Book Reviews

DOI: 10.1017/S1351324905213724

Knowledge and Learning in Natural Language presents Yang’s Variational Model of language acquisition. In overview, the book is composed of five sections: the first two are dedicated to the justification and detailed description of the Variational Model; in sections 4 and 5 the model is applied (making predictions about child language development and language evolution) and evaluated with respect to empirical evidence from child learners and historical texts; leaving section 3–perhaps mispositioned–a discussion of a competitive rule model for learning past-tense (similarly motivated to the variational acquisition model), that augments Pinker’s Words-and-Rules model, and perhaps surpasses it in its ability to account for empirical developmental evidence.

Yang’s work is from the school that believes that Gold’s (1967) paradigm of identification of a target in the limit is misleading when applied to natural language acquisition; successful acquisition cannot be defined as the convergence upon a target grammar, since linguistic environments containing productions exclusively from a single grammar do not exist. At this point, note that if we re-define a ‘target’ to refer to the grammar that most closely describes the language of the environment, then Gold’s paradigm still holds. Yang’s approach, however, is not to converge upon a target at all; instead arriving at a probability distribution for a grammar space. Acquisition becomes the process of associating probabilities with grammars. For such an approach to be feasible a finite grammar space is required; Yang cites the Argument from the Poverty of Stimulus (if we know X, and X is undetermined by learning experience, then X must be innate) as evidence for a Universal Grammar and thus an innately defined grammar space.

Yang sets out to provide a model of acquisition that is rigorous over three criteria: formal sufficiency, developmental compatibility, and explanatory continuity. He argues the inadequacy of previous methods with respect to these criteria:

– Yang classifies generalised statistical methods of learning to be those in which the learner has very little prior language knowledge and where learning is achieved by approximating the input distribution of the adult language. He argues that such methods are inadequate as models of learning since they cannot account for structures appearing in the developing language of a child, yet not occurring in the adult language: for example, optional infinitives and subject drop in children whose environmental language is English.
– The UG-based transformational methods of learning assume that the learner is endowed with a Universal Grammar which defines a finite hypothesis space
of possible languages. Learning is then the process of updating hypotheses; moving from one hypothesis to next on the basis of the input examples from the adult language. Yang rules out these methods; the developmental language of a child does not show any evidence of undergoing direct change when one hypothesis is replaced by another.

In particular, Yang reviews the inadequacies of the Trigger theory for parameter setting. Parameter setting is a method for moving between hypotheses. Chomsky has suggested that the finite hypothesis space is defined by a Universal Grammar consisting of principles and parameters (principles being a set of properties that are constant to all languages and parameters being language variables). For illustration: if the Universal Grammar contains N binary valued parameters the hypothesis space will be of size 2^N. The language of the environment (the target grammar) can be described by a particular setting of the parameters. Hence, by adopting the Principles-and-Parameters framework, learning becomes the process of finding the correct parameter settings for the environment.

The function of a triggering algorithm is to modify parameter values every time the learner cannot parse the current input language example (utterance) using the current settings. This method has been well analysed and has known problems such as:

1. local maxima – referring to the phenomena where a non-target grammar is reached from which the learner can never reach the target grammar;
2. ambiguous triggers – some input examples can be parsed by more than one grammar in the hypothesis space (i.e. there is more than one way to configure the parameter settings to achieve a successful parse), in these situations which grammar should be chosen? Choosing unwisely could lead to a local maxima;
3. noise – the algorithm is deterministic and will attempt to modify parameters even when the input example is erroneous.

The Structured Trigger Learner (Sakas & Fodor 2001), which addresses some of these shortcomings, is also excluded by Yang as a possible learning model as it requires the learner to make an active choice between competing grammars; Yang believes that a learner must be 'dumb', since acquisition is possible regardless of general intelligence.

