Review

CrossMark
click for updates

Author for correspondence:
Gary S. Dell
e-mail: gdell@illinois.edu

Royal Society Publishing

The P-chain: relating sentence production and its disorders to comprehension and acquisition

Gary S. Dell[1] and Franklin Chang[2]

[1]Beckman Institute, University of Illinois, Urbana-Champaign, 405 North Mathews Avenue, Urbana, IL 61820, USA
[2]School of Psychology, University of Liverpool, Liverpool, UK

This article introduces the P-chain, an emerging framework for theory in psycholinguistics that unifies research on comprehension, production and acquisition. The framework proposes that language processing involves incremental prediction, which is carried out by the production system. Prediction necessarily leads to prediction error, which drives learning, including both adaptive adjustment to the mature language processing system as well as language acquisition. To illustrate the P-chain, we review the Dual-path model of sentence production, a connectionist model that explains structural priming in production and a number of facts about language acquisition. The potential of this and related models for explaining acquired and developmental disorders of sentence production is discussed.

1. Introduction

Language production is the least studied of the three components of psycholinguistics—acquisition, comprehension and production. In the *Cambridge Handbook of Psycholinguistics* [1], 16 chapters are devoted to comprehension processes, six to acquisition and only three to production. Despite this imbalance, production research is playing an increasingly central role in psycholinguistic theory, because production-like processes are now seen as important for understanding comprehension and acquisition, and even linguistic theory. This article describes the role of these processes, first by introducing a psycholinguistic framework that we call the *P-chain*, and then by reviewing a model of sentence production that reflects that framework [2,3]. To conclude, we consider the ramifications of the model for an understanding of production disorders. We begin, though, with some history.

When psycholinguistics was in its classical period (the 1960s and 1970s), the purpose of production research was to reveal the grammatical structure of language. For example, pauses during speaking were collected to see whether deep- and surface-structure clause boundaries were reflected in the pause durations and distribution [4]. The search for the psychological reality of linguistic notions in performance data was not limited to production; comprehension and acquisition research were carried out for the same reason. In this respect, there was a fundamental unity among the components of psycholinguistics.

The classical period ended when the field concluded that performance data did not mesh well with the then current grammatical theory [5]. As a result, psycholinguistics lost its unifying theme. The 1980s and 1990s saw vigorous research on a number of issues, such as lexical and structural ambiguity, and rule versus network accounts of morphology and reading. But production, comprehension and acquisition drifted apart, each component focusing on internal issues. For example, in production, many experiments investigated the nature of the 'lemma', a hypothesized lexical representation that is more closely linked to grammatical than to phonological representations (for a review, see [6]). The lemma was hardly ever discussed in research on comprehension or acquisition.

Our impression is that, over the last 10–15 years, psycholinguistics has regained the unity that it lost after the classical period. Production, comprehension and acquisition researchers have increasingly more to say to one another. This is not because of a renewed interest in grammar, but instead owing to the development of

computational—often connectionist—models that specify the mutual influences among these psycholinguistic components. In the following section, we introduce the *P-chain*, which is our characterization of an emerging computational framework in psycholinguistics, a framework that reconnects production, comprehension and acquisition.

2. The P-chain

The P-chain (figure 1) is a set of propositions about the hypothesized interrelations among psycholinguistic concepts

— **P**rocessing involves **P**rediction,
— **P**rediction is **P**roduction,
— **P**rediction leads to **P**rediction error,
— **P**rediction error creates **P**riming,
— **P**riming is im**P**licit learning,
— im**P**licit learning is the mechanism for acquisition/adaptation of **P**rocessing, **P**rediction and **P**roduction, and
— **P**roduction provides the input for training **P**rocessing.

The first six propositions form a chain that leads from language comprehension ('processing') through other P-concepts, including importantly, production, ultimately to learning mechanisms that underlie language acquisition. The final proposition, introduced by MacDonald [7–9], puts a loop in the chain from production back to processing, further demonstrating the central influence of production. In the following, we provide further characterization of each link.

(a) **P**rocessing involves **P**rediction

As we comprehend, we generate expectations about upcoming material at multiple linguistic levels—anticipated meanings, prosody, words and sounds. The idea that speech perception and word recognition involve 'top-down' anticipation is not new [10], but these ideas have been recently formalized in information-theoretic accounts [11]. At the same time, on-line comprehension measures have documented just how powerful predictive processes are. For example, using the visual-world paradigm [12], Altmann & Kamide [13] monitored listener's eye fixations as they understood sentences such as 'The girl ate the cake'. The listeners looked at edible objects such as a cake in a display before hearing any part of the word 'cake', thus demonstrating anticipation of the semantic properties of upcoming words. Comprehenders also predict lexical items, and even phonemes, as shown through another online measuring tool—event-related brain potentials (ERPs). Delong *et al.* [14] collected ERPs as readers saw sentence fragments like 'The day was breezy and so the boy went outside to fly a/an . . .'. Clearly, 'kite' is expected. But if the actual phonological form of 'kite', that it is consonant-initial, is predicted, then the comprehension system further expects to see 'a' rather than 'an'. The ERPs demonstrated just this. There was a greater negativity after the unexpected 'an' compared to 'a' over electrode sites often associated with the N400, an ERP component that reflects violated lexical-semantic expectations.

Connectionist models explain prediction during comprehension from the models' ability to generalize their experience with whole sentences to sentences that are in progress. St. John & McClelland [15] trained a model that anticipated sentence
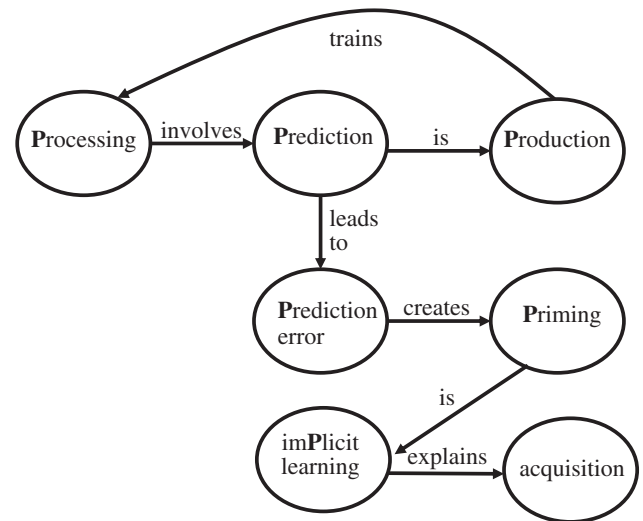

Figure 1. The P-chain framework for psycholinguistics.

meaning from words presented serially. As it processed the phrase 'The bus driver ate . . .', the model learned to activate semantic output that was consistent with its experience of what bus drivers eat, before the direct object was actually input. Elman's [16,17] simple recurrent network (SRN) model was trained by giving it sentences from a simple English-like grammar, where the task was to predict the words in the sentences one at a time. After training, it revealed its knowledge of the grammar by its predictions. When given a novel sentence, the model could predict the words that would be grammatical at each point in the sentence.

