THE ENCYCLOPEDIA OF LANGUAGE AND LINGUISTICS

Volume 3

Editor-in-Chief
R. E. ASHER
University of Edinburgh, UK

Coordinating Editor
J. M. Y. SIMPSON
University of Glasgow, UK

PERGAMON PRESS
OXFORD • NEW YORK • SEOUL • TOKYO

TOPIC (133)
yinu bunya buga·-dyi-ŋ
this ABS woman ABS eat-ANTIPASSIVE-PRES
'this woman is eating'

TOPIC (134)
wagu·dya bambi·-dyi-ɲu
man ABS cover-ANTIPASSIVE-PAST
'the man has covered himself'

bliography

linger D 1986 *Intonation and Its Parts: Melody in Spoken English.* Edward Arnold, London

esnan J, Mchombo S 1987 Topic, pronoun and agreement in Chicheŵa. *Lg* **63**: 741–82

own G, Yule G 1983 *Discourse Analysis.* Cambridge University Press, Cambridge

hafe W 1976 Giveness, contrastiveness, definiteness, subjects, topics, and point of view. In: Li C (ed.) *Subject and Topic.* Academic Press, New York

omrie B 1979a Russian. In: Shopen T (ed.) *Languages and Their Status.* Winthrop Publishers, Cambridge, MA

omrie B 1979b Definite and animate direct objects: A natural class. *Linguistica Silensiana* **3**: 13–21

ixon R M W 1979a Ergativity. *Lg* **55**: 59–138

ixon R M W 1979b *A Grammar of Yidiŋ.* Cambridge University Press, Cambridge

oley W A, Van Valin R D Jr 1985 Information packaging in the clause. In: Shopen T (ed.) *Language Typology and Syntactic Description. Vol. 1: Clause Structure.* Cambridge University Press, Cambridge

ivón T 1979a *On Understanding Grammar.* Academic Press, New York

ivón T (ed.) 1979b *Syntax and Semantics. Vol. 12: Discourse and Syntax.* Academic Press, New York

alliday M A K 1967 Notes on transitivity and theme in English. *JL* **3**: 37–81, 199–244

awkins J 1979 *Definiteness and Indefiniteness: A Study in Reference and Grammaticality Prediction.* Croom Helm, London

oue K 1979 Japanese: A story of language and people. In: Shopen T (ed.) *Languages and their Speakers.* Winthrop Publishers, Cambridge, MA

irsner R 1976 On the subjectless "pseudo-passive" in standard Dutch and the semantics of background agents. In: Li C (ed.) *Subject and Topic.* Academic Press, New York

uno S 1973 *The Structure of the Japanese Language.* MIT Press, Cambridge, MA

ambrecht K 1981 *Topic, Antitopic and Verb Agreement in Nonstandard French.* Benjamins, Amsterdam

i C N (ed.) 1976 *Subject and Topic.* Academic Press, New York

i C N, Thompson S A 1975 The semantic function of word order: A case study of Mandarin. In: Li C N (ed.) *Word Order and Word Order Change.* University of Texas Press, Austin, TX

rince E F 1978 A comparison of WH-clefts and *it*-clefts in discourse. *Lg* **54**: 883–906

rince E F 1981 Topicalization, focus-movement and Yiddish-movement: A pragmatic differentiation. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* **7**: 249–64

ilverstein M 1976 Hierarchy of features and ergativity. In: Dixon R M W (ed.) *Grammatical Categories in Australian Languages.* Australian Institute of Aboriginal Studies, Canberra

Thompson S A 1978 Modern English from a typological point of view: Some implications of the function of word order. *Linguistische Berichte* **54**: 19–35

Thompson S A 1987 The passive in English: A discourse perspective. In: Channon R, Shockey L (eds.) *In Honour of Ilse Lehiste.* Foris, Dordrecht

Weiner E J, Labov W 1983 Constraints on the agentless passive. *JL* **19**: 29–58

W. A. Foley

Information Theory

Information Theory (also Communication Theory) is a mathematical theory which allows quantitative analysis of uncertainty and reduction in uncertainty. It was developed in the 1940s by Claude Shannon and Warren Weaver (1949) and has been applied quite extensively to the study of language and language processing in a variety of ways. The technical notion of 'information,' as a measure of uncertainty, should be sharply distinguished from the intuitive notion of information, as concerning significance or meaning. A randomly chosen (and hence meaningless) stream of letters will generate more information than a stream of letters composing a meaningful text, since it is far more uncertain (that is, it is far less predictable). Information theory concerns amounts of information carried by a signal, and not the content, if any, of that signal. Nonetheless, Information Theory has found wide application in the study of natural language, being used to measure the amount of structure in written text, to model the transmission of linguistic and other material across a communication channel and in 'information-based' approaches to semantics.

