Competence and performance in computational linguistics

Autor(en): **Hacken, Pius ten**

Objekttyp: **Article**

Zeitschrift: **SPELL : Swiss papers in English language and literature**

Band (Jahr): **11 (1998)**

PDF erstellt am: **27.04.2024**

Persistenter Link: https://doi.org/10.5169/seals-99964

Competence and Performance in Computational Linguistics

Pius ten Hacken

In this paper I explore the relationship between Computational Linguistics (CL) and the competence-performance distinction familiar from linguistic theory. I start with an introductory analysis of CL (section 1) and a discussion of different understandings of the competence-performance dichotomy (section 2). I then relate competence and performance to the processing of language (section 3) and consider their relevance in each of the mappings distinguished in CL (sections 4-5). Against the background of this analysis, I identify two basic approaches to CL based on competence and performance respectively (sections 6-7). Finally, I evaluate the role of competence and performance in the development of different types of CL (section 8).

1. Computational Linguistics

Computational Linguistics (CL) is concerned with performing certain tasks relating to human language on a computer. Since language is often considered the most basic property of human beings, distinguishing them from animals and machines, it is a particularly great challenge to process it on a computer. Taking up this challenge can be rewarding because computer programs developed in CL can be and are in most cases meant to be of practical use, i.e. as solutions to real-life problems. Moreover, formulating the knowledge involved in dealing with human language in such a way that a computer can use it may contribute to our understanding of (aspects of) human language, provided a proper context for CL research is created.

Two examples of applications developed in CL are dialogue systems and machine translation systems. A dialogue system may be used to make data stored in a large computer database available to people who do not know the structure of the database. Depending on the complexity of the query, the computer may produce the answer immediately or ask for more details. A machine translation (MT) system translates from one language into another. As shown by Hutchins & Somers, most MT systems start from and produce written text, but Kay et al. describe a system intended to take spoken language as input.

It would be misleading to say that a successful CL system *understands* human language. Whereas the human translation process can be described as understanding the source text and expressing its meaning in a different language, the operation of a computer involved in MT can be described more accurately as calculating the correct substitution of one set of symbols for another. The following analogy may offer an impression of how far MT is removed from human translation. Suppose you are asked to translate from Vietnamese into Swahili and you do not know either language. Having at your disposal a Vietnamese to Swahili dictionary only helps to a limited extent. Problems not solved by the usual type of bilingual dictionary include the fact that not every (inflected) word-form in the text appears as a dictionary entry, that for many words you have to choose one of several alternative translations, and that you cannot assume that each word in Vietnamese can be replaced by a word in Swahili without further changes in the text. You need explicit instructions on how to identify the kind of substitutions to be made. If these instructions are adequate, you can produce a Swahili text corresponding to the Vietnamese original, without knowing the meaning of the text. In this way one can imagine how a computer can translate without understanding and why formulating the instructions to be followed is such a challenge.

In the light of this example it is reasonable to characterize CL as the mapping between different representations of a message. These representations may be of three types: speech, text, and abstract representations. Speech is a linear, acoustic representation in terms of a continuous flow of sounds. Text is a linear, visual representation in terms of discrete symbols. Abstract representations are typically non-linear, structured in complex ways, and not intended for processing by human end-users of the programs developed in CL. I will use *information* to designate the most abstract of these representations. At the interface between the computer program and the human user, speech and text are appropriate representations.[1]

If we have three types of representations, in principle six types of mappings between them are possible. In practice, however, text can be taken to

[1] The presentation is simplified by not including sign language, which can be said to have a visual phonology, cf. Uyechi, and handwriting, which need not have discrete symbols. They are of minor importance in CL in terms of the amount of work devoted to them.

be in between speech and information, so that the direct mappings between speech and information do not occur, reducing the number of mappings to the four in Fig. 1:

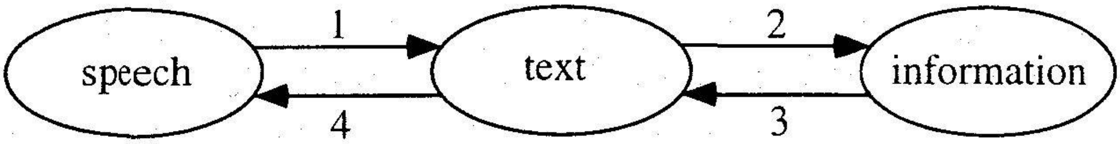

Fig. 1: Generic mappings in CL.

