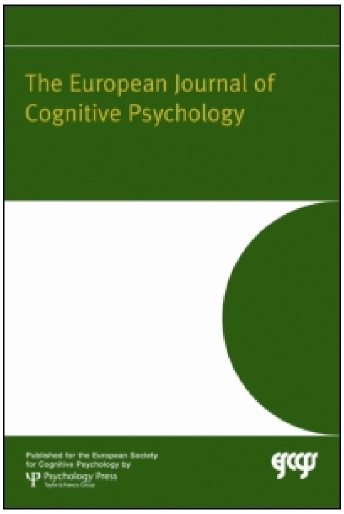

European Journal of Cognitive Psychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713734596

Modelling word recognition and reading aloud

Johannes C. Ziegler[a]; Jonathan Grainger[a]; Marc Brysbaert[b]
[a] Aix-Marseille University and Centre National de la Recherche Scientifique, Marseilles, France [b] Ghent University, Ghent, Belgium

Online publication date: 21 July 2010

To cite this Article Ziegler, Johannes C. , Grainger, Jonathan and Brysbaert, Marc(2010) 'Modelling word recognition and reading aloud', European Journal of Cognitive Psychology, 22: 5, 641 — 649

To link to this Article: DOI: 10.1080/09541446.2010.496263

URL: http://dx.doi.org/10.1080/09541446.2010.496263

EUROPEAN JOURNAL OF COGNITIVE PSYCHOLOGY
2010, 22 (5), 641–649

Psychology Press
Taylor & Francis Group

Modelling word recognition and reading aloud

Johannes C. Ziegler and Jonathan Grainger
Aix-Marseille University and Centre National de la Recherche Scientifique, Marseilles, France

Marc Brysbaert
Ghent University, Ghent, Belgium

Computational modelling has tremendously advanced our understanding of the processes involved in normal and impaired reading. The present Special Issue highlights some new directions in the field of word recognition and reading aloud. These new lines of research include the learning of orthographic and phonological representations in both supervised and unsupervised networks, the extension of existing models to multisyllabic word processing both in English and in other languages, such as Italian, French, and German, and the confrontation of these models with data from masked priming. Some of the contributors to the Special Issue also address hotly debated issues concerning the front-end of the reading process, the viability of Bayesian approaches to understanding masked and unmasked priming, as well as the longstanding debate about the role of rules versus statistics in language processing. Thus, the present Special Issue provides a critical analysis and synthesis of current computational models of reading and cutting edge research concerning the next generation of computational models of word recognition and reading aloud.

The present Special Issue provides a collection of papers that are partially based on communications given at a symposium at the 15th meeting of the European Society for Cognitive Psychology held in Marseille, France. We invited all participants of the symposium to submit papers and in doing so to reexamine their contributions in the light of the stimulating discussions that took place during the symposium. In addition, we solicited papers from

Correspondence should be addressed to Johannes C. Ziegler, Laboratoire de Psychologie Cognitive, Aix-Marseille University, 3 place Victor Hugo, 13331 Marseilles, France. E-mail: ziegler@up.univ-mrs.fr

Preparation of this Special Issue was partially funded by an ERC grant No. 230313 to JG and an Alexander-von-Humboldt fellowship to JCZ.

http://www.psypress.com/ecp DOI: 10.1080/09541446.2010.496263

the wider community of computational modellers. The primary goal was to provide an outlet for original and innovative work as well as an opportunity to discuss recent developments in computational modelling of word recognition and reading aloud. In the following sections, we discuss the importance of computational modelling in this area and we present some of the main issues that are covered in this Special Issue.

WHY CARE ABOUT MODELLING?

Every model is false . . . at least at some level! So why should we keep on modelling word recognition and reading aloud, or anything else for that matter? Despite being false, incomplete, and oversimplified, computational models have allowed us to make tremendous progress over the past decades in terms of understanding the basic processes involved in normal and impaired reading (e.g., (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Grainger & Jacobs, 1994, 1996; McClelland & Rumelhart, 1981; Perry, Ziegler, & Zorzi, 2007, 2010; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989; Zorzi, Houghton, & Butterworth, 1998). A computational model is a model that is implemented as a computer program. As such, computational models *behave*, that is, they produce behaviour that can be measured and evaluated qualitatively and quantitatively (Sun, Fum, Del Missier, & Stocco, 2007). Computational modelling is particularly useful when the system to be investigated is too complex, too interactive, or too difficult to deal with directly. Skilled reading is such a situation because highly specialised visual information needs to be integrated with the spoken language system (phonology, meaning) at various grain sizes and possibly different time scales. These core processes are fairly similar across languages but the language structure and the transparency of the mapping between spelling and sound affect the granularity of certain computations (Ziegler et al., 2010; Ziegler & Goswami, 2005). The amount of interactivity involved makes it difficult to gain a deeper understanding of such a dynamical system without the help of simulations (Rueckl, 2002).