(b) **P**rediction is **P**roduction

Predicting that we will hear 'kite' after 'the boy flied a' is a top-down process; knowledge, generated at the 'top', is used to anticipate what will happen at a lower level in the system. Production is also a top-down process because processing flows from intended meaning—the message—to levels that encode linguistic forms. It is thus not surprising that prediction and production are related: if you can predict, you can use that knowledge to produce. Hsu *et al.* ([18], p. 46), for example, stated that 'a learner, given an entire history of linguistic input, can eventually 'join in' and start *saying* its predictions'. The links between prediction and production go back to the analysis-by-synthesis theory of speech perception [19], but were not developed for sentence processing until recently. For example, Pickering & Garrod [20] claimed that, during language processing, the production system constructs a forward model, which can anticipate upcoming linguistic elements. Federmeier [21] distinguished between prediction, that is, anticipation of the future, and integration, which is combining the present with the past, as processing mechanisms in comprehension, and proposed that prediction during comprehension is carried out in the left hemisphere [22]. To explain why prediction is associated with the left hemisphere only, Federmeier noted that production is strongly left lateralized. If prediction is production, the lateralization of prediction is then explained. Chang *et al.*'s [3] model was a computational implementation of the idea that production abilities arise directly from learning to predict. Because prediction is production, prediction ability seamlessly transfers to production skill, a finding
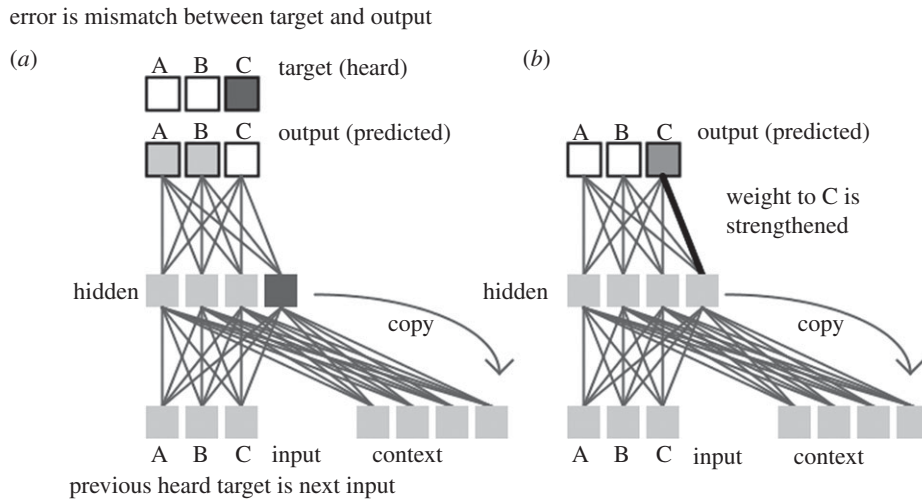
error is mismatch between target and output

Figure 2. A simple recurrent network model of sequence prediction (based on Cleeremans & McClelland [32]). (*a*) The network is predicting A or B as the next output, but instead, C is the actual output. This results in prediction error, which drives learning and leads, for example, to an increase in connection weights from units representing the current context to the output for C (*b*).

supported in a study of young children's abilities to predict and produce [23]. Finally, recent accounts of production based on the notion that speakers carefully control the predictability of what they say (e.g. uniform information density [24]) imply a strong connection between prediction and production.

(c) Prediction leads to Prediction error, which creates Priming, which is a form of imPlicit learning

Here, we collapse three links of the P-chain into a super-link that is the heart of the chain. Prediction during comprehension is necessarily error prone. One rarely hears exactly what one expects. The deviation between words that are expected and those that are actually processed constitutes an error signal, which then primes or changes the system so as to diminish future error in similar situations.

Priming is often thought of as applying to words. Processing a word, particularly when it is not expected, primes the system to respond more effectively to the word in the future [25]. Much recent research, however, has concerned structural priming in production. Structural priming is the tendency for speakers to reuse recently experienced structures [26]. For example, Bock *et al.* [27] gave speakers pictures that can be described with either of the two kinds of dative structures (double objects, 'The woman handed the boy the ice cream', versus prepositional datives, 'The woman handed the ice cream to the boy'). Participants described these pictures after either saying or hearing an unrelated prime sentence that was a double-object or prepositional dative. Priming was seen in the tendency for speakers to use the structure of the prime when describing the picture. Similar effects were seen for other structural alternations such as active transitive sentences ('The truck hit the nurse') versus passives ('The nurse was hit by the truck'). Importantly, the priming was equally strong for heard and produced primes. Thus, comprehending a structure later affects the production process. This finding can be explained by the P-chain. Comprehension of the prime entails prediction, and hence prediction error. As a result, the system that made the prediction changes so that future error is reduced. For example, if a listener does not expect the passive, but hears one, the listener's expectation for the passive is increased. These changes affect the

production of later sentences precisely because the prediction system *is* the production system. Thus, Bock *et al.*'s finding of robust priming from comprehension to production can be explained by the P-chain. But what is the nature of the priming? Bock *et al.* further claimed that priming is not just a temporary change to the system, but instead is a form of implicit learning, akin to the connection weight changes that characterize learning in connectionist models. The evidence that priming is learning is that the effect of a prime persists undiminished over time and over unrelated sentences (see [28] for review). The evidence that the learning is implicit is that it occurs in amnesic speakers who have no explicit memory of the prime sentence [29]. Finally, the evidence that the implicit learning that characterizes structural priming is based on prediction error comes from demonstrations that less common (more surprising) prime structures lead to more priming than common ones [30,31]. In sum, much of what we know about structural priming can be mapped onto the propositions of the P-chain.