In Fig. 1, the mappings are numbered for convenience of reference. Mapping 1 is called speech recognition. It can be used in isolation as an automatic dictating machine. Mapping 2 is often referred to as natural language processing (NLP), but this term is also sometimes used as a synonym of CL. Mappings 1 and 2 are analysis components. Mapping 3 is text generation and mapping 4 speech generation. Mapping 4 is used in isolation in a reading machine, reading for instance newspapers to visually impaired people. In most CL systems, two or more mappings are combined. Many MT systems thus combine mappings 2 and 3 with a further intermediate mapping from the information level of the source language to that of the target language. In a dialogue system, the information level corresponding to a question will be used as input to a look-up device or a more sophisticated information processing device such as an inference mechanism to produce the information level from which generation of the answer takes place.

2. Competence and Performance

The terms *competence* and *performance*, used in opposition to each other in linguistics, have given rise to a great deal of confusion. In order to prevent confusion here, I will present and justify the senses in which I want to use them and briefly discuss some of the alternative interpretations.

The first time competence and performance were introduced in the relevant sense in linguistics seems to be in Chomsky's contribution to the Ninth International Congress of Linguists in Cambridge (Mass.) in 1962, published separately as *Current Issues*. In the introduction, specifying what should be studied in linguistics, the terms are introduced without special emphasis, more or less as normal words (7-11). In *Aspects*, however, they are treated as technical terms, defined as follows:

> We thus make a fundamental distinction between *competence* (the speaker-hearer's knowledge of his language) and *performance* (the actual use of language in concrete situations). (4)

The knowledge referred to here is the knowledge underlying grammaticality judgements. As explained by Newmeyer, misconceptions associated with the contrast between competence and performance have often been a basis for criticism of Chomskyan linguistics and related approaches (*Grammatical Theory* 35-38). Three objections discussed by Newmeyer are particularly worth mentioning here.

First, it has sometimes been claimed that, by using the contrast, one is committed to the view that everything systematic about language is covered by competence. Second, it has been claimed that the opposition commits one to considering all phenomena outside competence as uninteresting. Since these two points are closely connected, I will discuss them together. Part of the misunderstanding seems to be due to the association of competence and performance with Saussure's dichotomy of *langue* and *parole*. As Saussure states:

> En séparant la langue de la parole, on sépare du même coup: 1° ce qui est social de ce qui est individuel; 2° ce qui est essentiel de ce qui est accessoire et plus ou moins accidentel. (30)

Although the Saussurean and Chomskyan opposition pairs share many properties, the above quotation highlights some clear differences. First, whereas competence as understood by Chomsky is individual, Saussurian language is social. Second, as Chomsky pointed out immediately (*Current Issues* 23), competence, unlike *langue*, is not an inventory of elements but primarily a set of rules. Since Saussure did not foresee the possibility of formulating syntactic rules as part of the *langue*, he classified syntax as part of *parole*. Thus, in Saussure's division, all rules are together and only *langue* is "essential." In the Chomskyan dichotomy, however, both sides may be described in terms of rules. Far from implying that rules are absent from performance, Chomsky encouraged the study of rules in performance (*Aspects* 15).

A third unwarranted objection to the contrast as defined by Chomsky points to the absence of hard-and-fast criteria for drawing the dividing line between them. Performance is the result of the interaction of a number of cognitive modules, an important one being competence. Which aspect of performance should be covered by which module is an empirical question, related to the overall simplicity of the description. The state of development of the theory for each of the modules determines what is sensibly included in competence and what is in the domain covered by other modules.

