretending to be the grocer, could discuss the merits of eggs and artichokes. If processing costs are an issue we might expect the child’s utterances to be more elaborate and more grammatical when the required words are more familiar.

To summarize, the processing literature points us to new language assessment tasks. In addition to identifying forms that a child is or is not using, speech-language pathologists need to determine the relative ease or difficulty of successful productions and the possible contributors to grammatical error or simplification. We can pursue these assessment goals by manipulating familiarity and activation levels in one language domain and looking for systematic performance variation in the remaining domains.

We turn now to consider two applications of processing load to intervention programming. The notion of total processing load not only helps us understand the nature of language production difficulties, it also provides the tools to ameliorate them. Throughout this review, we have used the notion of total processing load to refer to the sum of all the mental work required at the time of speech. Some of the effort in that moment relates directly to the utterance under construction, some of it relates to potentially competing tasks. Whichever the case, total load is a composite entity and as such can be decomposed. We can reduce the total cost of an utterance by removing or otherwise containing any of its component processes. Our purpose in doing so is not merely to maintain grammaticality and fluency in the moment, but to make permanent changes in the cost of one or more of the component processes. Mental operations that cost less will consume a smaller portion of the pool of resources and hence be both less vulnerable and less likely to hinder other contemporaneous activities.

(4) Free up resources for acquiring new forms by using only familiar forms elsewhere.

The literature reviewed above suggests that individual language processes draw on a common pool of resources and thus can support or constrain each other. A child’s success in using a new form will depend on whether or not he or she has the necessary cognitive resources to support that use, which in turn is determined by the total costs of everything else that is going on at the same time. One way to insure that a child has sufficient resource to attend to the details of a new syntactic pattern would be to present it with very familiar words; similarly, one way to ensure that a child has sufficient resources to attend to and produce the meaning and form of a new word would be to introduce it in a familiar sentence pattern. By simplifying the context as much as possible, we increase a child’s chance of acquiring and using a new form.

Simplifying the learning context, of course, can be carried out to an extreme, for example by using telegraphic speech in intervention. We note that to do so is to provide ungrammatical language models to the language learner. Furthermore, researchers who caution against the use of such telegraphic input point out that it actually deprives the language learner of important prosodic cues to language structure, and gives children less input experience with just those forms (such as grammatical morphemes) that they may actually need more exposure to (see Fey, Long, & Finnestack, 2003; van Kleeck et al., 2009). Readers are directed to a recent systematic review of telegraphic speech (van Kleeck, et al., 2009) for more information.

In the initial stages of learning, it may be best to focus on one goal at a time, because children may not be able to free up sufficient resources to attend to more than one unfamiliar form. Language learning may be unconscious and incidental, but this does not mean it is guaranteed. Utterances that require more resource than is available may not be practiced, and children whose mental resources are entirely consumed by the demands of their own utterance may be less likely to notice new language features. Focusing on one goal at a time may at first seem inefficient, but the evidence from adult dual task experiments, as well as acquisition data from normally developing children indicate the importance of attending to total processing load (Adamson, 2004; Kemper et al., 2006).

Having a single target in each activity does not, however, mean that only a single goal is targeted across an entire therapy session or across a period of time until mastery is achieved. While research in this area is very limited, two recent studies suggest that learning outcomes are better when the learning target is addressed in a distributed fashion (Ambridge, Theakston, Lieven, & Tomasello, 2006; Tyler, Lewis, Haskell, & Tolbert, 2003). Ambridge et al. (2006) found that preschoolers were more likely to produce a new sentence pattern if they were taught the pattern in a series of shorter sessions over multiple days than if they received the same number of teaching trials massed within a single session, while the study by Tyler et al. (2003) demonstrated better morphology intervention outcomes when morphology targets alternated with phonology targets than when morphology was targeted exclusively. Finally, it is important to note that while we suggest a high degree of focus within a single activity when new forms are being introduced, we do not expect this degree of focus to be necessary at all stages of learning.

(5) Provide graded opportunities for the use of new forms which will increase familiarity and thus reduce costs.

In the same way that we can engineer the language spoken to a child so that patterns may be more quickly discovered and newly discovered patterns may be produced, we can engineer the child’s language experience so as to improve the reliable use of the language representations...
he or she already knows. Although we know less than we would like to know about how to effect changes in the state of the child’s knowledge, one message comes through this literature loud and clear. Access to and coordination of knowledge schemes improves with experience. Or, familiarity reduces cost. Our role as therapists then is to provide opportunities for the repeated use of a given word or sentence pattern. And, though this is more speculative, by providing this practice while systematically varying the contexts of use we may help children become more able to access and deploy their language knowledge within available resources.