The advantages of computational modelling in the area of reading are obvious. In order to implement a model, one has to specify all details about how information is represented and transformed. If this is not done properly, the model will not run or it will produce inappropriate behaviour. Thus, although verbal theories can live with "comfortable inconsistencies" (Jacobs & Grainger, 1994), computational models cannot. Once computational models are properly implemented, they can be used to simulate data from small-scale experiments and large-scale studies (Ferrand et al., 2010; Yap & Balota, 2009), which together provide strong constraints for testing computational models. Computational models can deal with highly dynamic and interactive processes that could not be assessed in closed-form analyses; they can include random and stochastic processes and the development of fast computers makes it possible to run thousands of simulations to explore entire parameter spaces (Adelman & Brown, 2008; Ziegler, Perry, & Zorzi, 2009) or to investigate interindividual differences (Zevin & Seidenberg, 2006; Ziegler et al., 2008). Most importantly, computational modelling might lead to new ways of understanding the processes of interest, such as when models make novel predictions that have not yet been observed or when modellers are confronted with unexpected results. A particularly successful approach has been associated with nested incremental modelling (Jacobs & Grainger, 1994; Perry et al., 2007). The key idea is that a new model should account for the crucial effects accounted for by the previous generations of the same or competing models, which means that a new model needs to be tested against the data sets that motivated the construction of the old models before it is tested against new data sets. In successful nested modelling, the new model often includes the direct predecessor as a special case (see Zorzi, this issue 2010, for an example of the nested modelling strategy).

However, there are also some dangers associated with computational modelling. The first is the overspecification of irrelevant details (Lewandowsky, 1993). That is, in order to make the model run, modellers might need to apply certain implementation tricks that are of no great theoretical interest or justification but that might improve the performance of the model. It is crucial that modellers are aware and upfront about how much of the job is done by such possibly idiosyncratic implementation tricks. The second problem has to do with overfitting. Overfitting is said to occur when a model perfectly captures selected data sets but does not generalise to new ones. Overfitting can occur when a model does not only capture the major effects but also the noise that is inherent in the data. For example, by correlating response latencies to the same items across various large-scale databases, Perry, Ziegler, and Zorzi (2010) showed that even when the items means from one database were used to predict the item means in another, the amount of variance that was explained never went over 50%. This suggests that, if the same parameters are always used in a model, a model that accounts for more than 50% of the item-specific variance on a large-scale database may have been overfitted. The final problem is also known as *Bonini's paradox*: "As a model of a complex system becomes more complete, it becomes less understandable. Alternatively, as a model grows more realistic, it also becomes just as difficult to understand as the real-world processes it represents" (Dutton & Starbuck, 1971, p. 4). Clearly, some of the recent models have reached a level of complexity that makes them more difficult to understand than some of the earlier models of word recognition (e.g., McClelland & Rumelhart, 1981). At the same time, they have become a lot more powerful, accounting for up to 50% of the item-specific variance on large-scale databases (Perry, Ziegler, & Zorzi, 2010). The present Special Issue is a nice demonstration that the advantages of computational modelling clearly outweigh the disadvantages. In what follows, we will summarise some of the topics that are discussed in the Special Issue.