The confusion about competence and performance increased due to the way the domain of theory of grammar was treated in generative semantics. As described by Newmeyer, generative semanticists had a tendency to include ever more types of knowledge in grammar (*Linguistic Theory in America* 118-125). Thus, Hymes conflates competence as intended by Chomsky with various other types of knowledge relevant to performance, producing the concept of *communicative competence* (12).

In view of this confusion, Chomsky avoids using the terms *competence* and *performance* in some of his more recent works. This does not imply, however, that the meaning of the concepts has lost importance. In his fairly technical book *Knowledge of Language*, Chomsky creates the new terms *I-language* and *E-language* corresponding to competence and performance. In the more popular presentation of his theory in the *Managua Lectures,* he uses the contrast between "knowledge of language" and "the ability to use it" (9-12), taking up almost literally the definitions of *competence* and *performance* in *Aspects* without using the terms.

It should also be noted that the distinction is not restricted to Chomskyan linguistics. In the context of Lexical-Functional Grammar, Bresnan & Kaplan refer to the *competence hypothesis* as one of the central issues of their theory of grammar. According to their interpretation of this hypothesis, a grammar of a language must be a psychologically realistic model of the speaker-hearer's knowledge of the language (xvii). In their presentation of Head-driven Phrase Structure Grammar, Pollard & Sag point out how certain concepts they introduce can be used to state the distinction between I-language and E-language. As I show elsewhere ("Progress and Incommensurability"; "Chomskyan Linguistics and HPSG"), these theories belong to research programmes other than Chomskyan linguistics. Therefore we can conclude that the relevance of the concepts of competence and performance is not restricted to the Chomskyan framework.

Since this discussion shows that the concepts associated with the terms *competence* and *performance* in *Aspects* are still important in various theories of grammar, despite criticism and misunderstandings, I will use the terms in their *Aspects* senses here.

3. Processing

The discussion in the preceding section provides a background for stating the relationship between competence, performance, and information content in the context of human processing more precisely. As a basis for exposition I take the highly idealized model of generation, shown in Fig. 2:

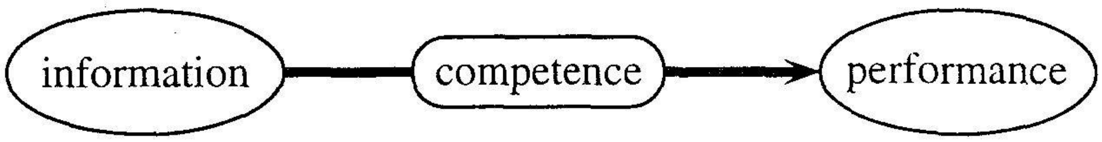

Fig. 2: Idealized interaction of competence and performance.

In Fig. 2, performance is represented as the result of the application of competence to the intention of conveying information. The reason for distinguishing competence and performance, however, is that performance is not the simple result of a process as in Fig. 2. Additional factors influencing performance are, on the one hand, other types of knowledge than competence, and on the other hand, restrictions on realization, limiting the correctness of performance with respect to competence and other knowledge modules. The other modules of knowledge include various types of pragmatic knowledge, interacting with competence in formulating the linguistic equivalent of the information to be conveyed. The restrictions on realization include memory limitations, imperfect concentration, etc., and may result in false starts or constructions which, upon reflection, are judged ungrammatical by the same speaker.