The language intervention literature has seldom made a distinction between facilitating new acquisitions and providing practice for mastery. Inconsistency in the use of a form has been ignored or interpreted as failure to generalize (Johnston, 1988). Studies of language processing, suggest instead that inconsistent use is to be expected during the interval between acquisition and mastery whenever the cost of a form remains high enough that the total processing load sometimes exceeds capacity. This view of language production suggests that the goals of therapy are not achieved at the emergence of a form, nor at the point when a form is used reliably in a structured context. Instead, therapy should aim to reduce the cost of a form enough to ensure its production in a wide range of real life situations.

What would this therapy look like? One possible answer to this question can be found in an article by Culatta and Horn (1982). The authors suggest that in the end stages of a therapy program, the therapist should systematically reinstitute any of the normal complexities of the communication task that had initially been removed to facilitate learning. Culatta and Horn’s (1982) intervention began by providing many opportunities to produce the target form during an intervention session, within a fairly repetitive and predictable speaking turn, coupled with frequent clinician models. As children progressed, the authors systematically reduced the frequency with which a target form was used by the child in a given activity, a change that in the current framework can be seen as moving from initially providing many opportunities for support from recent activation to fewer such opportunities. Likewise, the frequency of clinician models systematically decreased. At the same time, the range and complexity of uses of a form was gradually increased, from repetitive uses to child-initiated uses in natural contexts that were not engineered to elicit the target form with any exceptional frequency.

Following Culatta and Horn (1982), incremental moves towards normalcy could include activities that prompt the use of a targeted form in increasingly varied patterns, while decreasing the level of clinician support. Instead of a game in which the same sentence pattern is repeated over and over, the game could require repeated choices among two or three different sentence patterns. Instead of using a set of new verbs in a single sentence frame (e.g., “He wants to ______”), they could be used in many different frames (e.g., “He wants to ______”, “Now he is ______ing”, “______!” , “The teacher said no more ______”). The goal of such programming would be to provide the child with many opportunities for the use of newly learned words and sentence patterns in contexts that increasingly, and gradually, approach the challenges of everyday speech. The goal of such practice would be to increase familiarity and reduce cost, and reduce the likelihood of excessive total processing loads.

In summary, the processing load perspective presented in this paper suggests that we can facilitate the learning of new forms by limiting the cost of the context in which the form is first presented and practiced. It also suggests that language intervention should include supports for both initial acquisition and mastery learning of a form, and that different intervention approaches are needed for these two phases of learning.

**Opportunities for Clinician-Led Investigation**

We return in conclusion to the issue raised at the opening of our discussion of clinical applications, namely the scant data in support of the ideas expounded in this paper. Johnston (1999) offers various strategies to help with the decision of how to proceed in the face of limited clinical evidence — whether to act or wait for more evidence. One strategy has to do with theory, one has to do with the weight of available evidence, and one has to do with costs and benefits. Johnston notes that when an idea is consistent with a theoretical framework, when ancillary evidence points in a particular direction, and when potential benefits exceed potential costs, there is room for action. With the literature that we have just reviewed, we have developed the outline of a processing framework that has led to coherence across diverse observations. This outline has also directed us to new questions and ideas about the possible hidden costs of success and the importance of mastery learning. This is the power of theory: to take us to new places (Johnston, 1983). Turning to the weight of evidence, we have noted that there is scant evidence on the nature of production processes with children, especially in clinical contexts. But as we also noted, there is considerable ancillary evidence for the ideas generated by a processing load perspective. Finally, we can weigh potential costs and benefits to a decision. As an example, our discussion emphasizes the potential for hidden costs in language use, and describes assessment strategies to identify those costs. An important assumption behind this type of assessment is that it will lead to interventions to reduce those costs, and thus increase success in language domains that are affected by processing resource tradeoffs. This strategy assumes cross-domain intervention effects. Given the current state of evidence, additional consideration of potential costs and benefits can inform the decision of whether or not to act on this assumption in clinical practice. Consider a school-aged child with a history of language impairment who continues to omit grammatical morphemes despite apparent improvements in other areas of language. A clinician who takes a processing load perspective might suspect that
lexical and/or syntactic processing continue to be costly for this child, and that resource allocation ‘decisions’ sometimes leave the child with insufficient resources for morphological success. Choosing to act on the possibility of cross-domain effects, the clinician might explore lexical processing effects on morphological success by contrasting production in contexts of high and low lexical familiarity or presence/absence of recent processing support. If lexical influences appear, the clinician can provide treatment aimed at continuing to develop vocabulary and reduce lexical processing costs. The potential benefit of this decision is to obtain improvement in both lexical and morphological domains by focusing in particular on one domain. The potential cost, if cross-domain effects are in fact not present, is that time is devoted to lexical assessments and possibly interventions without any extra benefit to morphological success. Alternatively, a clinician may choose to not act on the possibility of cross-domain influences. The potential benefit of this decision is time saved by not devoting assessment and/or intervention time to lexical factors. However, if the client’s morphological success is in fact affected by lexical processing, the cost of this decision might be to continue morphological interventions without achieving the desired improvement, due to the interference from lexical processing. There are potential costs to both decisions. However, in the latter instance, the costs might include not knowing that an opportunity for improvement had been missed.

