Computational Exploration of Lexical Development in Down Syndrome

Angel Eugenio Tovar (eugeniot@comunidad.unam.mx)
Department of Psychology, UNAM
Avenida Universidad 3004, Ciudad Universitaria, Mexico

Gert Westermann (g.westermann@lancaster.ac.uk)
Department of Psychology, Lancaster University
Lancaster LA1 4YF, United Kingdom

Abstract

Research on lexical development in Down syndrome (DS) has emphasized a dissociation between language comprehension and production abilities, with production of words being relatively more impaired than comprehension. Current theories stress the role of associative learning on lexical development. However, there have been no attempts to explain the atypical lexical development in DS based on atypical associative learning. The long-term potentiation (LTP) and long-term depression (LTD) of synapses, underlying associative learning, are altered in DS. Here we present a neural network model that instantiates notions from neurophysiological studies to account for the disparities between lexical comprehension and production in DS. Our simulations show that an atypical LTP/LTD balance affects comprehension and production differently in an associative model of lexical development.

Keywords: Down syndrome; lexical development; associative learning; comprehension/production asymmetries; neurocomputational model.

Down syndrome (DS) is the most common genetic cause of intellectual disability. There has been extensive research in behavioral and neurophysiological sciences to understand how DS affects cognitive development.

One of the behavioral domains that has attracted particular attention in DS is language development, and specifically, lexical development. This is interesting because lexical development has been argued to be based on associative learning mechanisms (McMurray, Horst, & Samuelson, 2012), while studies on the neurophysiology of DS have consistently described an altered mechanism for synaptic adaptation (Begenisic et al., 2014; Scott-McKean & Costa, 2011) which lies at the core of associative learning. Nevertheless, the role of atypical associative learning in lexical development in DS has not been explored.

In this paper, we address this gap by describing a neurobiologically informed computational model that implements an altered associative learning mechanism described in DS to account for the atypical lexical development in DS. Our focus is on explaining an apparent dissociation between lexical comprehension and production in DS. We want to address to what extent this observed dissociation is based on general atypical associative learning mechanisms. Our hypothesis is that interactions between experience and the neurophysiological constraints of DS are sufficient to account for the differences in performance between lexical comprehension and production in this population.

This hypothesis is in accordance with a *domain-general* view of cognitive development, where the process of associative learning is affected overall, but depending on the demands of the task (i.e., comprehension or production) the observed outcomes are qualitatively different. We test this hypothesis in our computational model of lexical development. Therefore, a second aim of this paper is to provide a computational model of atypical lexical acquisition, biologically informed.

Lexical development in Down Syndrome (The Process)

Language development in DS, as in other developmental disorders, has attracted considerable attention for both theoretical and practical reasons. On the one hand, descriptions emphasizing a relatively greater impairment in language abilities in DS (Chapman & Hesketh, 2000; Rice, Warren, & Betz, 2005; Vicari et al., 2004) have motivated theoretical debate on the nature of language as a process resulting from a cognitive system with domain specific vs. domain general components (Marcus & Rabagliati, 2006; Stojanovik, 2014; Thomas & Karmiloff-Smith, 2005). On the other hand, there is interest in understanding atypical language trajectories in DS to develop better interventions and minimize dysfunction in these patients. Among the different domains of language development, in this review we focus on lexical development.

Lexical acquisition is traditionally studied through the number of words produced and number of words comprehended in a certain age range. These numbers are lower in DS when compared to typically developing children (TD) of the same chronological age, but the discrepancy between DS and TD diminishes when DS individuals are compared with TD children of the same mental age (i.e., level of non-verbal cognitive ability) (Galeote, Soto, Sebastián, Rey, & Checa, 2012).

In DS it is commonly reported that language comprehension abilities exceed language production abilities (Galeote et al., 2012; Kay-Raining Bird, Chapman, & Schwartz, 2004; Vicari et al., 2004). This pattern replicates a canonical finding in research of lexical development in TD: the number of words comprehended initially exceeds the number of words produced (McMurray

et al., 2012). However, critically, a number of studies have found that the discrepancy between the comprehension and production of words in DS is greater than expected on the basis of mental age, with comprehension at or near mental-age-typical levels, but production selectively impaired (Kay-Raining Bird et al., 2004; Vicari et al., 2004).