1. Amounts of Information

Information Theory treats the amount of information associated with some event with the reduction in uncertainty, roughly the reduction in the number of possible states of affairs, given that event. Suppose that the relevant states of affairs are the heads/tails orientation of three coins, after they are flipped. Assuming that the coins are unbiased, there are eight equally likely possible outcomes: HHH, HHT, HTH, HTT, THH, THT, TTH, TTT. If one coin is flipped and falls 'heads,' then the number of possible outcomes is reduced to four—the remaining two coins may still be either 'heads' or 'tails': HHH, HHT, HTH, HTT. Hence, uncertainty about the states of the coins has been reduced by the first coin's falling heads from 8 to 4. The second coin is then flipped and falls 'tails.' Now there are only two possible outcomes remaining—HTH, HTT. After the third coin has fallen 'tails' there is only one possibility remaining: HTT. There is no longer any uncertainty about the heads/tails states of the three coins. The flipping of each coin reduces the number of (equally likely) outcomes by $\frac{1}{2}$, and is said to generate 1 bit (binary digit) of information. The flipping of all three coins reduces the number of outcomes by $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}$, and generates 3 bits of information. In general, however, the information carried by a signal need not be an integer, but can be any positive real number.

Turning to a more general and precise formulation, let a source be a specification of a set of mutually exclusive and exhaustive possibilities each with an associated probability. For a source S we shall denote its n possible states, $S_1, \ldots, S_i, \ldots, S_n$. Each state S_i has an associated probability $p(S_i)$.

The amount of information associated with S being in state S_i is $-\log p(S_i)$ bits of information, where $p(S_i)$ is the probability that S is in state S_i. Thus each state, S_i, of a source is associated with an amount of information, known as its surprisal, which is related to how unlikely that state is (a state which has probability $\frac{1}{2}$ has a surprisal of 1 bit, a state with probability $\frac{1}{4}$ a surprisal of 2 bits, and so on).

The information, $I(S)$, generated by a source S is the expected surprisal of that source, $\Sigma_{i=1}^{n} p(S_i) \log p(S_i)$. For a source with a fixed number of states, the maximum amount of information is generated when each state is equally likely, rather than some states being very likely and others extremely unlikely. For although very unlikely states contribute a large amount of information when they do occur, since they occur very rarely they contribute little to $I(S)$.

2. Early Applications to Linguistic Material

An early application of Information Theory to language was Shannon's analysis of the uncertainty of sequences of words in English sentences. To the extent that a sequence is predictable rather than random, it is said to be 'redundant.' The extent to which a sequence of words is predictable depends, of course, on the amount of information that is utilized in making that prediction. If, for example, the only basis for the prediction is the previous word, then prediction may be relatively poor. If prediction is based on the previous two, three, and so on words, then better prediction will be possible (of course, better prediction still may be possible, given knowledge of the topic, the speaker, the audience, and so on). Thus, redundancy is not measured absolutely but relative to some particular model of the structure of the sequence. One class of interesting models that Shannon applied was based on the 'n-gram' or 'order' statistics of sentences. An 'n-gram' is a set of n adjacent linguistic items (which may be words, letters, phonemes). The n-gram statistics of a language (or more precisely, of some particular corpus which is under analysis) measure the frequencies with which each possible n-gram occurs. For entirely random sequences, at each particular value of n, all n-grams are equally likely; to the extent that the language is structured and therefore predictable (for a particular n), the n-gram statistics will be highly nonuniform.