More research is needed. Two areas seem to be particularly ready for more work: cross-domain influences in language intervention, and the effects of production-specific practice on intervention outcomes. As noted, we know little to date about where and when cross-domain effects occur. Of two large scale studies to look at grammatical morphology intervention effects on phonological outcomes for children with morphology and phonology impairments, one study found significant cross-domain effects (Tyler, Lewis, Haskill, & Tolbert, 2002), while the other did not (Fey et al., 1994). Looking at morphology outcomes, in contrast, Tyler et al. (2003) found that alternating morphological treatment with phonological treatment resulted in a two-fold increase in morphology improvements over morphological treatment alone. These results could indicate that reductions in phonological processing costs had freed up resources for children to address morphology. These studies all focused on interactions between phonology and other language domains. To our knowledge, no intervention studies have examined cross-domain effects among the lexicon, syntax and/or grammatical morphology, despite evidence for interactions among these domains from non-intervention research (Bloom et al., 1975; Boyle & Gerken, 1997; Leonard et al., 2000; Streim & Chapman, 1987). Moreover, these studies focused interventions on new acquisitions or still-nascent forms rather than addressing mastery learning to reduce costs after forms had been acquired. In our view, cross-domain intervention effects would be a promising area for clinical investigation.

As a second example, the literature reviewed in this paper raises the possibility that production-specific practice may be particularly important for reducing production costs. In the studies that have compared outcomes between interventions that did and did not contain an explicit production component, there is evidence that production practice may be particularly effective. Ellis Weismer and Murray-Branch (1989) reported data from a single-subject alternating-treatment design to suggest that the learning trajectory may be more stable when intervention includes a specific production component. Connell and Stone (1982) reported that children with SLI made greater gains in a novel morpheme production task following elicited imitation than with input alone, even though these children showed no difference between the two teaching conditions in their ability to perform a test of their comprehension of these new morphemes. Connell and Stone argued that production practice allowed children to not only build representations of the new morphemes, but to also obtain specific practice accessing their output phonological forms. Again, this would seem to be a promising area for clinical investigation. As is always true in the initial stages of a research program, small-N studies are critical for indicating directions for future, large-scale research. The study by Ellis Weismer and Murray-Branch (1989) provides an excellent example of small scale, single-subject research that we can all do to answer questions about effectiveness of a given course of action for an individual client. Using research of this type as a guide, we can contribute to the larger evidence base.

Conclusion

The purpose of this article was to review studies from both the adult and the developmental literature that could inform our thinking about the nature and effects of processing load on children’s language production. The research reviewed indicates that language-external mental work, language complexity, and the effort required for specific language operations all constitute identifiable sources of load, the sum of which can affect the form and content of children’s utterances and their overall communication success. This view of load provides heuristics for observing children’s language use, thinking about the nature of their difficulties, and planning interventions. In this paper, we have outlined a framework for thinking about the nature and effects of processing load in production. More research into both the sources and the effects of processing load in production is needed.

Acknowledgments

Monique Charest was supported by a CIHR Strategic Training Program Fellowship in the Canadian Child Health Clinician Scientist Program, in partnership with SickKids Foundation, Child & Family Research Institute (BC), Women & Children’s Health Research Institute (Alberta) and Manitoba Institute of Child Health. We thank Jeff Small for providing helpful comments on a previous version of this manuscript.
References


---

Endnote

1This description of Levelt and colleagues’ model of production is intended to provide a broad outline of the activities involved in sentence production, and for further details, readers are directed to Bock and Levelt (1994), Bock (1995), Ferreira and Slevc (2007), and Levelt (1989). In addition, we note that a central debate about language production processing concerns whether lexical processing only feeds forward through the system (e.g., Levelt et al., 1999), or whether there are bidirectional connections that allow feedback between processing levels (e.g., Dell, 1986). This debate, although of great importance, is beyond the scope of the current discussion, and will not be addressed.

2 Crystal (1987), Masterson (1997), and Bernhardt, Stemberger and Charest (2010) review studies of tradeoff effects involving phonology and articulation that are not addressed in the current review.

Author's Note

Correspondence concerning this article should be addressed to Monique Charest, Friedman Building 407, 2177 Wesbrook Mall, Vancouver, BC V6T 1Z3. Email: mcarest@interchange.ubc.ca.

Received: April 17, 2009
Accepted: July 8, 2010