Some studies also suggest that production and comprehension of words in DS follow qualitatively different developmental trajectories (Chapman, Hesketh, & Kistler, 2002; Galeote et al., 2012), with one study reporting that comprehension of words in DS even exceeded the level of non-verbal mental age (Glenn & Cunningham, 2005). However, in contrast to these results, other evidence has suggested that in DS both expressive and receptive language are significantly more impaired than what is expected on the basis of mental age (Bello, Onofrio, & Caselli, 2014).

Due to conflicting results it has been difficult to characterize a unique profile of cognitive and linguistic abilities in DS. High inter-individual variability in the DS population (Karmiloff-Smith et al., 2016), along with methodological constraints including small sample size and the use of different measures and procedures, may explain some disparities between studies. In an effort to analyze a larger sample of DS individuals in verbal skills, Næss and colleagues (2011) meta-analyzed data reported by different research groups between 1988 and 2009, and found that performance of children with DS is in line with TD children, matched by mental age, in receptive vocabulary but is significantly impaired in measures of expressive vocabulary.

A number of questions arise from this apparently uneven profile between lexical comprehension and production, and its failed predictability from the overall level of cognitive development: is lexical development in DS only delayed or deviated from the TD pattern? Is there a dissociation between lexical comprehension and production in DS? Is it possible to account for these results with a domain general approach?

Lexical Acquisition and Associative Learning (The Theory)

There is a vast literature on lexical acquisition and the study of word learning is at the core of this field. Word learning is viewed as the process by which we learn to link a phonological representation with a category of objects. Word learning involves a sequence of complex processes; the learner faces the challenge of selecting discrete phonological representations, picking a specific object in a cluttered visual scene, and creating meaningful representations linking the sounds and the visual objects.

Attempts to explain how the cognitive system deals with such a complexity have been based on three theoretical accounts. First, under the *lexical constraints account*, word learning is guided by a set of default assumptions (i.e., constraints) on hypotheses (Woodward & Markman, 1998). For example, the mutual exclusivity constraint describes the process of inferring which word corresponds with which object on the basis of knowing already the names of the other objects present in the visual scene.

Second, the *social-pragmatic accoun*t argues that children use cues such as the speaker's (e.g., caregiver) gaze or intention to learn the correspondences between sounds and objects (see Ambridge & Lieven, 2011). Third, the *associative learning account* explains word learning as a process governed by the domain-general rules of learning. The focus is on the linkages created between sounds and objects without appealing to any other prerequisites such as lexical constraints or social cues, even when these can exert a modulatory role on word learning. In recent years, this account has been formalized and tested through computational models (Mayor & Plunkett, 2010; McMurray et al., 2012, Westermann & Mareschal, 2014). Computational simulations have provided precise descriptions on how the qualitative properties of lexical development, empirically observed, as is the initial asymmetry between comprehension and production, the vocabulary spurt, and mutual exclusivity, emerge in a system that operates by establishing associations with language-like inputs (e.g., McMurray et al., 2012).

In this paper, we focus on analyzing the disruptive effects that atypical mechanisms of associative learning have on word learning for the DS population. For this reason, our approach is based on the model proposed by McMurray, Horst and Samuelson (2012), we call this the MHS model from here on. We selected this model for the following reasons: first, the theoretical account underlying this model distils the process to its basic computational components and develops an approach focused on the role of associative learning, and this is convenient for our purpose of analyzing atypical forms of associative learning on lexical development. Second, the architecture of this model is well suited to incorporating our computational formalization of biological descriptions of atypical learning in DS. Third, by building on previous work, we extend this previous and well accepted model to account for atypical behavior and in this extension (in terms of behavior, and populations) additional evidence is provided for the associative account of word learning.