So, what is Yang’s model of acquisition? Yang’s Variational Model combines elements of both statistical and UG-based models. The general principle involves pairing all possible grammars with a weight representing the prominence of that grammar in the learner’s language faculty; in other words, assigning a probability to every grammar in the hypothesis space. Grammar selection for parsing (and also production) is a function on the probability distribution over the grammars. The selected grammar is either rewarded or penalised depending on whether or not it is able to parse the incoming utterance. Rewards and penalties are calculated in accordance with the Linear Reward-Penalty Scheme (Bush & Mosteller 1951); a successful grammar has its probability increased while all other grammars’ probabilities are decreased; an unsuccessful grammar has its probability decreased while all other grammars’ probabilities are increased. Acquisition is complete when the grammar weights become constant. The method is passive and requires the learner to do
nothing other than select a grammar from the distribution and then reward or punish as appropriate.

There are several questions over such an approach: The first problem is that it is not entirely obvious that the weights will ever become constant. As Yang points out in his arguments over the failure of Trigger Learners, there is a great deal of noise inherent in speech. Surely then, without an ability to differentiate noise from well-formed utterances, the learner might penalise a grammar undeservedly. Since noise is unpredictable by definition the weights are not guaranteed to settle to constant values; they may, however, fluctuate within a tolerance threshold. Yang has a ‘batch’ mechanism in place that alleviates this problem without complicating the elegant maths of a Linear Reward Penalty Scheme. The batch approach requires that a grammar is confirmed or denied B times before the weights are recalculated. This initiative will definitely help to counteract the effects of noise – despite Yang’s motivation for implementing it being to approximate a variable learning rate, such that rewards become greater the closer the grammar is to the target.

The second problem with the Variational Model is that it is not altogether clear that the problem of local maxima has been solved. Yang’s model selects a grammar for parsing according to the current probability distribution, if that particular grammar is successful it alone is rewarded while all others are penalised; his learner has no idea whether the input utterance may be parsed by any other grammar in the grammar space. So what happens if a grammar that happens to represent a superset of the actual target, takes an early lead in the probability distribution? This grammar will be selected more often than any other grammar as a possibility for parsing – it will always be rewarded since it is capable of parsing all sentences that the actual target could – the more it is rewarded, the greater the rewards become, until finally the learner converges on the wrong target. During the entire learning process the learner was unaware that the utterances he received were ambiguous; and since the superset grammar can never be penalised, he could never realise his mistake.

The remaining problem is that the model does not reward individual parameters in the grammars but rather the entire grammar; leading to problems of interference between parameters. Consider a Universal Grammar with N binary valued parameters, let grammar A differ from the target grammar by 1 parameter value and grammar B differ by N parameter values: using Yang’s model, grammars A and B receive equal punishment when unable to parse the current input example. The fact that grammar A was almost totally correct goes unnoticed. Furthermore, knowing that even 1 parameter is correct reduces the hypothesis space by a factor of 2 (from 2^N to 2^(N-1)). By not considering parameters individually Yang’s method never reduces the hypothesis space. Yang admits that his approach is naive but hopes that ‘in the long run, the correct parameter values will prevail.’ The problem with this approach is that the specification of a Universal Grammar composed of parameters is now redundant, other than to define the size of the hypothesis space.

These problems aside, Yang has created a formally specified model capable of making quantitative predictions over developmental language patterns as well as giving an explanation for continuity between adult and child language. Furthermore, he is able to ‘formalise historical linguists’ intuition that grammar competition is
a mechanism for change’; an approach which has scope to be extended to a formalisation of wave theory.

Finally, a word about section 3 – Yang’s competitive rule model for learning past-tense. Here, Yang extends the concept of grammar competition to that of rule competition for generative phonology. A variation on Pinker’s (1999) Words-and-Rule model is suggested. To recap, the Words-and-Rule model requires a default rule for past tense formation; past tenses are formed using this rule together with a memorised word stem. The learner arrives at irregular past tenses by blocking default rule formation with a more specific form. Yang believes that past tenses are formed by applying phonological rules to verb classes; his Rules-and-Competition model, associates a probability distribution to a rules-space for each class of verbs. Now, errors in verb tense formation are no longer memory lapses, as Pinker suggests, but a result of the wrong rule being selected because of the probability distribution associated with the rules-space.

The formalisation of the Rules-and-Competition model allows Yang to discuss his theory with relevance to empirical data. The argument is convincing although not without problems, for instance the method by which we arrive at the division of verbs into classes is certainly not clear.