A striking point in Fig. 2 not highlighted in the discussion so far is the fact that competence and performance are not entities of the same type, differing only in one particular feature. On the contrary, competence is a module of knowledge and performance a level of representation. As a result, the ways in which the two are used in CL are different in nature. Performance is seen as the input or output of a mapping, in the same way as information. Competence is used in specifying how the mapping should be carried out. Another point which becomes more relevant when applying the terms to CL is their relationship to text and speech. The most straightforward initial assumption here is that both are performance, though of a different kind. In linguistic theory, following e.g. Bloomfield (20-21), written text is often considered a derived product of language, studied only when no spoken material is accessible, as in historical studies. In CL, however, written text is an essential part of the definition of many of the real-life problems to be solved.

4. Competence and Performance in Generation

Let us now consider each of the mappings in Fig. 1 in relation to competence and performance. The mapping which can be compared most straightforwardly to the representation in Fig. 2 is the one from information to text. Here it is obvious that competence plays a central role, ensuring that the output is grammatical. Other modules of knowledge play a role which may be subordinate, but can by no means be thought of as negligible. The output of text generation will in the general case constitute a set of alternatives which are synonymous as far as competence is concerned. The task of the other modules is to choose the best one from this set. Thus they should ensure that the output is contextually and emotionally acceptable. In a dialogue system, this knowledge can be directly encoded in the system by adapting the vocabulary and rules to a single application, i.e. avoiding messages which would be felt to be offensive or not to the point. In MT, the problem is in principle to reproduce the style of the original message. In generation, this problem is usually reduced in the same way as for dialogue systems, by restricting the subject field and text type. Lehrberger & Bourbeau explicitly propagate this strategy as the only one which can be expected to provide high-quality translations for a long time to come. Rosetta interprets its task as the production of the set of all possible translations for a sentence, leaving it to other modules to select one.

The mapping from text to speech has speech in the position of performance in Fig. 2 and text as information. This may come as a surprise, for we have regarded text as performance so far. If we restrict our attention to speech generation as an isolated system, however, all information available to the system is encoded in the text. It is obvious that a simple mapping from letters to phonemes is false, but even a full look-up in a pronunciation dictionary is insufficient. Recognizing phonological constituents necessary for an adequate pronunciation requires competence. A fully acceptable pronunciation can only be achieved by also taking into account other knowledge modules determining, for instance, intonation contours. In a system where both generation mappings are combined, the grammatical structure underlying the text may also be used to derive phonological constituents. The reason why Fig. 1 does not represent this as a direct mapping from information to speech is that the text representation is (almost) automatically available as a side effect.

5. Competence and Performance in Analysis

Turning now to the analysis components in CL, we find that the relationship between competence and the entire body of knowledge encoded in the CL system is of a different nature to that in generation. In generation, we can start with modelling competence so as to produce all grammatical messages corresponding to the input and improve the performance of the system by gradually including those parts of the other knowledge modules involved which are understood well enough to be formalized in a CL system. In analysis, on the other hand, we are confronted with performance. We cannot choose which part of the underlying modules we would like to take into account, but have to adapt to the input.

The optimal approach to analysis would be to model the full system of interacting modules underlying performance. If this were possible, we could make the computer analyse performance in the same way as human beings do. The problem is, however, that we lack theories for many relevant modules of knowledge. Moreover, an adequate coordinating theory of which modules are relevant in analysis and how they interact is beyond the horizon. What we do have is, on the one hand, a number of alternative linguistic theories describing competence more or less successfully and, on the other hand, performance, which is given as input. This provides us with two possible starting points.

First, we could take a linguistic theory, adapt its description of competence as a CL system and use it to analyse the input. If for the moment we disregard the question of which theories are more adequate than others, we will still find that they all fail in view of the task they are given here. Even a perfect description of competence would show mismatches with performance.

These mismatches do not influence the evaluation of the linguistic theories under consideration as theories of science. As I show in "Progress and Incommensurability," each research programme in linguistics has its own goals, leading to independent evaluation criteria. These goals are not directly approached by the application of the research results in CL, although successful application may have indirect beneficial effects by increasing the availability of research funding.

The second approach is to devise a CL system on the basis of performance. The problems facing us here are firstly that existing linguistic theories are of little use because they have a different goal, and secondly that positing a single source of performance requires a rule system of high complexity.