Associative Learning in Down Syndrome (The Underlying Mechanism)

From a neurobiological perspective, associative learning results from the adaptation of synaptic connections between neurons. Such adaptations are activity dependent; following Hebbian descriptions high co-activation between pre- and post-synaptic neurons lead to a strengthening of the synaptic connection. Complementary to the Hebbian account, empirical research has shown that decays in the efficacy of synaptic connections are also triggered by the co-activation between the pre- and post-synaptic units. A co-activation threshold is assumed to exist (Bienenstock, Cooper, & Munro, 1982) so that below-threshold co-activation values produce decays in the synaptic efficacy (i.e. long-term depression or LTD) and above-threshold co-activation

values lead to increase the synaptic efficacy (i.e., long-term potentiation or LTP).

A vast literature on the biological bases of associative learning in DS has described an atypical balance between LTP and LTD in different mouse models of this syndrome. When compared with euploid control mice, LTP is limited and LTD is increased in DS (Begenisic et al., 2014; Scott-McKean & Costa, 2011; Siarey, Villar, Epstein, & Galdzicki, 2005). This pattern of synaptic adaptation functionally corresponds with an increased co-activation threshold, where the same level of stimulation produces limited gains and increased decays in the connection strengths in DS relative to TD. An increased co-activation threshold has been proposed for other populations that show cognitive impairment (Meredith & Mansvelder, 2010).

While considerable progress has been made in the study of LTP/LTD in DS, with an emphasis on the design of pharmacological interventions (e.g., Begenisic et al., 2014), building the bridge from the basic level of altered neurophysiology to the high level of cognitive function has seen less progress. For example, it is not clear what is the role of the altered LTP/LTD balance on language development in DS. Descriptions of the exact way by which biological differences contribute to language impairments in different populations (e.g., TD, Williams syndrome, fragile X syndrome) will inform us on what is common across populations, the nature of language impairments, and how the language capacity is vulnerable (Rice et al., 2005).

Given the evidence from two fields of research, one informed by behavioral studies suggesting a preserved and marked asymmetry between comprehension and production of words, and another informed by neurophysiological studies describing an altered mechanism for associative learning, and in the context of an associative learning account to word learning, in this paper our focus is on exploring, the role of atypical associative learning mechanisms in word learning in DS.

Computational Model

Overview and Architecture The present model is based on the MHS model. It is designed to analyze the role of associative learning in the establishment of correspondences between auditory word forms and visual objects. In the following, we describe our model and we indicate the differences between the present model and the MHS model.

The present model is composed of a neural network with three layers of units. Two of these layers represent processing in the auditory and visual systems. These layers are used to present input patterns to the network and to collect responses. These layers are not directly connected with each other; instead they are indirectly connected through a third layer of "lexical units" (see Figure 1).

One assumption of this approach is that the auditory and visual systems can already categorize objects and select discrete elements from the environment. The units in the visual and auditory layers are localist; each unit represent only one stimulus.

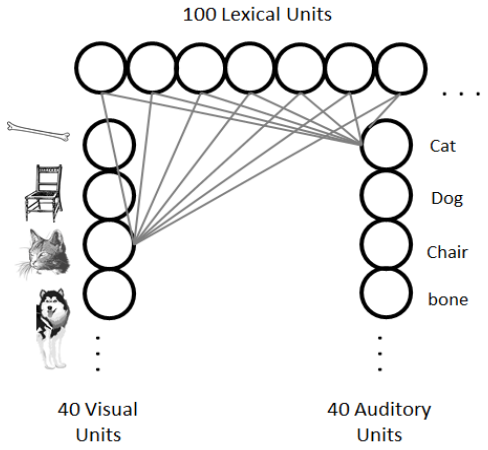

Figure 1: Architecture of the neural network with the visual, auditory and lexical layers. Only a few connections are shown to represent connectivity from auditory and visual units to lexical units.

The auditory and visual layers have 40 units each. Thus, 40 is the total number of words that the network is able to learn. The lexical layer contains 100 units. There are more lexical units than would be needed to learn 40 words –this allows for better learning (McMurray et al., 2012). Since the model could initially randomly associate two different inputs with the same lexical unit, increasing the number of lexical units prevents mismappings and increases discrimination of words (McMurray et al., 2012).