Shannon measured such statistics for English, and then generated strings of items according to such statistics. A '1-gram' sequence consists of an unstructured jumble of words chosen according to their individual frequencies; in a '2-gram' sequence successive pairs of words respect the bigram statistics of the language; and so on. With increasing n, such strings increasingly model the structure of English. Shannon applied this method both at the level of letters and at the level of words. For example, at the letter level, in a first-order approximation letters are chosen independently, but respecting their frequencies in the language: 'OCRO HLI RGWR NMIELWIS.' A second-order approximation, mirroring observed '2-gram' frequencies, produces strings like 'ON IE ANTSOUTINYS ARE T INCTORE ST BE S.' A third-order approximation, respecting trigram frequencies, yields sequences like 'BIRS GROCID PONDENOME OF DEMONSTURES OF.' With increasing n these sequences (and their word level counterparts) become successively better approximations to English text.

At the word level, in some early work the hope was expressed that sufficiently high order statistics would be able to fully capture the structure of language. This view comported well with 'response-chaining' views of language production from behaviorist psychology, according to which the sequential structure of sentences is a product of associations between previous words acting as controlling stimuli for subsequent words (Skinner 1957). Chomsky's (1957) insistence that the structure of language is not finite state dampened enthusiasm for this approach to syntax. Order statistics are, by their nature, able only to pick up local rather than unbounded dependencies, and are hence unable to capture the structure of languages with more than finite state power. Nonetheless, the use of frequency statistics in studying syntax is still pursued, though using rather more sophisticated techniques, under the title of stochastic parsing (see also *Stochastic Techniques for Automatic Speech Recognition*). Furthermore, in domains in which dependencies are in general quite local, information theoretic measures and related statistical methods are used extensively (for example, Hidden Markov Modeling in the study of low-level speech perception).

3. The Transmission of Information

So far information properties of a single informational source have been considered. Information transmission involves specifying the connection between two such sources, designated as the Source (S), which transmits information, and the Receiver (R) which picks up information from the Source. The set of conditional probabilities of each state of the Receiver given each state of the Source defines an information channel. If the state of the Receiver is informationally related to the state of the Source, then the state of the receiver will be informative about, i.e., will reduce the uncertainty, at the source. The expected value of this reduction of uncertainty is the amount of information carried. A usual metaphor is to think of information 'flowing' from Source to Receiver along the information channel. The direction of information flow need not correspond in any straightforward way to the direction of causality: smoke may carry the information that there is fire, though fire causes smoke.

The amount of information transmitted by a channel, $T(R \mid S)$, is the average amount to which the uncertainty of the source is reduced, given that the state of the receiver is specified. Suppose that the receiver is in a particular state R_j. Then the information generated at the source is calculated as before, but with probabilities conditional on R_j: $\Sigma_{i=1}^{n} p(S_i \mid R_j) \log p(S_i \mid R_j)$. The expected or average value of uncertainty at the source, over all states of the receiver is just $\Sigma_{j=1}^{m} \Sigma_{i=1}^{n} p(R_j) p(S_i \mid R_j) \log p(S_i \mid R_j)$. The average reduction in uncertainty is calculated by subtracting this quantity from the information, $I(S)$, generated by the source. This reduction gives $T(R \mid S)$, the amount of information that the receiver carries about the source (equally, the average amount of information transmitted by the channel).

Two other important quantities, noise and equivocation, are straightforwardly related to $T(R \mid S)$. Uncertainty at the Receiver which is not conveyed by the information channel is termed 'noise,' $N = I(R) - T(R \mid S)$, and uncertainty at the Source which is not transmitted by the information channel is called 'equivocation,' $E =$

$I(S) - T(R \mid S)$. Considering a phone line as supporting an information channel, frequencies of the speaker's voice that cannot be picked up by the microphone will give rise to equivocation; crackle on the line unrelated to the speaker counts as noise.

It is important to stress again that the amount of information transmitted by a signal is not (straightforwardly) related to the meaning of that signal. As noted above, the transmission of a meaningless sequence may carry more information than a meaningful, but highly redundant, string of natural language, since it may be far less predictable.