The implicit-learning theory of structural priming owes a debt to computational models of sequence learning. Cleeremans & McClelland [32] modelled sequence learning through a chain of prediction, prediction error and priming. In sequence learning studies, participants experience sequences of stimuli that follow sequential patterns [33]. Eventually, they become sensitive to the pattern, but at the same time lack awareness of its structure. Cleeremans and McClelland trained an SRN model to predict each element of a sequence from a representation of the immediately prior element plus a learned representation of what happened before that. The subsequent occurrence of the actual next element—the target—was then compared to this prediction, thus generating prediction error, specifically a prediction-error vector. For example, suppose that there are three possible elements, A, B and C, and that at a certain point in the sequence the model predicts that A or B, but not C will happen (figure 2). Thus, the model might generate output activations for A of 0.6, B of 0.6 and C of 0.2. Next, suppose that the model experiences the actual target and (surprise!) it is C. That is, the actual event can be described as A = 0, B = 0 and C = 1. All of the predicted outputs are thus wrong to various degrees, with the resulting prediction error vector being -0.6 for A, -0.6 for B, and $+0.8$ for C. The weights of the model's connections are then changed in
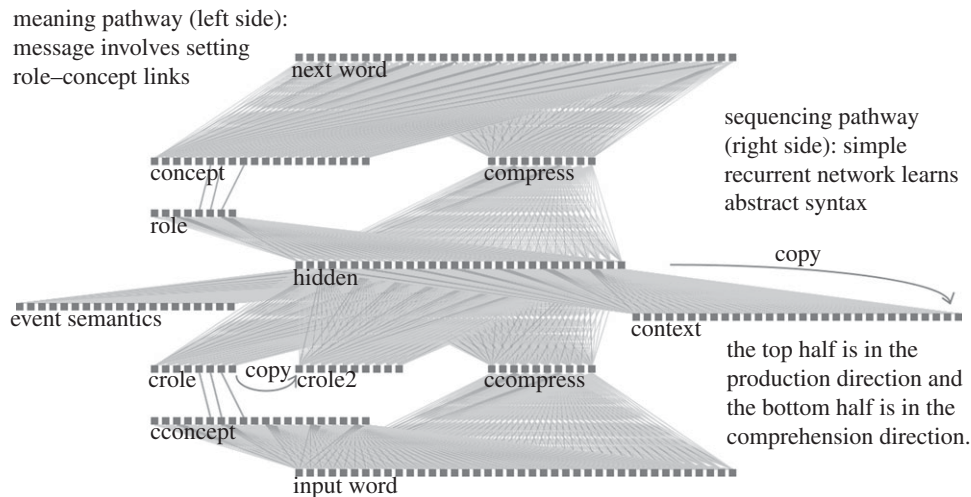
Figure 3. The Dual-path model. Units are represented by small boxes, and connections by lines. All connection weights are learnable, except for the connections between roles (or croles) and concepts (or cconcepts), whose weights are temporary and are set to represent the message. The prefix 'c' indicates unit layers that function in comprehension, rather than in production. For further information, see Chang *et al*. [3].

response to this error, in accordance with connectionist learning rules that attempt to minimize future error. These weight changes accumulate, with the result that the system encodes the structure of the sequence. The implicit nature of the learning is captured in the model by the fact that the learning inhabits the model's weights and influences its processing predictions, but there is no explicit representation that can be accessed.

(d) ImPlicit learning is the mechanism for acquisition and adaptation of Processing, Prediction and Production

The implicit learning that is revealed in structural priming and sequence learning studies is claimed to be an important mechanism for how linguistic skill changes over time [31,34,35]. It is the mechanism for how the mature system adapts to specific contexts and speakers and, it is claimed, ultimately for how the child acquires the system. We shall show in detail how this claim is realized in the model of Chang *et al*. [3], which explains structural priming and language acquisition by the same mechanisms.

(e) Production provides the input for training Processing

The processing system experiences distributions of linguistic elements, which it learns. But where do these distributions come from? As MacDonald [9] noted in her Production–Distribution–Comprehension proposal, they come from the *production* systems of other speakers. Processes intrinsic to production make some structures easier to say than others. For example, it is easier to produce sentences in English when long noun phrases can be shifted to the end (heavy NP shift). 'I fed him a large bowl of chicken soup that I'd just warmed up', would be easier to say than, 'I fed a large bowl of chicken soup that I'd just warmed up to him'. These production processes affect how likely speakers are to produce particular structures [36] and, hence, the distributions of structures that comprehenders experience. In this way, the production systems of other speakers 'train' one's processing system, making it sensitive to production-biased distributions.

To summarize, the P-chain framework offers a proposal for how language processing, production and acquisition are connected. In the following section, we make these ideas concrete by reviewing a sentence production model known as the *Dual-path model* [37].

3. The Dual-path model

The Dual-path model is a connectionist model of language acquisition and sentence production (figure 3). The central component of the model is a SRN which tries to predict the next heard word from the word that preceded it and a representation of prior linguistic context [16]. It then compares the predicted next word (technically, a prediction vector across all possible words) with the actual next word. The resulting prediction error is used to change the model's internal representations, thus enabling the model to acquire the syntactic knowledge that helps it to make these predictions (e.g. that 'the' is followed by nouns). But to do production, the model must also link this syntactic knowledge to the message, the meaning that one wants to convey. That is, to turn an SRN into a model of production, one has to include a message, and the model must learn how to use this message to guide the prediction process. Moreover, if it is to be a realistic model of language use, its ability to predict must function for novel and even unusual sentences. The Dual-path model learns about the relation between messages and sentences from comprehension. It assumes that a comprehender often can infer the meaning of what she is about to hear from the situation. Situational meaning is available in routine activities (e.g. boiling water is followed by making tea) where it is possible to partially predict what people will say next ('Do you want milk?'). The inferred meaning is represented in the model within a *meaning pathway* that is separate from the *sequencing pathway* represented by the SRN. In figure 3, the sequencing pathway is along the right side of the diagram, and the meaning pathway is on the left. The separation of the pathways is incomplete, though. They come together, once at the level of the model's hidden-units, and again at the output word level. The two pathways constitute the Dual-path architecture. The model learns representations in both pathways

and uses them in next-word prediction. The sequencing pathway helps the model predict grammatically appropriate words at each sentence position, and the meaning pathway ensures that the chosen words reflect the meaning. After these pathways have been fully trained, they combine to support processing for both previously experienced and novel sentences. As comprehenders can also be speakers, the same representations that support prediction during comprehension can support production of a self-generated message, including messages that encode completely novel sentences. For example, after learning by prediction error from training sentences, the model was tested by giving it random novel messages and allowing it to produce, that is, to 'predict' a sequence of words. It produced a grammatically correct sentence that accurately expressed the message 91% of the time [2]. So, if the model had been trained predicting the sequence 'The dog bit the man', it would be fully capable of later producing sentences from a novel message such as 'The man bit the dog' or even 'The man bit the vase' [38]. This implements the idea that prediction is production.

