

Michelle Morales & Felix Grezes

Professor Rosenberg

NLP/ML/Web

October 21, 2013

Review of Article: "Formal grammar and information theory: together again?"

Fernando Pereira 2000

In this article Pereira (2000) supports the idea that research on models of language is split between two conflicting approaches: Chomsky's formal linguistics approach, and Shannon's information theory approach. He also posits that within recent years (1990s-2000) information theory is thriving with successful implementation in the fields of speech recognition, machine translation, language analysis, and information retrieval. He attributes this success to the increase of storage capacity and speed of computers, which are now able to automatically estimate occurrences of linguistic events in large corpora of text and speech. Pereira's major claim builds off this success arguing that "progress in the design and analysis of such models is not only playing a central role in those practical advances but also carry the promise of fundamentally deeper understanding of information-theoretic and computational-complexity constraints on language acquisition" (Pereira, pg.2).

Pereira begins by giving a synopsis of the history behind the split of formal linguistics and information theory. A quote from Chomsky's book Syntactic Structures (1957) is used

to demonstrate what Pereira describes as the commencement of the divide. The quotes shown below in (1) and (2) capture well the two ideas Pereira pins against each other.

(1) Colorless green ideas sleep furiously

(2) Furiously sleep ideas green colorless

(Chomsky 1957)

As Chomsky describes it, these sentences show how a statistical model for grammaticality would rule out both sentences on the same grounds: remoteness. However, these sentences differ: (1) is grammatical and (2) is not. So where information theory would view these as 2 equally rare sentences formal linguistics would show how one is superior due to grammaticality. Contrary to Chomsky's view, linguist Zellig Harris advocated for an alliance between the two ideas. Like Harris, Pereira believes that both theories are beneficial and essential to each other. Pereira describes Harris' framework to show this alliance. Harris's framework includes 4 constraints: partial order, likelihood, reduction, and linearization. As Pereira states Harris's constraints are successful in "how they explicitly link linguistic structure with distributional regularities involving the relative frequencies of different structural configurations" (Pereira, pg.3). Harris's framework incorporates ideas from both a formal linguistic and information theory perspective; a clear alliance between information theory and formal linguistics can be seen.

Pereira then continues to discount Chomsky's 1957 claim about sentences (1) and (2). Chomsky's argument relies on the idea that unseen events are given a zero probability. It is now known that this naïve method overfits to training data and Pereira shows how

advances have solved this issue making Chomsky's point moot. He shows how forms of regularization, such as smoothing, can be used to constrain statistical models and lead to better generalization of unseen data. Thus showing the strength of one perspective: information theory. He also shows how these models now make more accurate and more formal linguistically aligned decisions than before.

Pereira then shows how hidden variables are useful tools to incorporate in a model. The Markov model of language is able to encode state information as the "externally observable sequence of past linguistic behavior" (Pereira pg. 6). Although linguistically simplistic, Pereira argues these examples are great in testing ideas for statistical modeling and constraints on model structure. Pereira shows how hidden random variables can be valuable in representing a user's uncertainty about the best response to an event, given a particular amount of observed events. Pereira argues that the prior distribution of probabilities in these hidden models can be seen as a form of 'innate knowledge' that implicitly biases the learner towards a particular grammar. However, he states that those interested in language acquisition have ignored this idea. But he continues on to argue that new theoretical approaches show that these models may play a regularizing role as well.

In sections 5-8 Pereira gives more examples to support the argument that formal linguistic ideas are supported throughout statistical information-theoretic models. He discusses the connection between information theory and formal linguistics in lexicalized models, in the correlation of a token and its environment, in local structure and global distribution, and in decision making.

Overall, Pereira succeeds in showing how recent advances in statistical modeling and machine learning have not only succeeded in practice but have also begun to capture

the productive and recursive nature of language. He shows how lexical models using information, such as phrase head, argument, and adjunct, are able to categorize complex linguistic events. He shows how machine learning may be used to create novel events by observing previous sequences of decisions. He provides many examples to show that these two perspectives, Chomsky's and Shannon's, are not quite as black and white as they may seem. And like Harris he demonstrates how an alliance may and should exist, importantly noting how beneficial this alliance is. His argument is clear and his paper does well to support it.

I wonder whether this paper written in 2000 still captures the existing debate between information theory and formal linguistics. I have noted many linguistically based approaches used in statistical models to help find solutions to the big questions in Natural Language Processing. I've seen tools like WordNet, FrameNet, and the Stanford dependency parser being used in conjunction with machine learning to accomplish tasks. The creation of these tools and the subsequent application of them are intuitively linguistic in nature. Therefore, I do believe NLP researchers understand the importance and influence of linguistic theory. I believe that two questions remain: how is linguistic theory defined? And do linguists value the research being done with computational models? I believe the former is the easier to answer: it depends. It depends on if you are functionalist or a generativist, if you believe in minimalism or not, and so forth. The answer to the second question follows from the first, your views as a linguist will influence how much or how little you value the Shannon theory inspired research. As computational linguists both are valued with a large interest in how they can be used in conjunction.