For a long time it has been considered impossible to write grammars of performance. With the increase of computer power, it is now possible to have the computer calculate a "grammar" on the basis of statistical operations on large quantities of performance data.

The evaluation of these two options is not the same for the two analysis mappings. A number of factors differentiating text analysis and speech analysis play a role in the extent to which each of the strategies is likely to be successful. A first difference is that whereas text consists of letters, i.e. discrete minimal units whose recognition is trivial, the minimal units of speech are much more difficult to recognize. Phonemes as produced and recognized by human language users are not acoustically present and cannot be recorded in any simple way by a computer. A single phoneme has a whole range of possible acoustic realizations, and the ranges of different phonemes show large areas of overlap. Overlap is reduced somewhat by concentrating on a single speaker or on a small domain to be talked about. In the latter case, the number of words to choose from is reduced, so that overlap in phonemes is less likely to result in ambiguity at word level. The problem can never be reduced in such a way, however, that phoneme recognition becomes as trivial as the recognition of letters.

Another factor which influences the evaluation of the two strategies for analysis is the influence of interfering factors on the production of the input to the CL system. These factors include how carefully the input has been formulated, to what extent reflection and correction are possible, etc. In general, text will be closer to grammaticality than speech. This is not a matter of two points on a simple cline. Rather, speech and text each present a cline from more to less grammatical, and these clines may overlap. Thus, the speech of an actor on stage or a professional news reader is likely to contain a higher rate of grammatical sentences than a corpus of quickly written e-mail messages. Even in this example, however, the two clines are separate in the sense that the written medium offers more opportunities to reduce ungrammaticality. When writing an e-mail message one can stop, consider what one has written, and insert or delete a few words. What is spoken cannot be corrected.

Both factors conspire to bring text closer to grammaticality in terms of competence than speech. As expected, there is a tendency for speech analysis to employ statistical knowledge based on performance and for text analysis to turn to linguistic theories of competence.

6. Competence-Based CL

As shown in the previous section, there are two broad approaches to CL, which can be labelled competence-based CL and performance-based CL. In competence-based CL, the theoretical basis of work in CL is a theory of linguistics. Obviously, the degree of success of such an approach depends on three factors: the task of the CL system, the linguistic theory chosen, and the application of the theory to the task. The task of the CL system determines how close its input and output are to what is accepted by competence. This has been dealt with in sections 4 and 5. Here we will concentrate on the other two factors.

The choice of a linguistic theory is relevant in different respects. The first is the perspective of explanation chosen in the research programme. The purpose of a scientific theory is to describe and explain the system underlying a certain class of empirical observations. The research programme in which the theory is embedded determines which perspective is chosen for explanation and what kind of entity is supposed to underlie the data. As I show in "Progress and Incommensurability", Chomskyan linguistics and Lexical-Functional Grammar (LFG) both take a grammar to be a description of competence. The perspective of explanation in Chomskyan linguistics is learnability, in LFG human language processing. Whereas Chomsky repeatedly formulates the goal of linguistic theory as explaining how a child can acquire its native language (e.g. *Aspects* 25-26; *Government & Binding* 3-4), Bresnan & Kaplan aim to explain how language users establish the link between a string of words and its analysis.[2] Since processing is more closely related to the tasks in CL, it is to be expected that theories in the research programme of LFG are more easily adapted to CL systems than theories in Chomskyan linguistics.

A second respect in which the choice of a linguistic theory is relevant to the success of a CL system is the type of formalization chosen in the linguistic theory. The formalism is in principle independent of the research programme. It has a major influence on the content of actual discussion and work in linguistic research, however, and determines how results are presented. The abundant use of movement, functional projections, and empty categories in Chomskyan linguistics makes it difficult to implement such

[2] Head-driven Phrase Structure Grammar (Pollard & Sag) fails to choose a perspective for explanation, as I show in "Chomskyan Linguistics and HPSG." Generalized Phrase Structure Grammar (Gazdar et al.) does not choose competence as the entity described by a grammar, as I show in "Research Programmes."

theories on a computer. In LFG, by contrast, Kaplan & Bresnan proposed a working processor based on unification at an early stage in the development of the research programme.