The model explains structural priming as implicit learning that results from prediction error during prime comprehension. After the model has been trained to 'adulthood', its learning algorithm continues to function with the result that the language acquisition process never really stops. As the model hears the prime sentence (e.g. 'the man sent the package to the woman'), it tries to predict each of its words. The resulting prediction error leads to small changes to the model's syntactic representations, specifically the network weights within the sequencing pathway. Then, when the model is given the target message to produce, for example GAVE (TEACHER, BOOK, STUDENT), these changes bias the model to use the primed structure (e.g. 'the teacher gave the book to the student' rather than 'the teacher gave the student the book'). This implicit learning mechanism can explain several aspects of structural priming. It has been found that priming is mostly structural; what persists from prime to target is the abstract syntactic structure of the sentence [39,40]. This has been shown by using primes that have different meanings and thematic roles than the target utterances. For example, the locative prime 'the 747 was landing by the control tower' is similar in structure to the passive, 'The nurse was hit by the truck'. They differ, however, in how the structure conveys meaning (e.g. the truck is doing the hitting, but the control tower is not doing the landing). Nonetheless, structural priming studies have found that locatives prime passives as much as passives do, thus demonstrating that priming can be based on the structures themselves (e.g. sequential and hierarchical relations among syntactic categories), rather than the functions associated with the structures. The model effectively reproduces this effect, as well as other demonstrations of the structural nature of priming [3]. The model's priming can be purely structural because syntactic categories arise out of the prediction process in SRNs [16], because these categories are often the best predictors of the next word when meaning is not available. The Dual-path model maintains this bias for syntactic categories, because it isolates the meaning pathway from the sequencing pathway. Thus, not only does prediction-based learning create priming, but it also helps to explain the syntactic nature of priming.

Because the model uses its prediction mechanism to acquire its language representations, it should also explain acquisition phenomena. One such phenomenon is seen in preferential looking studies, in which children exhibit syntactic knowledge in their looking behaviour before they exhibit this knowledge in production. Typically in these preferential looking tasks, children view two videos: a causative video, where one entity acts on another entity, and a non-causative video, where two entities do the same action together [41]. It has been found that children around 2 years of age tend to look at the causative video when they hear a transitive sentence with a novel verb like 'the boy pilked the girl' (causative–transitive pairing), but at the non-causative video when they hear an intransitive sentence like 'the boy pilked with the girl' (non-causative–intransitive pairing). This finding was simulated in the Dual-path model by comparing the prediction error from causative or non-causative messages that were experienced with either transitive or intransitive utterances. Like children, the model's 'preferences' (lower prediction error) occurred for the appropriate pairings (causative message with transitive sentence and non-causative message with intransitive sentence) compared to the reverse [3]. Importantly, the model exhibited these preferences even though it was not yet able to fully produce these structures. Moreover, the model exhibited a relatively stronger knowledge of the causative–transitive mappings than non-causative–intransitive mappings. This is because the components of the transitive mapping are present in various structures (e.g. dative and locative, as well as transitive sentences), and hence these mappings are learned more quickly than the intransitive mapping. The early strength of the causative–transitive link constitutes a prediction that has been confirmed in experimental studies in English [42–44] and in Japanese [45]. Thus, the word-based prediction error, which the model used for learning the language, is able to explain structural biases in toddler comprehension processing.

The implemented Dual-path model used prediction error to acquire its English representations. But if this mechanism is a general account of language acquisition, it should be able to acquire typologically different languages, for example Japanese. Japanese differs from English in several ways, two of which are important for our purposes. First, English uses word order to denote who did what to whom, while Japanese indicates this function with case marking on nouns (e.g. -ga, -ni, -o). The case marking makes it possible to scramble the order of noun phrases and maintain the same meaning. Second, unlike English, Japanese verbs occur at the end of clauses. These two differences can be seen in the Japanese version of 'John gave the book to Mary', which would be 'John-ga Mary-ni book-o gave' (when expressed with English content words). Chang [2] trained the Dual-path model on either English or Japanese input, and the model was able to acquire both languages to a similar degree. More importantly, the model was able to explain cross-linguistic differences in production biases in the two languages. As we mentioned earlier, a classic bias is heavy NP shift and, in English, it is seen as a tendency to place long or 'heavy' phrases later in sentences [46]. For example, speakers tend to convey the message in 'the man gave *the girl that was walking in the forest* the book' as 'the man gave the book to *the girl that was walking in the forest*'. Surprisingly, in Japanese, speakers have the opposite bias; they place long phrases before short phrases [47]. The Dual-path model was able to explain the differences in the biases because the languages' contrasting grammars lead to differences in their position-specific predictions. In English, the choice of where to put the heavy noun phrase usually occurs right after the verb (e.g. after 'gave'). At that sentence position, the model tends to produce short phrases for the simple reason that short unmodified phrases (e.g. 'the book') are more

6

frequent at that post-verbal position in the model's experienced input. In Japanese, however, the shift necessarily occurs earlier in the sentence, because all of the potentially shifting NPs occur before the verb. When the shift is earlier, there is less lexical or structure information on which to base predictions and, consequently, message information becomes crucial in predicting word ordering. As heavy NPs have more message information associated with them—they are heavy because they are modified and, in Japanese, modifying terms are placed before what is being modified—the model will often start sentences with these long, heavy NPs. This is how the surprising long-before-short bias in Japanese is explained in the model. In general, cross-linguistic differences in processing can arise from the fact that prediction error is computed at each word and different types of information are important at each position in different languages. And because prediction is production, these differences then create production word-order biases.