The architecture of the model is similar to the one presented by McMurray and colleagues (2012), but a key difference is in the number of units. The MHS model has 35 input units in the auditory and visual layers, and 500 lexical units. Our model incorporates more input units and fewer lexical units; thus our model requires less computational power to simulate the learning of a higher number of words.

Each unit in the input layers is connected to all the units in the lexical layer. These connections are bidirectional and their weights are initially randomized. In the MHS model, connections are not functionally bidirectional, since they use a different *temperature* parameter for feed-forward and feed-back connections.

Activation values of units range between 0 and 1. The activation values of the lexical units are initially normalized, such that the sum of all activation values equals 1. When an auditory or visual stimulus is presented to the input layers, the unit that represents this stimulus is activated with a value of 1, and all remaining inputs are set to 0. The activation flows through the connections and reaches the lexical layer, which then computes the net input as the sum of activations coming from the auditory and visual inputs weighted by the corresponding connection values. The activation values in the lexical layer then go through a process of normalization (Equation 1), during 7 cycles. In our model 7 cycles are optimal to stabilize 100 lexical units. It is not clear how many cycles the MHS model requires.

$$a_i^{(t+1)} = \frac{(a_i^{(t)})^2}{\sum_i^N (a_i^{(t)})^2} \quad (1)$$

The activation of the lexical units then feeds back to the auditory and visual layers; these units then sum the net input coming from the lexical layer with the activation from direct stimulation. This process allows integration of bottom-up with top-down information. Then, the connection weights are updated according to the rule described below.

Learning The MHS model incorporates a Hebbian learning algorithm that strengthens connections between co-acvtive units. The decay terms in the MHS model weaken the connections when either the lexical-, or the input units are inactive. In our model, the learning algorithm is designed to capture the functional differences in synaptic adaptation between TD and DS, as informed by studies with mouse models. Thus, both strengthening and weakening of connections result from the co-activation of units. Our algorithm incorporates a co-activation threshold (θ). Those co-activation values that surpass θ lead to gains in the connection weights, and co-activation values below θ lead to decays in connection weights. The simulations of DS use a relatively higher value for θ than simulations of TD (i.e., θ = 0.9 for DS and 0.7 for TD). Higher values of θ restrict connection strengthening and increase connection decay; in this way we simulate the atypical pattern of increased LTD and limited LTP that has been described in DS.

To stabilize changes in connection weights we also include a self-adjusting parameter called lambda (λ). It keeps weights between 0 and 1, by reducing changes as weights approach 1. As shown in Equation 2, for above-θ values, λ depends on the difference between the co-activation and the current connection weight. It is computed by subtracting the value of the current weight from the current co-activation. For below-θ values, lambda acquires a negative value proportional to the current weight.

$$\begin{array}{ll} \text{If } (a_i * a_j) > \theta, & \text{Then } \lambda = (a_i * a_j) - W_{ij} \\ \text{Else} & \lambda = -W_{ij} \end{array} \quad (2)$$

Lambda is a multiplicative parameter in the final learning algorithm (Equation 3).

$$W_{ij}(t+1) = W_{ij}(t) + \lambda\, ß\, (a_i * a_j) \quad (3)$$

Changes in weights (W_{ij}) then depend on the co-activation value ($a_i * a_j$) modulated by the interaction between the current state of the connection and the co-activation computed by λ, and a learning rate (ß). We ran two sets of simulations for DS. In the first set (DS-1) we used a relatively lower ß in DS compared to TD simulations to capture additional neurophysiological abnormalities in DS with impact on computing power, namely, a reduction of synapse density and inhibitory predominance (Dierssen, 2012). In the second set of simulations of DS (DS-2) we kept the same value ß as the one used in TD. We did this to be able to compare and explore the effects of an increased θ alone vs. increased θ and lower ß. (ß = 0.001 for TD and DS-2; and ß = 0.0005 for DS-1).