If a theory is chosen which adapts well to implementation in a CL environment, the adaptation itself is less important than in the case of a theory requiring more adaptation work. For theories taken from Chomskyan linguistics, it is often necessary to specify choices which are left underspecified in the theory, because in the framework in which the theory was proposed, these choices are irrelevant. Without affecting the evaluation of such a theory as (part of) a theory of grammar, this reduces its suitability for use in CL.

The advantages and problems of competence-based CL are strongly interrelated. If competence is described and implemented well, the system will generate grammatical sentences and, in analysis, recognize where performance deviates from it. In recognition especially, we can see this alternatively as an advantage or as a problem. The problem is what is usually called a lack of robustness, a tendency not to give any answer when the input does not correspond exactly to the system's expectation. The advantage is that the system can be made aware of these cases. This awareness can be used as a basis for improvement and extension and as a basis for explaining the behaviour of the CL system, thus going beyond mere technology and aspiring to the status of applied science.

7. Performance-Based CL

In performance-based CL, the information theory developed by Shannon provides a theoretical background. This theory was first devised for the reconstruction of spoken messages down a telephone line of uncertain quality. In its most general form, the underlying model can be represented as in Fig. 3.

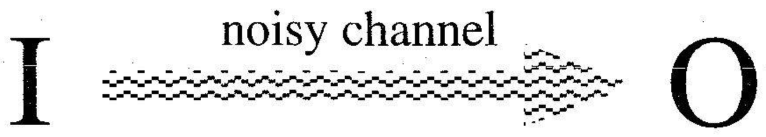

Fig. 3: The Noisy Channel Model.

The noisy channel transforms the original message I without any underlying intention, but not in an entirely random way either. The regularity can be approximated by collecting a corpus of messages and their corresponding output. The most probably intended message element In corresponding to an output element O_n is found by multiplying, for each possible input element I_x, the probability that I_x is intended, written as *Pr* (I_x), with the probability that I_x will appear as O_n after transmission, written as *Pr* ($O_n | I_x$). The I_x for which the product of the two probabilities is maximal is considered the best guess. The probabilities *Pr* (I_x) and *Pr* ($O_n | I_x$) are estimated on the basis of statistical operations over the corpus (cf. Charniak).

The application of the model in Fig. 3 to tasks other than reconstructing telephone conversations requires some stretching of the idea of *channel*. In speech recognition, the text corresponds to I, the speech to O, and the speaker to the noisy channel, imperfectly transmitting the text. In MT, the channel is a translator in the direction opposite to the direction of translation performed by the MT system. Strange though these applications may seem, the method can be and has been applied in these ways. The limiting condition on applicability of this model is the availability of a sufficiently large aligned corpus of input and output as a basis for the statistical calculations. The size of the corpus influences the success rate. The alignment of the input and output ensures that it is possible to know which element I_n corresponds to O_n in a particular position. The requirement that large aligned corpora are available restricts the applicability of the method in practical terms. Complex representations of information, as involved in mappings 2 and 3 in Fig. 1, should be avoided by going from one text representation to another directly. Dialogue systems can hardly be covered in this way.

The advantages and problems of performance-based CL are close to the mirror image of those of competence-based CL. Since all possible *Pr* (I_x) and *Pr* ($O_n | I_x$) have positive values, a performance-based CL system is entirely robust. It will always come up with a result, no matter how bad the input. This is often seen as an advantage, but it has a price. Errors in the input are not recognized as different from normal input, and the internal structure of the task has no correlation with the way the task is carried out by humans. As a result, these systems are purely technological without any explanatory element which might lead the way to applied science.