It is possible, however, to envisage Yang’s model extending to other problems in acquisition; in fact, the model could be used in any situation where learning can be expressed as a competition between forms. For an example, consider the acquisition of a phoneme set: adults identify and use only the phonemes of their own language despite being born able to recognise the phonemes of all languages; such a learning process is easily modelled with a distribution on a phoneme-space.

Yang’s ‘Knowledge and Learning in Natural Language’ provides a formal yet readable account of language acquisition, development and change. He combines the naturalistic theory of a Universal Grammar with a learner that is mathematically rigorous; allowing numerical predications to be forecast and analysed with respect to empirical data. Unfortunately, the first edition contains several typos (even in equations), but none so bad as to considerably alter the understanding. Also, Yang’s implementation of the Variational Model is in places, in the authors own words, “naive”; however, the theory influencing the model, that of a competition between weighted grammars, is convincing both logically and empirically.

**References**


Paula Buttery

Computer Laboratory, University of Cambridge, UK
This volume is a collection of ten original contributions on several areas of linguistics and language technology of interest to potential language engineers including topics such as grammar writing, knowledge representation and statistical modeling. The backcover states that the volume is intended for “linguists hoping to enter this burgeoning field of language processing [Language Engineering]”. Consequently it is assumed that the reader has some general knowledge of the field of linguistics but not previous knowledge in Natural Language Processing or Computer Science is assumed. The range of contributors, which includes academics, researchers from industrially sponsored labs and engineers in companies producing Language Engineering products, reflects the applied nature of Language Engineering. The volume is edited by Ali Farghaly, a linguist at SYSTRAN.

The volume begins with an introductory chapter by the editor. Farghaly argues that there is an increase in the need for linguists who can apply their skills to practical language processing problems and this can be seen from an increase in the number of jobs in industry advertised on the LINGUIST email list. But Farghaly also warns that the field is more complex than it may first appear, citing the ambiguity on natural language and lack of world knowledge available to systems as potential problems. These issues will be all too familiar to those already working in Language Engineering but may not be obvious to those new to the field.

One strategy for reducing the problems raised by ambiguity is to restrict attention to sublanguages. Chapter 2 “Domain Analysis and Representation” (Ali Farghaly and Bruce Hedin) describes the differences in linguistic behavior which one may expect to observe when working with text in a particular domain and point out that taking this information into account can make it significantly easier for a Language Engineering application to correctly interpret text from that domain. For example, it is known that domain knowledge can help to resolve word sense ambiguities but it can also help with noun bracketing and anaphora resolution. Farghaly also describes how domain-dependent dictionaries are used to supplement standard ones in the SYSTRAN machine translation system.

Chapter 3 “Language and the Internet” (Naomi S. Baron) attempts to provide comprehensive coverage of the implications for the Language Engineer arising from the rise of the internet as a mode of communication. There is a description of the linguistic phenomena, such as ungrammatical text and novel lexical items, which may be encountered in internet communications. The main contributions of this chapter are to provide an outline of the various modes of communication which have been created by new technologies (email, SMS text messaging on mobile phones and web pages), an outline of markup languages, an overview of search technologies with mention of the Semantic Web and agents and, finally, some of the
issues related to searching in a multilingual environment and the use of machine translation.

The next chapter, “Grammar Writing, Testing, and Evaluation” (Miriam Butt and Tracy Holloway King), is a practical guide to constructing natural language grammars. Much of the current research on parsing focuses on the automatic learning of syntactic rules from parsed corpora and the fact that most practical parsers rely on hand-crafted grammars is sometimes overlooked. This sentiment is reflected by the authors in their introduction: “Grammar writing is a difficult and often underappreciated task in natural language processing.” (p. 129) The authors present the main features of some well-known grammar formalisms, make a distinction between deep and shallow syntactic analysis and give practical advice on managing the process of writing and testing large grammars.

Chapter 5 “Ontologies” (Natalya Noy) describes the role of ontologies in language engineering. Noy also describes some of the most widely used ontologies and tools which can be used for ontology development. Of particular interest is a step-by-step outline of a process which could be applied to the construction of any given ontology which is motivated using an example ontology which describes credit card transactions.