The word-by-word computation of prediction error is a central assumption of the Dual-path model and it influences the syntactic representations that the model learns. Consider the case of centre-embedded sentences. Although sentences with embedded relative clauses challenge the processing system by virtue of the long-distance dependencies that must be tracked, subject relatives for example 'the boy that chased the girl was sleepy' are easier to process than object relatives for example 'the boy that the girl chased was sleepy' [48]. MacDonald & Christiansen [49] showed that an SRN model could exhibit this difference. Fitz *et al*. [50] extended this work by manipulating the frequency of different structures within a version of the Dual-path model that learned to produce subject and object relatives. They showed that the overall frequency of the structure was not the main force in creating the processing difference, but rather the frequency of substructures. The simplest substructure is just the transition between one word or category and another. Subject relatives with transitive verbs like 'chased' have a substructure where the word 'that' transitions to a verb (THAT VERB) and the object relatives have a substructure with 'that' followed by a noun (THAT NOUN). Although subject and object relatives with transitive embedded clause verbs were equally frequent in the input that trained the model, the existence of relatives with intransitive verbs (e.g. 'the boy that ran was sleepy') uniquely supported the substructure (THAT VERB). It is this extra experience with the (THAT VERB) substructure that makes the Dual-path model favour subject relatives. Given that, in the P-chain, experience in predicting is supposed to train production, we expect that there would be a preference for the subject relative over the object relative in production as well as in comprehension and, specifically, that the differences in difficulty could be detected (e.g. in pausing or other dysfluencies) right after THAT. In support, we know that speakers tend to be dysfluent after THAT in object relative clauses [51] and that they generally avoid producing these kinds of sentences altogether [52].

4. Applying the model to acquired and developmental disorders of sentence production

Acquired language disorders result from brain trauma that impairs language processing or production, resulting in aphasia, dyslexia, dysgraphia or other language-related impairments.

Aphasia is nearly always associated with some production deficit. It is most commonly caused by stroke that damages perisylvian areas of the left hemisphere. By contrast, developmental disorders are disorders in language that emerge during development and manifest as a non-normal level of language ability for the child's age group. These include specific language impairment (SLI, e.g. [53]) and the language difficulties associated with autism spectrum disorder [54].

Connectionist models are uniquely positioned to explain both acquired and developmental disorders, because they model both the mature system and its development. Acquired disorders can be modelled by lesioning particular units or links in the network after it has been trained [55,56]. Developmental disorders can be modelled by changing general characteristics of the model before it learns [57–59]. These characteristics could include distortion of the input, changes in the inherent connectivity in the network or characteristics of the model's learning mechanism (e.g. a lower learning rate). These differences can be assumed to apply globally or to specific parts of the network.

There are many connectionist models of acquired word production deficits [56,60–63]. Few such models, however, deal with aphasic *sentence* production. Earlier versions of the Dual-path model have simulated aphasia [37,64] by lesioning either the meaning or the sequencing pathways of the fully trained model. For example, Chang [37] found that lesions of the Dual-path model could simulate a double dissociation associated with aphasic sentence production (e.g. [65]), namely the difficulty that some agrammatic patients have with function words (determiners and prepositions) and other semantically light words for example light verbs (e.g. 'give' and 'go'), in comparison to the difficulty that anomic patients have with semantically heavy content words (e.g. most nouns). After training the model to adulthood, Chang randomly removed connections either from the hidden units to output words (sequencing-pathway lesion) or connections from concepts to words (meaning-pathway lesion). The sequencing-pathway lesion led to errors on function words or light verbs, which often yielded ungrammatical utterances. The lesion in the meaning pathway led, instead, primarily to omission and substitution errors on content words (e.g. 'girl'). This double dissociation involving function and content words arose in the model because, during training, the sequencing pathway assumes much of the responsibility for retrieving words with little semantic content. In essence, words such as 'the' or 'of' become more strongly associated with sentence position (represented in the sequencing pathway) than with sentence meaning (the meaning pathway), while the reverse is true for content words. Importantly, the model's distinction between function and content words is not specified *a priori*, but arises from the learning process and its architectural distinction between a meaning and a sequencing pathway.

The aphasic behaviour of the lesioned Dual-path model is an example of the way that models with multiple pathways learn to 'divide the labour' by placing different linguistic representations in different pathways [64,66]. This division of labour arises from using prediction error in learning, because the model tries to use the best predictors to guide its use of each linguistic category, with the result that the different categories depend differentially on syntactic, lexical, semantic and sound constraints. The result is a system where different linguistic categories have different connection weights from

different parts of the network. As a result, particular categories can be impaired by localized lesions. For example, lesions in the sequencing pathway led to loss of some function words because the strong sequential and weak semantic properties of those words led to their representations inhabiting weights in that pathway. Similarly, nouns and verbs will inhabit different connections in the system because verbs have more syntactic constraints on their usage than nouns do. These differences could yield dissociations when the network is damaged, thus enabling the model to mimic patients who are relatively impaired on one category, but not another (e.g. verbs, but not nouns, see [67]). In this way, connectionist models of acquired disorders can explain category-specific dissociations, without assuming that these categories are separately represented owing to innate architectural constraints.

In contrast with acquired disorders, developmental spoken-language disorders have been less well studied within the connectionist literature. The most commonly considered of these disorders is SLI [53]. Because SLI is a cover term for potentially a wide range of deficits, there may not be a unified basis for the disorder, aside from the fact that in SLI, language is delayed relative to that of typically developing children. An important feature of SLI compared to childhood aphasia is stability of the disorder [68]. The language ability of SLI children is less than normally developing children and this difference remains even into adulthood (Conti-Ramsden *et al.* [69] report that 44% of SLI children graduate compared with 88% of typically developing children). By contrast, studies of childhood aphasia show that many children recover to near-normal levels within months to a few years [70,71]. This difference is important, because it can be difficult to create stable persistent developmental disorders in connectionist models. Plaut [72] lesioned a connectionist model of reading such that it was only 20% accurate at word reading and then trained the model normally. Even though the model was missing 50% of its links, it was able to recover after 50 epochs to nearly 100% accuracy. Thomas & Karmiloff-Smith [59] tested a range of manipulations to create developmental disorders in a connectionist model of the past tense; these included changing learning rate, lesioning links, adding noise and changing the size of layers. Although they were able to create some delays in acquisition, most of the models achieved near-normal levels of accuracy at the end of training, particularly for regular forms that they experienced in training. This is in contrast with the behaviour of SLI children for whom linguistic exposure does not always improve performance. Oetting [73] reported that SLI children could use sentence structures to infer the meaning of novel verbs, but they did not retain this even after repeated exposures (typically developing children showed retention after one exposure). Similarly, a novel-verb training study showed that retention was poor in SLI children relative to controls [74]. Thus, connectionist models can easily explain recovery after acquired disorders, but modelling developmental disorders is more challenging.