Simulations

Training One auditory object was presented during each training trial along with many visual objects (usually five). These presentations simulate natural scenes where, in a discrete moment, one auditory word form is presented (spoken) to the child in the presence of a cluttered visual scene. For example, the first time a child hears the word /cat/, she can observe a visual scene that contains a cat, but also contains a dog, a container with milk, a ball of yarn, etc. Thus, the word /cat/ could initially refer to any of these visual objects. This problem of referential ambiguity needs to be solved by the child across many trials. Let's consider a second trial when the word /cat/ is presented again, but now the visual scene contains the cat, the container with milk, a pillow, and a table. If the child is sensitive to the environmental regularities, across many trials she will learn the correct correspondences between auditory words and visual objects (Smith & Yu, 2008). But this is a slow process that requires numerous trials. To capture this process, in our simulations, each time that an auditory word was presented, the correct visual object was presented with another 4 different visual objects. The additional visual objects changed for every trial. We simulated the learning of 40 words, by presenting each auditory-visual pairing a total of 20000 times.

Testing We presented trials to evaluate comprehension and production of words. Tests for comprehension were designed, as in the MHS model, to simulate a traditional test of lexical comprehension, The N-alternative forced choice, where a number of different visual objects are presented to the child and she is asked to point or select one in particular (e.g., where is the pencil? which one is the pencil?). In our simulations one auditory stimulus (e.g., pencil) was active, as well as 4 visual objects (e.g., pencil, cat, table, glass) in the visual layer. Activation flowed from inputs to the lexical layer and back. Then the unit in the visual layer with the highest activation (e.g., pencil) was taken as the response of the model. In this way, comprehension was conceptualized as the correct activation of the visual object in the presence of one particular auditory word form.

Following again simulations in the MHS model, tests for production of words were designed to simulate the "child says" measures of the MacArthur-Bates Communicative Development Inventory. In these trials one single visual object was active and all possible auditory word forms were active. Activation flowed from inputs to the lexical units and back, then the auditory unit with the highest activation value was taken as the response of the model. Production then corresponded to evaluating the activation of auditory word forms in the presence of a particular visual object.

The comprehension and production test trials were run after every 50 training epochs (each epoch was composed of the presentation of the 40 training trials). A total of 400 measures of comprehension and production were obtained for each simulation. We ran 20 simulations of TD, 20 of DS-1 and 20 of DS-2.

Results and Discussion

Figure 2 shows the mean values of words comprehended and produced for TD, DS-1, and DS-2. The standard deviation values are shown in the error bars (gray areas).

Our simulation of TD (Fig. 2A) shows that comprehension surpassed production in the early stages of learning; then, from the test trial 51 until the end of the simulation, comprehension and production were matched, and show complete learning of vocabulary.

The simulations of DS-1 and DS-2 (Fig. 2 B and C) show a qualitatively different trajectory of lexical acquisition. Some aspects shown by these simulations are of particular interest in the context of our theoretical and empirical review. First, performance in the comprehension task is always above the performance in the production task. Moreover, production of words never reaches the maximum possible value of 40 words. Second, DS-1 is more affected than DS-2. DS-1 used a higher co-activation threshold with a lower learning rate, while DS-2 used the higher co-activation threshold with a high learning rate. Data from DS-2 suggests that the atypical synaptic learning process in DS has a direct consequence on lexical development on its own, and the difference between DS-1 and DS-2 suggests that the learning rate has an additional effect. Third, the standard deviations show that the performance in the DS groups was more variable than the performance in TD. DS-1 showed the highest variability. These patterns replicate the high inter-individual variability usually observed in DS compared with TD (see Karmiloff-Smith et al., 2016).

Comprehension and production tests were different tasks in our simulations. Comprehension required the selection of a visual stimulus from a sample of a few objects, while production, a more demanding task, required the selection of an auditory stimulus from the total number of auditory word forms. These tasks were designed to reproduce the top-down and bottom-up interactions that a child processes when she produces names vs. when she comprehends auditory words. Then, in our model, the asymmetries between comprehension and production are (partially) explained by the properties of the tasks. Remarkably, the disparity between comprehension and production in TD was overcome as training continued, but this disparity persisted for the DS simulations, thus pointing to the atypical associative learning mechanism as an explanation for the persistence and more marked disparity between comprehension and production of words in DS.

Other factors may as well contribute to the lexical comprehension/production asymmetry in DS, such as an atypical physical development that affects correct articulation of words and therefore restricts experience with lexical production. Our model, however, shows that the atypical pattern of synaptic strengthening directly affects lexical development.