4. Information and Semantics

Despite the apparent mismatch between the informational–theoretic notion of information, and the everyday notion concerned with meaning or significance, Dretske (1981) has attempted to use information-theoretic ideas to ground a notion of semantic content. He considers representational structures to be Receivers which carry information about environmental Sources. He defines what it is to carry a particular piece of information, rather than simply to carry a certain amount of information. According to this definition, each state of the Receiver will typically carry many pieces of information about the Source. A particular piece of information (very roughly, the most general piece of information, in which all other pieces of information carried are included) is defined to be the digital content of a state of the representation structure, a notion that Dretske closely identifies with the intuitive conception of meaning. However, most work in information-based semantics (e.g., Barwise and Perry 1983), while committed to the idea that semantic notions must somehow be explicated in terms of informational dependencies, does not employ information theory directly.

Bibliography

Barwise J, Perry J 1983 *Situations and Attitudes.* MIT Press, Cambridge, MA

Chomsky N 1957 *Syntactic Structures.* Mouton, The Hague

Dretske F I 1981 *Knowledge and the Flow of Information.* Blackwell, Oxford

Shannon C E, Weaver W 1949 *The Mathematical Theory of Communication.* University of Illinois Press, Urbana, IL

Skinner B F 1957 *Verbal Behavior.* Appleton Century Croft, New York

N. Chater

Informative Analysis

A speech act is created when speaker S makes an utterance U to hearer H in context C. In every speech act S conveys a message by means of an illocutionary act, e.g., S makes a statement, a request, a promise, an apology, issues an order, asks a question (see *Speech Act Theory—an Overview*; *Speech Acts Hierarchy Locutions, Illocutions and Perlocutions*). There is often more than one H, and S may have different intentions toward (messages for) different Hs (see *Speech Act Theory—an Overview* on addressees and other kinds of hearers). When S is speaking, all those who can reasonably consider themselves Hs are expected, as part of the cooperative endeavor, to keep tabs on what is said, so that if called upon to participate they may do so appropriately. Clark and Carlson account for these facts in what they call 'the informative analysis':

> the speaker performs two types of illocutionary act with each utterance. One is the traditional kind, such as an assertion, promise, or apology; this is directed at the addressees. The other, called an informative, is directed at all participants in the conversation—the addressees and third parties alike. It is intended to inform all of them jointly of the assertion, promise, or apology being directed at the addressees.
> (1982: 332)

This entry considers whether it is useful to believe that *every* utterance is informative, along with having some other illocutionary point.

Suppose S utters (1) to H_1, H_2, and H_3:

Will the last one of you to go to bed please turn out the light? (1)

The informative analysis postulates the following illocutions (2):

S *informs* H_1, H_2, and H_3 that S *requests* the last one of them to go to bed to please turn out the light. (2)

The claim is that all Hs are informed of the request, but for the last one of them to go to bed (suppose it is H_2) the illocutionary point is that of a request—no request is made of H_1 or H_3. The traditional and simpler view is that the request is made of all three Hs; because at the time of utterance it is unknown who will eventually comply with the request (perhaps none of them). This view seems to be justified when one considers a realistic true report of (1): *S asked that the last one of us to go to bed should turn out the light.* It would be distinctly peculiar for the report to go, *S informed us that the last one of us to go to bed should turn out the light.* It would also be odd for any of H_1, H_3, or even H_2 to report *S asked me to turn out the light*: even though this is logically correct, it is pragmatically false because it ignores the true implicatures of (1).

[Traffic cop addresses Ed in the presence of passenger Jo.] (3)
Please show me your driver's license.
[Ed ignores the request, so the cop addresses Jo.]
You heard me ask him to show me his driver's license, didn't you?

It would not have been natural for the cop to say to Jo, *I informed you that I asked him to show me his driver's license, didn't I?*; though s/he might say, *I am asking this man* [Ed] *to show me his driver's license.* Here the cop is describing what s/he is doing and, perhaps, indirectly remaking the request of Ed. However, there is no confirmation of the informative analysis.

All entailments and implicatures of a proposition within an utterance are communicated, and these give rise to indirect (and occasionally unintended) illocutions. Strictly speaking S does not *tell H* these things, but *lets H know* them.

[Mom to Susan's fiancé Fred in front of Susan] (4)
Why are you playing around with that Mabel?

Example (4) entails that Mom believes Fred is two-timing Susan. Her question directly challenges him to explain his disgraceful behavior. But if there has been no admission by any participants hitherto of Fred's affair, it also succeeds in indirectly (and tactlessly) informing Susan of the content