One reason that it is difficult to model a stable deficit in connectionist models is that the models typically use error-based learning. Any manipulation that initially reduces the model's ability to predict the target will lead to higher levels of error, which will then drive the model even harder to reduce that error by further learning. Because the amount of change is tied to error, the models' natural tendency is to effectively compensate for any error increasing property. Given this, one way to explain the stability of developmental deficits

for example SLI within an error-based learning system is to assume that affected children have learned cues that can support more or less correct prediction, but these cues are the wrong cues to support generalization. For example, the voicing feature of the English past tense properly comes from the preceding consonant (e.g. /d/ is used for 'penned', but /t/ is used for 'sequenced'). If a model happens to learn that words that start with 'pe' append /d/ and words that start with 'seq' append /t/, then its predictions would be correct for these words, but the model's knowledge would be useless for generalizing to the past tense of a word like 'sequester', which uses /d/. As the known words do not generate error, no learning takes place. If these miscued, but correct, items are common, that would lead to an SLI-like result; the model would fail to generalize well, and it would be difficult for it to recover from that pathological state. One way to cause such miscued learning in connectionist models is to provide MORE information or resources, which is in line with the hypothesis that SLI children possess more exemplar knowledge [75]. Chang [37] showed that the generalization ability of the Dual-path model was permanently reduced by including a set of connections from the hidden units to the concepts, creating what was called the Linked-path model. In this model, the sequencing system has access to message information to help it to predict. It was able to produce trained utterances at 100% accuracy, but it was not able to generalize on a novel test set to the same level as the Dual-path model, because its syntactic representations were dependent on lexical-semantic regularities in the input. In this way, this model exhibited SLI-like symptoms that persisted over development owing to its encoding of input-specific information in its sequencing representations.

Prediction-based learning in connectionist models can explain recovery from acquired damage that increases the error signal. To explain developmental disorders that do not recover quickly, these models would have to learn miscued rules which allow correct prediction without error. If so, SLI interventions may want to focus on creating more error by decreasing the predictability of the items used.

5. Conclusion

The P-chain is an emerging framework in psycholinguistics that links processing, production and acquisition. Its key notion is that both language acquisition and the subsequent tuning of linguistic knowledge arise from acts of prediction that occur as language is incrementally processed in context. Prediction is assumed to be carried out by the production system, thus making production into a central component of psycholinguistics. Much of the P-chain framework is inspired by connectionist models of language processing and acquisition. In this article, we reviewed a model that shows how prediction-based learning can be used to explain structural priming in production and language development. We claim that these and related models also have the potential to provide a new perspective on acquired and developmental language disorders, a perspective that emphasizes how prediction shapes the structure of what is learned and its potential for recovery if the learned knowledge is deficient or damaged.

References

1. Spivey M, McRae K, Joanisse M, editors. 2012 *The Cambridge handbook of psycholinguistics*. New York, NY: Cambridge University Press.

2. Chang F. 2009 Learning to order words: a connectionist model of heavy NP shift and accessibility effects in Japanese and English. *J. Mem. Lang.* **61**, 374–397. (doi:10.1016/j.jml.2009.07.006)

3. Chang F, Dell GS, Bock K. 2006 Becoming syntactic. *Psychol. Rev.* **113**, 234–272. (doi:10.1037/0033-295X.113.2.234)

4. Ford M, Holmes VM. 1978 Planning units and syntax in sentence production. *Cognition* **6**, 35–53. (doi:10.1016/0010-0277(78)90008-2)

5. Fodor JA, Bever TG, Garrett MF. 1974 *The psychology of language: an introduction to psycholinguistics and generative grammar*. New York, NY: McGraw-Hill.

6. Levelt WJM, Roelofs A, Meyer AS. 1999 A theory of lexical access in speech production. *Behav. Brain Sci.* **22**, 1–38. (doi:10.1017/S0140525X99001776)

7. Gennari SP, MacDonald MC. 2009 Linking production and comprehension processes: the case of relative clauses. *Cognition* **111**, 1–23. (doi:10.1016/j.cognition.2008.12.006)

8. MacDonald MC. 1999 Distributional information in language comprehension, production, and acquisition: three puzzles and a moral. In *The emergence of language* (ed. B MacWhinney), pp. 177–196. Hillsdale, NJ: Lawrence Erlbaum Associates.

9. MacDonald MC. 2013 How language production shapes language form and comprehension. *Front. Psychol.* **4**, 226. (doi:10.3389/fpsyg.2013.00226)

10. Morton J, Long J. 1976 Effect of word transitional probability on phoneme identification. *J. Verbal Learn. Verbal Behav.* **15**, 43–51. (doi:10.1016/S0022-5371(76)90005-0)

11. Levy R. 2008 Expectation-based syntactic comprehension. *Cognition* **106**, 1126–1177. (doi:10.1016/j.cognition.2007.05.006)

12. Tanenhaus MK, Spivey-Knowlton MJ, Eberhard KM, Sedivy JC. 1995 Integration of visual and linguistic information in spoken language comprehension. *Science* **268**, 1632–1634. (doi:10.1126/science.7777863)

13. Altmann GTM, Kamide Y. 1999 Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* **73**, 247–264. (doi:10.1016/S0010-0277(99)00059-1)

14. DeLong KA, Urbach TP, Kutas M. 2005 Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* **8**, 1117–1121. (doi:10.1038/nn1504)

15. St. John MF, McClelland JL. 1990 Learning and applying contextual constraints in sentence comprehension. *Artif. Intell.* **46**, 217–257. (doi:10.1016/0004-3702(90)90008-N)

16. Elman JL. 1990 Finding structure in time. *Cogn. Sci.* **14**, 179–211. (doi:10.1207/s15516709cog1402_1)

17. Elman JL. 1993 Learning and development in neural networks: the importance of starting small. *Cognition* **48**, 71–99. (doi:10.1016/0010-0277(93)90058-4)

18. Hsu AS, Chater N, Vitányi P. 2013 Language learning from positive evidence, reconsidered: a simplicity-based approach. *Top. Cogn. Sci.* **5**, 35–55. (doi:10.1111/tops.12005)

19. Stevens KN, Halle M. 1967 Remarks on analysis by synthesis and distinctive features. In *Models for the perception of speech and visual form* (ed. W Wathen-Dunn), pp. 88–102. Cambridge, MA: MIT Press.