BEYOND READING ALOUD SINGLE SYLLABLES

A number of papers in the Special Issue deal with extending existing models or modelling approaches to reading multisyllabic words. By doing so, the papers also tackle some novel challenges, such as syllabic segmentation, stress assignment, effects of syllable frequency, and, most importantly, the scaling-up of these models to a size that approximates the human lexical system. The paper by Sibley, Kello, and Seidenberg (this issue 2010) shows how simple recurrent networks can be used to overcome the problem of representing words of variable length. A first sequence encoder model learned to map letter inputs onto fixed-width representations; a second sequence encoder model then learned to map these fixed-width representations onto phoneme outputs. The model was benchmarked against the mono- and disyllabic data from the English Lexicon Project (Balota et al., 2007), where it accounted for up to 25% of the item-specific variance. It is noteworthy that the model correctly simulated effects of stress assignment and syllable length. The paper by Pagliuca and Monaghan (this issue 2010) also extended the parallel distributed processing (PDP) approach to reading multisyllabic words, this time in a transparent language (Italian). The model was successful in learning to read 98% of the Italian lexicon accurately, even though words varied in length from one to three syllables. In particular, they assessed whether the transparency of the spelling-to-sound mapping of Italian would nevertheless allow the model to become sensitive to effects of larger grain sizes, such as morphemes. This was the case. Despite the regular one-to-one correspondence between letters and phonemes in Italian, the model learned a whole range of grain sizes, in some cases as large as the morpheme. This way the model was also able to simulate morphological effects in nonword reading without recourse to semantics or explicit representations of morphemes. The paper by Conrad, Tamm, Carreiras, and Jacobs (this issue 2010) is also concerned with multisyllabic processing in a language other than English, here Spanish. In contrast to the PDP models described earlier, Conrad et al. stayed within the localist interactive activation framework by simply adding an explicit syllable level of representations to the multiple readout model of Grainger and Jacobs (1996). They show that such a model was able to simulate the inhibitory syllable frequency effect, an effect that might well become a challenging benchmark for all other models.

LEARNING ORTHOGRAPHIC REPRESENTATIONS IN UNSUPERVISED NETWORKS

One area of computational modelling of reading that has been largely neglected until recently is the potential role of unsupervised learning algorithms (e.g., Li, Farkas, & MacWhinney, 2004; Zorzi, Perry, Ziegler, & Coltheart, 1998). Historically, the supervised backpropagation algorithm has dominated modelling efforts with networks that learn to read words (as opposed to stable-state networks such as the Interactive-Activation network). However, it is possible that part of the learning process proceeds in an unsupervised way, via passive exposure to printed words. While the beginning reader is focused on learning how to translate print-to-sound, unsupervised learning mechanisms could in parallel develop higher level orthographic representations of the printed words that the child is exposed to. The paper by Dufau et al. (this issue 2010) addresses this by investigating whether a self-organising neural network, trained on a realistic developmental corpus, would capture the developmental trajectory of two marker effects of visual word recognition, the effects of frequency and neighbourhood size. These results highlight the potentially important role of unsupervised learning for the development of orthographic word forms.

NONWORD READING: MORE THAN WORDS CAN SAY

The Special Issue highlights that nonword reading performance, that is, the ability to generate plausible pronunciations to novel items, remains not only the hardest test for computational models of reading aloud but also an area that provides invaluable insights into the way the brain maps spelling onto sound. The paper by Sibley et al. (this issue 2010) shows that while previous sequence encoder models had serious problems with nonword generalisation performance (Kello, 2006), nonword performance can be improved when the sequence elements consist of vowel-consonant clusters rather than single letters or phonemes. Thus, there seems to be interesting solutions to what appeared to be fundamental limitations of earlier PDP models. The paper by Perry et al. (this issue 2010) further investigates the variability of human nonword pronunciations with regard to vowel length in German. The data show that people are sensitive to the statistical distribution of vowel length in the mental lexicon and that they use surrounding context to determine vowel length. The authors demonstrate that a German version of the connectionist dual process model (CDP+) does a better job in capturing the variability of nonword pronunciations than a rule-based model, here the German version of the dual-route cascaded (DRC) model (Ziegler, Perry, & Coltheart, 2000).

FRONT-END OF READING

One area of research on visual word recognition and reading that has attracted increasing attention in recent years, is the front-end of the reading process—that is, the very first processing stages involved in transforming the visual stimulus into a linguistic code (for reviews, see Grainger, 2008; Grainger & Dufau, in press). More specifically, this change in focus towards early stages of processing has highlighted one fundamental question for research on printed word identification in languages that use alphabetical orthographies—how the positions of the word's component letters are encoded. Indeed, the activation of different letter identities by visual features extracted from the printed word must retain some information about the positions of these letters in order to distinguish between anagrams such as "trail" and "trial". How exactly such letter position coding is achieved is still a matter of debate, but there is at least a general consensus that the coding system must be fairly flexible in order to account for a number of key empirical phenomenon, such as transposed-letter priming effects (Schoonbaert & Grainger, 2004) and relative-position priming effects (Grainger, Granier, Farioli, van Assche, & van Heuven, 2006; Peressotti & Grainger, 1999). The paper by Davis (this issue 2010) provides a comprehensive comparison between two of the most influential models of orthographic coding, processing, the Self-Organising Lexical Acquisition and Recognition model (SOLAR; Davis, 1999) and the Sequential Encoding Regulated by Inputs to Oscillating Letter Unit model (SERIOL; Whitney, 2001).