8. Evaluation

On the basis of the above discussion, we may be inclined to divide the mappings in Fig. 1 between the two approaches to CL so that mapping 1 is attributed to performance-based CL and mappings 2–4 to competence-based CL. Although this is not entirely contrary to fact, it is a clear oversimplification. Church & Mercer claimed in 1993 that "Over the past 20 years, the speech community has reached a consensus in favor of empirical methods." Here empirical methods are what we have called performance-based CL. The consensus does not seem to be so general, however. Even in the selection of texts by Waibel & Lee, which Church & Mercer quote as evidence for their claim, "knowledge-based approaches" are represented by several papers from the second half of the 1980s. On the other hand, performance-based CL is less confined than the simplified generalization suggests, as exemplified by part-of-speech tagging and MT.

Part-of-speech tagging is the classification of words occurring in a text in terms of labels such as noun, verb, etc. or a more fine-grained variant of such a taxonomy. In a language such as English, with extensive ambiguity in this respect, the task is far from trivial. According to Sampson and Church & Mercer, this area of CL has been taken over almost entirely by performance-based approaches, and success rates of over 98% have become common. In Fig. 1, tagging is a mapping of type 2. Compared to other mappings of this type, it is relatively open to a statistical approach because the tags are a relatively small set associated with words in a simple way.

At first sight, MT might seem the most obvious area for competence-based CL. A problem, however, is the absence of a translation theory to go with a description of competence. Already in 1949 Weaver proposed to use statistical techniques for MT, but the amount of work involved required more powerful computers than would be available for a long time. In 1990 Brown et al. proposed a statistical MT system developed at IBM. As mentioned in section 7, a purely performance-based MT system does not have an information level, so that it defies the model in Fig. 1.

In general, the choice between a competence-based and a performance-based approach for a particular task in CL depends on the type of resources available and the type of solution desired. Performance-based CL requires large quantities of data in an aligned corpus of input and output, competence-based CL the availability of a theory for the relevant parts of competence. Performance-based CL offers a robust system without relevant internal structure, competence-based CL a more transparent, less robust system, which can in principle know its own performance.

For practical purposes, it may be useful to combine competence-based and performance-based modules in a single CL system. As long as the interfaces between the modules are well-defined, this is possible without losing any of the desirable properties of the individual modules. Thus, though fully performance-based MT systems are exceptional, the lack of a formal translation theory has given rise to various hybrid systems with statistical modules for particular tasks.

At various points in this paper I have hinted at the possibility of making CL scientific rather than just technological. For reasons of space, I cannot develop this issue here, but the following comparison is suggestive. In evaluating a competence-based CL system, we can collect errors in the output and analyse which features of the system are responsible for them. In evaluating a performance-based CL system, we can only register errors and calculate their frequency as a percentage of the input. Therefore, only competence-based systems can be improved locally and their performance can be explained in terms of the underlying linguistic theory. Explanation is an essential property of science.

Works Cited

Bloomfield, Leonard. *Language.* London: Allen & Unwin, 1933.

Bresnan, Joan and Ronald M. Kaplan. "Introduction: Grammars as Mental Representations of Language." *The Mental Representation of Grammatical Relations.* Ed. Joan Bresnan. Cambridge, MA: MIT Press, 1982. xvi-lii.

Brown, Peter, John Cocke, Stephen Della Pietra, Vincent Della Pietra, Fredrick Jellinek, John D. Lafferty, Robert L. Mercer and Paul S. Roossin. "A Statistical Approach to Machine Translation." *Computational Linguistics* 16 (1990): 79-85.

Charniak, Eugene. *Statistical Language Learning.* Cambridge, MA: MIT Press, 1993.

Chomsky, Noam. *Current Issues in Linguistic Theory.* Den Haag: Mouton, 1964.

———. *Aspects of the Theory of Syntax.* Cambridge, MA: MIT Press, 1965.