20. Pickering MJ, Garrod S. 2007 Do people use language production to make predictions during comprehension? *Trends Cogn. Sci.* **11**, 105–110. (doi:10.1016/j.tics.2006.12.002)

21. Federmeier KD. 2007 Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology* **44**, 491–505. (doi:10.1111/j.1469-8986.2007.00531.x)

22. Federmeier KD, Kutas M. 1999 A rose by any other name: long-term memory structure and sentence processing. *J. Mem. Lang.* **41**, 469–495. (doi:10.1006/jmla.1999.2660)

23. Mani N, Huettig F. 2012 Prediction during language processing is a piece of cake: but only for skilled producers. *J. Exp. Psychol. Hum. Percept. Perform.* **38**, 843–847. (doi:10.1037/a0029284)

24. Jaeger TF. 2010 Redundancy and reduction: speakers manage syntactic information density. *Cogn. Psychol.* **61**, 23–62. (doi:10.1016/j.cogpsych.2010.02.002)

25. Scarborough DL, Cortese C, Scarborough HS. 1977 Frequency and repetition effects in lexical memory. *J. Exp. Psychol. Hum. Percept. Perform.* **3**, 1–17. (doi:10.1037/0096-1523.3.1.1)

26. Bock K. 1986 Syntactic persistence in language production. *Cogn. Psychol.* **18**, 355–387. (doi:10.1016/0010-0285(86)90004-6)

27. Bock K, Dell GS, Chang F, Onishi KH. 2007 Persistent structural priming from language comprehension to language production. *Cognition* **104**, 437–458. (doi:10.1016/j.cognition.2006.07.003)

28. Bock K, Griffin ZM. 2000 The persistence of structural priming: transient activation or implicit learning? *J. Exp. Psychol.* **129**, 177–192. (doi:10.1037/0096-3445.129.2.177)

29. Ferreira VS, Bock K, Wilson MP, Cohen NJ. 2008 Memory for syntax despite amnesia. *Psychol. Sci.* **19**, 940–946. (doi:10.1111/j.1467-9280.2008.02180.x)

30. Bernolet S, Hartsuiker RJ. 2010 Does verb bias modulate syntactic priming? *Cognition* **114**, 455–461. (doi:10.1016/j.cognition.2009.11.005)

31. Jaeger TF, Snider NE. 2013 Alignment as a consequence of expectation adaptation: syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition* **127**, 57–83. (doi:10.1016/j.cognition.2012.10.013)

32. Cleeremans A, McClelland JL. 1991 Learning the structure of event sequences. *J. Exp. Psychol.* **120**, 235–253. (doi:10.1037/0096-3445.120.3.235)

33. Newport EL, Aslin RN. 2004 Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cogn. Psychol.* **48**, 127–162. (doi:10.1016/S0010-0285(03)00128-2)

34. Chang F, Janciauskas M, Fitz H. 2012 Language adaptation and learning: getting explicit about implicit learning. *Lang. Linguist. Compass* **6**, 259–278. (doi:10.1002/lnc3.337)

35. Warker JA, Dell GS, Whalen CA, Gereg S. 2008 Limits on learning phonotactic constraints from recent production experience. *J. Exp. Psychol. Learn. Mem. Cogn.* **34**, 1289–1295. (doi:10.1037/a0013033)

36. Stallings LM, MacDonald MC, O'Seaghdha PG. 1998 Phrasal ordering constraints in sentence production: phrase length and verb disposition in heavy-NP shift. *J. Mem. Lang.* **39**, 392–417. (doi:10.1006/jmla.1998.2586)

37. Chang F. 2002 Symbolically speaking: a connectionist model of sentence production. *Cogn. Sci.* **26**, 609–651. (doi:10.1207/s15516709cog2605_3)

38. Fodor JA, Pylyshyn ZW. 1988 Connectionism and cognitive architecture: a critical analysis. *Cognition* **28**, 3–71. (doi:10.1016/0010-0277(88)90031-5)

39. Bock K, Loebell H. 1990 Framing sentences. *Cognition* **35**, 1–39. (doi:10.1016/0010-0277(90)90035-I)

40. Bock K. 1989 Closed-class immanence in sentence production. *Cognition* **31**, 163–186. (doi:10.1016/0010-0277(89)90022-X)

41. Naigles LR. 1990 Children use syntax to learn verb meanings. *J. Child Lang.* **17**, 357–374. (doi:10.1017/S0305000900013817)

42. Arunachalam S, Waxman SR. 2010 Meaning from syntax: evidence from 2-year-olds. *Cognition* **114**, 442–446. (doi:10.1016/j.cognition.2009.10.015)

43. Noble CH, Rowland CF, Pine JM. 2011 Comprehension of argument structure and semantic roles: evidence from English-learning children and the forced-choice pointing paradigm. *Cogn. Sci.* **35**, 963–982. (doi:10.1111/j.1551-6709.2011.01175.x)

44. Yuan S, Fisher C. 2009 Really? She blicked the baby? *Psychol. Sci.* **20**, 619–626. (doi:10.1111/j.1467-9280.2009.02341.x)

45. Matsuo A, Kita S, Shinya Y, Wood GC, Naigles LR. 2012 Japanese two-year-olds use morphosyntax to learn novel verb meanings. *J. Child Lang.* **39**, 637–663. (doi:10.1017/S0305000911000213)

46. Arnold JE, Losongco A, Wasow T, Ginstrom R. 2000 Heaviness versus newness: the effects of structural complexity and discourse status on constituent ordering. *Language* **76**, 28–55.