FAST PHONOLOGY AND MASKED PRIMING

The masked priming paradigm has flourished over the last decade to become one of the dominant paradigms for research on single word reading. However, many of the key phenomena observed with this paradigm have received little or no attention from the computational modelling community. The present Special Issue rectifies this state of affairs to some extent as three contributions are directly concerned with modelling masked priming. The paper by Diependale, Ziegler, and Grainger (this issue 2010) tackles fast phonological priming effects (*bloo* primes *BLUE*). Indeed, this phenomenon was an important demonstration of the automatic activation of phonological representations during the process of silent reading for meaning. Critically, a study by Rastle and Brysbaert (2006) not only showed that this phenomenon was robust, but also suggested that the phonological assembly process in the classic DRC model (Coltheart et al., 2001) was too slow or too weak to allow the model to capture the effect. Diependale et al. show that the Bimodal Interactive Activation Model (BIAM), an extension of the interactive activation model, can capture the effect due to its fast parallel mapping of letters onto input phonemes. The paper by Mousikou, Coltheart, and Saunders (this issue 2010) addresses a second priming phenomenon that stimulated a lot of research in the past, the masked onset priming effect (MOPE). They provide a comprehensive review of the empirical findings on the MOPE in the English language, and they show how DRC simulates the various facets of the effect. Finally, the paper by Bowers (this issue 2010) provides a critical assessment of one of the core assumption of the Bayesian Reader model, namely that masked priming can be explained as optimal decision making to a target when the prime and target are (mistakenly) treated as a single perceptual object (Norris & Kinoshita, 2008). Bowers shows how interactive-activation type models can account for the data that were thought to challenge this class of models and highlights some priming results that seem to be difficult to account for by the Bayesian Reader.

CONCLUSION

The present Special Issue provides a perfect illustration of the strength of computational modelling at this moment in time for our understanding of the core processes involved in visual word recognition and reading aloud. Many of the models discussed here are publically available and will hopefully motivate decisive experiments and further developments.

REFERENCES

Adelman, J. S., & Brown, G. D. A. (2008). Methods of testing and diagnosing model error: Dual and single route cascaded models of reading aloud. *Journal of Memory and Language*, *59*, 524–544.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445–459.

Bowers, J. S. (2010). Does masked and unmasked priming reflect Bayesian inference as implemented in the Bayesian Reader? *European Journal of Cognitive Psychology*, *22*(5), 779–797.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. C. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256.

Conrad, M., Tamm, S., Carreiras, M., & Jacobs, A. M. (2010). Simulating syllable frequency effects within an interactive activation framework. *European Journal of Cognitive Psychology*, *22*(5), 861–893.

Davis, C. J. (1999). *The self-organising lexical acquisition and recogniti on (SOLAR) model of visual word recognition*. Unpublished doctoral dissertation, University of New South Wales, Australia.

Davis, C. (2010). SOLAR versus SERIOL revisited. *European Journal of Cognitive Psychology*, *22*(5), 695–724.

Diependale, K., Ziegler, J. C., & Grainger, J. (2010). Fast phonology and the Bimodal Interactive-Activation Model. *European Journal of Cognitive Psychology*, *22*(5), 764–778.

Dufau, S., Lété, B., Touzet, C., Glotin, H., Ziegler, J. C., & Grainger, J. (2010). A developmental perspective on visual word recognition: New evidence and a self-organising model. *European Journal of Cognitive Psychology*, *22*(5), 669–694.

Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Meot, A., et al. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods*, *42*, 488–496.

Grainger, J. (2008). Cracking the orthographic code: An introduction. *Language and Cognitive Processes*, *23*, 1–35.

Grainger, J., & Dufau, S. (in press). The front-end of visual word recognition. In J. S. Adelman (Ed.), *Visual word recognition: Vol. 1. Models and methods, orthography and phonology*. Hove, UK: Psychology Press.