Our approach supports a domain-general view of cognitive development, and we argue that it also strengthens the associative learning account to lexical development, since it explains a pattern of uneven development of lexical abilities in Down syndrome as a result of an altered domain-

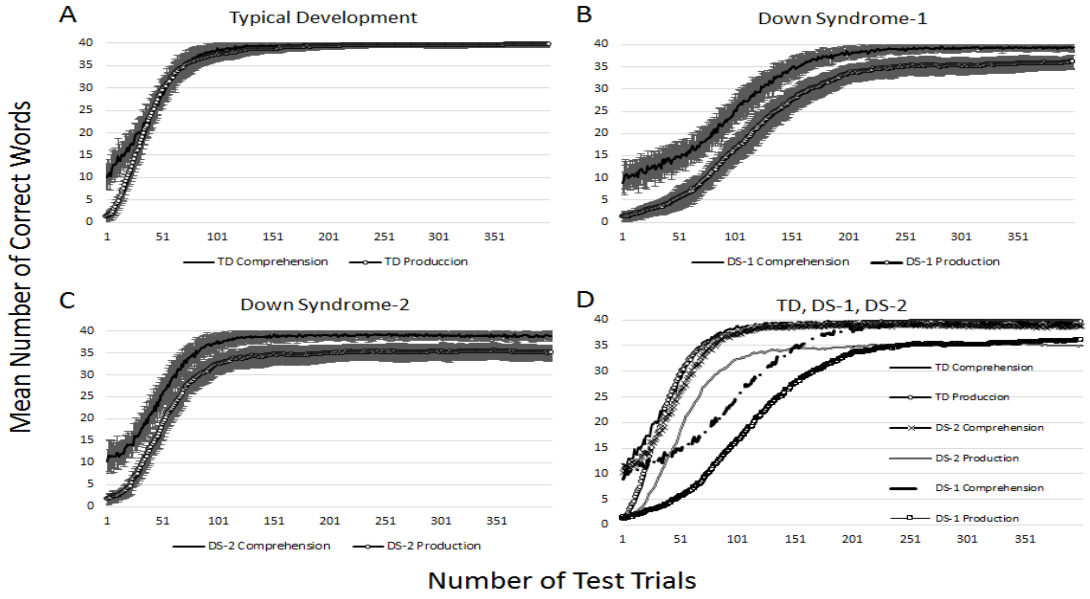

Figure 2: Mean values of comprehension and production across the 400 test trials for TD (Panel A), DS-1 (Panel B) and DS-2 (Panel C). The values from the three populations appear for comparison purposes in Panel D. Gray areas in Panels A, B and C show the standard deviation.

general mechanism in combination with the properties of the behavioral task.

Acknowledgments

AET was funded by Secretaría de Ciencia Tecnología e Innovación-Ciudad de México. GW is funded by the ESRC International Centre for Language and Communicative Development (LuCiD) at Lancaster University (ES/L008955/1) and by a British Academy/Leverhulme Trust Senior Researcher Fellowship.

References

Ambridge, B., & Lieven, E. V. (2011). *Child language acquisition: Contrasting theoretical approaches*. UK: Cambridge University Press.

Begenisic, T., Baroncelli, L., Sansevero, G., Milanese, M., Bonifacino, T., Bonanno, G.,… Sale, A. (2014). Fluoxetine in adulthood normalizes GABA release and rescues hippocampal synaptic plasticity and spatial memory in a mouse model of Down Syndrome. *Neurobiology of Disease*, *63*, 12–19.

Bello, A., Onofrio, D., & Caselli, M. C. (2014). Nouns and predicates comprehension and production in children with Down syndrome. *Research in Developmental Disabilities*, *35*(4), 761–775.

Bienenstock, E. L., Cooper, L. N., & Munro, P. W. (1982). Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *2*(1), 32–48.

Chapman, R. S., & Hesketh, L. J. (2000). Behavioral phenotype of individuals with Down syndrome. *Mental Retardation and Developmental Disabilities Research Reviews*, *6*(2), 84–95.