———. *Lectures on Government and Binding.* Dordrecht: Foris, 1981.

———. *Knowledge of Language: Its Nature, Origin, and Use.* Westport, CT: Praeger, 1986.

———. *Language and Problems of Knowledge; The Managua Lectures.* Cambridge, MA: MIT Press, 1988.

Church, Kenneth W. and Robert L. Mercer. "Introduction to the Special Issue on Computational Linguistics Using Large Corpora." *Computational Linguistics* 19 (1993): 1-24.

Gazdar, Gerald, Ewan Klein, Geoffrey Pullum and Ivan Sag. *Generalized Phrase Structure Grammar*. Oxford: Blackwell, 1985.

ten Hacken, Pius. "Progress and Incommensurability in Linguistics." *Beiträge zur Geschichte der Sprachwissenschaft* 7 (1997): 287-310.

———. "Research Programmes in Linguistics." *Tendenzen europäischer Linguistik, Akten des 31. Linguistischen Kolloquiums, Bern 1996*. Ed. Jürg Strässler. Tübingen: Niemeyer. 67-70. (In print)

———. "Chomskyan Linguistics and HPSG as Competing Research Programmes." *Actes du 16e Congrès International des Linguistes (Paris 20-25 juillet 1997)*. Ed. Bernard Caron. Oxford: Elsevier Sciences. (In print)

Hutchins, W. John and Harry L. Somers. *An Introduction to Machine Translation*. London: Academic Press, 1992.

Hymes, Dell. "Competence and Performance in Linguistic Theory." *Language Acquisition: Models and Methods*." Eds. Renira Huxley and Elisabeth Ingram. London: Academic Press, 1971. 3-24.

Kaplan, Ronald M. and Joan Bresnan. "Lexical-Functional Grammar: A Formal System for Grammatical Representation." *The Mental Representation of Grammatical Relations*. Ed. Joan Bresnan. Cambridge, MA: MIT Press, 1982. 173-281.

Kay, Martin, Jean Mark Gawron and Peter Norvig. *Verbmobil: A Translation System for Face-to-Face Dialog*. Stanford, CA: CSLI, 1994.

Lehrberger, John and Laurent Bourbeau. *Machine Translation: Linguistic characteristics of MT systems and general methodology of evaluation*. Amsterdam: John Benjamins, 1988.

Newmeyer, Frederick J. *Grammatical Theory: Its Limits and Its Possibilities*. Chicago: University of Chicago Press, 1983.

———. *Linguistic Theory in America*. Second edition, New York: Academic Press, 1986.

Pollard, Carl and Ivan A. Sag. *Head-Driven Phrase Structure Grammar*. Chicago: Chicago University Press & Stanford, CA: CSLI, 1994.

Rosetta, M.T. *Compositional Translation*. Dordrecht: Kluwer Academic, 1994.

Sampson, Geoffrey. "Probabilistic models of analysis." *The Computational Analysis of English: A Corpus-Based Approach*. Eds. Roger Garside, Geoffrey Leech and Geoffrey Sampson. London: Longman, 1987. 16-29.

Saussure, Ferdinand de. *Cours de linguistique générale*. [1916] Eds. Charles Bailly & Albert Sechehaye. Édition critique préparée par Tullio de Mauro. Paris: Payot, 1981.

Shannon, Claude S. "The mathematical theory of communication." *Bell Systems Technical Journal* 27 (1948): 379-423 and 623-656.

Uyechi, Linda. *The Geometry of Visual Phonology*. Stanford, CA: CSLI, 1996.

Waibel, Alex and Kai-Fu Lee, eds. *Readings in Speech Recognition.* San Mateo, CA: Morgan Kaufmann, 1990.

Weaver, Warren. "Translation." Ms. [1949] Reprinted *Machine Translation of Languages: Fourteen Essays*. Eds. William N. Locke and A. Donald Booth. Cambridge, MA: Technology Press of MIT & New York: Wiley, 1955. 15-23.