47. Yamashita H, Chang F. 2001 'Long before short' preference in the production of a head-final language. *Cognition* **81**, B45–B55. (doi:10.1016/S0010-0277(01)00121-4)

48. Traxler M. 2002 Processing subject and object relative clauses: evidence from eye movements. *J. Mem. Lang.* **47**, 69–90. (doi:10.1006/jmla.2001.2836)

49. MacDonald MC, Christiansen MH. 2002 Reassessing working memory: comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychol. Rev.* **109**, 35–54. (doi:10.1037/0033-295X.109.1.35)

50. Fitz H, Chang F, Christiansen MH. 2011 A connectionist account of the acquisition and processing of relative clauses. In *The acquisition of relative clauses* (ed. E Kidd), pp. 39–60. Amsterdam, The Netherlands: John Benjamins.

51. Jaeger TF. 2006 Redundancy and syntactic reduction in spontaneous speech. PhD thesis, Linguistics Department, Stanford University, CA, USA.

52. Gennari SP, Mirković J, MacDonald MC. 2012 Animacy and competition in relative clause production: a cross-linguistic investigation. *Cogn. Psychol.* **65**, 141–176. (doi:10.1016/j.cogpsych.2012.03.002)

53. Bishop DVM. 2006 What causes specific language impairment in children? *Curr. Dir. Psychol. Sci.* **15**, 217–221. (doi:10.1111/j.1467-8721.2006.00439.x)

54. Norbury CF. 2005 Barking up the wrong tree? Lexical ambiguity resolution in children with language impairments and autistic spectrum disorders. *J. Exp. Child Psychol.* **90**, 142–171. (doi:10.1016/j.jecp.2004.11.003)

55. Plaut DC, Shallice T. 1993 Deep dyslexia: a case study of connectionist neuropsychology. *Cogn. Neuropsychol.* **10**, 377–500. (doi:10.1080/02643299308253469)

56. Ueno T, Saito S, Rogers TT, Lambon Ralph MA. 2011 Lichtheim 2: synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron* **72**, 385–396. (doi:10.1016/j.neuron.2011.09.013)

57. Joanisse MF, Seidenberg MS. 2003 Phonology and syntax in specific language impairment: evidence from a connectionist model. *Brain Lang.* **86**, 40–56. (doi:10.1016/S0093-934X(02)00533-3)

58. Thomas M, Karmiloff-Smith A. 2002 Are developmental disorders like cases of adult brain damage? Implications from connectionist modelling. *Behav. Brain Sci.* **25**, 727–750. (doi:10.1017/S0140525X02000134)

59. Thomas MSC, Karmiloff-Smith A. 2003 Modeling language acquisition in atypical phenotypes. *Psychol. Rev.* **110**, 647–682. (doi:10.1037/0033-295X.110.4.647)

60. Dell GS, Schwartz MF, Martin N, Saffran EM, Gagnon DA. 1997 Lexical access in aphasic and nonaphasic speakers. *Psychol. Rev.* **104**, 801–838. (doi:10.1037/0033-295X.104.4.801)

61. Laine M, Tikkala A, Juhola M. 1998 Modelling anomia by the discrete two-stage word production architecture. *J. Neurolinguistics* **11**, 275–294. (doi:10.1016/S0911-6044(98)00006-2)

62. Nozari N, Dell GS, Schwartz MF. 2011 Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cogn. Psychol.* **63**, 1–33. (doi:10.1016/j.cogpsych.2011.05.001)

63. Oppenheim GM, Dell GS, Schwartz MF. 2010 The dark side of incremental learning: a model of cumulative semantic interference during lexical access in speech production. *Cognition* **114**, 227–252. (doi:10.1016/j.cognition.2009.09.007)

64. Gordon JK, Dell GS. 2003 Learning to divide the labor: an account of deficits in light and heavy verb production. *Cogn. Sci.* **27**, 1–40. (doi:10.1207/s15516709cog2701_1)

65. Breedin SD, Saffran EM, Schwartz MF. 1998 Semantic factors in verb retrieval: an effect of complexity. *Brain Lang.* **63**, 1–31. (doi:10.1006/brln.1997.1923)

66. Plaut DC, McClelland JL, Seidenberg MS, Patterson K. 1996 Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychol. Rev.* **103**, 56–115. (doi:10.1037/0033-295X.103.1.56)

67. Rapp B, Caramazza A. 2002 Selective difficulties with spoken nouns and written verbs: a single case study. *J. Neurolinguistics* **15**, 373–402. (doi:10.1016/S0911-6044(01)00040-9)

68. Bottari P, Cipriani P, Chilosi AM, Pfanner L. 2001 The Italian determiner system in normal acquisition, specific language impairment, and childhood aphasia. *Brain Lang.* **77**, 283–293. (doi:10.1006/brln.2000.2402)

69. Conti-Ramsden G, Durkin K, Simkin Z, Knox E. 2009 Specific language impairment and school outcomes. I: identifying and explaining variability at the end of compulsory education. *Int. J. Lang. Commun. Disord.* **44**, 15–35. (doi:10.1080/13682820801921601)

70. Woods BT, Carey S. 1979 Language deficits after apparent clinical recovery from childhood aphasia. *Ann. Neurol.* **6**, 405–409. (doi:10.1002/ana.410060505)

71. Woods BT, Teuber H-L. 1978 Changing patterns of childhood aphasia. *Ann. Neurol.* **3**, 273–280. (doi:10.1002/ana.410030315)

72. Plaut DC. 1999 Connectionist modeling of relearning and generalization in acquired dyslexic patients. In *Neuronal plasticity: building a bridge from the laboratory to the clinic* (eds J Grafman, Y Christen), pp. 157–168. Berlin, Germany: Springer.

73. Oetting JB. 1999 Children with SLI use argument structure cues to learn verbs. *J. Speech Lang. Hear. Res.* **42**, 1261.

74. Riches NG, Faragher B, Conti-Ramsden G. 2006 Verb schema use and input dependence in 5-year-old children with specific language impairment (SLI). *Int. J. Lang. Commun. Disord.* **41**, 117–135. (doi:10.1080/13682820500216501)

75. Hsu HJ, Bishop DVM. 2010 Grammatical difficulties in children with specific language impairment: is learning deficient? *Hum. Dev.* **53**, 264–277. (doi:10.1159/000321289)