Chapman, R. S., Hesketh, L. J., & Kistler, D. J. (2002). Predicting longitudinal change in language production and comprehension in individuals with Down syndrome: hierarchical linear modeling. *Journal of Speech, Language, and Hearing Research*, *45*(5), 902–915.

Dierssen, M. (2012). Down syndrome: the brain in trisomic mode. *Nature Reviews. Neuroscience*, *13*(12), 844–858.

Galeote, M., Soto, P., Sebastián, E., Rey, R., & Checa, E. (2012). Vocabulary acquisition in children with Down syndrome: Normative data and developmental trends. *Infancia y Aprendizaje*, *35*(1), 111–122.

Glenn, S., & Cunningham, C. (2005). Performance of young people with Down syndrome on the Leiter-R and British picture vocabulary scales. *Journal of Intellectual Disability Research*, *49*(4), 239–244.

Karmiloff-Smith, A., Al-Janabi, T., D'Souza, H., Groet, J., Massand, E., Mok, K., … Strydom, A. (2016). The importance of understanding individual differences in Down syndrome. *F1000Research*, *5*.

Kay-Raining Bird, E., Chapman, R. S., & Schwartz, S. E. (2004). Fast mapping of words and story recall by individuals with Down syndrome. *Journal of Speech, Language, and Hearing Research*, *47*(6), 1286–1300.

Marcus, G., & Rabagliati, H. (2006). What developmental disorders can tell us about the nature and origins of language. *Nature Neuroscience*, *9*(10), 1226–1229.

Mayor, J., & Plunkett, K. (2010). A neurocomputational account of taxonomic responding and fast mapping in early word learning. *Psychological Review*, *117*(1), 1–31.

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831–877.

Meredith, R. M., & Mansvelder, H. D. (2010). STDP and mental retardation: dysregulation of dendritic excitability in Fragile X syndrome. *Frontiers in Synaptic Neuroscience*, *2*.

Næss, K.-A. B., Lyster, S.-A. H., Hulme, C., & Melby-Lervåg, M. (2011). Language and verbal short-term memory skills in children with Down syndrome: A meta-analytic review. *Research in Developmental Disabilities*, *32*(6), 2225–2234.

Rice, M. L., Warren, S. F., & Betz, S. K. (2005). Language symptoms of developmental language disorders: An overview of autism, Down syndrome, fragile X, specific language impairment, and Williams syndrome. *Applied Psycholinguistics*, *26*(1), 7–27.

Scott-McKean, J. J., & Costa, A. C. S. (2011). Exaggerated NMDA mediated LTD in a mouse model of Down syndrome and pharmacological rescuing by memantine. *Learning & Memory*, *18*(12), 774–778.

Siarey, R. J., Villar, A. J., Epstein, C. J., & Galdzicki, Z. (2005). Abnormal synaptic plasticity in the Ts1Cje segmental trisomy 16 mouse model of Down syndrome. *Neuropharmacology*, *49*(1), 122–128.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.

Stojanovik, V. (2014). Language in genetic syndromes and cognitive modularity. En L. Cummings (Ed.), *The Cambridge handbook of communication disorders* (pp. 541–558). Cambridge: Cambridge University Press.

Thomas, M., & Karmiloff-Smith, A. (2005). Can Developmental Disorders Reveal the Component Parts of the Human Language Faculty? *Language Learning and Development*, *1*(1), 65–92.

Vicari, S., Bates, E., Caselli, M. C., Pasqualetti, P., Gagliardi, C., Tonucci, F., & Volterra, V. (2004). Neuropsychological profile of Italians with Williams syndrome: an example of a dissociation between language and cognition? *Journal of the International Neuropsychological Society*, *10*(6), 862–876.

Westermann, G. and Mareschal, D. (2014) From perceptual to language-mediated categorization. *Philosophical Transactions of the Royal Society B*, 369, 20120391

Woodward, A. L., & Markman, E. M. (1998). Early word learning. In *Handbook of child psychology: Vol. 2: Cognition, perception, and language* (pp. 371–420). Hoboken, NJ: John Wiley & Sons Inc.