Grainger, J., Granier, J. P., Farioli, F., van Assche, E., & van Heuven, W. J. (2006). Letter position information and printed word perception: The relative-position priming constraint. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(4), 865–884.

Grainger, J., & Jacobs, A. M. (1994). A dual read-out model of word context effects in letter perception: Further investigations of the word superiority effect. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(6), 1158–1176.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, *103*(3), 518–565.

Jacobs, A. M., & Grainger, J. (1994). Models of visual word recognition: Sampling the state of the art. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(6), 1311–1334.

Kello, C. T. (2006). Considering the junction model of lexical processing. In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing*. New York: Taylor & Francis.

Lewandowsky, S. (1993). The rewards and hazards of computer simulations. *Psychological Science*, *4*, 236–243.

Li, P., Farkas, I., & MacWhinney, B. (2004). Early lexical development in a self-organizing neural network. *Neural Networks*, *17*(8–9), 1345–1362.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: 1. An account of basic findings. *Psychological Review*, *88*(5), 375–407.

Mousikou, P., Coltheart, M., & Saunders, S. (2010). Computational modeling of the masked onset priming effect in reading aloud. *European Journal of Cognitive Psychology*, *22*(5), 725–763.

Norris, D., & Kinoshita, S. (2008). Perception as evidence accumulation and Bayesian inference: Insights from masked priming. *Journal of Experimental Psychology: General*, *137*(3), 434–455.

Pagliuca, G., & Monaghan, P. (2010). Discovering large grain sizes in a transparent orthography: Insights from a connectionist model of Italian word naming. *European Journal of Cognitive Psychology*, *22*(5), 813–835.

Peressotti, F., & Grainger, J. (1999). The role of letter identity and letter position in orthographic priming. *Perception and Psychophysics*, *61*(4), 691–706.

Perry, C., Ziegler, J. C., Braun, M., & Zorzi, M. (2010). Rules versus statistics in reading aloud: New evidence on an old debate. *European Journal of Cognitive Psychology*, *22*(5), 798–812.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, *114*(2), 273–315.

Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the connectionist dual process (CDP++) model. *Cognitive Psychology.* doi:10.1016/j.cogpsych.2010.04.001.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*(1), 56–115.

Rastle, K., & Brysbaert, M. (2006). Masked phonological priming effects in English: Are they real? Do they matter? *Cognitive Psychology*, *53*, 97–145.

Rueckl, J. G. (2002). The dynamics of visual word recognition. *Ecological Psychology*, *14*(1–2), 5–19.

Schoonbaert, S., & Grainger, J. (2004). Letter position coding in printed word perception: Effects of repeated and transposed letters. *Language and Cognitive Processes*, *19*, 333–367.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*(4), 523–568.

Sibley, D. E., Kello, C. T., & Seidenberg, M. S. (2010). Learning orthographic and phonological representation in models of monosyllabic and bisyllabic naming. *European Journal of Cognitive Psychology*, *22*(5), 650–668.

Sun, R., Fum, D., Del Missier, F., & Stocco, A. (2007). The cognitive modeling of human behavior: Why a model is (sometimes) better than 10,000 words. *Cognitive Systems Research*, *8*, 135–142.

Whitney, C. (2001). How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin and Review*, *8*(2), 221–243.

Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language*, *60*, 502–529.

Zevin, J. D., & Seidenberg, M. S. (2006). Simulating consistency effects and individual differences in nonword naming. *Journal of Memory and Language*, *54*, 145–160.

Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., et al. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science*, *21*, 551–559.

Ziegler, J. C., Castel, C., Pech-Georgel, C., George, F., Alario, F. X., & Perry, C. (2008). Developmental dyslexia and the dual route model of reading: Simulating individual differences and subtypes. *Cognition*, *107*, 151–178.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3–29.

Ziegler, J. C., Perry, C., & Zorzi, M. (2009). Additive and interactive effects of stimulus degradation: No challenge for CDP. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 306–311.

Zorzi, M. (2010). The connectionist dual process (CDP) approach to modelling reading aloud. *European Journal of Cognitive Psychology*, *22*(5), 836–860.

Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1131–1161.

Zorzi, M., Perry, C., Ziegler, J. C., & Coltheart, M. (1998). Category-specific deficits in a self-organizing model of the lexical-semantic system. In D. Heinke, G. W. Humphreys, & A. C. Olson (Eds.), *Connectionist models in cognitive neuroscience* (pp. 137–148). London: Springer-Verlag.