Chapter 6 “Text Mining, Corpus building, and Testing” (Karine Megerdoomian) provides an introduction to corpus linguistics. Megerdoomian describes some of the best known corpora used in NLP research and the different types of linguistic annotation they contain. The chapter also describes some ways in which corpus linguistics could be used for Language Engineering by describing applications including lexicon acquisition, word sense disambiguation and machine translation. The chapter also includes a description of the corpus processing tools available in the UNIX operating system plus various concordancing and collocation analysis tools.

Chapter 7 “Statistical NLP” (Chris Callison-Burch and Miles Osborne) provides an overview of the statistical approaches. This methodology may appear strange to those from a linguistics background but the authors take time to justify statistical approaches in terms of the advantages they offer in providing rapid prototyping, robustness and cost efficiency. A useful aspect of this chapter is that it provides an excellent step-by-step explanation of the process involved when applying statistical techniques to language processing problems; modeling the problem, estimating the values of parameters, learning from data and evaluation. This sort of information is difficult to generalise from research papers and will be very useful for those new to this field.

Chapter 8 “Knowledge Representation for Language Engineering” (Matthew Stone) has the most philosophical outlook and is the least engineering focused of the contributions to this collection. It explains that the main role of knowledge representation in language engineering is to capture the connections between language and the world. Three knowledge representation schemes are introduced; description logic, first-order logic and modal logic, along with an explanation of their expressiveness and computational requirements.
Chapter 9 “Speech Recognition and Understanding” (Jan W. Amtrup) provides a comprehensive introduction to the field of speech processing. The main focus is on speech recognition although synthesis is described briefly. A large part of the chapter is given over to a description of how Hidden Markov Models can be applied to speech recognition including topics such as Viterbi decoding, model training and smoothing probability estimates.

The volume concludes with another contribution from the editor entitled “Language Engineering and the Knowledge Economy”. Farghaly describes the general trend towards a knowledge-based economy and points out the potentially important role of Language Engineering given that language is the most common medium for encoding knowledge.

Overall this volume certainly represents a useful practical introduction to Language Engineering. There are useful tutorial-like worked examples in several places and these will surely be invaluable to the intended audience. For example, Megerdchoomian’s overview of UNIX tools for corpus processing (Chapter 6) and Noy’s outline of a process for constructing ontologies (Chapter 5) would be useful for anyone starting a project in one of those areas with no previous experience. The coverage of topics in this collection is generally introductory but this is the correct level for this volume and is unlikely to cause problems for the reader in search of greater detail since each contribution provides comprehensive links to further online and print resources. The contributions also contain numerous examples of ways in which the technology could be applied to real Language Engineering applications.

Some of the chapters will require a reasonable degree of confidence in dealing with mathematical ideas and notation. In particular some of the sections in Chapters 7 and 9 (“Statistical NLP” and “Speech Recognition and Understanding”) may be a turn off for those coming from a linguistics background.

There are also places in which the same material is covered in more than one chapter. However, there should be clearer links between these. For example, sections 4.3.3 and 6.5.2 both discuss Abney’s approach to chunk parsing (Abney, 1996) but do not reference each other. The authors of these chapters chose different terminology to describe this work (“chunk parsing” and “partial parsing”) so those sections cannot be linked through the index. While this volume is not a tutorial introduction to Language Engineering, and it is sometimes appropriate for the same work to be described within the various chapters, it would be useful to see a clearer links between the contributions in some places.

This volume certainly forms a useful introductory text for the intended audience. It will also be useful for other groups such as those with a background in Computer Science who embark on a project in Language Engineering or Natural Language Processing and require an accessible introduction to a particular aspect of the field. The aim of the vast majority of publications related to Language Engineering is for scientists to communicate new results to their peer group. This volume has the unusual goal of synthesizing experience from practitioners in the field as an introduction to those with less experience.

It will be useful to anyone in that position and I shall recommend selected chapters as introductory reading for students new to language processing.
References

Mark Stevenson
*University of Sheffield*
marks@dcs.shef.ac.uk